Machine Learning Methods for Human Identification from Dorsal Hand Images

Mona Saleh Alghamdi, BSc (Hons), Msc
School of Computing and Communications
Lancaster University

A thesis submitted for the degree of
Doctor of Philosophy

July, 2023

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography. A rough estimate of the word count is: 31735

Mona Saleh Alghamdi

Machine Learning Methods for Human Identification from Dorsal Hand Images

Mona Saleh Alghamdi, BSc (Hons), Msc.
School of Computing and Communications, Lancaster University
A thesis submitted for the degree of *Doctor of Philosophy*. July, 2023

Abstract

Person identification is a process that uniquely identifies an individual based on physical or behavioural traits. This study investigates methods for the analysis of images of the human hand, focusing on their uniqueness and potential use for human identification. The human hand has significant and distinctive characteristics, and is highly complex and interesting, yet it has not been explored in much detail, particularly in the context of the contemporary high level of digitalisation and, more specifically, the advances in artificial intelligence (AI), machine learning (ML) and computer vision (CV). This research area is highly multi-disciplinary, involving anatomists, anthropologists, bioinformaticians, image analysts and, increasingly, computer scientists. A growing pool of advanced methods based on AI, ML and CV can benefit and relate directly to a better representation of the human hand in computer analysis. Historically, the research methods in this area relied on 'handcrafted' features such as the local binary pattern (LBP) and histogram of gradient (HOG) extraction, which necessitated human intervention. However, such approaches struggle to encode the human hand in variable conditions effectively, because of the change in camera viewpoint, hand pose, rotation, image quality, and self-occlusion. Thus, their performance is limited. Recently, there has been a surge of interest in deep learning neural network (DLNN) approaches, specifically convolutional neural networks (CNNs), due to the highly accurate results achieved in many applications and the wide availability of images. This work investigates advanced methods based on ML and DLNN for segmenting hand images with various rotation changes into different patches (e.g., knuckles and fingernails). The thesis focuses on developing ML methods like pre-trained CNN models on the 'ImageNet' dataset to learn the underlying structure of the human hand by extracting robust features from hand images with diverse conditions of rotation, and image quality. Also, this study investigates fine-tuning the pre-trained models of DLNN on subsets of hand images, as well as using various similarity metrics to find the best match of

the individual's hand. Furthermore, this work explores different types of ensemble learning or fusions, those of different region and similarity metrics to improve human identification results. This thesis also presents a study of a Siamese network on sub-images or segments of human dorsal hands for identification tasks. All presented approaches are compared with the state-of-the-art methods. This study advances the understanding of variations in and the uniqueness of humans using patches of their hands (e.g., different types of knuckles and fingernails). Lastly, it compares the matching performances of the left- and right-hand patches using various hand datasets and investigates whether the fingernail produces better identification results than the knuckles. This research shows that the proposed framework for person identification based on hand components led to better person identification results. The framework consists of vital feature extractions based on deep learning neural network (DLNN) and similarity metrics. These elements enhanced the performance. Also, the fingernails' shape performed better than other hand components, including the base, major, and minor knuckles. The left hand can be more distinguishable to individuals than the right hand. The fine-tuning of the hand components and ensemble learning improved the identification results.

Publications

The following publication have been generated while developing this thesis, and to an extent has guided the thesis into what it has become:

M. Alghamdi et al. (2019). "Self-Organising and Self-Learning Model for Soybean Yield Prediction". In: *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019*. ISBN: 9781728129464. DOI: 10.1109/SNAMS.2019.8931888

Mona Alghamdi, Plamen Angelov, and Bryan Williams (2021). "Automated Person Identification Framework Based on Fingernails and Dorsal Knuckle Patterns". In: *2021 IEEE Symposium Series on Computational Intelligence, SSCI 2021 - Proceedings*. DOI: 10.1109/SSCI50451.2021.9659850

Mona Alghamdi, Plamen Angelov, and Lopez Pellicer Alvaro (2022). "Person identification from fingernails and knuckles images using deep learning features and the Bray-Curtis similarity measure". In: *Neurocomputing* 513, pp. 83–93. ISSN: 18728286. DOI: 10.1016/j.neucom.2022.09.123. URL: https://doi.org/10.1016/j.neucom.2022.09.123

Mona Alghamdi (2022). "A Multi-modal Biometric Approach Based on Score-level Fusion and Fine-tuning Deep Learning Features". In: *IEEE IS'22*

Acknowledgements

First and foremost, I want to thank the God Almighty for bestowing upon me the stamina, wisdom, and fortitude to endure all the difficulties I encountered during my studies and for enabling me to complete my thesis. My parents, Mr. Saleh and Mrs. Refah, who have helped me physically and spiritually through my journey, are the winners of this thesis and all of my accomplishments. I want to thank them for their love, prayers, encouragement, and belief in my ability to finish studying abroad.

My main supervisor, Professor Plamen Angelov, and my co-supervisor, Dr. Hossein Rahmani, deserve special recognition for their support and appreciation of all the work I have provided them. I have noticed that several of my findings have been mentioned in other research as a result of their encouragement to publish. My main supervisor is a wonderful example of a really good person because he understood every one of my circumstances.

I would especially like to thank my brothers and sisters for always being there for me through every step of my life. Without their assistance in my life, I am confident that this company would not have been. They have motivated me, and I want to thank them all for that.

I want to express my gratitude to my son, Faisal, and daughters, Elaine, and Nour, for their unconditional love and encouragement, experiencing the ups and downs in our life during this journey. I would like to thank all my relatives and friends who supported me overseas.

I want to extend a sincere appreciation for the collaboration and support I have received from my friends, colleagues, and faculty members in the School of Computing and Communications.

I genuinely appreciate the internal and external examiners, Dr. Richard Jiang and Professor Christopher Hinde, for agreeing to review my Ph.D. thesis and accepting the viva meeting discussion.

I am grateful for the scholarship from Imam Abdulrahman bin Faisal University in Jubail, Saudi Arabia, and the cultural Bureau of Saudi Arabia in London to study Ph.D. in Artificial intelligence and machine learning at the Lancaster university in the UK.

Contents

List of Figures

List of Tables

Nomenclature

11k Hands-L 11k Hands- Left Hands

11k Hands-R 11k Hands- Right Hands

2DLGF 2D Log Gabor Filter

AGAM Angular Geometric Analysis Method

AI Artificial Intelligence

AR Augmented Reality

ASSR Adaptive Single-Scale Retinex

ATM Automated Teller Machine

BC Bray-Curtis

BOP Bubble Ordinal Pattern

BoW Bag-of-visual-Words

BSIF Binarized Statistical Image Feature

CLAHE Contrast Limited Adaptive Histogram Equalization

CLBP Compound Local Binary Pattern

CLTP Completed Local Ternary Pattern

CMC Cumulative Matching Characteristic

CNN Convolutional Neural Networks

CRR Correct Recognition Rate

CV Computer Vision

DCDV Direction Convolution Difference Vector

DCTNet Discrete Cosine Transform Network

DDBFL Discriminative Direction Binary Feature Learning

DIP Distal Interphalangeal

DL Deep Learning

DLNN deep learning neural network

DLNNs Deep Learning Neural Networks

DNA Deoxyribonucleic Acid

DoG Difference of Gaussian

DRB Deep Rule-Based

EBM Energy-Based Models

EER Error Equal Rate

F-SIFT Fourier Scale Invariant Feature Transform

FBDT fuzzy binary decision tree

FBKS Finger Back Knuckle Surface

FC Fully Connected

FKI Finger Knuckle Image

FKIMNet A Novel Full Finger Image-Based Matching Network

FKP Finger Knuckle Pattern

FLDA Fisher Linear Discriminate Analysis

FRR False Rejection Rate

GANs Generative Adversarial Nets

GAR Genuine Acceptance Rate

GOP Gradient Ordinal Pattern

GORP Gradient Ordinal Relation Pattern

Graph CNN Graph Convolutional Neural Network

HCI Human-Computer Interaction

HOG Histogram of Gradient

HSV Hue-saturation-value

HUQ H-Unique

ICA Independent Component Analysis

IEEE SSCI IEEE Symposium Series on Computational Intelligence

IIT Indian Institute of Technology

IP Interphalangeal

IR Infra-red

IRT Image Ray Transform

KNN K-Nearest Neighbours

KWT Kekre Wavelet Transform

L Left

LBP Local Binary Pattern

LDA Linear Discriminate Analysis

LFI Local Feature Integration

LFW Labeled Faces in the Wild

LI Left Index

LIF Left-Index-Finger

LM Left Middle

LMF Left-Middle-Finger

LOOCV Leave-One-Out Cross-Validation

LTP Local Ternary Pattern

MCP Metacarpophalangeal

ML Machine Learning

MV Majority Voting

MW Matcher-Weighting

NIR Near-infrared

NN Nearest Neighbor

NZSL New Zealand Sign Language

OTPs One-time Passwords

PAFs Part Affinity Fields

PCA Principal Component Analysis

PCANet Principal Component Analysis Network

PIFK Person Identification Based on Fingernails and Dorsal Knuckle Patterns

PIFK^{+} Developed PIFK

PIP Proximal Interphalageal

PNN Probabilistic Neural Network

PolyU Hong Kong Polytechnic University Contactless Hand Dorsal Images Database

POM Product of Matching

R Right

R-CNN Region-based Convolutional Neural Network

ReNet Recurrent Neural Network Based Alternative to Convolutional Networks

ReSeg Recurrent Neural Network-Based Model for Semantic Segmentation

RGB Red, Green, and Blue

RI Right Index

RIF Right-Index-Finger

RLTP Relaxed Local Ternary Pattern

RM Right Middle

RMF Right-Middle-Finger

ROI Region of Interest

ROIs Regions-of-interests

SD Standard Deviation

SGOP Star GOP

SGORP Star Gradient Ordinal Relation Pattern

SIFT Scale Invariant Feature Transform

SOM Sum of Matching

SOP Star Ordinal Pattern

SVM Support Vector Machine

SWM Weighted Sum of Matching

TCM Texture Code Matrix

TH Thermal Image

VAE Variational Autoencoder

VGG-19 Visual Geometry Group-19

VIS Visible Images

VR Virtual Reality

WA Weighted Average

YOLO You Only Look Once

Chapter 1

Introduction

1.1 Problem definition

Human identification is a field of science with many branches, including forensic, biometric, identity, and security, which have merged with advanced technologies to identify individuals for investigation related to general security issues. Passports and ID cards, including face photos and, in certain circumstances, fingerprinting, were the principal instrument for biometric recognition of individuals in the twentieth-century nation-state system. More recently, in certain circumstances, fingerprinting has become the principal technique for identifying persons, and identity documents have become a potent tool for various purposes, including taxes, movement restrictions, police monitoring, law enforcement, war, and segregation (Green, 2009).

The term 'biometric' is derived from the Greek terms 'bio' (life) and 'metrics' (to measure) (Stratton, 2015). Desmarais (Reid, 2004) defined biometrics as the use of personal traits to differentiate between persons for personal identification, verification, or authentication. Biometrics is also known as biometric recognition, which is the automated identification of persons based on physiological and/or behavioural traits. It is feasible to validate or establish an individual's identification using biometrics based on "who she is" rather than "what she owns", for example, an ID card or password. Such systems aim to ensure that the offered services are only accessible to authorised users. Secure access to buildings, computer systems, laptops, cellular phones, and automated teller machines (ATMs) are examples of such uses. Humans have used physical traits such as faces, voices, ears, and gait for thousands of years to identify one another. More recently, in the mid-nineteenth century, Alphonse Bertillon, chief of the police department's criminal identification section in Paris, invented and then put into effect the notion of utilising a variety of body measures to identify offenders. Unfortunately, just as his method was gaining popularity, it was overshadowed by the considerably more significant and practical discovery of the

uniqueness of human fingerprints in the late nineteenth century (A K Jain, Ross, and Prabhakar, 2004).

The advances in machine learning (ML), computer vision (CV), and deep learning neural network (DLNN) technologies have revolutionised the ability to recognise objects from images and differentiate amongst various classes and types of images. This can significantly improve and facilitate research in biometrics, forensic studies, and biomedical applications. In particular, the human hand provides a significant amount of information potentially specific to each person and can thus help to identify individuals or certain medical conditions. For example, recognising individuals based on images and videos in criminal investigations have become increasingly important. As mentioned above, the traditional techniques used to identify individuals rely on fingerprints (A. Jain, Hong, and R. Bolle, 1997) and deoxyribonucleic acid (DNA) (Kurosaki, Matsushita, and Ueda, 1993). However, these resources are often unavailable, while the quantity of simple images and videos captured on camcorders has increased. The challenge is to study the unique anatomical features of hands to make the identification of individual hands possible.

The human hand has significant and distinctive characteristics, and is a highly complex part of the anatomy that has not been explored in much detail, particularly in the context of the contemporary high level of computational intelligence and, more specifically, the advances in artificial intelligence (AI), ML and CV. This topic is highly multi-disciplinary, as it is studied by anatomists, anthropologists, bioinformaticians, image analysts and, increasingly, by computer scientists. A growing pool of advanced methods based on AI, ML and CV can benefit and relate directly to a better representation of the human hand in computer analysis.

Historically, the research methods in this area relied on 'handcrafted' features such as the local binary pattern (LBP) (Yiding Wang et al., 2010) and Histogram of Oriented Gradient (HOG) (Qian, J. Yang, and Gao, 2013) extraction, which necessitated human intervention. However, such approaches struggle to effectively identify the human hand in variable conditions, because of the change in camera viewpoint, hand pose, image quality and self-occlusion. Thus, their performance is limited. In light of these advancements, this thesis investigates the dorsal hand image in developing human identification frameworks using advanced ML and DLNN that include CNN models, to learn the underlying structure of the human hand by extracting robust features from hand images with diverse conditions of camera viewpoint, hand rotations, pose, and image quality.

A biometric system should meet the following criteria (A K Jain, Ross, and Prabhakar, 2004): Universality means the feature should be present in all individuals; Distinctiveness means any two individuals must be sufficiently distinct in terms of the trait; Permanence means the characteristic must be sufficiently invariant throughout time; and Collectability means the attribute may be quantifiable and measured

(A K Jain, Ross, and Prabhakar, 2004). In addition, biometric systems must overcome the following real-world challenges (Sundararajan and Woodard, 2018):

- Biometric recognition entails differentiating possibly millions of individuals, some of whom have subtle distinguishing features;
- More complicated models are required to identify huge numbers of individuals in a large-scale dataset;
- Multiple samples of an individual are typically necessary to obtain all variations in features. These intra-person variances can occasionally be more significant than inter-individual variations, for example, inter-person variations; and major knuckle of the thumb, respectively
- Because of noisy biometric sensors or other causes, biometric data acquired in real-world applications is quite noisy and distorted;
- Extracting useful biometric information from noisy input data involves extensive pre-processing (e.g., extracting a person's face from a complex background or a person's speech signal from a noisy background);
- Stability, which means biometric qualities may be inconsistent because they depend on human characteristics. Physiological biometrics change gradually over time, whereas socio-environmental variables also alter behavioural biometrics;
- Unpredictability, which means it is unknown if a single biometric feature can be used to identify a person uniquely. Most behavioral biometrics, in particular, are ineffective for identification and are only used for verification;
- Biometric system assaults, which are vulnerable to various attacks.

In recent years, considerable attention has been paid to biometric recognition and its suitability to play a major role in person identification. Biometrics augmented by advanced ML techniques is a field of study aiming to recognize and identify subjects based on their characteristics. Examples of biometric identifiers are the face (Sim et al., 2014), iris (Wildes, 1997), ear (Kamboj et al., 2021), hand (Ajay Kumar and Ravikanth, 2009), and footprint (Connor and Arun Ross, 2018). Each type has its strengths and weaknesses (Jaswal, Kaul, and Nath, 2016).

More recently, there has been growing attention to the investigation of the hand as a biometric identifier. The hand plays an important role in our daily activities, including the way we interact with the world (Simon et al., 2017). The hand has specific characteristics which can identify an individual (Kumar, 2012) and has many advantages as a biometric reservoir. The process of acquiring hand data is

user-friendly, simple, cost-efficient, and time-efficient (Jaswal, Nigam, and Nath, 2017a), (Ajay Kumar and Ravikanth, 2009). The hands have many characteristic components such as the fingernails, geometrical shape, fingerprints, hand veins and palm creases and knuckle creases (known in anatomy as the metacarpophalangeal (MCP) joints, proximal interphalangeal (PIP) joint, distal interphalangeal (DIP) joint, and interphalangeal (IP) joint), commonly known as the base, major, minor knuckle of the finger, and major knuckle of the thumb, respectively (Ajay Kumar and Ravikanth, 2009). The finger knuckle also has a specific pattern, which includes creases, curves, lines, and textures and is a discriminating feature that has the potential for identification. Utilising knuckle creases as a biometric trait could be of value for human identification in different scenarios, such as criminal investigations (Jaswal, Nigam, and Nath, 2017a).

The framework for automatic person identification is based on different hand components, including the knuckles and fingernails of the hands. This framework can comprise different stages, such as detection, segmentation, feature extraction, and matching of various regions. Different feature extraction methods can be considered, which might be categorised as so-called 'handcrafting' (Lin Zhang, Lei Zhang, David Zhang, and Zhu, 2010), DLNN (K. Sid et al., 2017), and hybrid handcrafted and DLNN feature extractions (e.g., using DLNN and handcrafted feature extractions for skin lesion classification (Majtner, Yildirim-Yayilgan, and Hardeberg, 2017)). In addition, the biometric system can be based on a single modality biometric (e.g., major knuckle of one finger) (Lin Zhang, Lei Zhang, David Zhang, and Zhu, 2010), or multi-modal biometric (e.g., knuckle of many different fingers) (K. Sid et al., 2017).

DLNN algorithms improve the recognition performance in different domains such as object localisation, pattern recognition, and image segmentation (Mona Alghamdi, P. Angelov, and B. Williams, 2021). Furthermore, transfer learning of a DLNN that has been trained in a specific domain can be employed for another task and used in biometric systems. For example, these systems can apply a fusion of two biometrics, which are finger veins, and finger knuckles (Daas et al., 2020), or different knuckle patterns for person identification (Vyas, Rahmani, et al., 2021). In (Minaee et al., 2019), many studies used DLNN in biometric systems for person identification. However, knuckle patterns or fingernails for biometric systems were not included. DLNN in the form of CNNs and transfer learning have been employed in various biometric systems, which used knuckle patterns and fingernails (Mona Alghamdi, P. Angelov, and B. Williams, 2021; Vyas, Rahmani, et al., 2021). These systems can be based on primary biometrics, such as the face (Vizilter et al., 2017), iris (Jha et al., 2020), fingerprint (Su et al., 2017), and finger-vein (Kuzu et al., 2020).

Unimodal biometrics examples include fingerprints, DNA, finger knuckles, fingernails, hand geometry, and so on. Nevertheless, in today's digital society, there is no one biometric feature that can meet the security and performance needs of all

applications (Hong, Anil K Jain, and Pankanti, 1999). Furthermore, the majority of them frequently exhibit some widely known problems such as non-uniqueness, noise in collected data, interclass similarities, non-universality, spoofing, intra-class differences, and poor discrimination capabilities. To address these issues, a multi-modal biometric system, which integrates two or more physiological or behavioural feature modalities to raise the chance of success or, more specifically, to improve the accuracy of any identification or verification application, may be a viable alternative method (Jaswal, Kaul, and Nath, 2016; Lahat, Adali, and Jutten, 2015).

The development of DLNN in recent years in numerous domains of pattern recognition and CV, such as image classification (Rawat and Z. Wang, 2017), object identification (Ren et al., 2017), and object tracking (C. Ma et al., 2015), has encouraged the community to employ deep networks to learn metrics or descriptors (Hanif, 2019). In the early 1990s, the Siamese network architecture (Bromley et al., 1994) was introduced, which created the groundwork for using neural networks to learn similarities from image patches. In its most basic form, a neural network consists of two identical sub-networks with the same weights to learn a suitable representation of the input data. Pairs of image patches often represent the data. The distance between the outputs of the two branches or a decision network is used to determine similarity. The conventional Siamese network and its newly suggested CNN-based variations have been widely used for applications such as signature verification (Bromley et al., 1994), face verification (Chopra, Hadsell, and LeCun, 2005), image recognition (Koch et al., 2015), and person re-identification (Ahmed, Jones, and Marks, 2015; Hanif, 2019).

Many studies investigate the 'handcrafting' feature extraction for specific knuckles, such as the minor or major dorsal finger knuckle patterns (Hammouche, A. Attia, and Akhrouf, 2022; Grover and Hanmandlu, 2015). Nevertheless, there need to be more studies on the methods of DLNN models, such as VGG16 or RestNet50, as well as other DLNN models, which are the core of the Siamese Networks. Also, there have been no studies investigating the building of the person identification system from all the knuckles and fingernails of both right and left dorsal human hands, which includes localisation, detection, segmentation, feature extraction, and matching of region of interest (ROI), for example, the fingernail. Most of the studies investigating the localisation and detection of the hand component are from a specific region of the hand, for example, the major knuckle or a particular finger, such as the left index, left middle, right index, or right middle finger (Grover and Hanmandlu, 2015), or the major knuckles and fingernails from the index, middle, or ring finger (Heidari and Chalechale, 2022). However, none have investigated the localisation and detection of components of the base, major, minor, and fingernails of all hand fingers.

Fusion methods in multimodal biometrics are divided into four levels: pixel level, feature level, score level, and decision level. Degraded image quality significantly

impacts pixel and decision levels, making them unstable, unfeasible, and inefficient, resulting in poor recognition performance. Since the feature and score levels are more stable and demonstrate effective recognition performance, they are the most common in multimodal biometrics fusion (Daas et al., 2020).

However, there are insufficient studies of the efficiency and performance of various distance metrics in the matching process of the biometric system. The utilisation of fusion or ensemble learning with all ROIs of the hand including base, major, minor knuckles, and fingernails, can significantly improve the performance, and reduce the limitation of using mono-modal biometric systems by combining more than two modalities. The ensemble learning provides a practical solution and several benefits, including high accuracy, resistance to spoofing, and dealing with the problem of non-universality (Hammouche, A. Attia, and Akhrouf, 2022).

1.2 Aims and objectives

1. Developing a new framework for person identification, including joints or keypoints detection, segmentation, feature extraction, and finding the best match using similarity metrics.

2. Utilising the base model and fine-tuning of DLNN feature extraction and various similarity matching to improve the framework.

3. Utilising the score-level fusion to improve the identification framework.

4. Employing an advanced approach, such as the Siamese network, to enhance the framework for identification.

1.3 Contributions of the thesis

The following are summaries of the thesis contributions:

1. A new method for segmenting the knuckle creases and fingernails for person identification: The study extended a well-known hand-detecting method known as the multi-view bootstrapping to label automatically and segment hand components.

2. Segmenting left and right human dorsal hands into 19 components, including all knuckle creases and fingernails. The hands detected locations include the MCP, PIP, DIP, and the fingernails of the five fingers on the hand's left and right dorsal surfaces.

3. Investigating many base models of DLNNs (including VGG16, ResNet, MobileNetV2, and others) and similarity metrics (including Cosine, City-block, and Bray-Curtis) for feature extraction and similarity matching in the proposed person identification framework called PIFK. The investigation showed that the base model of the DenseNet201 and Bray-Curtis metric obtained the best performance.

4. Improving the person identification framework's PIFK$^+$ performance by fine-tuning the DenseNet201 DLNN model using subsets from segmented hands in the '11k Hands' and 'PolyU'.

5. Developed the performance of PIFK$^+$ using fusion techniques. The fusions were conducted based on the component or region of all fingers, including the fingernails, minor, major, and base knuckle creases and on the whole hand components. Also, studying the fusion based on different similarity metrics: Bray-Curtis, Cosine, and Euclidean. Two main fusion rules were used, which are the majority voting (MV) and weighted average (WA).

6. A new framework for identifying individuals was presented based on an advanced deep learning technique known as the Siamese network, which included identifying the persons based on many sub-images of the human dorsal hand. Also, many types of deep learning CNNs were studied for feature extraction, such as DenseNet201, ResNet152, and VGG16. Fine-tuning was employed in the best model of deep learning CNN, which was the DenseNet201, to improve the identification result.

7. Interesting results were found: the fingernail components performed the best among other components. In general, the major knuckle creases obtained higher identification results than others. Also, the left segmented sub-images of the hands obtained better recognition performance than the right segmented sub-images of the hands.

1.4 Thesis structure

This section gives an overview of the thesis structure. The thesis includes seven chapters, with the Introduction chapter.

Chapter 2: The Literature Review

This chapter explores the state-of-the-art literature in person identification. The chapter includes an analysis, investigation, and examination of the previous decades

to the current state-of-the-art of research in the field of biometric recognition.

Chapter 3: Automated Person Identification Framework Based on Fingernails and Dorsal Knuckle Patterns

This chapter discusses the published work in the IEEE Symposium Series on Computational Intelligence (SSCI) conference, comprising a comprehensive framework of knuckle creases and fingernail keypoints detection, segmentation, feature extraction, and similarity estimation. More specifically, the chapter demonstrates the effectiveness of the original model of DLNN and the similarity metric for feature extraction and finding matching, respectively.

This chapter also shows a developed framework for person identification. The chapter focuses on the transfer learning approach, and fine-tuning the DLNN model for better feature extraction. It investigates different similarity metrics to find the best one for matching.

Chapter 4: A Multi-modal Biometric Approach Based on Score-Level Fusion and Fine-tuning DLNN Features

In this chapter, the developed approach was improved further by introducing the score-level fusion implemented using the identification scores employed in various hand joints. The chapter displays the validity of this method for person identification.

Chapter 5: A Method for Identification of Humans from Dorsal Hand Sub-images using a Siamese Network Model

This chapter investigates a DLNN model known as the Siamese network, including building the database into positive and negative pairs, constructing the Siamese network model, and using a distance metric to find the similarity. The identical Siamese network models build from DLNN models. The features of the pair sub-images are extracted using a high-level DLNN feature extraction. Lastly, Euclidean distance is used to find the similarity between pairs.

Chapter 6: Conclusion

This concluding chapter of this thesis identifies the interesting findings, potential study paths as well as the potential significance of each one.

Chapter 2

Literature Review

A person identification pipeline can include localising, detecting, segmenting, and extracting pattern information and finding valid individuals using a classifier, such as a support vector machine (SVM). A recognition system is a field of study that combines advanced ML, including DLNN, and biometrics from individuals based on their physical and behavioural traits (Anbari and Fotouhi, 2021; Daas et al., 2020). These traits can be single biometric features, such as the ear (Kamboj et al., 2021), face (Juefei-Xu, Verma, and Savvides, 2017), hand gesture (Wu, Ishwar, and Konrad, 2016), finger-vein (Kuzu et al., 2020), or multiple biometrics, such as fusion of finger vein and knuckle pattern (Daas et al., 2020).

More attention has been paid to the characteristics of the hand as it is critical in the human identification system and because it has many distinctive features, which can distinguish and identify the individual (Kumar, 2012). Recently, considerable attention has been paid to developing extraction techniques. These technologies aim to extract unique hand elements such as hand dorsal vein patterns (X. Li, D. Huang, and Yunhong Wang, 2016), palm vein patterns (W.-Y. Han and Jen-Chun Lee, 2012), palm creases (David Zhang et al., 2003), fingerprints (Alshehri et al., 2018), knuckles (Ajay Kumar and Zhihuan, 2016) and pigmentation (Romero Lopez et al., 2017), for several purposes.

The approaches to extracting features can be divided into two main categories. The first category includes traditional techniques that deal with many visual characteristics. However, the traditional feature descriptor method is problem-specific and usually requires intervention to choose the descriptor (Nanni, Ghidoni, and Brahnam, 2017; Yann, Yoshua, and Geoffrey, 2015). The second, recent DLNN techniques, outperform traditional techniques in many applications (Daas et al., 2020). In the following subsections, a revision of the stages of person identification and verification framework using human dorsal hand images will be provided.

2.1 Image capturing devices and image types

Different types of capturing devices and benchmark datasets exist for human dorsal hand images. For example, an image-capturing device can be a handheld camera device, which acquires red, green, blue (RGB) images, or a handheld infra-red (IR) thermal imaging camera, which captures infra-red images. The benchmark datasets demonstrated in table 3.1, only contain images of the palm, the dorsal side, or both sides of hands (left (L), right (R)), and lack the ground truth data. The ground truth data is the labeled or annotated data, which can identify the location of the knuckles and fingernails. However, as it is helpful to utilise the ground truth data of hand images to train DLNN models, this PhD thesis will focus on this aspect. In addition, there is a benchmark known as The Hong Kong Polytechnic University (PolyU) contactless finger knuckle images database (Version 1.0) (Ajay Kumar, 2014) for a dedicated area of the middle finger, and the Indian Institute of Technology (IIT) Delhi Finger Knuckle Database (Version 1.0) (Ajay Kumar and Yingbo Zhou, 2009; Ajay Kumar and Ravikanth, 2009) for the minor and major knuckles, which will be used in the thesis.

In (Choudhury, Amioy Kumar, and Laskar, 2019) work, low-resolution hand dorsal images of the index, middle, and ring fingers were collected from a database of 890 (5 each from 178 people aged 17 to 65 years). The setup included a low-cost camera and a wooden box with a top hole to fit the camera lens. The wooden box incorporated a fluorescent light source attached beneath its roof to give uniform illumination. As there were no pegs in the setup, the picture capture system was unconstrained and fully touchless. In addition, the imaging infrastructure enabled a wide range of orientation changes to be integrated into all images in the database (Choudhury, Amioy Kumar, and Laskar, 2019).

In the following subsection, brief description of the dorsal knuckle pattern and fingernails databases will be provided, including IIT Delhi Finger Knuckle Database (Version 1.0), Hong Kong Polytechnic University Hand Dorsal Images (Version.1), the Hong Kong Polytechnic University Contactless Hand Dorsal Images Database, 11k Hands, and Tecnocampus Hand Image Database.

2.1.1 IIT Delhi Finger Knuckle Database (Version 1.0)

The pictures collected from the IIT Delhi database make up the database of finger knuckle images. Using a digital camera, this database was collected on the IIT Delhi campus between August 2006 and June 2007 and acquired from 158 people. The database contains only subjects between the ages of 16 and 55. Every user's integer identification/number has been systematically numbered for this database of 790 photos. These photographs are available in bitmap format and have a resolution

of 80 x 100 pixels (Ajay Kumar and Yingbo Zhou, 2009; Ajay Kumar and Ravikanth, 2009). Samples of these images are shown in figure 2.1.

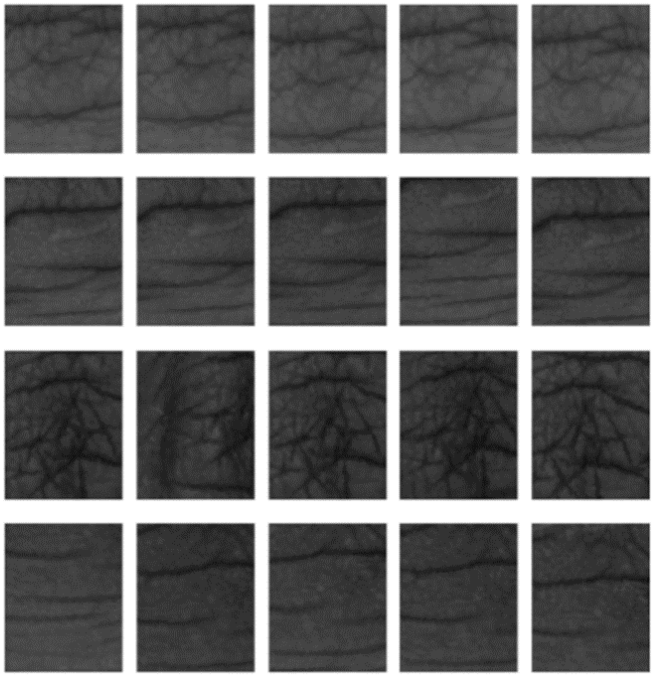

Figure 2.1: Samples of IIT Delhi campus knuckle pattern images from five subjects (Ajay Kumar and Yingbo Zhou, 2009)

2.1.2 Hong Kong Polytechnic University Hand Dorsal Images (Version.1)

Male and female volunteers contributed to the PolyU contactless finger knuckle images database (Version 1.0). Between 2006 and 2013, this database was mainly amassed at PolyU and the IIT Delhi campuses, utilising a contactless configuration that required a hand-held camera. The database contains 2515 bitmap (*.bmp) photos of the dorsum of the middle finger taken from 503 people. About 88% of the individuals in this dataset are under 30. In addition, to assess the stability of the knuckle crease and curved lines, this database also includes two sessions of finger knuckle images taken after a relatively long period (4 to 7 years) (Ajay Kumar, 2014). Examples of the finger images are displayed in figure 2.2.

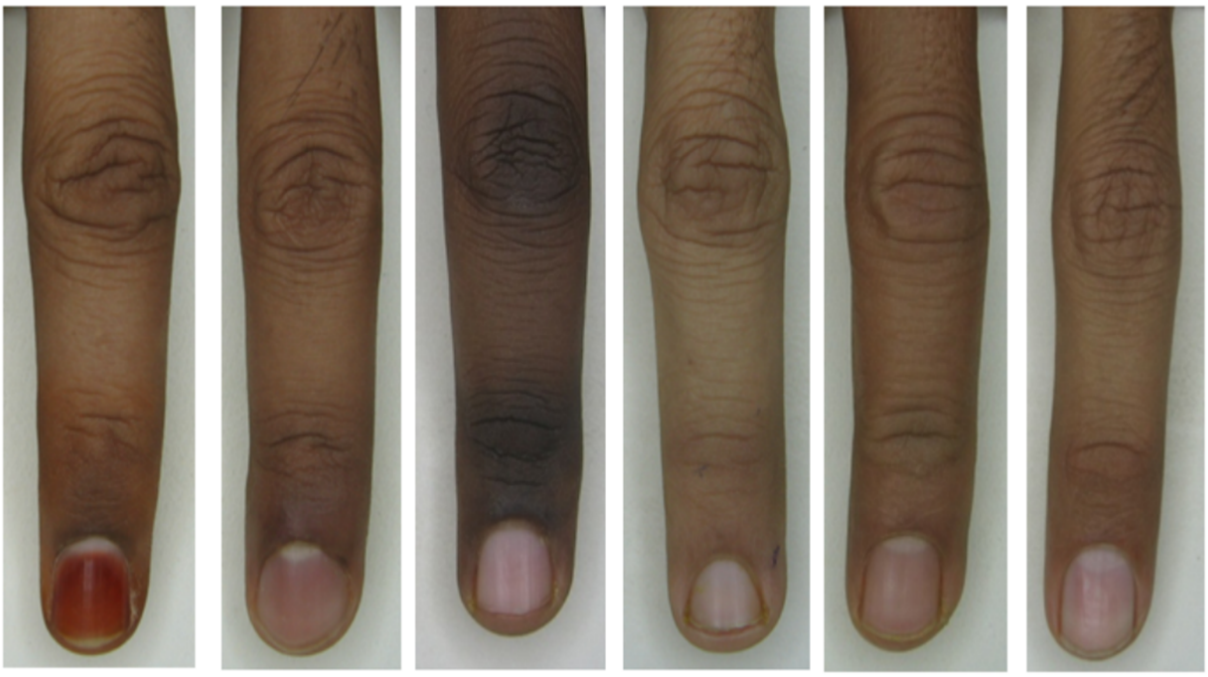

Figure 2.2: Samples of finger images (Ajay Kumar, 2014)

2.1.3 The Hong Kong Polytechnic University Contactless Hand Dorsal Images Database

The Contactless Hand Dorsal Images Database was primarily collected from male and female volunteers between 2006 and 2015 using a mobile phone and a hand-held camera on the campuses of the IIT Delhi, PolyU, and various Indian villages. This database contains 2505 photos of the right hand's dorsum from 501 distinct participants, showing three knuckle patterns on each of the subject's four fingers. Each picture is in a bitmap (*.bmp) format. Other hand dorsal photos from 211 participants are also available in this collection. However, they are not very clear or do not show second minor knuckle patterns. The combined database from the dorsal hand photographs of 712 people is available to the public. To help investigations linked to the stability of knuckle patterns, this database offers two session hand dorsal pictures, with numerous samples in different age groups that were taken after a very long interval (4 to 8 years). Additionally, employing fully automated segmentation, this database offers segmented/normalized images of the major, first minor, and second minor knuckles. The names of the corresponding photographs and folders in the database make it simple to locate these images for all subjects and different/respective

fingers (Ajay Kumar, 2014). Samples of these dorsal hand images can be seen in figure 2.3.

Figure 2.3: Samples of hand images from the Hong Kong Polytechnic University Contactless Hand Dorsal Images Database (Ajay Kumar, 2014)

2.1.4 11k Hands

In total, 11,076 hand photos with 1600 × 1200 pixels from 190 participants, ranging in age from 18 to 75, are included in this collection. Each participant was asked to open and close their individual fingers on both right and left hands. On a constant white background, every hand was photographed from the palmar and dorsal sides at roughly the same distance from the lens. Each image has a record of metadata that includes information about the hand photographed, including the right or left hand, the hand side (dorsal or palmar), and logical indicators indicating the presence of accessories, nail polish, or other irregularities. Figure 2.4 shows samples of these hand images.

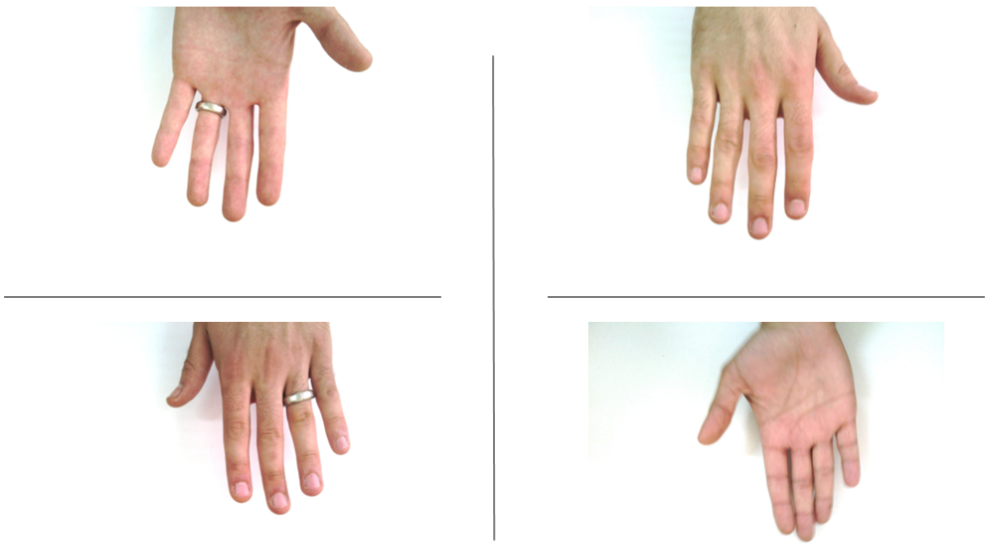

Figure 2.4: Samples of the dorsal and palm side from left and right hand images in '11k Hands' Database (Afifi, 2019)

2.1.5 H-Unique (HUQ)

H-Unique (HUQ) is a large dataset and was gathered as part of the H-Unique project. More than 3000 hand dorsal images from people of different ethnic backgrounds may be found in the dataset. This dataset has the most challenging settings because it is uncontrolled and does not allow for the modification of features like background, orientation, scale, or illumination (further information can be found at https://h-unique.lancaster.ac.uk/, viewed on October 19, 2022).

2.1.6 Tecnocampus Hand Image Database

In the Tecnocampus Hand Image Database (Font-Aragones, Faundez-Zanuy, and Mekyska, 2013), the TESTO 882-3 thermal camera captured a visible (VIS) and a thermal (TH) image. Except for the final user, who was added one month after the others, the data from 104 people in this database were collected in five separate sessions, each lasting one week. Four VIS plus four TH pictures were collected per participant and per acquisition session. The other two photos are of the dorsum, while the other two are of the palm. Figure 2.5 displays examples of VIS and TH images.

Figure 2.5: Samples of VIS and TH images for the right dorsal hands.

In the coming section, we will review the literature on the typical framework of biometric recognition, including detection, localisation, pose estimation, feature extraction (handcrafting and DL) approaches, ensemble learning, and similarity metrics for the hand in general and the knuckle patterns and fingernails, in particular.

2.2 Pre-processing: Detecting and localisation of the hand's joints

Detecting and localisation of human hand joints are important topics in the CV field and have undergone much study over time (Grzejszczak, Kawulok, and Galuszka, 2016). Many techniques have been developed (Grzejszczak, Kawulok, and Galuszka, 2016) for identifying static (Erol et al., 2007; Vančo, Minarik, and Rozinaj, 2014) or dynamic (Mitra and Acharya, 2007) hand gestures, facial expressions (Y. Sun, Reale, and Yin, 2008), or body actions (Poppe, 2010). Finding the rectangle that encloses the finger knuckle print (FKP) pattern region is the prime goal of FKP component extraction, which considers a pre-processing step (A. Attia, 2022). In (A. Attia, 2022) study, a Gaussian smoothing operation was applied to the input palm print and FKP images. First, the procedure found the X-axis of the coordinate system that was fixed from the finger's bottom edge, then used the Canny edge detector to extract the finger's bottom border. Next, the method established the region's Y-axis for the coordinate system based on the Canny edge detector, which was extracted from the X-axis of the original image. The location of the convex direction was revealed in this stage. Finally, the procedure defined the ROI coordinate system, where the rectangle represents the ROI information area (A. Attia, 2022).

However, the study mentioned above utilised calculating shapes and lines to extract the required elements. Also, the knuckle element extraction process was

taken from a small specified knuckle region instead of extracting from the dorsal hand images.

In (Tarawneh et al., 2022) study, the Delhi Finger Knuckle Database (Ajay Kumar and Yingbo Zhou, 2009; Ajay Kumar and Ravikanth, 2009) was used because the images were already extracted in segments of major and minor knuckles, and there was not much work need for ROI extraction or pre-processing. However, anyhow it is not challenged to extract the ROI of the knuckle pattern from the image.

In a paper by (Ajay Kumar and Zhihuan, 2016), the authors identified the localisation of keypoints of knuckles in the finger. Also, histogram equalization is applied to the acquired images. In binarisation, Otsu's method is used and followed by eliminating isolated noisy pixels used to create hand contour images. However, the study did not measure the performance using the rank-1 accuracy.

The ROIs of improved FKP images were recovered using a modified Gabor filter approach in a study by (Tarawneh et al., 2022). The ROI data were further altered by utilising two innovative Bubble Ordinal Pattern (BOP) and Star Ordinal Pattern (SOP) approaches to strengthen the image representations. First, a hierarchical deep matching technique was used to match the scale invariant feature transform (SIFT) image features. Deep matching was constructed from the top down and bottom up to handle weak textures and non-rigid deformable areas successfully (Jaswal, Nigam, and Nath, 2017a).

The dorsal finger image was binarised using Otsu's thresholding method in (Hammouche, A. Attia, and Akhrouf, 2022). A binarised finger shape was used to predict the position of the fingertip from the convex hull to eliminate the background image over the fingertip. Moment techniques were used, and the binarised image was utilised to determine the orientation of the fingers (Ajay Kumar and Yingbo Zhou, 2009). Finally, the finger image was segmented using the coarse segmentation approach (Hammouche, A. Attia, and Akhrouf, 2022). In (Vyas, B. M. Williams, et al., 2022) work, the knuckle regions are localised using an ensemble approach that applied various object detector frameworks in a functionally relevant manner. The method employed each of the common object detectors' unique characteristics to deliver a more accurate localisation of the knuckle region. The authors considered the single-stage and two-stage detector classes (you only look once (YOLO) and faster region-based convolution neural network (R-CNN), respectively) for selecting the individual detectors (Vyas, B. M. Williams, et al., 2022). Also, They annotated two datasets with hand components: the '11k Hands' and 'PolyU'.

The work by (Anbari and Fotouhi, 2021) did not consider ROI extraction. Before the feature extraction step, image enhancement was required to emboss the main parts of the visual appearance, and counteract the influence of lighting changes, local shadowing, and highlights. As in the approach given for face recognition application in (Tan and Triggs, 2010), the pre-processing step involved gamma correction, difference

of Gaussian (DoG) filtering, and contrast equalisation.

However, manually analyzing knuckle creases for identification requires much time and might be subjective, necessitating the knowledge of skilled forensic anthropologists, whose availability is highly scarce. Therefore, a more efficient R-CNN model capable of producing accurate bounding boxes for the target objects in the input image. The result was a collection of knuckle bounding boxes, detection confidence ratings, and knuckle type classifications for each knuckle showed in (Vyas, Rahmani, et al., 2021).

In (A. Attia, Akhtar, and Chahir, 2021) study, the dorsal finger images were collected and all binarised. Otsu's thresholding approach was applied. Using moment-based approaches, it was possible to predict the finger orientation from the binarised image. The minor finger knuckle region was segmented using coarse segmentation, excluding the major finger knuckle region and most of the fingernail (A. Attia, Akhtar, and Chahir, 2021).

The study in (A. Attia, Akhtar, Chalabi, et al., 2020) chose the binarised statistical image features (BSIF) descriptor and Gabor feature extraction in this research since they have been widely used for object identification and recognition. The extraction of the ROI for FKP (A. Attia, Akhtar, Chalabi, et al., 2020) comprises of several phases. After applying the Gaussian smoothing process to the original image, the smoothed image was down sampled to 150 dpi. Second, the x-axis of the coordinate system fitted from the finger's bottom boundary was determined; it may be simply retrieved using a Canny edge detector. Third, the y-axis of the coordinate system was chosen by using a Canny edge detector on a cropped sub-image taken from the original image base on the x-axis, and the convex direction coding scheme was identified. Finally, the ROI can be calculated (A. Attia, Akhtar, Chalabi, et al., 2020).

In (Chlaoua et al., 2019) study, each FKP image underwent pre-processing to extract information about the finger knuckle region (ROI sub-image) exclusively. First, the pre-processing technique from (Zhang et al., 2003) was applied. In Zhang et al. (Zhang et al., 2003), the authors' main pre-processing steps were to extract the central part of a palmprint by obtaining the binary image using Gaussian smoothing. Next, using a boundary-tracking method, a threshold was employed to transform the convoluted image into a binary image. They then determined the boundaries of the gap between the fingers, defined a coordinate system, and extracted the central part as a sub-image (Chlaoua et al., 2019).

In (Thapar, Jaswal, and Nigam, 2019) work, the extracted ROIs of the major knuckle, minor knuckle, and nail features were trained using an R-CNN that utilised various bounding boxes as labels to identify and localise the ROIs across images (Thapar, Jaswal, and Nigam, 2019).

The ROI was retrieved in (Kim, Cho, and Park, 2019). Then, the extracted ROI was rotated and scaled at a given angle based on the ROI centre. To minimize the

lighting variance of the input images, a Retinex filtering technique was used. The acquired RGB ROI picture was transformed into a hue-saturation-value (HSV) colour space. Only the value channel (illumination) was subject to Retinex filtering among the three channels (Kim, Cho, and Park, 2019).

In the (Hom Choudhury, Amioy Kumar, and Laskar, 2019) study, the image acquisition system is unrestricted and entirely touchless. A Gabor filter was applied on the fingernails of the index, middle, and ring fingers and a Canny edge detector was utilised on the index, middle, and ring knuckles' ROI (Hom Choudhury, Amioy Kumar, and Laskar, 2019). In (Fei et al., 2019) experiment, extracting the ROIs from FKP images and enlarging them to 55×110 pixels using the enhanced approach proposed in (Lin Zhang, Lei Zhang, David Zhang, and Zhu, 2010) were employed.

The suggested FKP was derived from the PolyU FKP database in the study presented in (Muthukumar and Kavipriya, 2019). In this case, the ROI was used to find edges by examining the local maxima of the gradient for images. The gradient computations were based on the Gaussian filter's derivative (Muthukumar and Kavipriya, 2019).

To extract the ROI from FKP images, the (Zhai et al., 2018) work adopted a two-stage centre point detection approach (Yu et al., 2015). The procedure was divided into two stages: i) preliminary identification of the centre point and ii) accurate secondary positioning of the centre point. We can see from the FKP images that the brightest region was connected to the central point. The ROI extraction approach was based on the assumption that the brightest area's centre point would be the centre point of the FKP. This assumption, however, was incorrect in several circumstances, particularly when the image skewed vertically. As a result, (Zhai et al., 2018) presented a two-stage centre point recognition approach to handle the mentioned issue. The approach detected the centre point in two stages: they identified the brightest region in FKP images and approximately treated the centre of the area as the original centre point; and relocated the initial centre point (Yu et al., 2015; Zhai et al., 2018).

In (Gao et al., 2019) work, the Gabor filter responses were employed, and for ease of computation, the reconstruction coefficients were assessed using the regularised least squares formula (Gao et al., 2019). The ROI of the raw FKP image was cropped using the central knuckle point as a reference, and the contrast was increased using contrast limited adaptive histogram equalization (CLAHE) in (Jaswal, Nigam, and Nath, 2017b) study. The ROIs of raw images were retrieved by calculating the magnitude from the phalangeal joint with a modified Gabor filter to identify the middle knuckle line and middle knuckle point in this work (Jaswal, Nigam, and Nath, 2017a). The resulting ROI samples were then converted using two unique gradient ordinal pattern (GOP) and Star GOP (SGOP) approaches, yielding more distinct vcode and hcode, which means vertical code and horizontal code, picture representations. The SIFT

picture features were matched using a multiple hierarchical deep matching technique that was very sensitive to weak texture and non-rigid deformable areas (Jaswal, Nigam, and Nath, 2017b).

The ROI was extracted using the edge detection-based technique presented in (Waghode and Manjare, 2017) study. Pre-processing, such as extraction, skin filtering, gray-scale conversion and histogram equalization, was also included (Waghode and Manjare, 2017). The suggested study in (Khellat-Kihel et al., 2016) employed the same pre-processing on several modalities, using the Gabor filter 2D applied to various source images. This filter type is essentially a bi-dimensional Gaussian function (Khellat-Kihel et al., 2016).

In (Qian, J. Yang, Tai, et al., 2016) study, each image's face from the NUST-RWFR database was individually cropped and then normalised to 80×80 pixels. The images of cropped finger knuckle prints were downsized to 55×110 pixels. For the PolyU palmprint data, they only utilised the cropped image that has been downsized to 64×64 pixels.

In the (Usha and Ezhilarasan, 2016b) work, a sample region of 90×180 pixels was cut from the collected FKPs to obtain its sub-image. This pixel size was determined by empirical analysis. The Canny edge detection technique was applied to the generated sub-image.

The authors in (Usha and Ezhilarasan, 2016a), captured the back surface of the four fingers except the thumb finger. As a preliminary step in pre-processing, the obtained finger back knuckle surface (FBKS) was binarised, which converts each pixel of an FKP image into one bit information (0 or 1) to reflect the pixel's intensity at a specific position, then it is contour extracted. The ROI sub-images of the proximal and distal knuckles were produced from the knuckle image by extracting fixed size pixel values of 110×220 and 90×180, respectively.

A modified version of the Gabor filter was used in (Nigam, K. Tiwari, and Gupta, 2016) to estimate the centre knuckle line and central knuckle point in order to extract the FKP ROI. The middle knuckle point was constantly employed to segment the FKP The suggested ROI extraction technique consisted of three steps: knuckle area detection, central knuckle-line detection, and central knuckle-point detection.

The (Kulkarni, Raut, and Dakhole, 2015) study proposed a finger knuckle acquisition device, which was made from a finger supporter, a digital camera (SonyDSC-W380), and a light acrylic box with a white background. The device captured only the middle knuckle of the right index finger (RIF) and RMF . The lower, middle and upper knuckles are present only on the four fingers, not the thumb. The proposed device in (Kulkarni, Raut, and Dakhole, 2015) described the uncontrolled acquisition environment of hand images. These images were altered to improve the visibility of distinguishing characteristics such as ridges and folds surrounding the phalangeal joint of the finger knuckle surface. This was a critical step towards

improving recognition. As a pre-processing step, we applied the Wiener filter and reflection reduction to improve the quality of finger knuckle images obtained from the proposed device (Kulkarni, Raut, and Dakhole, 2015).

In (Ajay Kumar and B. Wang, 2015) study, ROI extraction was similar to work by (Ajay Kumar and Ravikanth, 2009), which was chosen to represent the major knuckle patterns. Kumar's (Ajay Kumar and Ravikanth, 2009) work can automatically construct a normalised and fixed area of interest from dorsal finger images. Each image was binarised to produce a silhouette depicting the finger form. Then, the length of the finger image and length of the finger form in the silhouette was utilised to extract a fixed length knuckle area image that was proportionate to the length of the finger (Ajay Kumar and B. Wang, 2015). However, the mentioned study only extracts the ROI from the dorsal finger rather than the dorsal hand.

The low-resolution hand pictures commonly obtained via webcam photography frequently show fractured hand borders or hand images that are relatively blurry. Therefore, they are inadequate for extracting the nail plate areas. However, although the picture quality or stability may be poor, the nail plate surface may still be retrieved as an ROI and used for personal identification, as this work explored. The fingernail plate surface segmentation method described in the (Amioy Kumar, Garg, and Hanmandlu, 2014) study can properly segment the ROI regardless of whether the fingernail plate had grown out or was covered in polish, as in the case of samples from female hands. The suggested method for segmentation classifies each pixel into a fingernail surface or non-nail surface region at the pixel level. Then the proposed Gabor filtering methodology was proposed for precise ROI extraction in (Amioy Kumar, Garg, and Hanmandlu, 2014). Nevertheless, the mentioned study only focused on the nail ROI, not knuckle patterns.

In general most of the above-mentioned methods did not consider extracting all the hand joints from a human dorsal hand image. In addition, handcrafted methods were utilised.

2.3 Pose estimations

Real-time hand pose estimation, including articulating object pose estimation, is a fundamental step for many practical applications, for example, human-computer interaction (HCI), virtual reality (VR), augmented reality (AR), etc. Hand pose estimation is used to identify 3D or 2D of the hand shape and pose from a 2D RGB image. This includes estimation of 2D or 3D locations of the hand keypoints, including fingernails, and knuckles (Ge et al., 2019). Until recently, hand pose estimation involved locating the hand object and then cropping the hand image considerably. It evolved with the introduction of commodity depth images for achieving accurate

hand pose estimation. Nevertheless, this process encountered challenges, such as the restricted sensing distance and low-resolution depth maps, which remain a general open problem. With the significant development of CNN, many researchers have discussed hand pose estimation from colour images (Simon et al., 2017). However, because of the flexible finger movements, self-occlusions and ambiguities in the appearance of colour images, there have been challenges (Yangang Wang, B. Zhang, and Peng, 2020).

The (Yangang Wang, B. Zhang, and Peng, 2020) investigation focused on the issue of estimating a 2D hand pose from monocular colour images with Simultaneous Region Localization called SRHandNet. The goal was to estimate 21 pre-defined keypoints of a 2D hand pose in real-time with monocular colour images input. SRHandNet obtained the encoder-decoder network architecture as the backbone of the network and considered a solution to estimate the pose for small hand images by using the hand bounding box to obtain the extracted hand ROIs to achieve hand pose estimation (Yangang Wang, B. Zhang, and Peng, 2020).

A method that can identify the location of hand joints from RGB images can be found in (Simon et al., 2017) study. The method would allow analyses of human motion, including the face, body, hand, and foot, on vast sources of visual data available on media such as YouTube and Netflix. The (Simon et al., 2017) paper also presented a method that permits real-time 2D hand tracking from single-view video and 3D hand motion capture.

As large datasets of annotated keypoints are not available for hand images, unlike the face and body, establishing this dataset remains a significant challenge because hands may have self-occlusions due to articulation, viewpoint, and even a grasped object. These factors make the annotation of hand parts difficult to achieve, even with manual annotation by an annotator, increasing time of annotation and cost while degrading accuracy (Simon et al., 2017). In (Simon et al., 2017) work, the authors offered a method to increase the precision of a given keypoint detector by utilizing a multi-camera setup. The method, referred to as "multi-view bootstrapping" depends on the following hypothesis: although an individual image of the hand has significant occlusion, a non-occluded view usually exists. Obtaining an non-occluded view of the hand can be accomplished by employing the multi-view bootstrapping method to generate a more powerful hand keypoints detector, which can identify the hand joints, namely: MCP, PIP, DIP, and IP as well as the fingernail from both the left and right hands.

In (Z. Cao et al., 2017) paper, the authors offered a sufficient method for multi-person pose estimation using the first bottom-up representation of association scores via Part Affinity Fields (PAFs), and encoding the location and orientation of limbs across the image area to obtain a set of 2D vector fields. The method showed the simultaneously inferring these bottom-up representations of detection and association,

which encoded satisfactory global context for a greedy parse and reached high-quality outputs, at a low computational cost. The overall pipeline of the proposed method involves passing the image through the first ten layers of a visual geometry group-19 (VGG-19) model to extract feature maps; these are treated with multiple steps CNN to produce a collection of part confidence maps (projections for body part locations; each location has a map). Also, there was a collection of PAFs consists of a set of 2D vector fields which represent the degree of association between parts) (Z. Cao et al., 2017).

Unlike most of the recent work that estimated the 3D location of hand keypoints from a monocular RGB image only, the paper by (Ge et al., 2019) estimated the 3D shape and pose of the full hand from a single RGB image. The authors proposed a method based on a Graph CNN to reconstruct a 3D mesh of the hand shape and pose. They created a large-scale synthetic dataset consisting of both ground truth 3D meshes and 3D poses to train a network with a supervision manner. They proposed a weakly-supervised method by presenting the depth map as weak supervision in training for fine-tuning the networks on real-world datasets without 3D ground annotation (Ge et al., 2019).

The work by (F. Zhang et al., 2020) offered MediaPipe, a cross-platform open-source framework. The framework included high-fidelity hand and finger detection methods using ML to identify 21 3D keypoints of a hand. The method involves a palm detector that locates palms using an aligned hand-bounding box and operates on a whole input picture. A hand landmark model generates high-definition 2.5D landmarks based on the palm detector's clipped hand bounding box. The in-the-wild dataset for the palm detector was utilised since it is sufficient for locating hands and provides the greatest range of aesthetic options. However, the hand landmark model is trained using all datasets. Twenty-one landmarks are added to the real-world photos as annotations and projected ground-truth 3D joints are used to create synthetic views. In order to sample the region, excluding annotated hand regions as negative instances for hand presence, we use a subset of real-world pictures as positive examples (F. Zhang et al., 2020).

Cross-modal learning applies to any learning that includes information from more than one modality. The term modality generally assigns to a sensory modality (Spurr et al., 2018). The paper by (Spurr et al., 2018) estimated 3D joints of a human hand using RGB image, which learned them from cross-modal RGB, depth images, and its ground truth using latent representation learning. The authors used Generative Adversarial Nets (GANs) were learned an underlying distribution of the data via an adversarial learning process. The Variational Autoencoder (VAE) optimized the log-likelihood of the data under a latent space manifold (Spurr et al., 2018).

Moreover, image pre-processing may include many techniques such as image enhancement, filtering, normalization, and illumination correction (Jaswal, Kaul,

and Nath, 2016). Also, image segmentation and localisation of ROIs may play an important role in the efficiency of the overall framework. In general, detection and region localisation methods of the hand elements can be divided into two main categories. In the first category, several algorithms have been designed to extract low-level features, such as edges, texture, convexity and colour (Lin Zhang, Lei Zhang, David Zhang, and Zhu, 2010), (Nigam, K. Tiwari, and Gupta, 2016), enhancing the colour and contrast information (Schaefer et al., 2011). In the second category, some research papers have been published which propose that CNN models are trained on manually annotated data of knuckles, and part of the trained CNN are used as abstract features. An example of this approach is the method proposed in (Ranjan Jha et al., 2017) to train a state-of-the-art region-based CNN using different bounding boxes as ground truths for segmenting the PIP (major knuckles). The same method was extended to segment the PIP and DIP (minor knuckles), and fingernails (Thapar, Jaswal, and Nigam, 2019).

Many papers have focused on the PIP (Ranjan Jha et al., 2017; Thapar, Jaswal, and Nigam, 2019). However, in real-world scenarios, knuckles must be extracted from the hand image with various rotations or quality variations (Vyas, Rahmani, et al., 2021). A framework to localise and recognise different regions of the hand for person identification was presented in (Vyas, Rahmani, et al., 2021) in which faster R-CNN and DenseNet201 models were trained on annotated data of DIP, PIP, and MCP (base knuckles) from hand images from the '11k Hands' and 'PolyU HD' datasets. The trained R-CNN model was used to localise and recognise different hand regions. However, the paper did not consider the fingernails or the thumb. The study (Thapar, Jaswal, and Nigam, 2019) reports the use of the DIP, PIP and fingernails based on the PolyU finger knuckle image (FKI) database (Ajay Kumar, 2014), which consists of raw images of middle fingers only. This data can then be passed through a Siamese network to identify individuals.

Most methods for DL-based segmentation require extensive labelling of data for training the network, which is often hard to fulfil for sizeable datasets (Neverova et al., 2015). It has to be stressed that to date, the study of fingernails from hand images has not attracted attention within biometric systems research. One reason may be that fingernails are prone to change.

Nevertheless, in the above section, most prior works for hand joint detection did not include segmenting the defined hand joints.

2.4 Feature extraction of finger knuckles patterns and fingernails

A large portion of the CV literature is devoted to identifying, comprehending, characterising, and improving features that can be extracted from images because many CV algorithms for image classification rely on detecting and extracting local characteristics in images. As a result, many elements mentioned in the literature have been hand-crafted or carefully developed to resolve problems such as occlusions and differences in scale and lighting. Finding the ideal balance between accuracy and computing economy is a common challenge in creating handmade features (Nanni, Ghidoni, and Brahnam, 2017). In a pattern recognition problem, feature extraction has a critical role (Jaswal, Kaul, and Nath, 2016). Coding methods (Lin Zhang, Lei Zhang, and David Zhang, 2009; Le-Qing and San-Yuan, 2010); subspace approaches (Guru, Nagasundara, and Manjunath, 2010); (Shariatmadar and Faez, 2011); texture analysis methods (Lei Zhang et al., 2011), and other image processing methods are the most used traditional knuckle-print (Jaswal, Kaul, and Nath, 2016) and fingernail algorithms, and more recently, other algorithms depend on DL of CNN methods. Therefore, feature extraction can be categorised into two main types: handcrafted and DL approaches. These categories can be employed in uni-modal or multi-modal recognition systems. The following subsections give an overview of them.

2.4.1 Handcrafting approach for uni-modal recognition system

A decade ago, research in the area of anatomical hand feature extractions relied on the handcrafted engineering method. The preponderance of state-of-the-art methods using this approach followed the classical pipeline (Uhl, 2019; Jun and Guoqing, 2017; Lee et al., 2014), which consists of image acquisition using optical cameras for collecting RGB images of hand knuckles, pigmentation, scars and other features. The pre-processing phase, for example, involved hand region segmentation, noise removal and contrast enhancement (Jun and Guoqing, 2017). The feature extraction phase, for example, entailed a Gabor filter. A feature vector, which is a digital representation was then obtained. Lastly, paired images were matched using similarity measurements.

The paper by (Ajay Kumar and Ravikanth, 2009) developed an automated method to obtain knuckle texture and geometrical features from the finger back surface and study its performance for a biometric system. Four fingers of each hand were used. The major knuckle was extracted from each image. Different methods were applied: principal component analysis (PCA), linear discriminate analysis (LDA),

and independent component analysis (ICA). However, the extraction process ignored other elements of the hand.

In a study by (Ajay Kumar and B. Wang, 2015), five images of the dorsum index finger from each of the 120 participants in this database are available. The thinned ROI images were utilised to identify minute patterns in the knuckle area using pre-designed masks that created particular patterns for the targeted singularities in the thinned image to extract the location and kind of minutiae. This convolution-based technique proved quick and accurate in localising numerous kinds of minutiae recovered in this work. The authors investigated many approaches for quantifying knuckle minutiae before settling on a combination of minutiae quality and local coherence of gradients (Ajay Kumar and B. Wang, 2015). However, this study only employed the major knuckle of the index finger and disregarded the other elements of the hand.

In (Qian, J. Yang, Tai, et al., 2016) study, three major steps were conducted to investigate deep gradient information: partial sub-images were created by decomposing the image using a local HOG orientation; the others were the original image's orientation map and magnitude map; using HOG to re-explore the gradient information on all of the sub-images, and dimensional-reduction technology was used to create a low-dimensional and discriminative feature vector. Fisher linear discriminate analysis (FLDA) was used to compress the feature space in order to increase discriminative power and computing efficiency. The paper used various modalities, but did not investigate fusion techniques.

The (Waghode and Manjare, 2017) study demonstrated another subspace mechanism for individual validation using the rear surface of the finger knuckle to compare the client and dataset knuckles from a captured image using a web camera. The framework's goal was to confirm or provide data to a person, and it accomplished this by employing subspace methodologies such as the Gabor channel, PCA, and LDA computations with a probabilistic neural network (PNN) classifier. A total of 114 volunteers who gave five images of each forefinger to build a FKP database, which has 570 images in total. However, the study did not specify from which finger or knuckle these images had been taken. The first step in the framework is to demonstrate that a recognition system based on local characteristics, utilising an extraction approach called the Compound Local Binary Pattern (CLBP), can perform well (Amraoui, Fakhri, and Kerroum, 2017). The second step consists of analysing and comparing the performance of recognition to other methods in order to demonstrate the usefulness of the proposed methodology (Amraoui, Fakhri, and Kerroum, 2017). However, the work did not show the results in terms of equal error rate (EER), false rejection rate (FRR), and genuine acceptance rate (GAR).

For FKP recognition, Fei et al. (Fei et al., 2019) proposed a discriminative direction binary feature learning (DDBFL) technique. A direction convolution

difference vector (DCDV) was developed in (Fei et al., 2019) study to better represent the direction information of FKP images. The DDBFL approach was then used to learn K-hash functions so as to translate a DCDV into a binary feature vector. Concatenating block-wise histograms of the DDBFL codes yields the final description for FKP recognition (Fei et al., 2019). The PolyU dataset was used to evaluate the proposed method. However, the identification accuracies, average and standard deviations, were measured from the index and middle fingers only.

2.4.2 Handcrafting approach for multi-modal recognition system

Handcrafting feature extracting methods can be applied in a multi-modal recognition system to improve the performance. Feature extraction techniques in the spatial and spectral domains were used for major and minor knuckles in (Ajay Kumar, 2014). Local binary patterns, enhanced local binary patterns, band limited phase only correlation, and 1D log-Gabor filter based matching were used in this research. Topothesy fractal dimension characteristics and structure function-based transform features were extracted from left index (LI), left middle (LM), RI, and right middle (RM) finger knuckle images in (Grover and Hanmandlu, 2015). Topothesy is a method of assessing surface roughness that is least affected by scale fluctuations. The authors of (Grover and Hanmandlu, 2015) work were also interested in determining structure function transform characteristics. However, the study did not demonstrate the rank-1 accuracy, or receiver operating characteristic (ROC) curve for performance evaluation.

In (Amioy Kumar, Garg, and Hanmandlu, 2014) study, the Haar wavelet was the most basic conceivable wavelet, capable of providing information at the highest possible resolution of an image. This was achieved by building a Haar wavelet pyramid technique. As a result, the technique was able to identify both global and local texture alterations. Typically, collected images can include duplicate information. This redundancy may be used to arrange data into factor representations. The relationships between image pixels in such representations may be isolated as independent components and employed as appearance features using ICA. The work considered the nail shape of the index, middle, and ring finger only and ignored the knuckle elements of the dorsal hand.

In (Kulkarni, Raut, and Dakhole, 2015) study, the Kekre Wavelet Transform (KWT) was utilised to extract localised spectral information from the finger knuckle images' ROI. The author calculated phase congruency, as well as local orientation and phase, using a local image patch. These three local characteristics were distinct from one another and represent various parts of image local information; they were merged with global features such as Fourier transform coefficients. This is known as local feature integration (LFI) and the proposed method was tested on the right index and

middle finger knuckles (Kulkarni, Raut, and Dakhole, 2015).

In (Nigam, K. Tiwari, and Gupta, 2016) work, the FKP texture was improved for richness and discriminative strength. The retrieved FKP ROI was partitioned into blocks, with the mean of each block serving as the coarse illumination for that block. This mean was multiplied by the original block size. contrast limited adaptive histogram equalization (CLAHE) was used to improve the contrast of the resulting ROI. The normalised and improved FKP data were changed with the use of two suggested encoding techniques, Gradient Ordinal Relation Pattern (GORP) and Star GORP (SGORP). These transformations can accommodate variations in light. The extraction of corner features has a significant derivative in two orthogonal dimensions and can give adequate information for tracking. The study proposed in (Usha and Ezhilarasan, 2016a) described the development of an integrated methodology based on angular geometry analysis and multi-resolution transform-based texture analysis methods for extracting geometric and texture characteristics from the proximal and distal finger knuckle areas, respectively. The PCA on knuckle contours' contourlet coefficients provided fewer dimensions by detecting orthogonal linear combinations with higher variances among the coefficients. Nevertheless, the raw images were acquired from a pre-defined dedicated area around the major knuckle, which caused a high correct recognition rate (CRR).

Convex curves obtained from the FKP pictures were utilised in an angular geometric analysis method (AGAM) to extract accurate form-oriented characteristics in the proposed study in (Usha and Ezhilarasan, 2016b). The retrieved ROI image of an FKP was additionally subjected to a rapid discrete curvelet transform to obtain the texture of the FKP. This image analysis technique is a multi-resolution, band-pass, direction, and functional analysis approach that is extremely useful in portraying curved singularities in an image in a computationally efficient manner.

Image filtering was accomplished using Gabor bank filters in this study (Khellat-Kihel et al., 2016) to compute the magnitude responses of the filtered images by applying all filters to the original input images. They presented a feature fusion concatenation employing three feature sets (finger vein, fingerprint, and FKP) to create a new feature vector. Because feature reduction can increase feature discrimination, they recommended feature selection of the best qualities by decreasing the feature vectors derived from many current approaches, such as PCA and LDA.

However, the raw data were collected from a specific region around the major knuckle, and the performance using the CRR metric was not provided in these studies (Usha and Ezhilarasan, 2016b; Khellat-Kihel et al., 2016).

In (Jaswal, Nigam, and Nath, 2017b) a multi-fusion biometric system was presented. The work employed a unique feature matching approach inspired by the operation of DLNNs. Although the matching method did not generate DLNN models for learning, it was a hierarchically inspired matching paradigm. The most common

local descriptors, such as SIFT, SURF, and others, focused exclusively on prominent feature locations and fitted deformable objects poorly. As a result, they were unable to compute dense point wise correspondences. Deep-matching techniques, on the other hand, relied on a quad tree-based multilayer network to filter an image cell at multiple scales, and matched non-rigid deformable surfaces as well as significant displacements extremely successfully (Jaswal, Nigam, and Nath, 2017b). However, the knuckle pattern data were acquired from a small region around the major knuckle. In (Jaswal, Nigam, and Nath, 2017a) study, the additive noise was considered to be multiplicative in image enhancement, which has resulted in substantial improvement. To extract robust edge information, innovative image transformation algorithms based on the ordinal connection of two neighbouring pixel gradient values, namely BOP and SOP, were presented.

Usha et al. (K. Usha and M. Ezhilarasan, 2018) proposed a strategy for improving performance by simultaneously extracting and incorporating geometric and texture features of the finger knuckle by score level fusion on the entire dorsal surface of the finger. A completed local ternary pattern (CLTP), 2D log Gabor filter (2DLGF) and Fourier-SIFT (F-SIFT) were implemented to extract the local texture features of an acquired finger back knuckle surface. However, only the index, middle, ring, and little fingers' knuckle surfaces and corresponding proximal and distal knuckle regions were obtained. Also, the thumb finger and the nail elements were ignored. In (Muthukumar and Kavipriya, 2019) work, short and long Gabor features were used to analyse the different frequency content information around the image's localised region. These Gabor features are multidimensional and have a high level of redundant features, which aided in the reduction of dimension when employing this Gabor filter.

A local textural descriptor was a BSIF used in (A. Attia, Akhtar, Chalabi, et al., 2020). This feature descriptor essentially employs a series of fixed-size filters to represent the neighbourhood layout of the core pixel. Another powerful tool employed in this work was the Gabor wavelet, which has optimum localization capabilities in both the spatial and frequency domains. In (A. Attia, Akhtar, and Chahir, 2021), Using ICA, a group of convolution filters were learned from the source images using BSIF. A histogram of the binary codes for each pixel represented the outcomes of BSIF features. The textural elements of the photographs of the dorsal finger patterns may be well described by these histograms. One significant feature vector was created by combining the histograms obtained from the minor, major, and dorsal finger pattern pictures that were encoded using the BSIF descriptor. Prior to the matching stage, dimensionality reduction was applied to these big, high-dimensional vectors using a common and straightforward method called PCA.

However, the raw images were taken from a dedicated region of the major knuckles from the middle and index fingers of the left and right hands in these studies (A. Attia, Akhtar, Chalabi, et al., 2020; A. Attia, Akhtar, and Chahir, 2021).

Cascading feature extraction was applied using local ternary pattern (LTP) in the first stage in (Anbari and Fotouhi, 2021). The features from the relaxed local ternary pattern (RLTP) were then extracted, and a histogram of those features was created. To solve the issue of low-frequency, a uniform pattern was used to refine feature vectors. The work evaluated the proposed approach using rank-1 accuracy; however, the EER, FRR, and GAR were not provided.

In (Hammouche, A. Attia, and Akhrouf, 2022) investigation, the suggested approach started by separating two components: light and reflectance, applied to the major and minor knuckles. The adaptive single scale retinex (ASSR) technique was used to compute these components from the major or minor ROI images. Using ICA, BSIF trained an ensemble of convolution filters from original images. All pixels of the major and minor ROI images taken from dorsal finger images, as well as reflectance and illumination images, were coded using these learnt filters. As mentioned above, the proposed method only extracted features from the major and minor knuckles and excluded the base knuckles.

The suggested approach by (Jaswal, Nigam, and Nath, 2017a) was based on the fusing of several texture properties at the score level. The ROIs of improved FKP images were first retrieved using a modified Gabor filter approach. To produce robust image representations, the collected ROI data were further altered utilising two innovative BOP and SOP approaches. A hierarchical deep matching technique was used to match the SIFT image features. Deep matching was constructed from the top down and bottom up to handle weak texture and non-rigid deformable areas successfully. However, the knuckle patterns were obtained from a pre-defined raw image of the major knuckles.

In general, handcrafted features can require little computation time, a small number of data samples and are easy to understand (Lee et al., 2014). The feature descriptor is problem-specific, and requires human intervention (Nanni, Ghidoni, and Brahnam, 2017; Yann, Yoshua, and Geoffrey, 2015). However, CNNs outperform the handcrafted features (Taigman et al., 2014) and have demonstrated success and high performance in dealing with different problems (Schmidhuber, 2015).

2.4.3 DLNN

The DLNN is composed of multiple processing layers that allow computational models to learn representations of data with various levels of abstraction. These methods have significantly enhanced the state-of-the-art in speech recognition, object recognition, object detection and many other areas, such as genomics. Deep learning discovers complex structures in large data sets by utilising the back-propagation algorithm to identify how a model should adjust its internal parameters used to compute the representation in each layer from that in the prior layer. Furthermore, DLNNs have

brought about advances in processing images, video, speech and audio (Lecun, Bengio, and Hinton, 2015). This technology influences many aspects of everyday life and is present in products such as cameras and smart devices (Lecun, Bengio, and Hinton, 2015); ML frameworks are used to distinguish objects in images, and match two or more objects. Progressively, these applications benefit from the aforementioned DL mechanism. On the other hand, classical ML techniques are restricted in their ability to process natural data in their raw format.

Decades ago, constructing a pattern-recognition or ML system demanded careful engineering and significant field expertise. It involved designing a feature extractor that transformed an image's raw data (pixel values) into a proper internal representation or feature vector. In this case, the learning subsystem, usually a classifier, could recognise patterns in the input image. Representation learning of an image is a collection of techniques that permits a machine to receive raw images and to identify automatically the representations required for detection or classification (C. Zhang et al., 2020), learn representation methods with multiple levels, and compose non-linear modules. Each module transforms the representation at one level (beginning with the raw input image) toward a representation at a somewhat abstract high-level. Using the composition of such transformations, highly complicated functions can be learned. In the classification problem, higher layers of representation increase aspects of the input that are essential for discrimination and to overcome unnecessary variations. For example, the raw image forms an array of pixel values, and the learned features in the first layer of representation usually describe the existence or absence of edges at particular directions and positions in the image. The second layer commonly discovers motifs by locating patterns of edges, despite small differences in the edge regions. The third layer may collect motifs into more comprehensive combinations that match components of familiar objects, and consequent layers would identify objects as compounds of these parts. The fundamental aspect of DL is that human engineers do not design these layers of features: they are learned from data by a general-purpose learning system (Goodfellow, Bengio, and Courville, 2016; Lecun, Bengio, and Hinton, 2015). The following subsections will review the DL approach for both uni-modal and multi-modal recognition systems.

2.4.4 DL approach for uni-modal recognition system

In reviewing the literature, it was observed that there are few papers dealing with uni-modal recognition systems using DL approaches. In (Zhai et al., 2018) study, the authors improved the texture image distribution by augmenting the FKP ROI images with the histogram equalisation approach. Histogram equalisation at various levels was employed to enrich the FKP data, considerably improving the classification results. In general, the FKP database after histogram equalisation will meet the data

quantity need while avoiding the over-fitting issue (Zhai et al., 2018). The study designed a novel CNN structure for FKP recognition. However, the authors did not attempt to retrain a pre-trained CNN in the original dataset "ImageNet", such as AlexNet.

Two images were sent into the Siamese neural network in (A. Attia, Akhtar, Chalabi, et al., 2020) study. The weights its sub-networks were identical, and updated across both branches simultaneously. For each input image, the sub-networks generate a feature vector for a later matching procedure.

In a study by (Tarawneh et al., 2022), layers F6 and F7 of the VGG-19 were employed for feature extraction, and dimensionality was reduced using PCA, in addition to computing the average, maximum, and minimum of the merging of in-depth features. The study used accuracy, precision, recall, and F-score to evaluate the performance, However, EER, FRR or GAR evaluation metrics were not provided for the authentication. Also, the work utilised the VGG-19 DLNN model to obtain deep features, even though others could have been tried, such as RestNet, and VGG-16.

2.4.5 DL approach multi-modal recognition system

Authors chose the shape of the outer part of the finger knuckle as a biometric feature in the work (K. B. Sid et al., 2017). These characteristics are permanent and stable throughout life, making them unique to each person. The work aimed to achieve uni-modal and multi-modal biometric systems based on multi-sample FKP images, and from feature extraction, they used a DL discrete cosine transform network (DCTNet). The goal of multi-sample images was to increase the performance of the biometric system and the value of safety and trust in security systems based on biometric technologies; and the DCTNet was used to view different representations of several levels together to give upper-level characteristics (K. B. Sid et al., 2017). However, the proposed approach was not evaluated using another dataset like '11k Hands'.

In (Gao et al., 2019) work, image reconstructions from six orientations fully exploited the directional information captured by the Gabor filter. Then, for matching with the binary Gabor response, a mask was built that takes into account the reconstructive mistake and the Gabor response at a given place. The specified mask highlighted the importance of a specific spot along a particular orientation. However, the work did not attempt to evaluate the proposed method in various benchmarks, including knuckles or fingernails of the hand.

In (Hom Choudhury, Amioy Kumar, and Laskar, 2019), AlexNet was retrained using datasets of knuckles and fingernails from the index, middle, and ring fingers; an FC7 layer was employed for extracting features. For an input image, the Siamese network returned a 128-D feature embedding in (Thapar, Jaswal, and Nigam, 2019). The network determined the L2 distance between two input images and the

appropriate feature embedding. The triplet loss function was used to train the network. The authors strengthened the network by implementing a dynamically adjustable margin, hard negative mining, and data augmentation for efficient and successful training. Although the authors employed transfer learning to adapt the pre-trained AlexNet to the selected database, the authors of the suggested method did not study the pre-trained models of other DLNN models.

The PCANet model in (Chlaoua et al., 2019) cascaded several filter bank convolutions extracted from input FKP using the PCA method with an intermediate mean normalisation step and two further phases, binary hashing and histogram composition. PCA filters can be applied in phases using the PCANet technique to retrieve higher-level feature vectors. The LIF, LMF, RIF, and RMFs were captured from the subject in the PolyU FKP Database. However, the PCANet model was not evaluated using other knuckles or fingernails.

The fully connected layers of a pretrained DenseNet201 were used to perform transfer learning in (Vyas, Rahmani, et al., 2021) for a person identification system. To accommodate the number of classes in the new task, the layers were replaced with new fully connected ones. The loss of multi-class cross-entropy was then computed. However, this framework did not detect or extract features, or identify the persons using the fingernails.

Using palmprint and FKP modalities at matching score-level fusion, a quick and easy hand-based recognition multimodal biometrics system was suggested in (A. Attia, 2022). The suggested method used a PCANet DL approach. The DL feature extraction in particular was utilised to extract features from each ROI of the FKP and palmprint images, then a PCANet was employed. This algorithm consisted of two stages: binary hashing and a PCA filter bank. In addition to being straightforward, PCANet techniques offer a manageable feature vector size at a quick processing rate, making them an effective and desirable option for a number of applications (A. Attia, 2022). However, the PolyU FKP database has images taken from a major knuckle of a finger per subject and dismissed other hand elements or full images of the hand.

2.4.6 Siamese network

A Siamese network is a type of DLNN technique that has been used in many applications. For example, in verification using signatures (Bromley et al., 1994). Various image matching algorithms have been presented in recent years, aiming to improve their accuracy and performance. These approaches may be divided into two general types. The first category comprises hand-crafted representations that forecast whether an image pair is positive (similar) or negative (dissimilar). For example, the two images act as pairs of the same view. These methods are examples of the bag-of-visual-words (BoW) (Csurka et al., 2004; Afifi, 2019). The second category is

based on DL models, namely DL of CNNs, which have been utilised successfully in image classification, object detection, and image retrieval. Image representation and similarity measures are becoming increasingly crucial in image retrieval tasks that try to discover matching images in a large dataset. This paper (Babenko et al., 2014) suggested and demonstrated using a pretrained CNN for a similar image classification issue in image retrieval, with promising results.

In (Melekhov, Kannala, and Rahtu, 2016) study, the authors suggested locating matching and non-matching pairs of images by expressing them using NN-based feature vectors, the similarity of which are quantified by Euclidean distance. CNNs were used to train the feature vectors from labelled examples of matching and non-matching image pairings in a Siamese network architecture using a contrastive loss function for a generic image matching task. However, this work did not attempt to use different DLNNs, such as VGG-19 or ResNet50, as the base structure for the Siamese network.

In (Tang et al., 2019) study, the authors offered an approach for extracting additional distinguishing characteristics from finger vein images. In response to a lack of training data, they employed a substantial image augmentation method and created a pretrained-weights-based CNN. The expanded network was built using the shallow layers of ResNet-50 and a few additional, specially constructed layers. They then built a Siamese structure combined with a modified contrastive loss function for training the CNN, which significantly increased its performance. Finally, they created a lightweight CNN with depth-wise separable convolution to make the CNN deployable on embedded devices. They used a knowledge distillation approach to learn the knowledge from the pretrained-weights-based CNN, making it small, yet effective (Tang et al., 2019). Nevertheless, the study did not show the performance using a CRR evaluation metric.

A multi-Siamese CNN architecture is proposed in this work (Lin and Ajay Kumar, 2017) for correctly matching contactless and contact-based fingerprint images. In addition to fingerprint images, the proposed architecture incorporates hand-crafted fingerprint properties such as minutiae and core points. The fingerprint photos and extracted characteristics are used to train this multi-Siamese CNN. As a result, the concatenation of deep feature vectors generated by many networks yields a more robust deep fingerprint representation. A publicly accessible database of contact-based and related contactless fingerprints is used to illustrate the usefulness of the suggested technique (Lin and Ajay Kumar, 2017).

Knuckle images may identify an individual and be used to discover similarities amongst knuckle images (Joshi et al., 2019). The authors suggested a Siamese CNN structure, and used the Euclidean distance to train the Siamese network to increase the distance or dissimilarity between images of various individuals based on their fingers, and reduce the distance between images of the same person or finger (Joshi et al.,

2019). This paper validated the Siamese network using the PolyU FKP database as a benchmark. The 165-person dataset was made up of 12 high-quality images of major knuckles of the hands from each of four fingers: the LI, LM, RI, and RM. In total there were 7920 images from 660 different fingers (Joshi et al., 2019). However, this study dismissed other fingers, knuckles and fingernails images.

The paper by (Bromley et al., 1994) described an algorithm for validating signatures. Two identical Siamese sub-networks were linked at their outputs to form this network. The joining neuron monitored the separation between the two feature vectors as the two sub-networks extracted features from the two signatures during training. Comparing an extracted feature vector to a feature vector stored for the signer constitutes verification. All other signatures were disregarded as forgeries unless they were closer to this stored representation than a set threshold (Bromley et al., 1994).

Another study used a similar approach to the one above. It employed a Siamese structure for the face verification application. The main difference was the nature of the loss function minimised by the training phase. The loss function was derived from the discriminative learning framework for energy-based models (EBM) (Chopra, Hadsell, and LeCun, 2005). A Siamese CNN model was trained using pre-processed knuckle ROI images. The model scored an excellent performance, better than the state-of-the-art when tested against the open-source PolyU FKP database. However, the images were obtained from the knuckle area only (Joshi et al., 2019).

A biometric system can use a Siamese network and Variational autoencoder (VAE). An autoencoder is a neural network that has been taught to attempt to replicate its input. It features an internal hidden layer that defines the code used to represent the input. The network may be divided into two parts: an encoder function and a decoder that generates a reconstruction (Goodfellow, Bengio, and Courville, 2016). The paper by (Agnihotri et al., 2019) showed a case study of palm and finger vein authentication using a convolutional auto-encoder and Siamese triplet loss network. The study used an auto-encoder by U-Net model for the segmentation of the vein. The autoencoder network learned how to transform the original image to a texture code matrix (TCM) domain. Following that, the second auto-encoder network learned how to generate an image ray transform (IRT) domain from the TCM domain. These two auto-encoders were fused. The advantages of using a Siamese network is that it requires less data to train, improves the accuracy, and its task is to report whether the two pairs of images are the same or not the same (M. Attia et al., 2019).

It is suggested in (Thapar, Jaswal, and Nigam, 2019) by Thapar et al. to match the entire dorsal finger as compared to merely the major and minor ROIs using the proposed FKIMNet (Thapar, Jaswal, and Nigam, 2019), which is a Siamese-based CNN matching framework. The framework used hard negative mining, data augmentation, and a dynamically adaptable margin for efficient and effective training.

The network then produces an encoding image with 128-dimension features. The paper by (Daniel III, Piraccini, and Tosti, 2004) explored cases of using the fingernail in forensic science. It can be observed from the previous studies, more studies need to be carried out using the Siamese network for person identification from the dorsal hand images.

2.5 Feature matching in a uni-modal recognition systems

This section will explore the feature matching in a uni-modal recognition systems. Feature extraction and selection play a fundamental role in pattern recognition problems to find the best match. In (Ajay Kumar and B. Wang, 2015), they examined three methods to match the recovered knuckle minutiae in this study: minutiae triangulation, minutiae cylinder-code, and spectral minutiae. The nearest neighbour (NN) classifier was employed for classification in (Qian, J. Yang, Tai, et al., 2016). Face, FKP, and palmprint data were used. A similarity score using the Euclidean distance was employed on faces from the Labeled Faces in the Wild (LFW) database (Qian, J. Yang, Tai, et al., 2016).

The study in (Waghode and Manjare, 2017), a PNN was inspired by Bayesian configuration and proven estimators for likelihood thickness capabilities. The PNN's primary function is to estimate the probability thickness capacity of components of each class from grave preparation tests using the Gaussian Kernel. The study in (Amraoui, Fakhri, and Kerroum, 2017), produced matching scores for recognition using the following distance-based classifiers: Jeffrey Divergence, City-block, and Euclidean distance.

In (A. Attia, Akhtar, Chalabi, et al., 2020) work, the Siamese network was trained to increase the distance (dissimilarity) between images from various people or fingers while minimising the distance between images from the same finger. The work presented in (Fei et al., 2019) calculated the similarity of two discriminative direction binary feature learning (DDBFL) based descriptors for FKP recognition using the simple and efficient chi-square distance. Multinomial logistic loss was calculated in this study (Kim, Cho, and Park, 2019) using the soft-max method.

2.6 Fusion methods

By merging data from several sources in a systematic way, a multi-biometric system can effectively address some of the drawbacks of a uni-biometric system. In addition, since the combined information is more distinctive to a person than the information

gained from a single source, using many sources frequently leads to higher recognition performance and increased system dependability. Moreover, the fusion level can be at a sensor-level, feature-level, score-level, rank-level, or decision-level. The sources of multi-biometric systems can be obtained from multi-sensor, multi-algorithms, multi-instance, multi-sample, and multi-modal (M. Singh, R. Singh, and Arun Ross, 2019). In the following subsections, different fusion techniques, including sensor-, feature-, score-, rank-, and decision- level, will be discussed.

2.6.1 Sensor-level fusion

In multi-sensor or multi-sample algorithms, fusion or data-level fusion typically applies when data is integrated right away after its acquisition. In other words, data fusion is done on the raw data itself before feature extraction. This relates to a direct pixel-level combination of images taken from a camera, for instance various poses of flat, and angled dorsal hand images (M. Singh, R. Singh, and Arun Ross, 2019).

2.6.2 Feature-level fusion

Systems process input samples using a variety of algorithms. A single sensor is used to collect data from a biometric modality, but various algorithms are employed to process it. For instance, a palmprint identification system may use Gabor, line, and appearance-based palmprint representations for matching. In contrast, a fingerprint recognition system may use minutiae and texture information to match fingerprints. These systems profit from the ability to collect and use many forms of data from the same sample. Multi-algorithm systems frequently perform better when two algorithms or feature sets give complementary information (M. Singh, R. Singh, and Arun Ross, 2019). Example of studies that have utilised feature-level fusion are (Kulkarni, Raut, and Dakhole, 2015; Khellat-Kihel et al., 2016; Jaswal, Nigam, and Nath, 2017b; A. Attia, Akhtar, and Chahir, 2021).

2.6.3 Score-level fusion

Fusion refers to algorithms that combine the match scores generated by various matchers (distance, similarity metrics, or classifiers). Mean score fusion, maximum score fusion, and minimum score fusion are some of the popular fusion algorithms used at this level, where the mean, maximum, or least score of many matches is taken into account as the final result. Because of the ease of obtaining scores produced by matchers, this form of fusion is most frequently documented in the literature (M. Singh, R. Singh, and Arun Ross, 2019), as due to its great performance and simplicity of use, it is the most often used strategy at the matching score level

(Ajay Kumar, 2014; Amioy Kumar, Garg, and Hanmandlu, 2014; Kulkarni, Raut, and Dakhole, 2015; Nigam, K. Tiwari, and Gupta, 2016; Usha and Ezhilarasan, 2016a; Usha and Ezhilarasan, 2016b; Amioy Kumar, Garg, and Hanmandlu, 2014; Grover and Hanmandlu, 2015; K. B. Sid et al., 2017; Jaswal, Nigam, and Nath, 2017a; Gao et al., 2019; Muthukumar and Kavipriya, 2019; Thapar, Jaswal, and Nigam, 2019; A. Attia, Akhtar, Chalabi, et al., 2020).

2.6.4 Rank-level fusion

After comparing the input probe to the templates in the gallery set, or the database, fusion is carried out. A ranked list of matching identities is frequently generated by the matcher when performing an identification task that involves comparing a given probe image against a gallery of photos. Rank lists from various matchers have been combined in the literature using logistic regression and the highest rank methods. Rank-level fusion is frequently considered efficient in settings with restricted access to features or match scores (M. Singh, R. Singh, and Arun Ross, 2019).

2.6.5 Decision-level fusion

Algorithms that accomplish fusion at the decision level are called fusion algorithms. One of the most popular fusion algorithms used at the decision level is majority voting. The final judgement is reached by combining decisions made by n classifiers or matchers based on a majority vote. Because only the final decisions are available in black-box systems, decision-level fusion has the benefit of functioning well with them. This is valid for many commercial systems where gaining access to features, ratings, and ranks might need to be more practical (M. Singh, R. Singh, and Arun Ross, 2019).

Different fusion strategies were conducted in (Ajay Kumar, 2014), such as max, product, weighted sum, and nonlinear fusion rules. The nonlinear fusion approach achieved the best performance. Major and minor knuckle patterns can be mixed simultaneously to increase matching accuracy for personal identification. The aforementioned study investigated match score combinations utilising linear and nonlinear algorithms and the score level combination strategy as one of several options for integrating minor and major knuckle patterns.

Features were retrieved from the index, middle, and ring fingernail plates and merged at the score level before being used as a biometric identification in (Amioy Kumar, Garg, and Hanmandlu, 2014). The authors in (Amioy Kumar, Garg, and Hanmandlu, 2014) provided two methods of nail-plate integration: (1) score-level rules were first employed for matching score fusion, and (2) classifier-based fusion of matching scores was used using a fuzzy binary decision tree (FBDT). The

PolyU FKP database was used to evaluate the adaptive fuzzy decision level fusion and hybrid fusion in (Grover and Hanmandlu, 2015). ROIs were acquired from left index, left middle, right index, and right middle finger knuckle images. In the (Grover and Hanmandlu, 2015) study, the authors investigated an extension of decision-level fusion known as adaptive fuzzy decision-level fusion, which brought uncertainty into consideration while making judgments. Fuzzification was required in (Grover and Hanmandlu, 2015) because it models uncertainty in data by selecting suitable membership functions. It was adaptive because the different biometric modalities' judgments were integrated adaptively, and a decision rule was chosen to fulfil the required security restriction.

In the study by (Kulkarni, Raut, and Dakhole, 2015), the phase congruency, local orientation and local phase using a local image patch, were conducted. These three local characteristics were distinct and represented various parts of the image's local information. These local characteristics were merged with global features such as Fourier transform coefficients; local feature integration (LFI) was the name given to this feature integration. The distance between training and test images was calculated using three nearby characteristics to compute these distances. The ultimate matching distance was calculated by fusing these three distances. Furthermore, these local characteristics were combined with global characteristics. The matcher-weighting (MW) rule was used to incorporate the characteristics.

The ultimate score of vertical and horizontal codes of GORP and SGORP can be calculated by combining basic sum rules in the presented study in (Nigam, K. Tiwari, and Gupta, 2016). Because some images have more discrimination in the vertical direction while others have it in the horizontal direction, such a fusion was highly advantageous. It may improve the performance of the suggested system.

Score level fusion was used to accomplish the integration process in (Usha and Ezhilarasan, 2016a). To assess the combined performance, three fundamental principles of matching score level fusion were used: sum of matching scores (SOM), a product of matching scores (POM), and a weighted sum of matching scores (SWM). The final matching score was derived by applying the SOM rule to the collected scores. The identification of a person is determined by combining the matching scores from two separate classifiers as illustrated in (Usha and Ezhilarasan, 2016b).

Features were retrieved from the index, middle, and ring fingernail plates and merged at the score level before being used as a biometric identification in (Amioy Kumar, Garg, and Hanmandlu, 2014). Two methods of nail-plate identification were integrated by employing score-level rules for matching score fusion and classifier-based fusion of matching scores using a FBDT. The PolyU FKP database was used to evaluate the adaptive fuzzy decision level fusion and hybrid fusion in (Grover and Hanmandlu, 2015).

The characteristics vectors acquired by the feature extraction step from three

modalities were the parameters of the fusion in (Khellat-Kihel et al., 2016). These modalities were finger vein, fingerprint, and finger-knuckle-print. They also suggested a feature selection of the best features. This space reduction can enhance feature discrimination by lowering the feature vectors generated by current approaches such as PCA and LDA. Finally, they got the outcomes using the SVM and the k-nearest neighbours (KNN). The fusion was performed using simple existing methods based on the FKP uni-modal system's decisions because the FKP uni-modal system outperformed the other two uni-modal systems. By varying the curvature and distance parameter values between two flipped curved filters, many curved Gabor filters may be created in (Jaswal, Nigam, and Nath, 2017b). Then, the combination of GOP and SGOP algorithms provided the highest recognition results in this work.

In (K. B. Sid et al., 2017), a model was a set of relevant information, discriminating and non-redundant characteristics of one or more persons with similarities. The authors would put in the same class, and these classes change depending on the kind of choice. SVM was a discrimination and classification approach based on the application of kernel functions that allowed for the best separation of plan points into multiple categories. Fusion at the matching score level was the most popular and widely utilised approach due to its high performance and ease of usage. The outputs of two or more matching modules, left index finger, left middle finger, right index finger, and right middle finger, were integrated at the matching-score level using a fusion of simple sum, product, minimum, and maximum rules.

Rather than utilising hamming distance between codes as dis-similarity, the authors in (Jaswal, Nigam, and Nath, 2017a) calculated and matched characteristics across images acquired after transformation. The suggested approach was based on fusing several texture properties at the score level. When evaluated on individual FKP data sets, multi-feature fusion beats individual feature-based approaches.

The study by (Gao et al., 2019) has two possible matching distances, which are before and after reconstruction. The matching distances before and after reconstruction are adaptively merged using a score-level adaptive binary fusion method.

In (Muthukumar and Kavipriya, 2019) study, the fusion performed using the Min and Max constraints for various sorts of arrangements applied on the right index (RI), RM, LI, LM, such as (RI + RM, RI + LI, RI + LM, RM + LI, RM + LM, LI + RM) had a greater recognition rate.

In (Hom Choudhury, Amioy Kumar, and Laskar, 2019) study, the similarity metric that applied to the extracted features was the euclidean distance. Then, the fusion of two biometric attributes, fingernail plates and finger knuckles, was applied. The suggested system in verification mode has been evaluated using four well-known score-level fusion rules: the sum rule, the product rule, the max rule, and the min rule.

In (Thapar, Jaswal, and Nigam, 2019) investigation, the L2 distance between two

input images was calculated. The study then provided a weighted sum score fusion of various finger features, including the major, minor, nail, and complete finger. The author in (Thapar, Jaswal, and Nigam, 2019) employed a multi-class linear SVM operation on each image's extracted feature vector. It was a one-against-others support vector classification technique that was employed. As a result, this technique uses binary SVMs, each of which differentiates a single class from all other classes in either a one-against-others or one-against-all fashion (SVM-pairwise scheme). Each node of the trees containing these classifiers corresponds to an SVM. Several fingers of the hand are then utilised in the biometric system. The system may operate in a uni-modal or multimodal manner for that. In this last method, the fusion is carried out at the level of matching scores using the same fusion principles.

In (A. Attia, Akhtar, Chalabi, et al., 2020) work, Deep Rule-Based (DRB) classification technique was employed. A DRB classifier is a process with four major phases that describe its global system mechanism (P. P. Angelov and Gu, 2018), which include initialisation, preparation, updating, and fuzzy rules generation. All FKP modalities were fused at the score level using BSIF and DRB. In (A. Attia, Akhtar, and Chahir, 2021) study, the closest neighbour classifier, which employed the cosine Mahalanobis distance, was used in the matching step of the proposed system. Minimising the distance (score) between the input query characteristic and the stored template was the standard for determining if two things were similar or dissimilar. Feature-level fusion of major and minor dorsal finger knuckles was utilised. Finally, feature-level fusion-extracting BSIF histograms were combined for the three modalities, which are major, minor, and finger dorsal patterns.

In (Vyas, Rahmani, et al., 2021) study, cosine distance is used to compare test features, with smaller values indicating a higher probability of a match and vice versa. Then, all three knuckles were calculated holistically using score-level fusion and averaging rule. In the proposed work in (Anbari and Fotouhi, 2021) the dominant features were obtained using an efficient three-layered model. Finally, an investigation was carried out using the extracted texture features from two, three, and four fingers combined.

In the presented study in (Hammouche, A. Attia, and Akhrouf, 2022), PCA and LDA approaches were employed for dimension reduction. The PCA algorithm was performed first, followed by the LDA algorithm on the PCA weights. In the matching stage of the proposed system, the NN classifier based on cosine Mahalanobis distance was used. Minor and major dorsal fingers were then used in the score-level fusion, and the symmetric sum-based was applied for the fusion rules. In the suggested approach by (Tarawneh et al., 2022), the authors developed a patch-level matching technique based on correlations. The algorithm began by computing correlation maps for tiny patches and progressed to more extensive patches by aggregating the smaller ones. The suggested system was then assessed using the score-level fusion rule.

However, there is generally a need for fusion investigation, including different fusion levels, such as sensor, feature, score, rank, or decision, in the current studies that can improve the person identification from human dorsal hands.

Table 2.1: The details of the available databases for palm and dorsal hand side

Dataset name	**Total no. of images**	**Population**	**Imaging type**	**Hand side**	**L/R hand**	**Ground truth?**
Hong Kong Polytechnic University Hand Dorsal Images (Ajay Kumar and Zhihuan, 2016)	2,505	501	RGB	Dorsal	Right	No
11k hands (Afifi, 2019)	11,076	190	RGB	Both	Both	No
H-Unique (HUQ) (*H-unique* 2019)	more than 3000	500	RGB	Both	Both	Yes
Hong Kong Polytechnic University Hand Dorsal Images (Version.1) (Ajay Kumar, 2014)	2,515	503	RGB	Dorsal middle finger	Not specified	No
IIT Delhi Finger Knuckle Database (Version 1.0) (Ajay Kumar and Ravikanth, 2009)	790	158	RGB	Major knuckle	Not specified	No
Tecnocampus Hand Image Database (Font-Aragones, Faundez-Zanuy, and Mekyska, 2013)	1,000	104	RGB and Thermographic (TH)	Both	Right	No
NCUT (Zhao, Y.-D. Wang, and Y.-H. Wang, 2008)	2,040	102	Near-infrared (NIR)	Dorsal	Both	No
Bosphorus Hand Vein Database (Yuksel, Akarun, and Sankur, 2011)	1,575	100	NIR	Dorsal	Left	No

Chapter 3

Person Identification from Fingernails and Knuckles Images using Deep Learning Features and the Bray-Curtis Similarity Measure

3.1 Introduction

The chapter presents a novel framework for automatic person identification based on different dorsal hand components, which has a total of 19 components, including the knuckles and fingernails of the hands, and the experimental results. This framework includes detection, and segmentation of the various regions. It involves feature extraction using the pretrained CNN model known as the DenseNet201, dissimilarity measuring using the Bray-Curtis distance metric to find the best match between individuals for person identification in a forensic application. The biometric framework for person identification typically consist of several stages including image acquisition, image pre-processing, image segmentation, feature extraction, similarity evaluation, and ranking the results. In the next two subsections, these stages will be presented in more detail and evaluated using popular databases.

On a brief review of the feature extraction methods of the hand, the studies did not consider feature extraction via deep learning without the need of retraining. Therefore, this experiment focuses on utilising a pre-trained deep learning model as a feature extractor.

The rest of this chapter is organized as follows: Section 3.2, and section 3.4 introduce the design of the proposed framework for ***p***erson ***i***dentification based on ***f***ingernails and ***k***nuckle patterns localisation from hand images (PIFK) and the extended one (PIFK^{+}), respectively. Section 3.3 and section 3.5 describe the

experimental results and evaluations of both approaches, and section 3.6 presents the conclusion and outlines future research.

3.2 The proposed PIFK framework

This section introduces the design of the proposed framework for ***p***erson ***i***dentification based on ***f***ingernails and ***k***nuckle patterns localisation from hand images (PIFK). Section 3.3 describes the experimental results and evaluation, and section 3.6 presents the conclusion and outlines future research. Figure 3.1 depicts the schematic diagram of the proposed PIFK framework. The proposed system consists of two phases: i) acquisition, and ii) identification. In the acquisition phase, a person presents the dorsal hand to capture images using the acquiring device. The captured images are then passed to pre-processing and generate cropped sub-images of the hand's keypoints as shown in figure 3.1. Each hand comprises 19 different components, namely:

1. 5 base knuckles or MCP,
2. 4 major knuckles or PIP,
3. 4 minor knuckles or DIP,
4. 1 major knuckle of the thumb or IP, as well as
5. 5 fingernails.

These are then stored in the relevant data-store d_n ($n \in \{1, 19\}$) as explained in subsection 3.3.2. A pre-trained deep neural network is then used to extract abstract high-level features from each finger's component, as described in more detail in subsection 3.2.2. These extracted features are sophisticated and instrumental to discriminate each sub-image.

The proposed PIFK framework consists of the following phases (see figure 1):

- pre-processing: the knuckles and fingernails localisation, detection, labelling, segmentation,
- feature extraction,
- (dis-) similarity evaluation and decision (matching).

These stages are detailed in the following subsections.

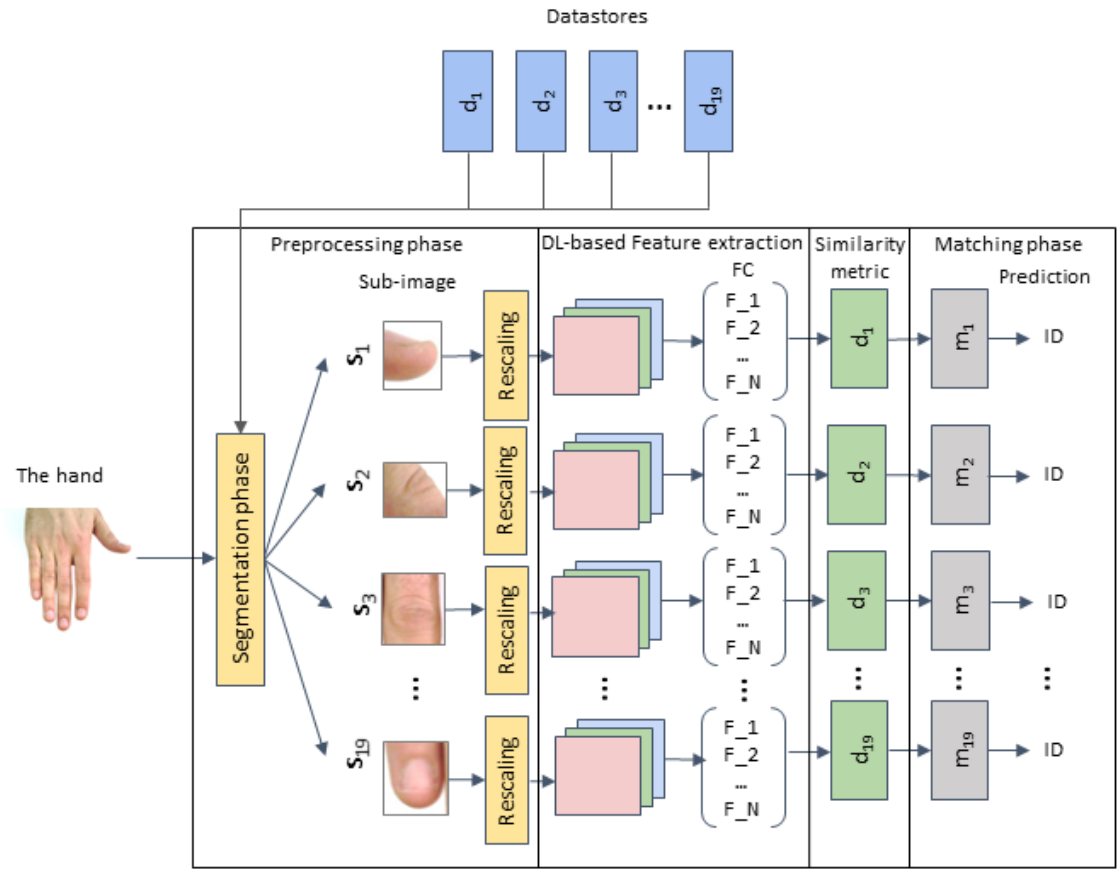

Figure 3.1: The schematic diagram of the proposed framework for ***p***erson ***i***dentification based on ***f***ingernails and dorsal ***k***nuckle patterns (PIFK) from the hand image.

3.2.1 Detection and segmentation of knuckle and fingernail keypoints

In the pre-processing phase, firstly, the original image of the hand was resized to 224×224 pixels after many experiments in image resizing. Although, the work in (Simon et al., 2017) resized the patch of the hand image to a square shape with a height and width equal to 368 pixels and added a margin of pixels around the hand image. Then the keypoints were projected/mapped to the original high-definition image, allowing the segmented image to be in suitable dimensions. Each image was then segmented based on the detected keypoints. Then re-scaling (normalization) of these sub-images was conducted to speed up the process of extracting features. The

original pixel values of the sub-images range from 0 to 255. By re-scaling, the pixel values are transformed to the interval from 0 to 1. In this work, unlike other published approaches, segmentation of five types of region of interest was considered, namely: the MCP, PIP, DIP, IP as well as the fingernail from both the left and right hands. The keypoints of these elements were obtained using the multi-view bootstrapping method for hand pose estimation (Simon et al., 2017).

The method reported in (Simon et al., 2017) was extensively trained on large datasets of annotated hand keypoints including challenge occlusion using a multi-camera setup to capture a multi-view of the hand. These datasets are the MPII human pose dataset (Andriluka et al., 2014), and images from the New Zealand sign language (NZSL) exercises of the Victoria university of Wellington Therefore, the method in this framework has not been trained on new data because it is already pre-trained for the hand regions localisation using the mentioned databases.

The basic concept of this method Simon et al., 2017 is to use multiple views of multiple frames of the hand. The method can detect and project the position of the 3D keypoints. The method utilises a keypoint detector $d(.)$, which converts a cropped input image patch of a hand $I \in \mathbb{R}^{w \times h \times 3}$ to P component and x locations as follows:

$$d(I) \mapsto \{(x_p, c_p) \text{ for } p \in [1...P]\}, \tag{3.1}$$

where confidence c_p is associated with each detection; a location point p matches one component, such as PIP or DIP; w denotes vertical and h denotes horizontal dimension of the image and 3 corresponds to the RGB channels.

Each location was detected using a pre-trained deep neural network. The detector was pre-trained on the previously mentioned databases with corresponding keypoint annotations, (I^f, y_p^f), where f indicates an image frame, and the annotated keypoints for the image I are in the set $y_p \in \mathbb{R}^2$.

The purpose of utilising this method was to estimate the approximate centre of the hand components. Each detected keypoint has a coordinate location x_p, a confidence score c_p, and an associated index. In this work, the associated index $p \in [1...P]$ was used to support the labelling of each of the 19 ROI automatically with the confidence measure $c_p \in [0, 1]$ ranging from the lowest to highest score.

In this work, the image of the hand was utilised to represent by one frame and one dorsal flat view of the hand: $f = 1$, and $v = 1$, respectively. The confidence score c for the view v and keypoint p is considered as the degree of certainty, combined with the pre-training of the keypoints in the work (Simon et al., 2017). The detection method was extended to segment each component of the hand. The segmentation was conducted by determining the dimensions of a bounding box of each region empirically based on the detected centre-point. Therefore, the segmentation of the required region can be conducted in two stages: firstly, identifying the centre of the segmented part using the multi-view bootstrapping method; and secondly, defining the bounding box

by specifying a fixed height and width per type of component in regards to the centre-point. The dimensions of the five types of components/ROI (MCP, PIP, DIP, IP and fingernails) were defined. Figure 3.2 displays the stages of the segmentation.

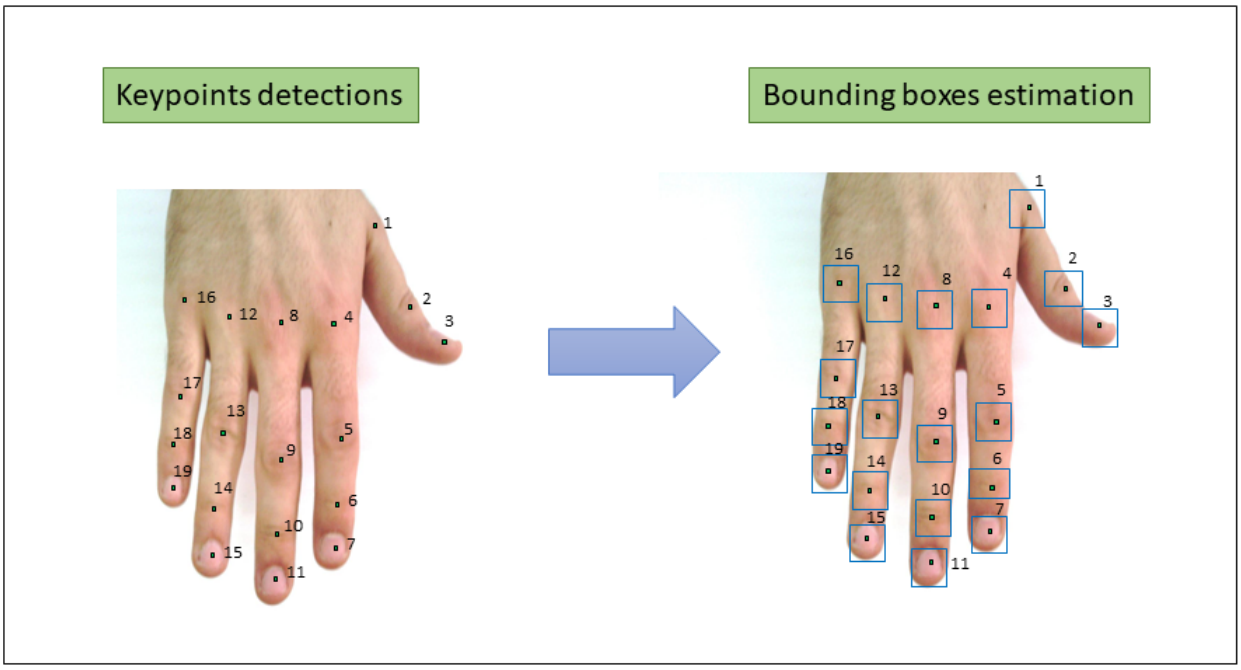

Figure 3.2: Example of the localisation including keypoints/landmarks detection and bounding boxes for all fingernail and knuckle crease regions.

The following steps demonstrate the detection of the ROI and related bounding box estimation for image segmentation. The work in (Simon et al., 2017) needs the input image of the hand with a defined dimension. They presumed that the input image patch $I \in \mathbb{R}^{w \times h \times 3}$ is a crop around the right hand for the keypoint detector. Therefore, they required a method to construct this bounding box to use the keypoint detector in any real-world scenario. Because the detection method requires a fixed size of the input image with dimensions 224×224 pixels, each detected keypoint assigned as (x', y') should be mapped to the same keypoint position denoted as (x, y) on the original high dimensional image. The image was resized and the coordinates transformed accordingly.

Thirdly, the bounding box was identified to generate a sub-image S_n of the region:

Suppose a centre location of a keypoint is (x'', y''), to identify a bounding box, an estimated margin of pixels defined by the user assigned as a and b is added to both positive and negative sides of the y-, and x-axis, respectively. The values of a and

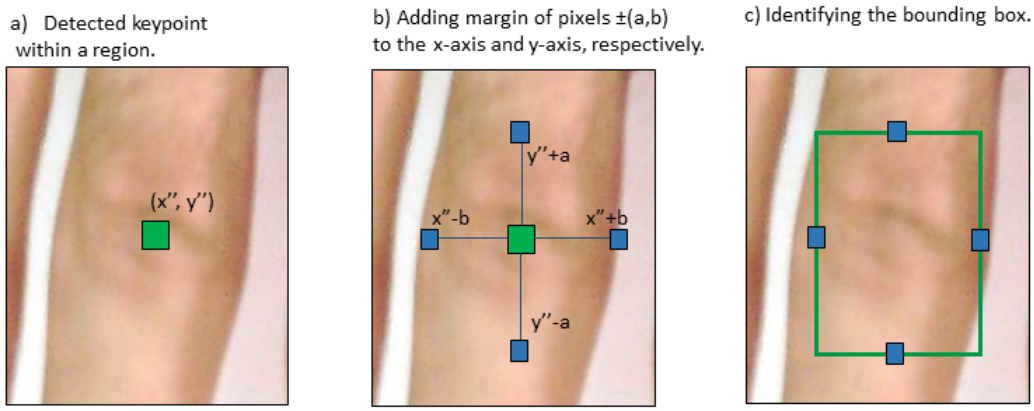

Figure 3.3: An illustration of forming the bounding box around an identified keypoint.

b (a stands for the width and b stands for the height) per component were as the following: (150, 168) pixels for the MCP. (150, 160) for the PIP and IP. (150, 140) for the DIP. (160, 184) for the fingernails.

Figure 3.3 illustrates this concept using a patch of the knuckle area. Therefore, to produce a segment/sub-image (S) the following equation can be used:

$$S = \begin{cases} \text{y''} + \text{b} \\ \text{'' - b} \\ \text{z''} + \text{a} \\ \text{z''} - \text{a} \end{cases} \tag{3.2}$$

3.2.2 Feature extraction phase by the base model of DLNN

Many image processing techniques consider low-level characteristics, such as edges, shapes, geometries, and lighting changes. However, these techniques often require extensive human intervention in the design stages. By contrast, deep learning, specifically CNNs, started to be widely used as a common mechanism for extracting features and learning the pattern representation from high dimensional data such as images (Khan, 2018). These are usually pre-trained on large image sets, such as the popular one known as 'ImageNet' (Jia Deng et al., 2009).

A deep neural network model pre-trained on one domain can transit the representation of learned data to another domain or task using a technique called transfer learning (Yann, Yoshua, and Geoffrey, 2015). It offers an architecture that enables extracting abstract and high-level features (Jia Deng et al., 2009), which

usually represent the outputs of the final fully connected (FC) layer in a vector form, the dimensions of which vary depending on the pre-trained model. In this section, fine-tuning for the original model of DenseNet201 was not considered. Instead, The original or base DenseNet201 model (G. Huang et al., 2017) that pre-trained on the famous 'ImageNet' dataset (Jia Deng et al., 2009) was considered a feature extractor. This model has a more complex architecture which makes it a more powerful feature extractor. The architecture of the model (G. Huang et al., 2017) considers specific layer connectivity for data flow. In traditional CNNs, each layer has l inputs, which are feature-maps from all previous convolutional blocks. Instead, in the DenseNet201 network, each layer l, receives data inputs from all previous layers, maps them into features, and forwards them to all succeeding layers. The output layer is denoted as x_l. A layer achieves a concatenation of features by applying $\frac{l(l+1)}{2}$ connections, compared to a typical CNN architecture that only uses a l layer network. The DenseNet201 network requires fewer parameters compared to other CNNs, enhancing data flow and gradients through the model. Figure 3.5 shows the network layout with three dense blocks.

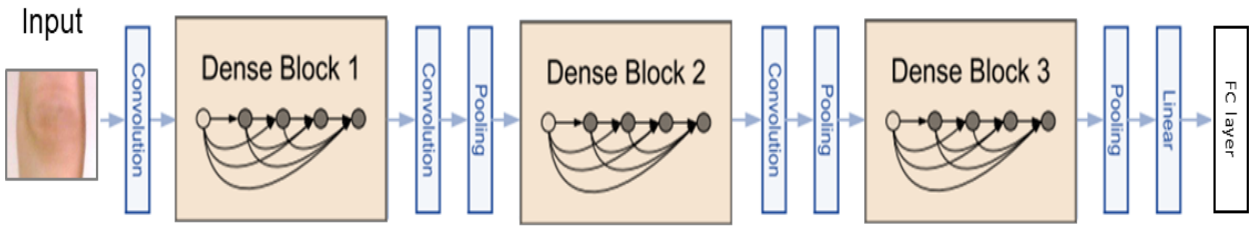

Figure 3.4: DenseNet201 architecture with three dense blocks (G. Huang et al., 2017).

The feature-maps from all preceding layers are received as input to the l^{th} layer as follows:

$$x_l = H_l([x_0, x_1, ..., x_{l-1}]), \tag{3.3}$$

where $[x_0, x_1, ..., x_{l-1}]$ denotes feature-map concatenation output in layers $0, ..., l-1$, and $H_l(.)$ refers to compound operation that comprises of batch normalization, followed by a rectified linear unit and a 3×3 convolution (G. Huang et al., 2017). The segmented sub-images were resized to 224×224 pixels to be compatible with this network for feature extraction. The sub-images were re-scaled to a new value domain ranging from 0 to 1. In this way, all features are given an equal weight, which is useful for feature extraction and the similarity metric (J. Han, Kamber, and Pei, 2012).

3.2.3 Similarity estimation and matching

Similarity metrics can play an essential role in biometric systems. In this work, various distance metric were used, which are the Bray-Curtis (BC) (Shyam and Y. N. Singh, 2015), Cosine (C), Euclidean (E), Cityblock (CB), to identify the similarity among different ROI as part of the decision phase of the person identification task. After exploring the matching results among other distances, the Bray-Curtis performed better. Figure 3.5 shows a diagram of the similarity measuring using the Bray-Curtis metric.

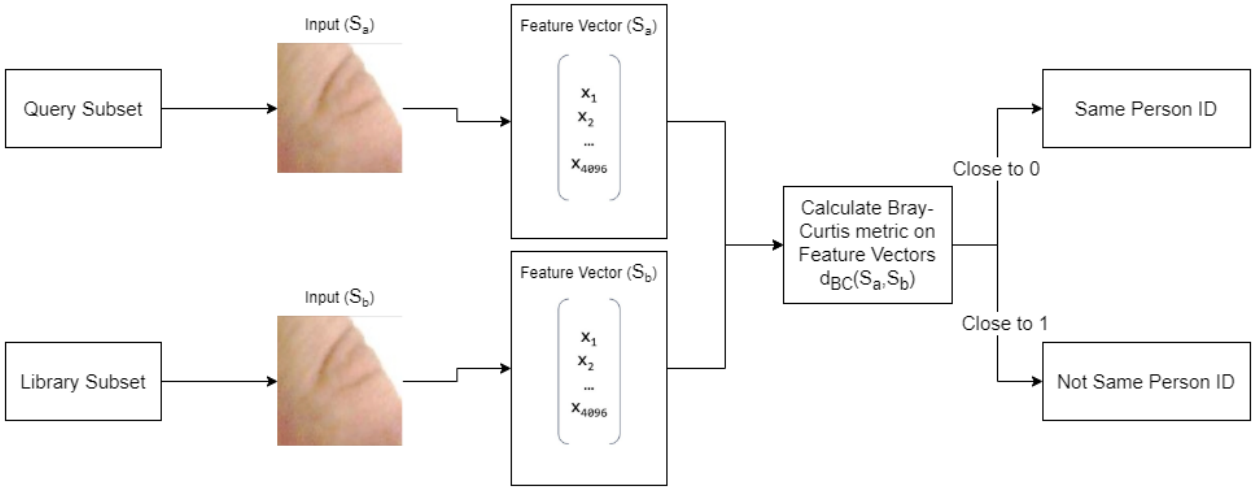

Figure 3.5: An illustrative diagram of the similarity measuring using the Bray-Curtis metric.

The Bray-Curtis (BC) metric is a distance of two vectors for two segments that can be indicated as $d_{BC}(S_a, S_b)$ and defined as follows:

$$d_{BC}(S_a, S_b) = \frac{\sum_{i=1}^{P} |a_i - b_i|}{\sum_{i=1}^{P} |a_i + b_i|} \tag{3.4}$$

The Cosine (C) metric can measure the distance between two vectors as follows:

$$C(S_a,\ S_b) = \frac{S_a.S_b}{\|S_a\|\|S_b\|} \tag{3.5}$$

The Cityblock (CB) metric may calculate the distance between two vectors as shown in 3.6. Suppose the vectors are composed of two dimensions (x, y), the CB can be defined as the following:

$$CB(S_a,\ S_b) = \|x_{S_a} - x_{S_b}\| + \|y_{S_a} - y_{S_b}\| \tag{3.6}$$

The similarity between two vectors is the opposite to the distance (the lower the distance between two vectors the higher the degree of similarity and vice versa) and can be defined as:

$$Sim_{BC}(a, b) = 1 - d_{BC}(a, b) \tag{3.7}$$

The ID is assigned to the closest match determined through the Bray-Curtis similarity as follows:

$$\hat{ID} = argmin\{d_{BC}\} = argmax\{Sim_{BC}\} \tag{3.8}$$

3.2.4 Experimental protocol

In the experiments, a valid person recognition, the one that matches two different sub-images taken from the same person, was considered and vice versa. The *rank*-1 recognition rate was employed, which is computed as follows:

$$rank\text{–}1 = \frac{N_i}{N} \times 100, \tag{3.9}$$

where N_i denotes the number of images correctly assigned to the right individual, and N indicates the overall number of images attempted to be identified.
In order to demonstrate the overall performance the cumulative matching characteristic (CMC) was used, which shows the accuracy performance in terms of $rank - n$ (Anil K Jain et al., 2008). The performance of a 1:m identification system is evaluated using the CMC. It measures an identifying system's ranking abilities. The FAR and FRR curves are the only information provided by the CMC. In addition, the FAR and FRR may be used to calculate the CMC, which is another way to visualise the data. The FAR and FRR are used to calculate the CMC, another data visualisation method. Additionally, the ROC is a widely used metric to represent the effectiveness of (1:1) matchers. For example, CMC describes how well biometric identification systems that yield ranked lists of subjects (1: m) perform. In particular, when an identification system consistently delivers the IDs connected to the K biometric samples from an enrollment database (gallery), that scored the highest (R. M. Bolle et al., 2005).

Let's assume we have large biometric samples (B_i) with corresponding ground truth IDs (B_i) to define CMC. The combination of both sets of samples is essential to measure a CMC curve associated with the following:

- First, a gallery set (G). The set $G = (B_1, B_2, ..., B_m)$ contains m biometric samples, which are (m) biometric IDs of various persons. The enrolled biometric identification database may be called the gallery set (G).

- Second, a collection of n "unknown" or query samples $(B_1^{'}, B_2^{'}, ..., B_n^{'})$, or $B_l^{'}, l = 1, ..., n$, linked to the n individuals is referred to as a probe set and is denoted by (Q).

  Any set of individual sub-images can make up the probe or query set Q. However, probe IDs are typically thought to be in gallery G. The probe set does not must include a sample from each subject in G, but it may have more than one biometric sample of a certain person.

A biometric matcher produces a similarity score $s(B_l^{'}, B_i)$ given a query biometric $B_l^{'} \in Q$ and a matching biometric $B_i \in G$. Each probe biometric is matched with every gallery biometric $n \times m$ in order to estimate the CMC. A total of n similarity scores are generated using the formula $S_l = s(B_l^{'}, B_1), s(B_l^{'}, B_2), ..., s(B_l^{'}, B_m), l = 1, ..., n$ (R. M. Bolle et al., 2005).

Moreover, the description of the datasets used to evaluate the proposed method, and the results, will be demonstrated in the next section.

3.3 Experimental results and evaluation of PIFK

This section reports the results of the evaluation of the proposed PIFK on the '11k Hands' dataset (Afifi, 2019).

3.3.1 Datasets description

First, the '11k Hands' dataset (Afifi, 2019) was used in the evaluation process of the proposed approach. Only RGB images of the dorsal surface of the right and left hands of 190 subjects were utilised in this study. Next, the Hong Kong Polytechnic University Contactless Hand Dorsal Images Database (Ajay Kumar and Zhihuan, 2016) was considered, which consists of 4650 surfaces of right-hand dorsal images in a flat position from 501 subjects. These images with the exact resolution (1600×1200 pixels) were captured using mobile and handheld cameras. More details of these two datasets can be found in section 2.1.

3.3.2 Pre-processing phase

The pre-processing phase, as shown in figure 3.1, was described in subsection 3.2.1. In this subsection, the specifics of PIFK regarding this phase will be detailed.

Step 1: Fingernail and knuckle crease detection: The input image is a human dorsal hand image. Samples of these images can be found in section 2.1. The automatic detection and localisation of the hand components were obtained from the multi-view bootstrapping method. First, the image was resized to 224×224. Next,

the resized image pass through the method. The result of this step can be illustrated using three different cases as shown in figures 3.6, 3.7, and 3.8, which indicate the case of a correct detection and localisation of all 21 keypoints/joints, missing one keypoint detection from the wrist, and misdetection keypoints in the middle, ring, and little fingers of the dorsal hand in the '11k Hands' dataset, respectively. The process of this step was described in subsection 3.2.1 and illustrated in figure 3.2.

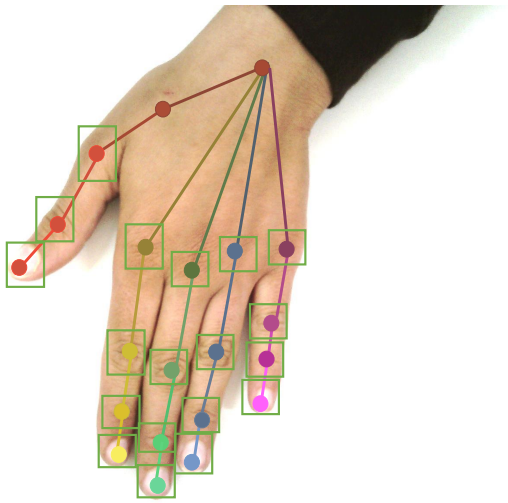

Figure 3.6: A sample of a correct detection with emphasis on localisation of all 21 keypoints/joints and bounding boxes from dorsal hand in '11k Hands' dataset.

A confidence score was associated with each keypoint/joint to indicate the accuracy of the detected joint.

Step 2: Recognition of knuckle creases and fingernails:

PIFK used automatic indexing of each ROI based on their locations; for example, index 1 and index 2 support labelling of the MCP and IP of the thumb, respectively (see figure 3.2). The pre-trained Multi-view bootstrapping method supports automatic labelling based on the ROI location.

Step 3: Bounding box estimation and segmentation:

The estimated dimension of the bounding box in terms of (width, height) was predefined in both datasets as follows: (150, 168) pixels for the MCP, (150, 160) for the PIP and for the IP, (150,140) for the DIP, and (160, 184) for the fingernails. These values were found through experimentation (see figure 3.2).

Step 4: Splitting the data into two data sets:

In this phase, the splitting of the produced sub-images into two data sets was conducted as the following:

- a query data set, and

Figure 3.7: A sample of a missing one keypoint detection from the wrist of the dorsal hand in the '11k Hands' dataset.

Figure 3.8: A sample of misdetection keypoints in the middle, ring, and little fingers of the dorsal hand in the '11k Hands' dataset.

- a library data set.

In this study, the Leave-One-Out Cross-Validation (LOOCV) evaluation method was used since the number of images is not very large. The method assessed the performance of the proposed algorithm. In LOOCV, each fold has only one sample, and random partitioning of the data into training and testing does not exist. Also, every prediction in the identification problem is independent of each other because the samples are independent (Wong, 2015). Therefore, a single individual subject was considered in the query at a time (and vary this averaging the results at the end) and the remaining individuals in the library set (also replacing one at a time). In addition, there were for each individual images of both, left and right hand and 19 ROI per image. In the '11k Hands' dataset, due to false detection of keypoints, the 87 images were ignored of left hands and 6 images of right hands. After segmentation due to the erroneous detection, 428 sub-images of left hands' components were removed and 573 of right hands' components. As a result, the total number of sub-images in the query set for all components of the left hands were 3,589 and 3,609 for the right hands. Finally, the total number of images in the library set for all components of the left and right hands were 48,780 and 50,795, respectively. In the 'PolyU HD' dataset, the total number of removed sub-images are $5,733$ out of $88,392$. Therefore, the total sub-images in the query set are $9,411$, whereas the library set is $73,248$.

3.3.3 Pre-trained DL based feature extractor

Regarding feature extractions, this study used a DenseNet201 neural network, which was pre-trained on the ImageNet dataset. The outputs of the last FC layer were considered, with dimension (1×1920), namely FC2 as a vector representing abstract high-level features.

The evaluation of the proposed algorithm is in a subject-independent manner. The data used to train the original DenseNet201 model is the 'ImageNet' dataset. Therefore, it differs from the data used to evaluate the proposed method, which was the '11khands' and 'PolyU' datasets.

3.3.4 Similarity estimation and matching

The matching process includes estimating the similarity between two sub-images of the same hand element (one from the query and one from the library). BC distance was used to estimate the similarity between pairs of sub-images as detailed in equations (3.4), (3.7) and (3.8) from subsection 3.2.3.

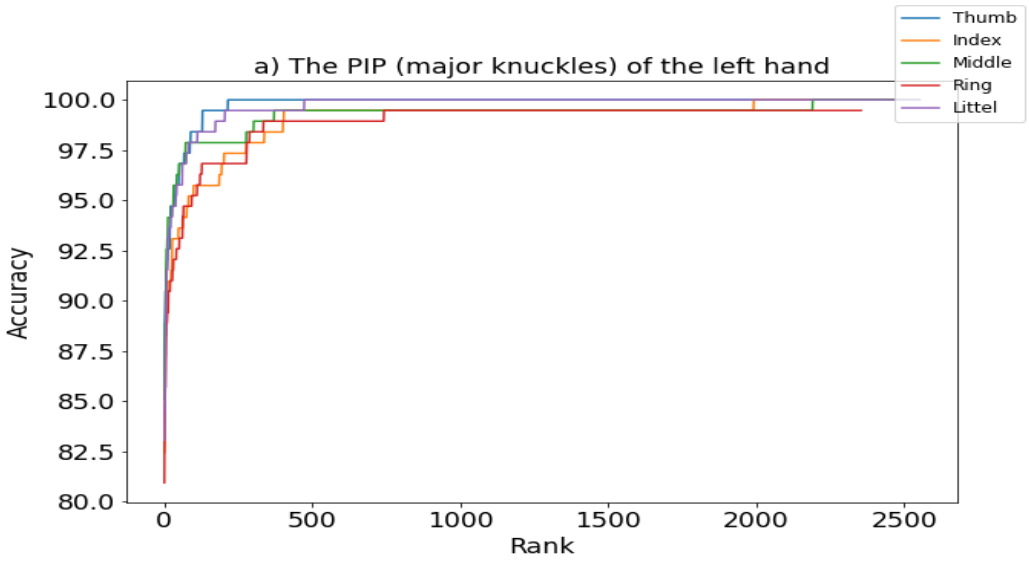

Figure 3.9: The CMC of the proposed PIFK; a) major knuckles of the left hands from the '11k Hands' dataset.

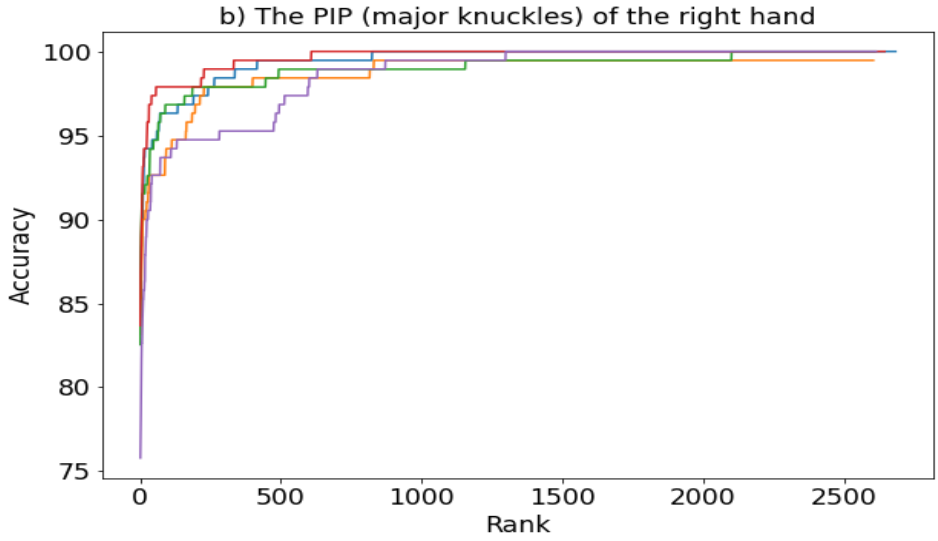

Figure 3.10: The CMC of the proposed PIFK; b) major knuckles of the right hands from the '11k Hands' dataset.

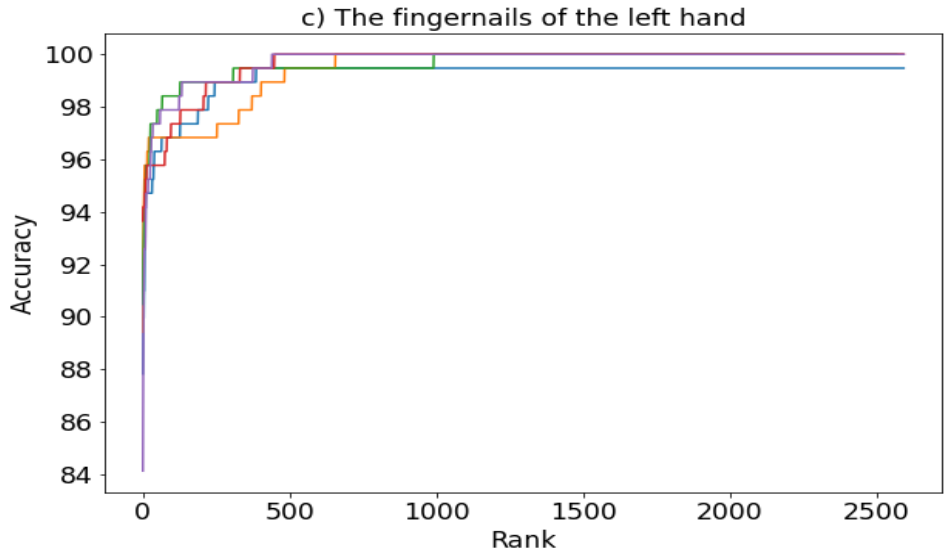

Figure 3.11: The CMC of the proposed PIFK c) fingernails of the left hands from the '11k Hands' dataset.

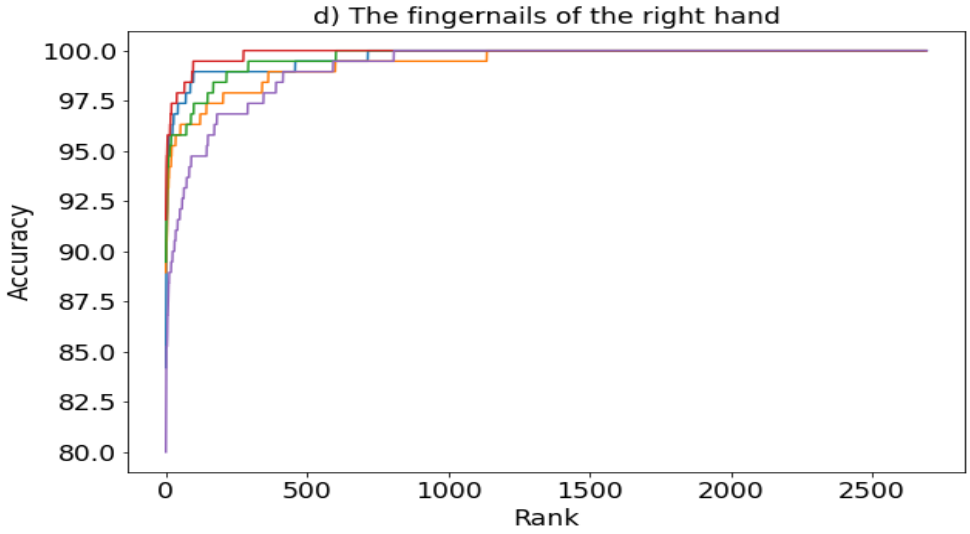

Figure 3.12: The CMC of the proposed PIFK d) fingernails of the right hands from the '11k Hands' dataset.

In figure 3.9, 3.10, 3.11, and 3.12 the CMC of the major knuckles of the left (chart a) and right (chart b) hands and fingernails of the left (chart c) and right (chart d)

Table 3.1: The rank-1 recognition rate (shown in %) using DenseNet201 as feature extractor and BC similarity metric for the '11k Hands' and 'PolyU HD' datasets

Region	Finger	11k Hands-L	11k Hands-R	PolyU-R
Fingernail	Thumb	87.83	84.21	93.81
	Index	89.42	88.95	90.40
	Middle	90.48	89.47	87.65
	Ring	93.65	91.58	85.10
	Little	84.13	80.00	87.30
Minor Knuckle	Index	84.66	76.84	72.47
DIP	Middle	85.19	82.11	68.92
	Ring	84.13	77.89	71.46
	Little	81.48	76.84	78.43
Major Knuckle	Thumb	85.71	84.21	76.83
PIP	Index	82.45	83.16	79.71
	Middle	85.11	82.54	76.64
	Ring	80.95	83.68	83.23
	Little	83.07	75.79	82.06
Base Knuckle	Thumb	87.30	85.26	58.35
MCP	Index	78.84	81.58	64.34
	Middle	80.95	77.37	67.27
	Ring	80.42	77.37	66.47
	Little	84.13	80.53	67.94

hands from the '11k Hands' dataset are shown, respectively.

Interestingly, as indicated by the results (see table 3.1, figure 3.9, 3.10, 3.11, and 3.12), the left hand—including the fingernails and knuckles from '11K Hands'- is more identifiable than the right hand for the majority of the fingers.

It can be also observed that the recognition using fingernail sub-images from both

hands and both datasets is higher than that of all knuckle sub-images, which can clearly be seen in table 3.1 for both data sets and visualized in figures 3.13 and 3.14 for the '11K Hands' dataset.

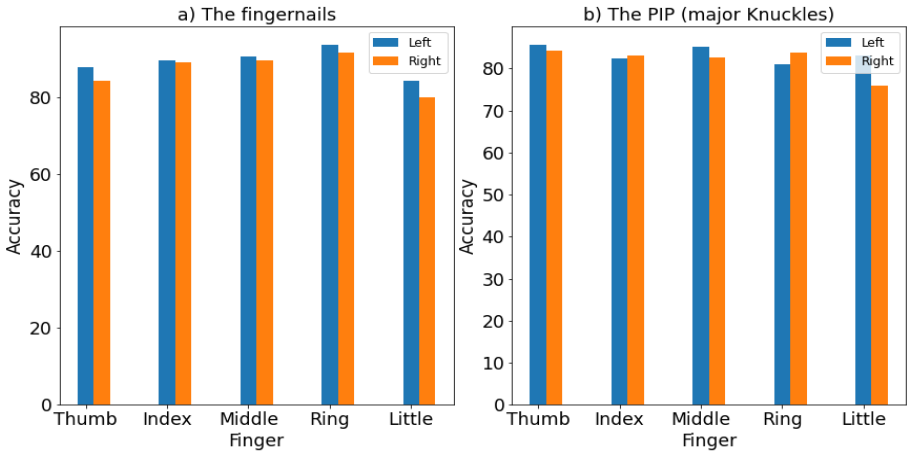

Figure 3.13: The recognition accuracy for the left and right hands from '11K Hands' dataset based on fingernails (left) and the major knuckles, PIP (right).

In particular, the best performance results were achieved on the left fingernails, with rank-1 accuracies of 93.65% on the ring finger, 90.48% on the middle, and 89.42% on the index from the '11K Hands'.

Among the right fingernails, the highest performance -as demonstrated by table 3.1- was achieved by using the ring, middle, and index, with accuracies of 91.58%, 89.47%, and 88.95%, respectively.

The best results for the 'PolyU HD' dataset were also achieved for the fingernails with rank-1 accuracies 93.81% and 90.40% for the thumb and the index finger, respectively. These results may be linked to the fact that the fingernails have a more rigid and defined shape (Thapar, Jaswal, and Nigam, 2019) than the knuckle creases.

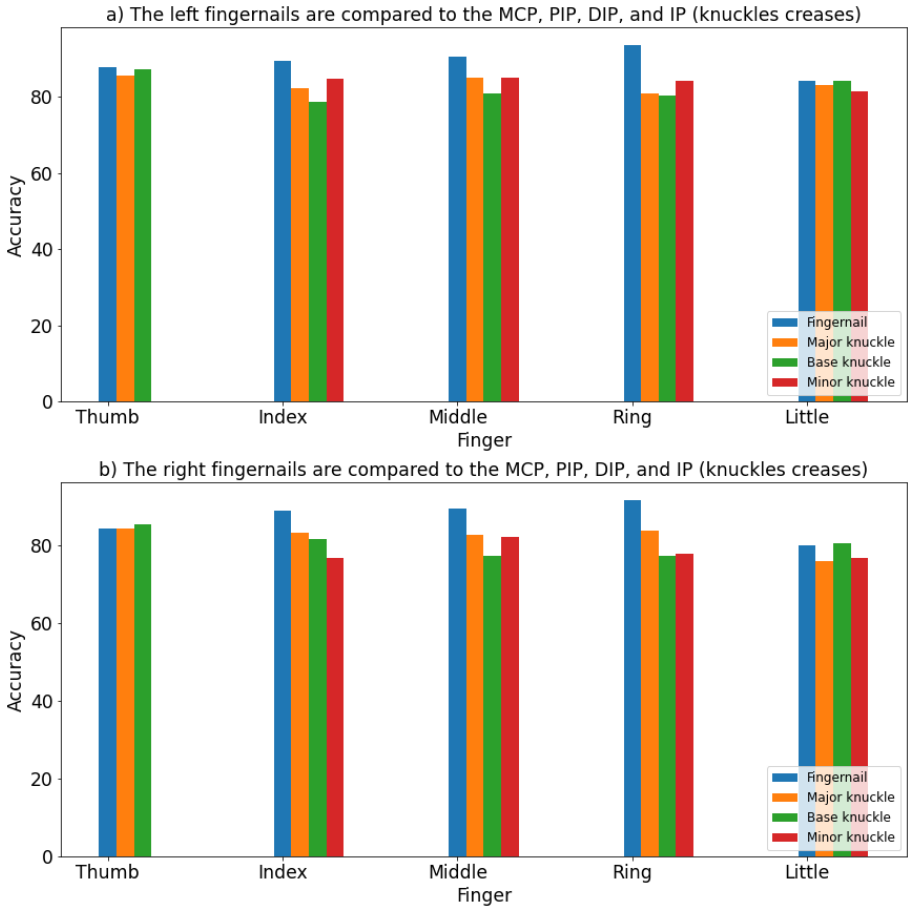

Figure 3.14: The performance of fingernail regions is compared to the knuckle regions for: a) the left, and b) the right hands in the '11k Hands'.

In addition, it is possible to observe that the MCP are relatively insignificant compared to the other components of the hand. The PIP perform slightly better than the DIP, particularly on the thumb and middle fingers of the left hands, with rank-1 accuracies 85.71% and 85.11%, respectively for the '11K hands' dataset. Among the right-hand parts, the IP and the PIP obtain better results, with rank-1 accuracies 84.21% and 83.68% respectively for the same dataset. This observation is also valid for the 'PolyU HD' dataset with rank-1 accuracies 83.23% and 82.06% for the ring and little fingers, respectively. This may refer to the fact that the PIP patterns contain more discriminative creases, which might increase the identification performance over the DIP patterns.

The results also revealed that the IP and MCP of both thumbs achieved higher results than the other fingers from the '11k Hands' dataset; the IP (major knuckles) of the left and right thumb obtained rank-1 accuracies 85.71% and 84.21%, respectively,

while the MCP of the left and right thumb obtained rank-1 accuracies 87.30% and 85.26%, respectively.

In this study, for the first time, 19 components of human hands were studied at the same time and the results are shown in table 3.1. The comparison of the results with existing state-of-the-art approaches is complicated by the fact that there are no publications that use the full variety of components. However, the approach was validated on two different well-known datasets.

Furthermore, a general comparison was made with another state-of-the-art approach, namely (Thapar, Jaswal, and Nigam, 2019) which only used the PolyU FKI (Ajay Kumar, 2014) and PolyU FKP (Ajay Kumar and Ravikanth, 2009) databases. In this study the PolyU FKI (Ajay Kumar, 2014) database was used to segment 15 components, namely the fingernails, PIP (major knuckles), and DIP (minor knuckles) from finger images. However, the study did not mention on which finger these components where located. Furthermore, the '11k Hands' dataset contains the hand images of both hands and is larger than the PolyU FKI database. The PolyU FKP (Ajay Kumar and Ravikanth, 2009) dataset only contains the right and left PIP and it is not clear from which finger or hand these images were captured.

The performance using the FKIMNet method (Thapar, Jaswal, and Nigam, 2019) achieved a rank-1 score 94.83% for the fingernail, 90.52% for the PIP and 88.73% for the DIP using the PolyU FKI dataset. These results are similar in magnitude to the results that were reported. However, these results are on a larger and more complex dataset and clearly indicate where the keypoints are located.

Furthermore, Interestingly observations were confirmed on two different datasets that fingernails provide better results than other hand components. In addition, it can be also observed that the left hands achieved higher results than the right hands.

3.4 The person identification based on fingernails and knuckles (PIFK$^+$)

The PIFK approach was further investigated by improving the feature extraction phase with the fine-tuning. An overview architectural diagram of the developed PIFK$^+$ framework is presented in figure 3.15. Figure 3.17, and 3.5 display the stages of the framework in more details. The PIFK$^+$ is expanded from the recently published method in (Mona Alghamdi, P. Angelov, and B. Williams, 2021). The first phase was concerning the detection, localisation, mapping, segmentation, and normalisation as explained in section 3.2.1.The second phase was the essential part of the developed framework. This phase involved feature extraction using fine-tuning or sometimes called transfer learning of the Densenet201 model. Fine-tuning is a significant difference between PIFK and PIFK$^+$. Fine-tuning can involve two various

tasks. The first task was the base model of, e.g. the Densenet201 learned from the popular dataset "ImageNet". The second task was to learn the identification of persons from a subset of hand segments for each component. In this experiment, fine-tuning was conducted per component per dataset.

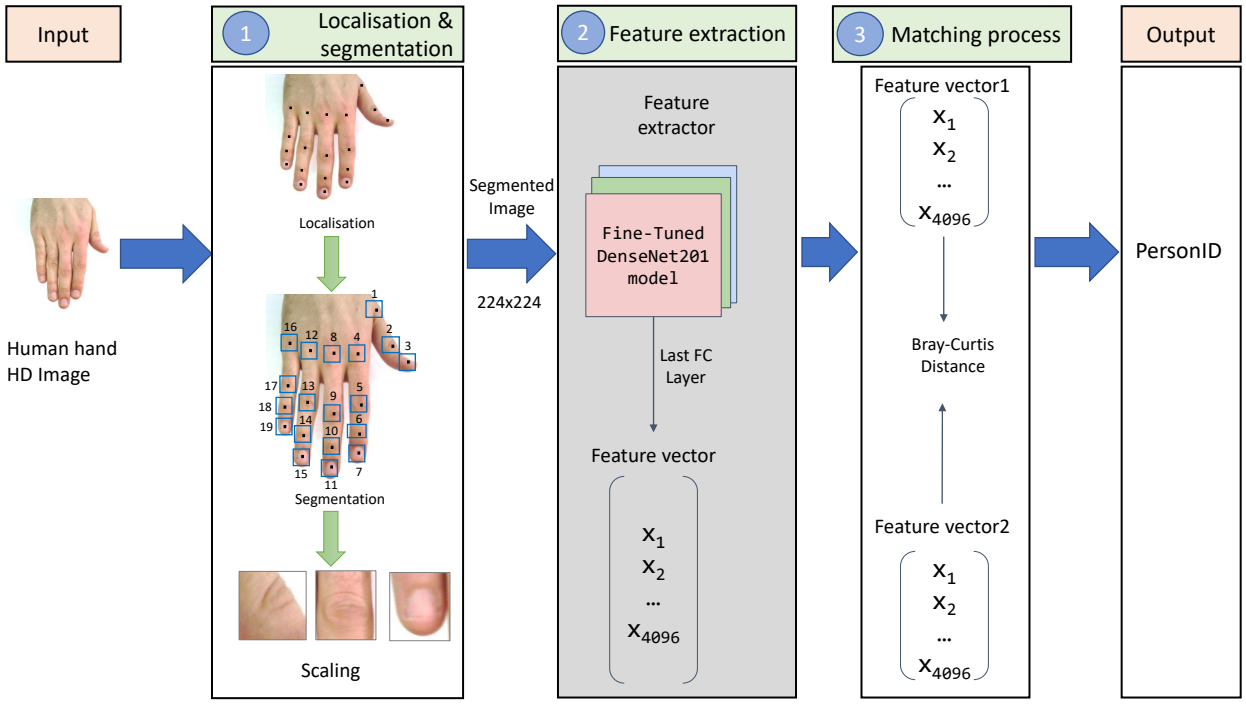

Figure 3.15: An overview schematic diagram of the developed framework for person identification based on dorsal fingernails and knuckle patterns (PIFK$^+$)

3.4.1 Feature extraction by fine-tuning DenseNet201

Traditional image processing methods have been popular in many computer vision applications for decades. However, these methods often require extensive human intervention (Nanni, Ghidoni, and Brahnam, 2017) in the design of the feature extractor. This work uses these abstract high-level features extracted from each segmented part of the hand. In the beginning, many base models of CNN were tried to evaluate the best-performing model in the '11k Hands' dataset. It was found that the base model of DenseNet201 and the Bray-Curtis for feature extraction and similarity measuring achieved excellent results in different parts of the hands.

Figure 3.16 presents the most base (pretrained) models with high performance in different sub-images from the right and left hands in the '11k Hands' dataset. Fine-tuning parts of the segmented images per ROI was then attempted from the '11k

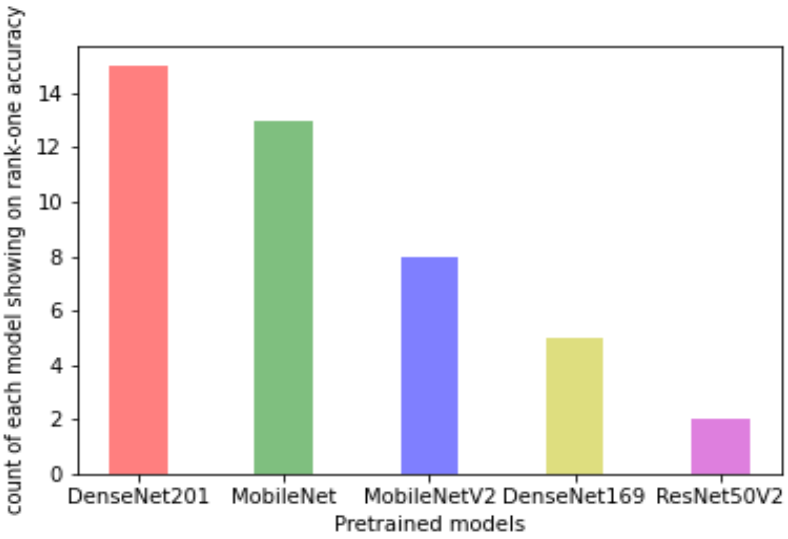

Figure 3.16: Count of each model showing on rank-1 accuracy for fingernails and knuckles-print regions the '11k Hands' dataset.

Hands' and 'PolyU' datasets using the DenseNet201 model as feature extractors.

The reason for choosing to fine-tune the DenseNet201 model among various DLNNs, such as VGG16, ResNet50V2, MobileNet, and others. Because this model showed high performance for extracting high-level and abstract features (Mona Alghamdi, P. Angelov, and B. Williams, 2021; Vyas, Rahmani, et al., 2021). Regarding the fine-tuning, the first 700 layers were frozen from the original model of DenseNet201. 2D global average pooling was dded , batch normalisation with 0.90 momentum, dropouts equal to 0.5, dense with 4096 vectors, a Relu activation function, 0.6 dropouts, batch normalisation with 0.9 momentum, and a classification layer with 170 dimensions and softmax activation function were added. The network was trained with 150 epochs. A stochastic gradient descent was used as the learning optimisation, the learning rate was 0.001, with a Nesterov momentum of 0.9, and the loss function was categorical cross-entropy as illustrated in figure 3.17. Finally, the same network was retrained with all its layers and settings with 100 epochs.

Several types of augmentations were applied to the training samples before fine-tuning to reduce the problem of small training data. The augmentation includes rescaling the pixels from 0 to 1, horizontal flipping, a randomised rotation range of 30, a randomised width and height shift range value of 30, and a randomised zoom in the range of 0.9 to 1.1. The training data was enriched with these augmentations, which enhanced model learning.

After the feature extraction stage, the matching process was considered to find

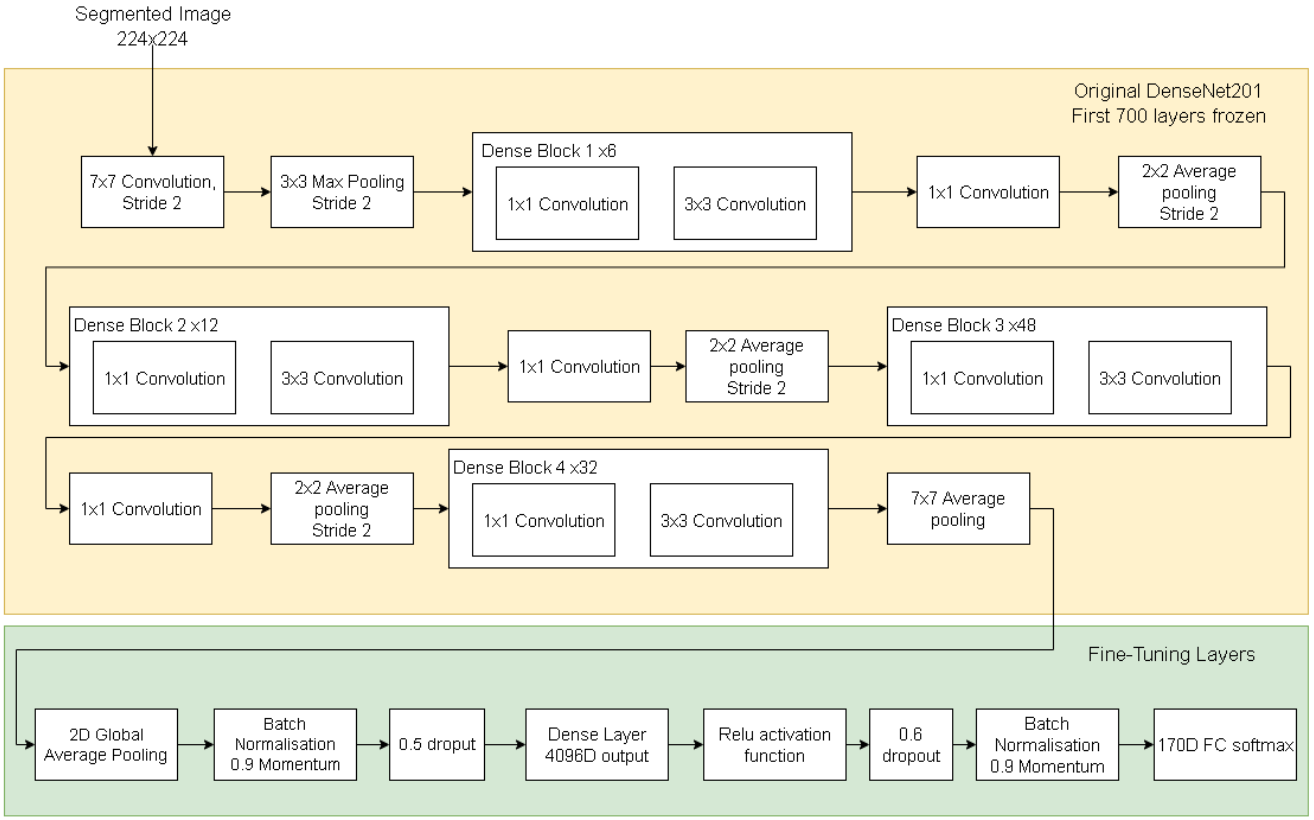

Figure 3.17: A demonstrative diagram of the fine-tuning of the DenseNet201 layers for feature extractor (G. Huang et al., 2017; Mona Alghamdi, P. Angelov, and Alvaro, 2022).

the best match. This phase is illustrated in the above section 3.3.4.

3.5 Experimental results and evaluation of PIFK$^+$

The experiment that was performed aimed to identify a person from a group of individuals based on the input of hand images. This section will evaluate the proposed framework PIFK$^+$ on the 11k Hands and PolyU datasets.

3.5.1 Datasets description

There are many benchmarks for hand images, such as CASIA (Z. Sun et al., 2005), Bosphorus (Yoruk et al., 2006), IITD (Yoruk et al., 2006), GPDS150hand (Ferrer

et al., 2007), however they consist of only the palm side of the hand. The 11k Hands and PolyU benchmark contains RGB images of the dorsal hands as explained in section 2.1. The number of the right-hand segments is $54,404$, and of the left hand is $52,369$. 47,719 and 46,890 segmented images were used from the left and right 11k Hands dataset, respectively, for the training and validation of the fine-tuned model. The dataset was also divided into 70% and 30% in training and validation. The testing part included different sets of sub-images. They were 380 sub-images in the query, and around $4,517$ sub-images in the gallery for the left-hands. Also, 380 sub-images in the query and around $4,528$ sub-images in the gallery for the right-hands.

The Hong Kong Polytechnic University Contactless Hand Dorsal Images Database (Ajay Kumar and Zhihuan, 2016) was also considered in this experiment as described in section 2.1. However, this dataset has only the right hands, which limits its use of the dataset. The training and validation total sub-images were $75,312$, and the data was divided into 70% and 30% in training and validation, respectively. The total testing included 916 and $6,426$ segmented images in the query and gallery sets, respectively.

3.5.2 Pre-processing phase

As shown in figure 3.15, the pre-processing phase was explained in subsection 3.2.1 and the outputs of this stage is the same as ones in the section 3.3.2. Therefore, the details of PIFK$^+$ about this phase will be covered in this subsection.

Step 1: *Detection and recognition of the knuckle creases and fingernails:* The detection of the knuckle creases and fingernails is previously described in subsection 3.2.1 and produces the same mentioned outputs.

Step 2: *Resizing and re-scaling:* The segmented images were also resized into 224×224 pixels and re-scaled. The resizing was important to be compatible with the feature extractor model as explained in section 3.3.2.

Step 3: *The query and library data sets:* The data was structured into the query and library sets in this step. An evaluation of the method was considered using Leave-One-Out Cross-Validation (LOOCV) to evaluate the proposed algorithm's performance. The LOOCV had only one fold. Furthermore, because the samples were independent, each prediction in the identification problem was independent of the others (Lin Zhang and H. Li, 2012). As a result, only one sub-image of a subject per ROI in the query was considered. As a result, the query set had 20 sub-images of 20 different subjects per component. The rest sub-images were 239 of the same subjects per component in the library. There were also 19 ROI per individual of the left and right hand, respectively.

87 images of left hands and 6 right hands were disregarded in the 11k Hands dataset due to inaccurate keypoint detection. Following segmentation due to erroneous

detection, 428 sub-images of left-hand segments and 573 of right-hand segments were eliminated. The total number of eliminated sub-images in the PolyU collection was 5,733 out of 88,392. Disregarded images and sub-images were not included in the framework evaluation, which has been considered a limitation of the study.

3.5.3 DL for feature extractor

For visual object recognition, deep learning has become the dominant machine learning approach (G. Huang et al., 2017). In this study, the recognition performance was evaluated using two different steps. In the first phase, many base models of CNN were investigated that pretrained on the popular database' ImageNet' to extract the features from sub-images. The architecture is varied among these pretrained models. However, they shared a significant and complicated structure that significantly impacted the model performance (G. Huang et al., 2017). Studying many models is to identify the most sophisticated ones in extracting discriminative features. This step is crucial and leads to better individual identification. Interestingly, the best performing base models in terms of rank-1 recognition accuracy, from highest to lowest, are DenseNet201 (G. Huang et al., 2017), MobileNet (Howard et al., 2017), MobileNetV2 (K. Sid et al., 2017), DenseNet169, and ResNet50V2 (He et al., 2016). Figure 3.18 and table 3.2 display the best base models in rank-1 recognition rate from the right and left hands in the '11k Hands' dataset.

Table 3.2: The rank-1 recognition accuracy of 11k-hands database (shown in %) using different pretrained models and similarity distances.

Side	Region	Finger	Rank-1	Distance	model
Left	Base Knuckle	Thumb	87.83	Bray-Curtis	DenseNet201
		Index	82.01	Bray-Curtis	ResNet50V2
		Middle	80.95	Cityblock	DenseNet201
		Ring	98.94	Cosine	DenseNet201
		Little	99.47	Cosine	MobileNetV2
	Major Knuckle	Thumb	86.77	Cosine	MobileNet
		Index	85.63	Cosine	MobileNet
		Middle	86.70	Bray-Curtis	MobileNetV2
		Ring	99.47	Cosine	MobileNet
		Little	100	Bray-Curtis	DenseNet169
	Minor Knuckle	Index	84.12	Cosine	MobileNet

		Middle	86.77	Bray-Curtis	DenseNet201
		Ring	99.47	Bray-Curtis	DenseNet201
		Little	99.47	Cityblock	DenseNet201
	Fingernail	Thumb	87.83	Bray-Curtis	DenseNet169
		Index	91.00	Cosine	MobileNet
		Middle	100	Bray-Curtis	DenseNet201
		Ring	99.47	Bray-Curtis	DenseNet201
		Little	89.00	Cosine	MobileNet
Right	Base Knuckle	Thumb	85.26	Bray-Curtis	DenseNet201
		Index	81.57	Bray-Curtis	DenseNet201
		Middle	80.00	Bray-Curtis	MobileNetV2
		Ring	80.00	Bray-Curtis	MobileNetV2
		Little	83.15	Bray-Curtis	DenseNet169
	Major Knuckle	Thumb	84.21	Bray-Curtis	DenseNet201
		Index	83.15	Bray-Curtis	DenseNet201
		Middle	85.18	Cosine	MobileNet
		Ring	83.68	Bray-Curtis	DenseNet201
		Little	80.00	Bray-Curtis	MobileNetV2
	Minor Knuckle	Index	80.00	Bray-Curtis	MobileNetV2
		Middle	85.18	Bray-Curtis	ResNet50V2
		Ring	83.15	Bray-Curtis	DenseNet169
		Little	80.00	Bray-Curtis	MobileNetV2
	Fingernail	Thumb	85.78	Cosine	MobileNet
		Index	90.00	Bray-Curtis	MobileNet
		Middle	91.05	Bray-Curtis	MobileNet
		Ring	93.68	Bray-Curtis	MobileNet
		Little	84.73	Cosine	MobileNet

This work evaluated the results of the first step in the experiment using the rank-1 recognition accuracy as displayed in the table 3.2 w.r.t. the '11k-hands' benchmark. This study considered several distance metrics and pretrained models to identify the best-performing ones. Table 3.2 demonstrates the best distances and models performing on the base, major, minor knuckles, and fingernails of five fingers on the left and right hands of the '11k-hands' database. The results of many experiments,

as shown in table 3.2, indicate that the distance of Bray-Curtis obtained the highest identification performance compared to other distances. Surprisingly, the results, as shown in table 3.2, indicate that the left components of the hand are more identifiable than the right ones. It can be also observed that the identifiability of fingernail patches from both hand sides is relatively higher than different knuckle patches. This result may refer to the nail being consistent, having a unique shape, and the proposed approach performing exceptionally well.

Figure 3.16 refers to the best models that appear on rank-1 accuracy among all regions of both hands. From the highest to the lowest, it can be observed that the best models are DenseNet201, MobileNet, MobileNetV2, DenseNet169, and ResNet50V2. In addition, many distance metrics are studied to examine their impact on the matching result. Indeed, many distances appeared on rank-1 accuracy; however, the most repeated distance between different regions were showed. Bray-Curtis, cosine, and city-block were the most achieved distances, from the highest to the lowest on rank-1.

In the second phase, fine-tuning the DenseNet201 model was considered for extracting features. Figure 3.3 displays the structure of the retrained DenseNet201 model. This model was considered because it showed the best performance, as demonstrated by the phase one results. The base model of DenseNet201 (G. Huang et al., 2017) was originally pretrained on the well-known ImageNet dataset. This model has many advantages: DenseNet201 naturally scales to hundreds of layers, providing no optimization challenges; DenseNet201 requires fewer parameters and processing time.

High performances were achieved in training and validation using the fine-tuned DenseNet201 in subsets of each ROI from 11k Hands and PolyU. The last fully connected (FC) layer was utilized from the fine-tuned model to extract the features per hand component. Finally, pairs of the query and library feature vectors were generated from the last FC layer to calculate the similarity using Bray-Curtis.

3.5.4 Similarity estimation and matching

Estimating the similarity between two sub-images of the same hand is a part of the matching process. In the matching process, one segmented image from the query can match with one or more corresponding segmented image(s) in the library. The similarity based on the feature vectors was estimated and gained from the last FC layers of the fine-tuned DenseNet201 model as described in subsections 3.4.1, and 3.3.4. The Bray-Curtis similarity metric was then utilised for the matching process. As a result, the proposed PIFK^{+} framework gained a high performance compared to the state-of-the-art methods. Table 3.3 shows proposed framework's rank-1 recognition rate and standard deviation (SD) and a comparison of the proposed

method and other state-of-the-art methods.

Table 3.3: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k-L	11k-R	PolyU
Fingernail	Thumb	The proposed method (Rank-1)	100	95.00	100
		The proposed method (SD)	0	0.22	0
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	87.83	84.21	93.81
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Index	The proposed method (Rank-1)	100	100	95.83
		The proposed method (SD)	0	0	0.2
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	89.42	88.95	90.40
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Middle	The proposed method (Rank-1)	100	100	91.83
		The proposed method (SD)	0	0	0.27
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	90.48	89.47	87.65
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Ring	The proposed method (Rank-1)	100	100	93.88
		The proposed method (SD)	0	0	0.24
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	93.65	91.58	85.10
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Little	The proposed method (Rank-1)	100	100	95.83
		The proposed method (SD)	0	0	0.2
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	84.13	80.00	87.30

Table 3.3: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k-L	11k-R	PolyU
		(Vyas, Rahmani, et al., 2021)	-	-	-
Minor Knuckle	Index	The proposed method (Rank-1)	100	95.00	91.67
DIP		The proposed method (SD)	0	0.22	0.27
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	84.66	76.84	72.47
		(Vyas, Rahmani, et al., 2021)	86.35	89.56	74.81
	Middle	The proposed method (Rank-1)	100	95.00	85.71
		The proposed method (SD)	0	0.22	0.35
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	85.19	82.11	68.92
		(Vyas, Rahmani, et al., 2021)	93.69	93.17	84.57
	Ring	The proposed method (Rank-1)	100	95.00	91.84
		The proposed method (SD)	0	0.22	0.27
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	84.13	77.89	71.46
		(Vyas, Rahmani, et al., 2021)	91.45	89.56	80.80
	Little	The proposed method (Rank-1)	100	95.00	91.67
		The proposed method (SD)	0	0	0.2
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	81.48	76.84	78.43
		(Vyas, Rahmani, et al., 2021)	83.91	80.83	73.25
Major Knuckle	Thumb	The proposed method (Rank-1)	95.00	95.00	91.11
PIP		The proposed method (SD)	0.22	0.22	0.28

Table 3.3: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k-L	11k-R	PolyU
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	85.71	84.21	76.83
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Index	The proposed method (Rank-1)	100	95.00	97.87
		The proposed method (SD)	0	0.22	0.14
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	82.45	83.16	79.71
		(Vyas, Rahmani, et al., 2021)	93.28	93.93	91.23
	Middle	The proposed method (Rank-1)	100	95.00	93.75
		The proposed method (SD)	0	0.22	0.24
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	85.11	82.54	76.64
		(Vyas, Rahmani, et al., 2021)	94.70	95.26	88.90
	Ring	The proposed method (Rank-1)	90.00	95.00	91.49
		The proposed method (SD)	0.30	0.22	0.28
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	80.95	83.68	83.23
		(Vyas, Rahmani, et al., 2021)	94.30	92.03	90.34
	Little	The proposed method (Rank-1)	95.00	100	97.87
		The proposed method (SD)	0.22	0	0.14
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	83.07	75.79	82.06
		(Vyas, Rahmani, et al., 2021)	88.59	86.34	88.57
Base Knuckle	Thumb	The proposed method (Rank-1)	100	100	79.59
MCP		The proposed method (SD)	0	0	0.40

Table 3.3: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k-L	11k-R	PolyU
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	87.30	85.26	58.35
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Index	The proposed method (Rank-1)	100	90.00	95.92
		The proposed method (SD)	0	0.3	0.19
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	78.84	81.58	64.34
		(Vyas, Rahmani, et al., 2021)	78.82	84.82	62.38
	Middle	The proposed method (Rank-1)	100	95.00	97.96
		The proposed method (SD)	0	0.22	0.14
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	80.95	77.37	67.27
		(Vyas, Rahmani, et al., 2021)	84.32	85.96	61.49
	Ring	The proposed method (Rank-1)	100	80.00	97.96
		The proposed method (SD)	0	0.22	0.14
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	80.42	77.37	66.47
		(Vyas, Rahmani, et al., 2021)	76.99	73.62	57.16
	Little	The proposed method (Rank-1)	100	100	97.96
		The proposed method (SD)	0	0	0.14
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	84.13	80.53	67.94
		(Vyas, Rahmani, et al., 2021)	83.71	81.78	59.16

The CMC of the knuckles and fingernails of the left (chart a-d) and right (chart

e-h) hands in the 11k Hands are illustrated in figure 3.18. The left hand—including the fingernails and knuckles from 11K Hands—is more recognisable than the right hand for most of the fingers, as shown by the results (see table 3.3 and figure 3.18). This result also complies with the study shown in (Mona Alghamdi, P. Angelov, and B. Williams, 2021). It can be also observed that fingernail sub-images from both left and right hands from the 11k Hands and PolyU datasets achieved better results than the recognition using all knuckle sub-images. When the SD is close to 0, then the rank-1 is the highest value and is more than 0, then the rank-1 is lower than 100. These results are shown in table 3.3 and the CMC chart 3.18.

In general, the proposed framework PIFK$^+$ outperformed the state-of-the-art methods (Mona Alghamdi, P. Angelov, and B. Williams, 2021; Vyas, Rahmani, et al., 2021; Thapar, Jaswal, and Nigam, 2019). It can be observed that the best results for both hands in almost all fingernails. The fingernails rank-1 accuracies were 100% in the 11k Hands dataset and 100% on the thumb and ring finger in the PolyU dataset. Overall, the result of this study can show that the DIP/minor knuckles got minimal performance compared to the other hand components in all datasets. However, the PIP components' performance was slightly better than the DIP components in both datasets. The MCP components were also higher than the PIP and DIP in the two datasets. The PIP obtained a rank-1 accuracy of 100% in the index and middle fingers in the left hands and right little finger of the 11k Hands dataset and 97.87% in the index and little fingers in the PolyU dataset, compared to the middle finger that achieved 94.70% and 85.11% in the left-hands of the 11k Hands dataset in (Vyas, Rahmani, et al., 2021), and (Mona Alghamdi, P. Angelov, and B. Williams, 2021), respectively.

The DIP component in each finger of the left 11k Hands dataset achieved a rank-1 accuracy of 100%, whereas the one in the ring finger obtained 91.84% in the PolyU dataset. This observation is also valid among the right-hand parts of the PolyU dataset. The PIP outperforms the DIP component, with rank-1 accuracies of 97.87% for the index and little fingers. For the MCP/base knuckle component, 100% was gained in most fingers in the 11k Hands dataset and 97.96% in the middle, ring, and little fingers from the PolyU datasets in comparison to (Mona Alghamdi, P. Angelov, and B. Williams, 2021), and (Vyas, Rahmani, et al., 2021) that got lower results using the same component as displayed in table 3.3.

Although there is a lack of studies similar to this approach, this study was compared with the previous most similar ones like (Mona Alghamdi, P. Angelov, and B. Williams, 2021; Thapar, Jaswal, and Nigam, 2019; Vyas, Rahmani, et al., 2021). In addition, the fingernails results were analysed with the method in this study (Thapar, Jaswal, and Nigam, 2019). The proposed framework PIFK$^+$ was achieved excellent performance for most fingers, with a rank-1 accuracy of 100%. However, this method (Thapar, Jaswal, and Nigam, 2019), which scored 94.83% in

the fingernail identification, did not specify which hand (right or left) or finger the component locate. The best performance was 98.60% using the fusion of the right hands in the major, minor, fingernail, and finger components in the (Thapar, Jaswal, and Nigam, 2019) method using the PolyU dataset. However, in this study, using only one component, such as the fingernail of the thumb finger, a higher result with a rank-1 score of 100% was achieved . The study in (Vyas, Rahmani, et al., 2021) gained 100% in the fusion of all components and fingers of a hand. However, a 100% score was achieved in most components of the fingernail in the index, middle, ring, and little finger of the 11kHands dataset and the thumb finger of the left hands in the 11k Hands and PolyU datasets.

3.6 Conclusion

PIFK$^+$ is a comprehensive framework developed from the older version of PIFK to identify a person automatically for forensic and security applications presented in this study, where the hand image plays a significant role in identifying a person. The framework uses the dorsal surface of the five human hand components (fingernails, MCP, PIP, DIP, and IP joints of the five fingers). The proposed approach PIFK$^+$ utilises all 19 hand components from two popular datasets: the 11k Hands and PolyU datasets. The study investigates many approaches, which are the base model of many DLNNs and fine-tuning of DenseNet201 in segmented subsets per component and applies the Bray-Curtis, the best similarity metric among many similarities to find the corresponding pairs' similarities. In addition, the framework automatically labels and generates bounding boxes around the detected hand components. The proposed approach was evaluated on these two most common datasets, and the presented results outperform the state-of-the-art methods. Furthermore, intriguing facts about the hands and their components were discovered. Such as the continuous difference in outcomes between the left and right hands, favouring the left. In addition, the fingernails were found to be the best-performing components in both datasets. Developing the framework using other fine-tuned CNN models for feature extraction may be considered for future studies. Furthermore, evaluating both the quality-segmented images and real-world poor-quality segmented images due to false ROI localisation in the recognition system may be relevant to future studies.

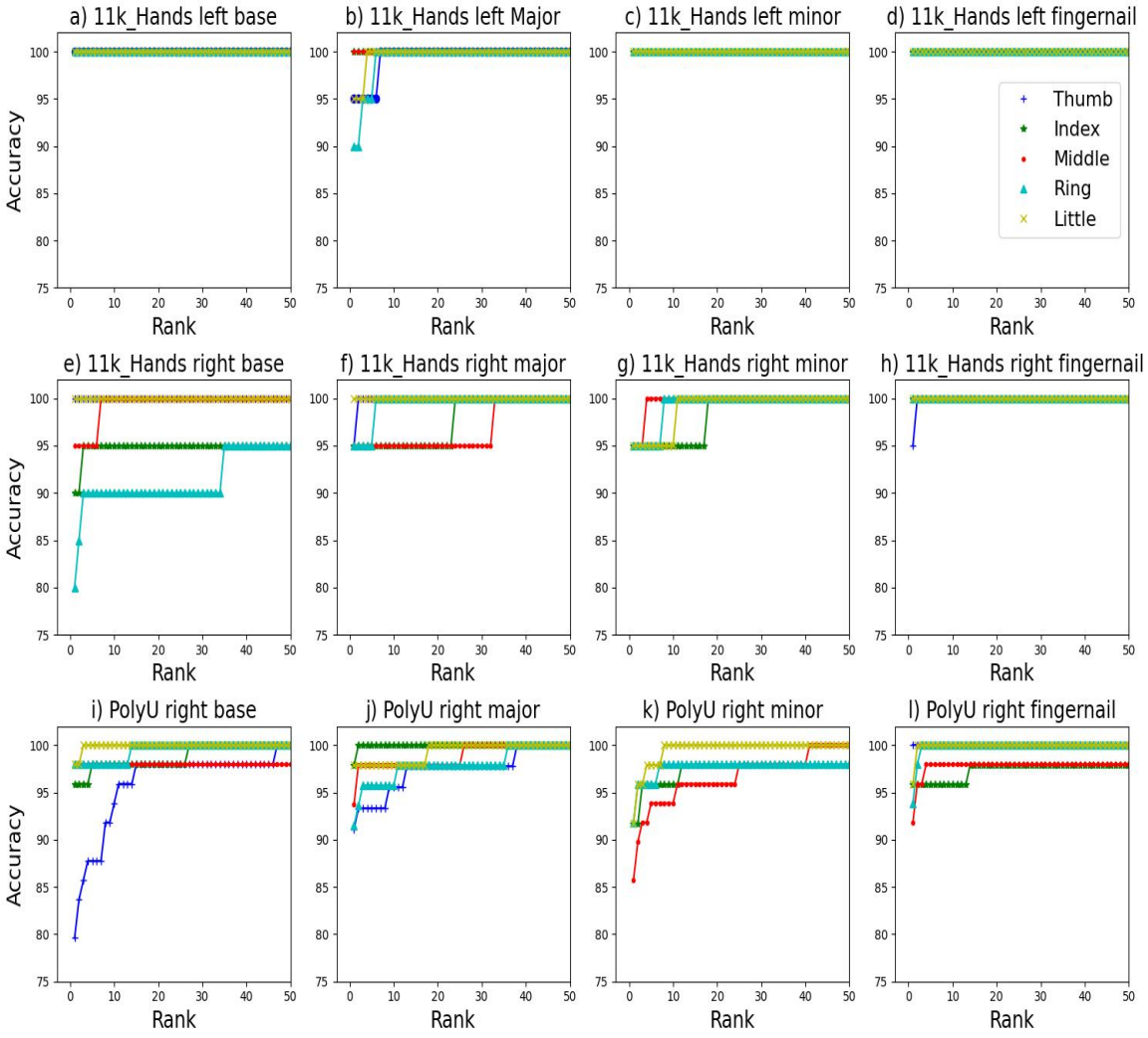

Figure 3.18: The CMC of the proposed PIFK^{+}; the a) base; b) major; c) minor knuckles; d) fingernails of the left hand; e) base; f) major; g) minor h) fingernails of the right hands from the 11k Hands dataset; i) base; j) major; k) minor l) fingernails of the right hands from the PolyU dataset.

Chapter 4

A Multi-modal Biometric Approach Based on Score-Level Fusion and Fine-tuning Deep Learning Features

4.1 Introduction

In this chapter, a multimodal biometric approach is proposed. The approach collects various features from different sub-images of human hands. These sub-images were acquired by employing the detection and segmentation method in the study (Mona Alghamdi, P. Angelov, and B. Williams, 2021). The sub-images include fingernails, and minor, major, and base knuckles, which are fused using two major fusion rules: the MV and WA. A literature review of the multi-biometric systems is in section 2.6. The remainder of this chapter is structured as follows: the novel multimodal biometric approach from fingernail and knuckle patterns is in section 4.2, the result and discussion are in section 4.3, and finally, the conclusion and recommendations for further work are in section 4.4.

4.2 The novel multimodal biometric approach

This section presents a multi-biometrics system that is efficient, based on the score-level fusion of the fingernails, minor, major and base of the dorsal finger knuckles from five fingers of the left and right hands. Figure 4.1 depicts an overview architecture of the suggested approach.

This approach was divided into four major phases: localization, detection,

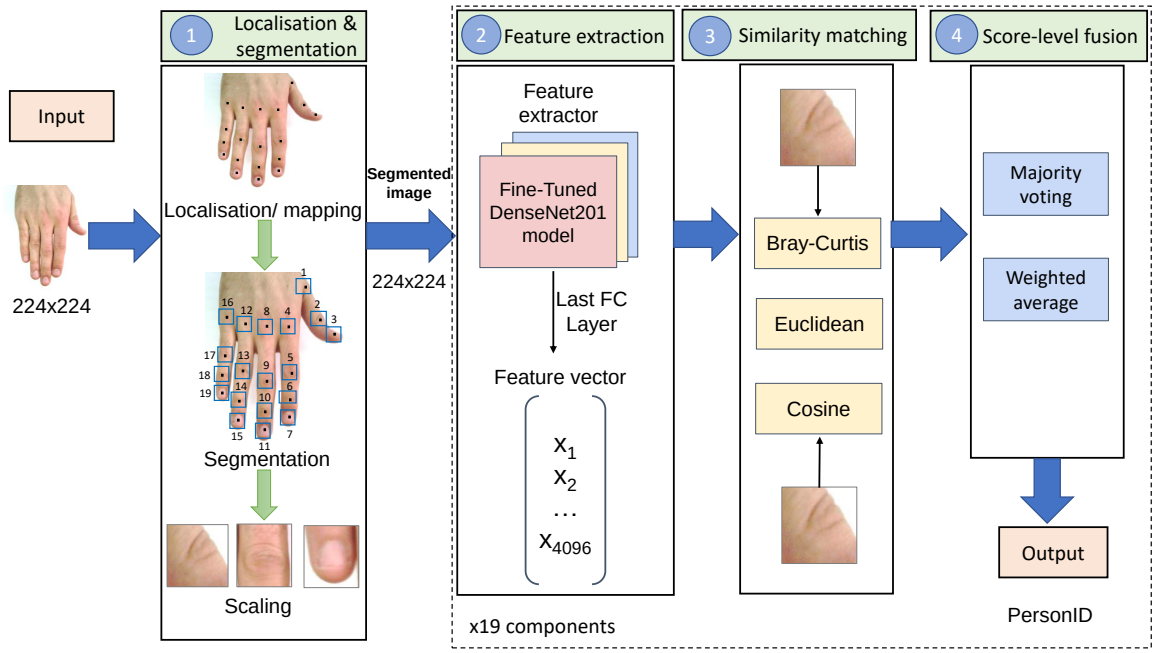

Figure 4.1: A schematic illustration of the proposed fusion-based approach for identifying people using dorsal fingernails and knuckle patterns

labelling, defining the region and segmentation; retraining of deep learning for feature extraction; employing similarity metrics to discover the best matching; and multi-modality score-level-based fusion. The pre-processing and feature extraction phase are previously explained in section 3.2.1 and 3.2.2.

4.2.1 Similarity metrics

The third stage was to distinguish between individuals based on similarity metrics, as displayed in figure 4.1. These metrics can be implemented using the extracted features from the segmented joints and are quite valuable in identifying individuals in multi-biometric systems. In this experiment, three metrics were employed, which are the Bray-Curtis, Cosine, and Euclidean distances. The definition of the Bray-Curtis and Cosine were illustrated in section 3.2.3. The similarity metric of Euclidean is the inverse of the Euclidean distance. Assume there are two segments, x and y. Their vectors are denoted by $x = (x_1, x_2, ..., x_p)$ and $y = (y_1, y_2, ..., y_p)$, respectively. The following section provides a brief description of the Euclidean similarity metrics:

Euclidean distance (E) is the most popular distance measure, which can measure the dissimilarity between two feature vectors of any pair of sub-images (J. Han, Kamber, and Pei, 2012). The Euclidean distance equation can be identified as follows

(5.1):

$$E(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (y_n - x_n)^2} \tag{4.1}$$

The inverse of the distance or dissimilarity between two feature vectors is the degree of similarity. The greater the degree of similarity (Sim), the shorter the distance (Dis) between two vectors, and vice versa, as indicated by:

$$Sim(x,y) = 1 - Dis(x,y) \tag{4.2}$$

4.2.2 The score-level fusion

"Ensemble" algorithms are widely used to optimise an overall performance in the multimodal paradigm (Tandel, A. Tiwari, and Kakde, 2021; J. Han, Kamber, and Pei, 2012). In this work, two ensemble rules were utilised: the MV and WV. These algorithms optimise the overall performance of several sub-models of DenseNet201 trained in various sub-images from different hand joints. The ensemble mechanism was applied in the similarity prediction of the sub-models. In the MV, the number of modalities from hand joints from each similarity metric should be odd to avoid a tie between the estimated subject (Tandel, A. Tiwari, and Kakde, 2021), otherwise the first vote has a major factor. In this work, 19 separate hand components and three similarity matches were used in the ensemble mechanism.

4.2.2.1 Majority voting (MV)

The MV was applied on various hand components and similarity metrics. The voting system was based on the models' estimated probability. The components representing the number of voters should be odd to avoid a tie between projected class labels (Tandel, A. Tiwari, and Kakde, 2021). As a result, an odd number of models (five) were employed. For example, five fingernails, five major knuckles or five base knuckles, except the minor knuckles, are four. In the latter case, when the votes are equal between voters, the first vote wins.

4.2.2.2 Weighted average (WA)

The WA (Lei Zhang et al., 2011) of matching data $(m_1, m_2, ..., m_n)$ with corresponding non-negative weights $(w_1, w_2, ..., w_n)$, where n indicates the total number of subjects, is formalised as:

$$WA = \frac{\sum_{p=1}^{n} w_i m_i}{\sum_{p=1}^{n} w_i} \tag{4.3}$$

which includes:

$$WA = \frac{w_1 m_1 + w_2 m_2 + ... + w_n m_n}{w_1 + w_2 + ... + w_n} \tag{4.4}$$

As a result, high-weighted samples contribute more to the WA than low-weighted data samples. Negative weights are not permitted.

When the weights are normalised, they add up to one $\sum_{n=1}^{n} w_i = 1$ (Tandel, A. Tiwari, and Kakde, 2021). The weight values were pre-defined based on the results of rank-1 for each component.

4.3 Experimental Results and Evaluation

This section presents the results of the proposed approach's evaluation on the datasets '11k Hands' and 'PolyU'. More details of these two datasets explained in section 2.1.

This study involved the identifiability of each finger in a holistic manner that integrates the performance of all fingernails and knuckles belonging to a particular finger. The fusion approaches were conducted in various regions and similarity metrics. This comprehensive computation is accomplished by score-level fusion, including the MV and WV. The proposed multimodal biometric approach was evaluated using both datasets, the '11k Hands', which was divided into '11k-Left' and '11k-Right', for the left and right hands, respectively, and the 'PolyU'. The performance was assessed and gained promising results using the rank-1 recognition rate as presented in table 4.1 (individual modalities), figures 4.8, 4.9, 4.10, and 4.11 (multimodal biometrics based on modality type), table 4.2 (fusion of the whole hand based on the similarity metrics), and the CMC chart in the figure 4.7, respectively. The preprocessing was conducted on the human hands, including resizing to 224×224, localisation of the ROI keypoints, mapping the keypoints to the image's original size, segmentation, and scaling as explained in subsection 3.2.1. In the second stage, the features were extracted by retraining the CNN model of DenseNet201, as explained in the subsection 3.4.1. Then, the dataset was divided into training, validation, and testing. The training, validation, and testing percentages were 70%, 20%, and 10%, respectively. The subject sub-images were included in all ROIs; therefore, the data was cleaned and discarded the subjects with missing data. The last FC layer with 4096 dimensions was then employed to extract abstract high-level features from 19 different modalities. The Leave-One-Out Cross-Validation (LOOCV) evaluation method in this study was used because the number of images was small. Each fold in the LOOCV has only one sample, and there is no random data partitioning into training and testing. Furthermore, because the samples are independent, each prediction in the identification problem is independent of the other (Wong, 2015). For each ROI, a single subject was considered in the query at a time (and vary this, averaging the

results at the end). The remaining individuals were in the library set (also replaced one at a time). Furthermore, there are individual images of both the left and right hand and 19 ROI per image. The LOOCV was employed with three different metrics: Bray-Curtis, Cosine, and Euclidean.

In general, single modalities achieved excellent results, and there were no significant differences in the performance between various metrics, as demonstrated in table 4.1.

At first glance, the cosine metric can demonstrate a reasonable recognition rate using various hand components from different datasets, including the left and right hands of the '11k Hands' and the right hands of the 'PolyU' datasets, as shown in figure 4.2.

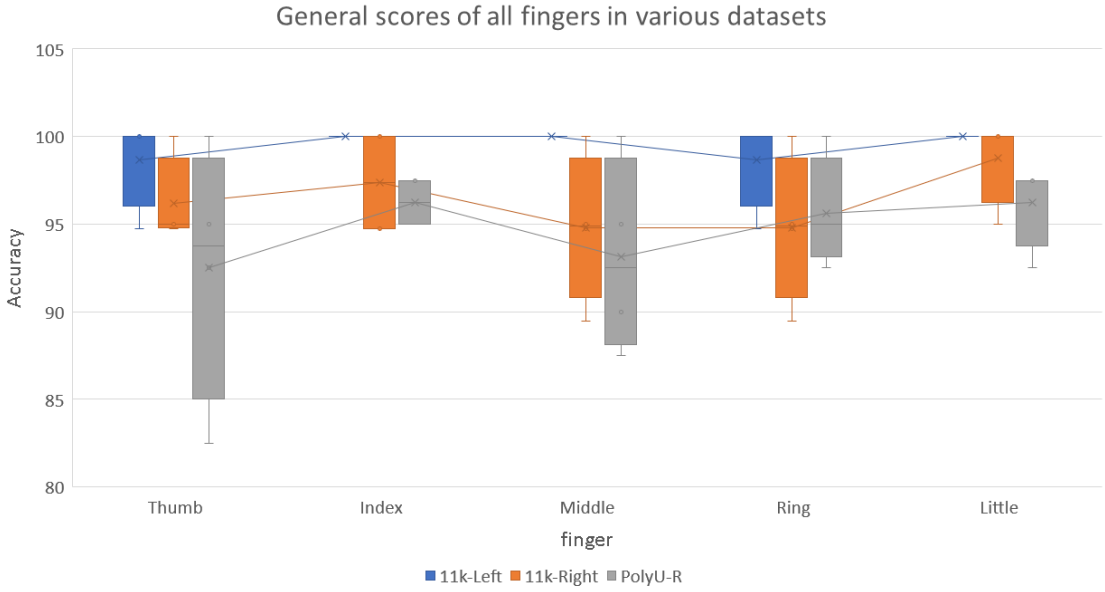

Figure 4.2: General scores of all fingers in various datasets using cosine metrics

The index, middle, and little fingers of the left hands of the '11k hands' dataset are in the top scores and the index and little fingers of the right hands in the same dataset achieved the best results. The index and ring fingers performed well in the 'PolyU' dataset. In general, the left hands performed better than the right hands. In general, the left hands performed better than the right hands. One reason could be that a human typically uses the right hand; thus, the left hand is less affected by natural circumstances. Another factor could be that the researcher started the experiment of components segmentation with the left hand and applied the same configurations to the right hand. Thus, the left hand obtained more accurate segmentation results,

which led to better identification results.

Table 4.1: The rank-1 recognition rate (shown in %) for individual modalities in the '11k Hands' and 'PolyU' datasets.

Region	**Finger**	**similarity metric**	**11k-Left**	**11k-Right**	**PolyU-R**
Fingernail	Thumb	Bray-Curtis	100	94.74	100
		Cosine	100	95	100
		Euclidean	100	95	100
	Index	Bray-Curtis	100	100	95
		Cosine	100	100	95
		Euclidean	100	100	95
	Middle	Bray-Curtis	100	100	90
		Cosine	100	100	90
		Euclidean	100	100	92.50
	Ring	Bray-Curtis	100	100	95
		Cosine	100	100	95
		Euclidean	100	100	95
	Little	Bray-Curtis	100	100	97.50
		Cosine	100	100	97.50
		Euclidean	100	100	97.50
Minor Knuckle	Index	Bray-Curtis	100	94.74	95
		Cosine	100	95	95
		Euclidean	100	94.74	97.50
DIP	Middle	Bray-Curtis	100	94.74	87.50
		Cosine	100	95	87.50
		Euclidean	100	94.74	87.50
	Ring	Bray-Curtis	100	100	95

		Cosine	100	95	95
		Euclidean	100	100	95
	Little	Bray-Curtis	100	94.74	92.50
		Cosine	100	95	92.50
		Euclidean	100	94.74	92.50
Major Knuckle	Thumb	Bray-Curtis	94.74	94.74	92.50
		Cosine	94.74	94.74	92.50
		Euclidean	100	94.74	92.50
PIP	Index	Bray-Curtis	100	94.74	97.50
		Cosine	100	94.74	97.50
		Euclidean	100	94.74	97.50
	Middle	Bray-Curtis	100	94.74	95
		Cosine	100	94.74	95
		Euclidean	100	94.74	97.50
	Ring	Bray-Curtis	94.74	94.74	92.50
		Cosine	94.74	94.74	92.50
		Euclidean	100	100	95
	Little	Bray-Curtis	100	100	97.50
		Cosine	100	100	97.50
		Euclidean	100	100	97.50
Base Knuckle	Thumb	Bray-Curtis	100	100	82.50
		Cosine	100	100	82.50
		Euclidean	100	100	82.50
MCP	Index	Bray-Curtis	100	94.74	97.50
		Cosine	100	94.74	97.50
		Euclidean	100	84.21	97.50
	Middle	Bray-Curtis	100	94.74	100

	Cosine	100	89.47	100
	Euclidean	100	94.74	100
Ring	Bray-Curtis	100	84.21	100
	Cosine	100	89.47	100
	Euclidean	100	89.47	95
Little	Bray-Curtis	100	100	97.50
	Cosine	100	100	97.50
	Euclidean	100	100	97.50

In the following subsections, a brief overview of person identification based on various hand regions using the cosine similarity metric will be given.

4.3.1 Fingernails

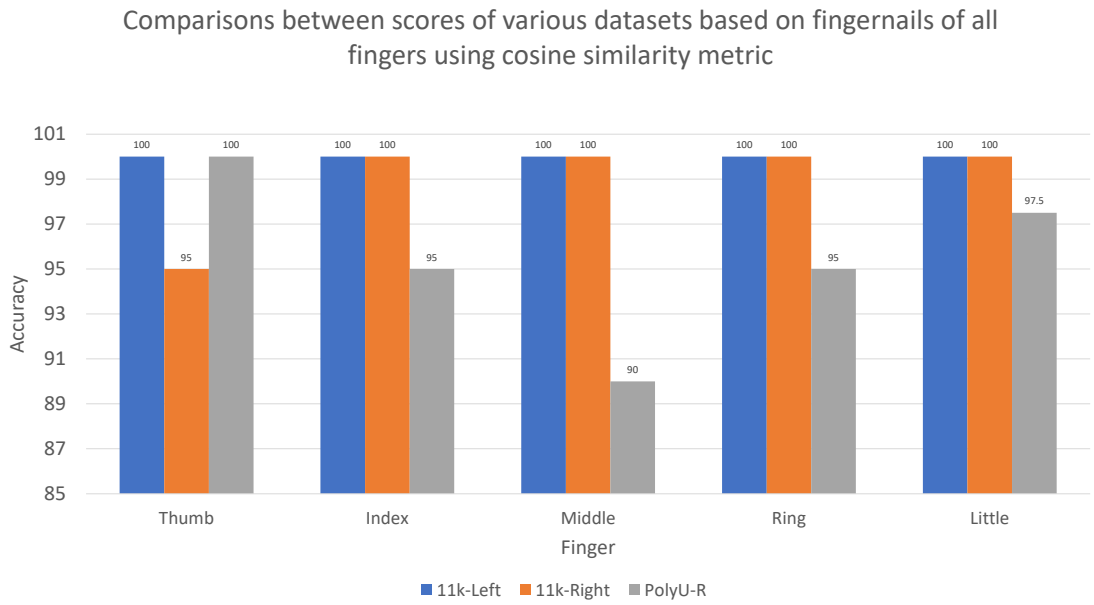

Figure 4.3: Recognition rate scores based on fingernails in all fingers of various datasets using cosine metrics

For fingernails, the left hand in the '11k hands' dataset achieved the highest results for all fingers using the cosine metric. Then, in second place, was the right hand; whereas the right hand in the 'PolyU' dataset achieved the lowest results, as illustrated in figure 4.3.

4.3.2 Minor knuckles

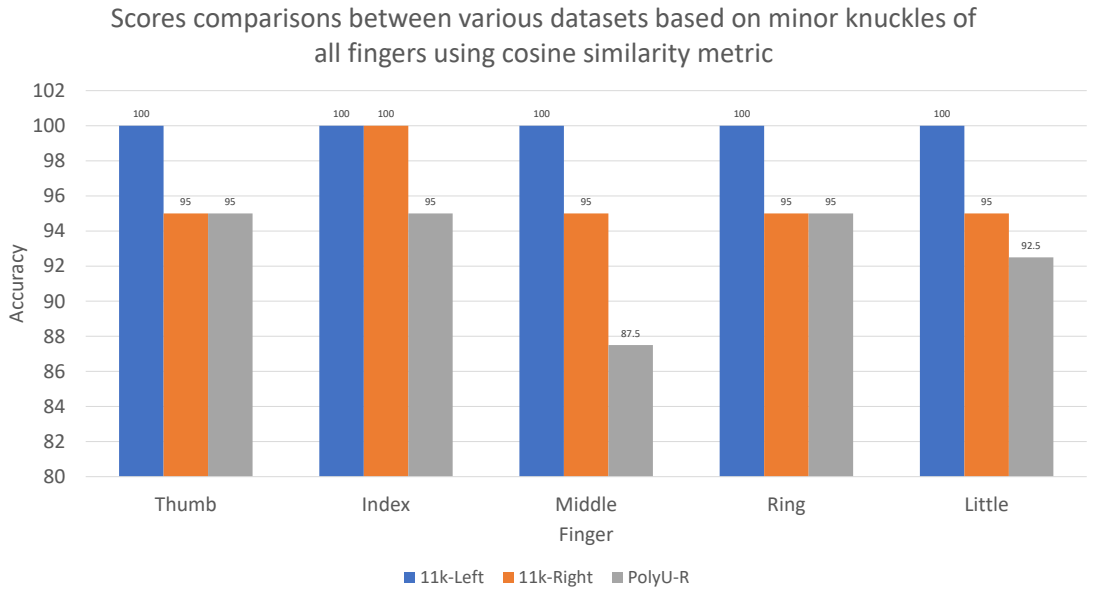

Figure 4.4: Recognition rates based on minor knuckles of various datasets using cosine metrics

Figure 4.4 shows that the left hands in the '11k hands' dataset performed the best compared to the minor knuckle of all fingers in other datasets. The right hands of the '11k hands' dataset came next, followed by, the 'PolyU' dataset.

4.3.3 Major knuckles

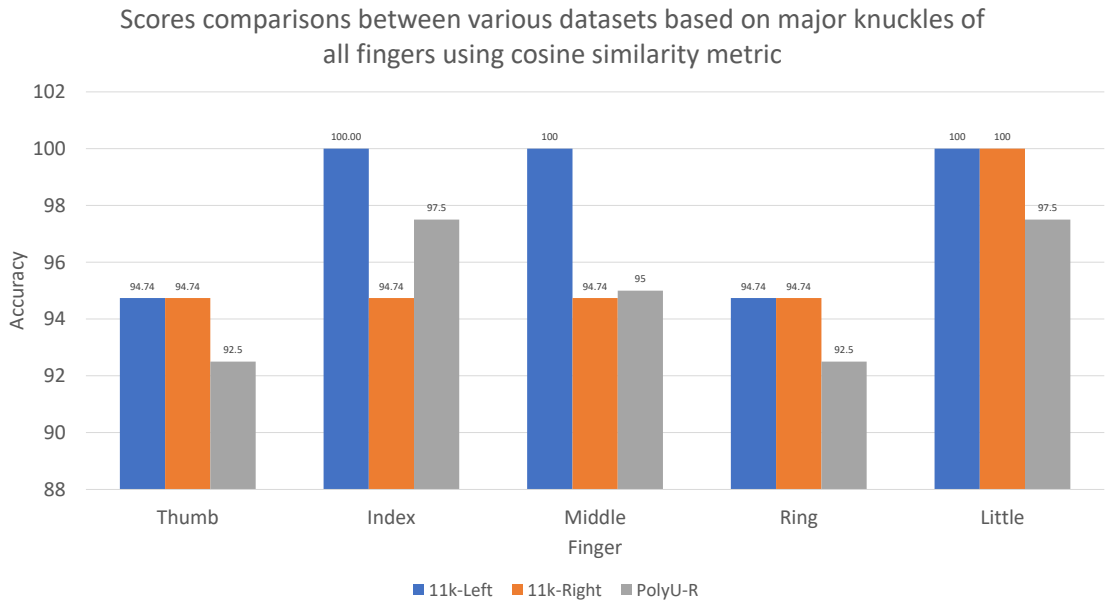

Figure 4.5: Recognition rate scores based on major knuckles in all fingers of various datasets using cosine metrics

Figure 4.5 shows that the left hand in the '11k hands' dataset performed the best compared to the major knuckles of all fingers in other datasets. The right hands of the '11k hands' dataset came next, then those of the 'PolyU' dataset.

4.3.4 Base knuckles

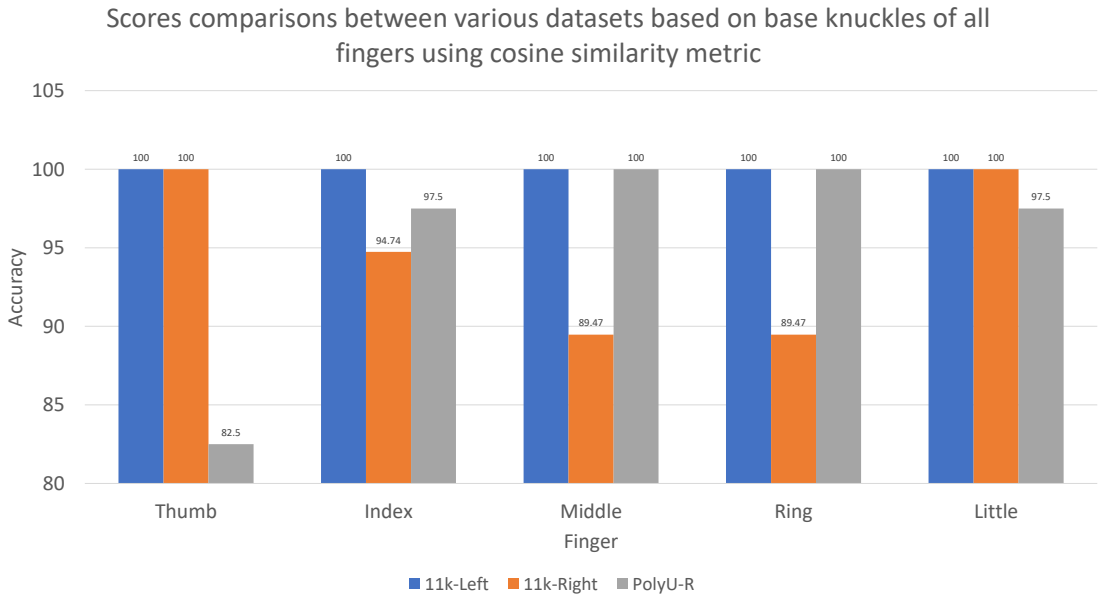

Figure 4.6: Recognition rate scores based on base knuckles in all fingers of various datasets using cosine metrics

Figure 4.6 illustrates how the base knuckles of the left hands of the '11k Hands' dataset performed better than any other dataset. Then, the 'PolyU' dataset performed better than the right hands of the '11k hands' dataset.

The proposed approach was also evaluated using the fusion of different multi-modalities. Two fusion rules were utilised: the MV and WA, which were employed with the metrics and improved the results as displayed in figures 4.8, 4.9, 4.10, and 4.11, and tables table 4.1 and 4.2. The weight was estimated for each component of the hand based on the rank-1 accuracy in table 4.1. In general, the MV performed better than the WA. For example, the fingernails-based fusion and the base-knuckle-based fusion achieved 100% in all datasets using the MV that applied on the Bray-Curtis similarity metric. Also, the cumulative matching characteristic (CMC) chart in figure 4.7 shows a sample result from the fingernails using the cosine metric and the MV applied in all datasets, including the holistic and individual performance of each hand's component. It is evident from the chart that the holistic technique outperforms the performance of individual modalities.

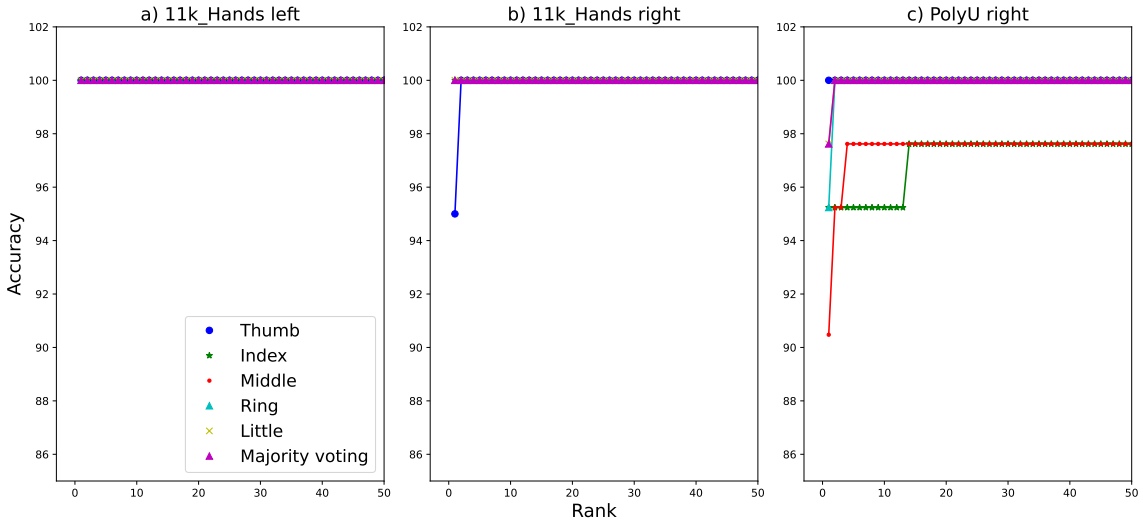

Figure 4.7: The CMC diagram of the proposed fusion technique applied in the fingernails of : a) the '11k Hands left'; b) '11k Hands right'; c) 'PolyU right'

4.3.5 Fusion based on fingernails

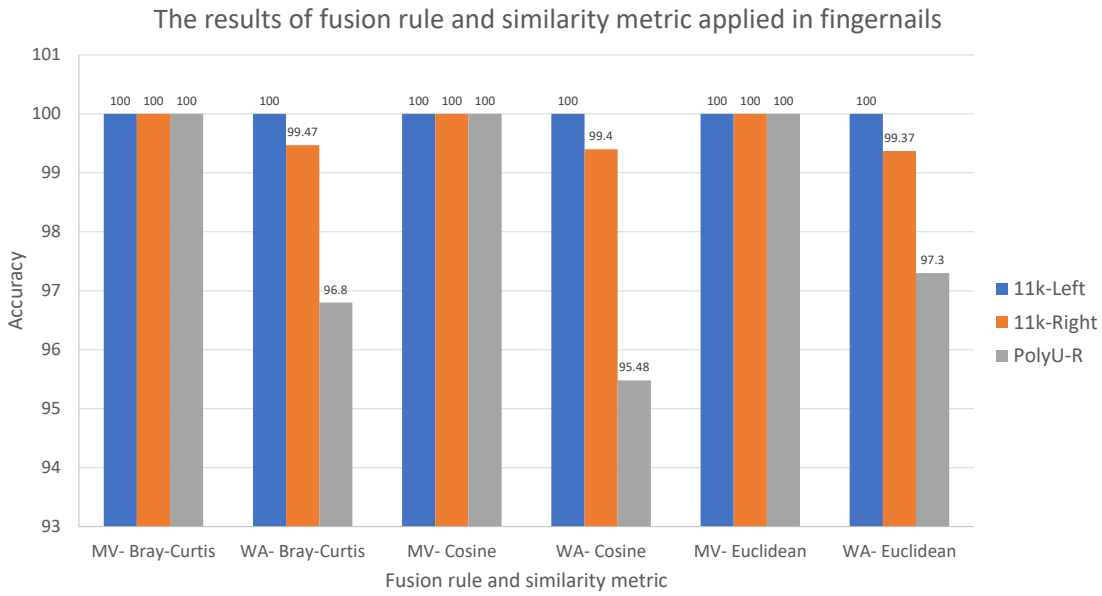

Figure 4.8: The results of MV and WA fusion rules, and various similarity metrics applied in fingernails of different datasets.

The chart in figure 4.8 displays two fusion rules, MV and WA, and the Bray-Curtis, Cosine, and Euclidean similarity metrics. The figure shows that the MV performed better than the WA fusion using all metrics. The WA fusion rule achieved the best result for the left hands of the '11k Hands', then the right hands of the '11k Hands', followed by the 'PolyU' dataset, from the highest to the lowest, respectively.

4.3.6 Fusion based on minor knuckles

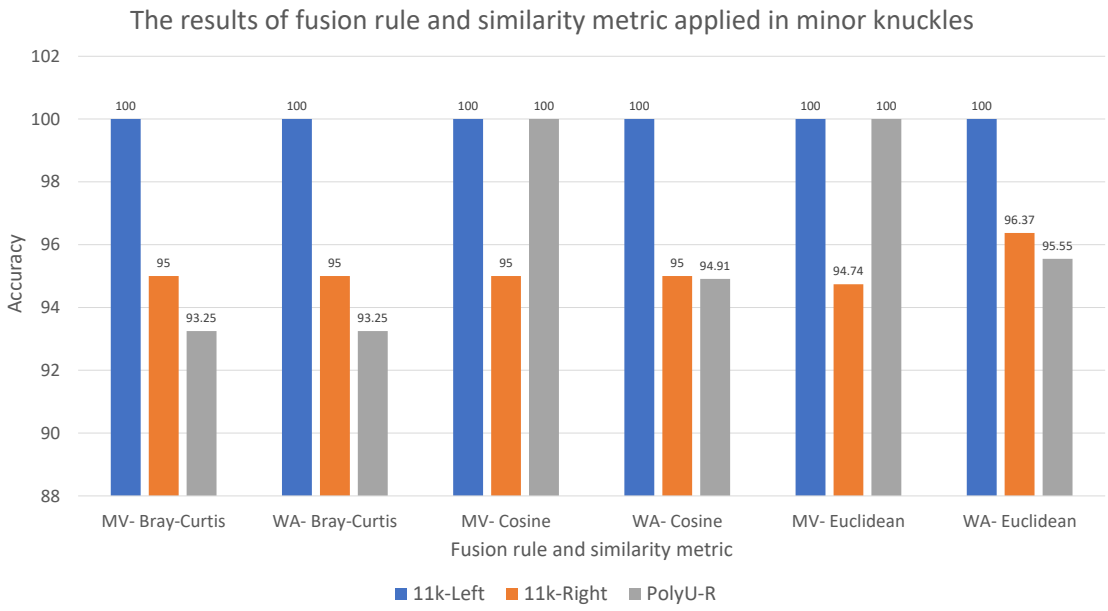

Figure 4.9: The results of MV and WA fusion rules, and various similarity metrics applied in minor knuckles of different datasets.

For the minor knuckles, the MV using all similarity metrics show excellent performance with a rate of 100% for the left hand of the '11k Hands', as shown in figure 4.9. Also, the MV rule with a cosine and Euclidean metric obtained a score of 100% in the 'PolyU' dataset.

4.3.7 Fusion based on major knuckles

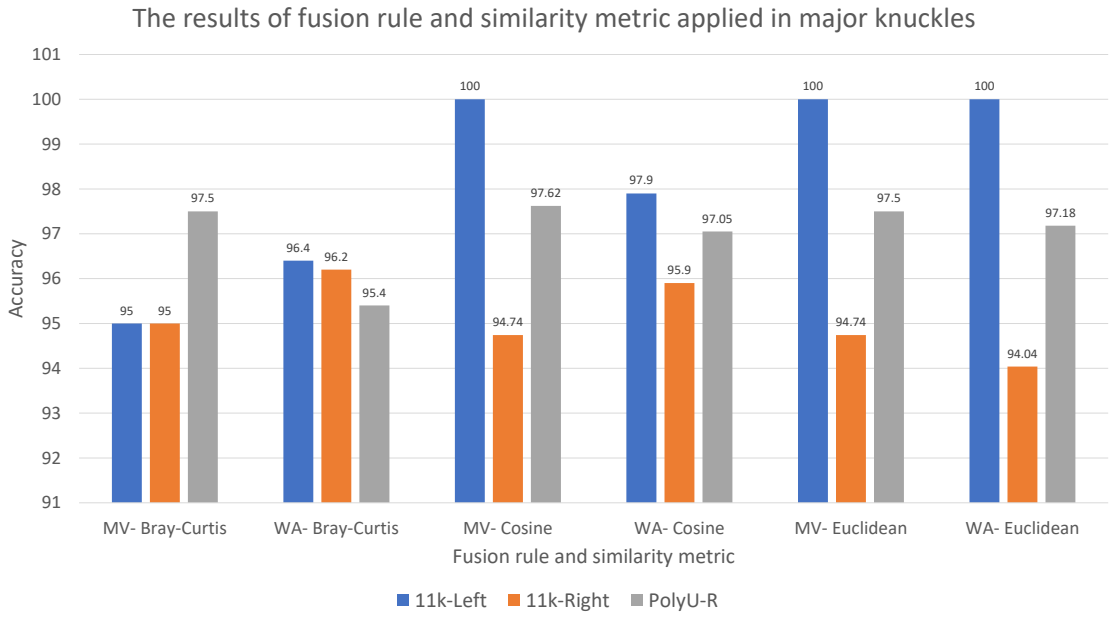

Figure 4.10: The results of MV and WA fusion rules, and various similarity metrics applied in major knuckles of different datasets.

Figure 4.10 shows the left hands of the '11k Hands' dataset achieved high results using MV with Cosine, with the Euclidean metric, and WA with the Euclidean metric, which was applied on the major knuckles. Also, this region shows good performance in the 'PolyU' dataset, with which most of the techniques obtained more than a 97% success rate.

4.3.8 Fusion based on base knuckles

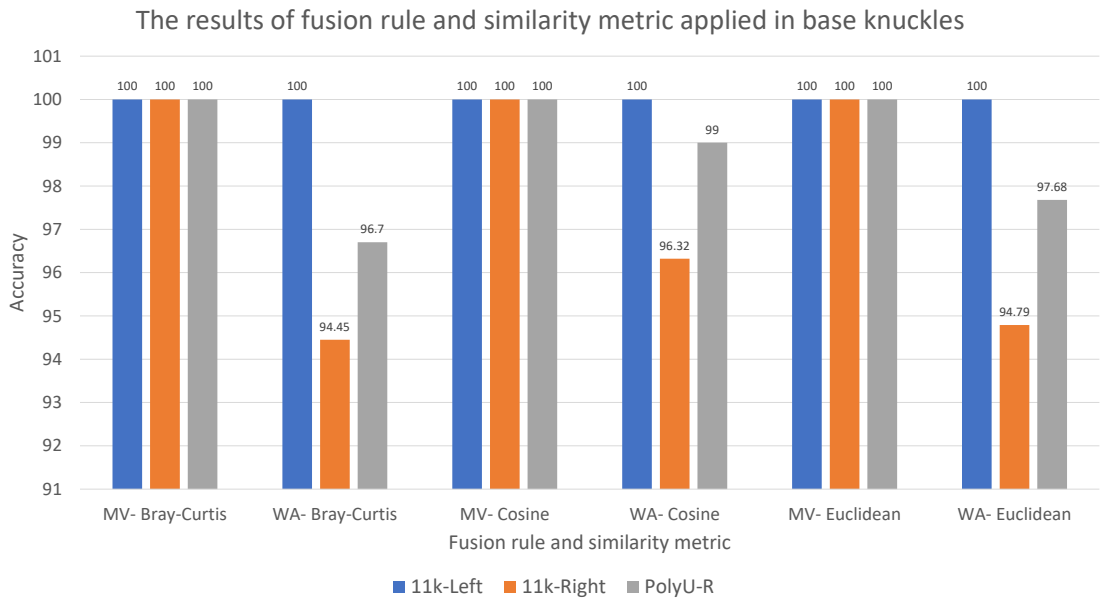

Figure 4.11: The results of MV and WA fusion rules, and various similarity metrics applied in base knuckles of different datasets.

For the base knuckles fusion, as displayed in figure 4.11, high results were achieved for the left hands in the '11k Hands' dataset using both fusion techniques with all metrics. The 'PolyU' dataset obtained scores above 96% using WA fusion. Also, MV fusion performed well using all metrics in all datasets.

4.3.9 Fusion of hand components

Table 4.2: The rank-1 recognition rate (shown in %) for fusion of hands components based on various similarity metrics in the '11k Hands' and 'PolyU' datasets.

Similarity metric	Fusion rule	11k-Left	11k-Right	PolyU-R
Bray-Curtis	MV	100	100	100
	WA	99.64	96.67	97.73
Cosine	MV	100	100	100
	WA	99.2	96.86	97.49
Euclidean	MV	100	100	100
	WA	100	97.60	97.29

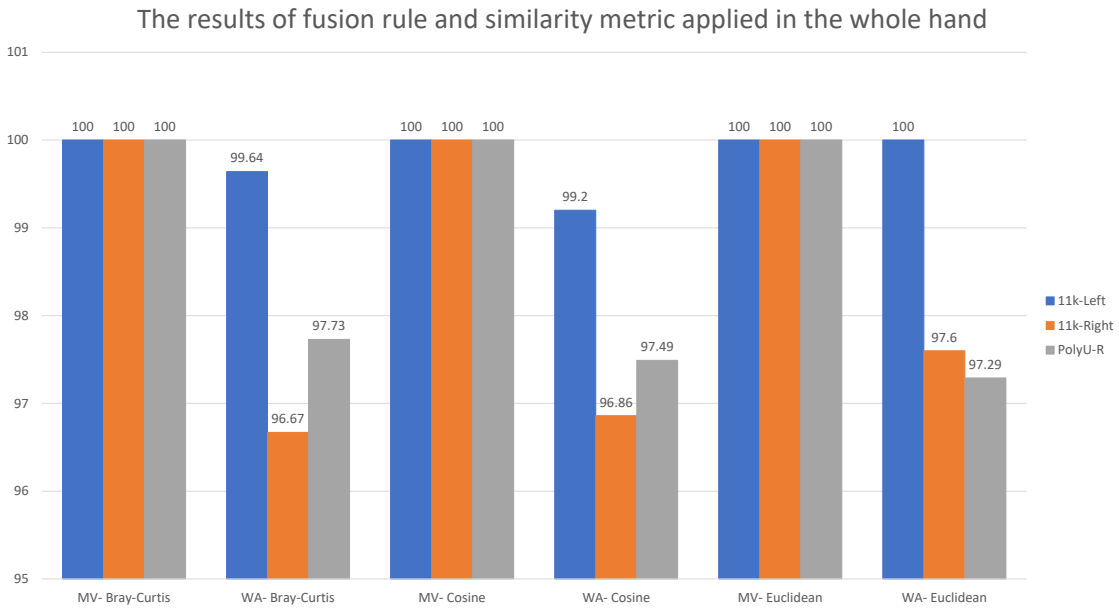

Figure 4.12: The results of MV and WA fusion rules, and various similarity metrics applied in whole hand components from different datasets.

The fusion rule and metric similarity results applied on the whole hands from

various datasets can be seen in figure 4.12 and table 4.2. The figure demonstrates that the fusion of all hand components using the MV fusion technique performed better than the WA fusion technique with a recognition score of 100%. Also, the WA fusion obtained promising results on the left hand of the '11k Hands' dataset with the rank-1 of 99.64%, 99.2% and 100% using the Bray-Curtis, Cosine, and Euclidean similarity metrics, respectively.

In the next section, a comparison against the state-of-the-art approaches will be presented, and this proposed multimodal biometric outperformed them.

4.3.10 Comparison between the proposed approach and the state-of-the-art

There is a minimal number of available studies that investigate the identification of persons using sub-images of their dorsal hands and the fusion of sub-images from the fingernails, knuckle creases, and whole hands using different similarity metrics and fusion rules. The proposed approach was compared with the recent study in (Vyas, Rahmani, et al., 2021). However, the (Vyas, Rahmani, et al., 2021) paper (Vyas, Rahmani, et al., 2021) assessed the Rank-1 identification accuracy using a fusion based on the entire finger. In this presented approach, the fusions were based on regions, e.g., the fusion of the base knuckles of fingers. In (Vyas, Rahmani, et al., 2021) study, the whole hand score-level fusion using averaging achieved 100%, 99.62%, and 99.33% in '11k-Left', '11k-Right', and 'PolyU', respectively. In comparison, in this experiment, the MV method with all metrics gained 100% using all datasets as shown in table 4.2.

4.4 Conclusion

To the author's knowledge, this is the first approach for person identification that uses all fingernails and knuckle creases from human fingers. Evaluating different similarity metrics for fusion allows for solid whole-hand identification with outstanding results. Through retraining a pretrained model of DenseNet201, the proposed approach extracts the distinctive properties of different fingernails and knuckles from the five fingers of both the right and left hands. These extracted traits are then employed with varying similarity metrics to determine the identity of all fingernails and knuckles from the five fingers and the hand as a whole. In our future work, the plan is to utilise various fusion levels, including feature, score, and decision level fusion, and different rules (e.g., sum, minimum, maximum, etc.).

Chapter 5

A Method for Identification of Humans from Dorsal Hand Sub-images using Siamese Network Models

5.1 Introduction

An approach for identifying individuals using Siamese networks, and knuckle crease and fingernail sub-images is presented in this chapter. It introduces a framework for automatic person identification that includes localization of many components' regions of interest (ROI) within hand images, recognition and segmentation of the detected components using bounding boxes, and similarity matching between two sets of segmented images. Feature extraction methods are major components of the proposed framework. Different deep learning neural networks (DLNNs) are used in this chapter to extract discriminative high-level abstract features. For the matching process, Euclidean distance is employed. The proposed method is validated using well-known benchmarks such as the 11k Hands dataset and the Hong Kong Polytechnic University Contactless Hand Dorsal Images, also known as 'PolyU'.

5.2 The Siamese network models for human identification

This section describes an efficient uni-modal biometric system based on segmented human dorsal hand images of the fingernails, minor, major, and base of the five fingers of the left and right hands. Three base models were investigated of DLNNs, including

the DenseNet201, VGG16, and ResNet152. Based on the results of evaluations, the worst performance was VGG16; thus, the latter was discarded, and the best base model was DenseNet201, which was then fine-tuned. The structure and different phases of the proposed approach are shown in the figure 5.1, and 5.2 shows a simple illustration of the Siamese network structure.

The proposed framework consists of many stages, as presented in the figure 5.1, including the localisation of joints in the human dorsal hand, and segmentation of the ROIs in the first stage; feature extraction using the base models of several deep learning CNNs, and fine-tuning of the DenseNet201 in the Siamese network including the training and validation in the second stage; and finding the best match including the testing using Euclidean distance, in the third stage. In the next subsections, these phases will be explained in detail.

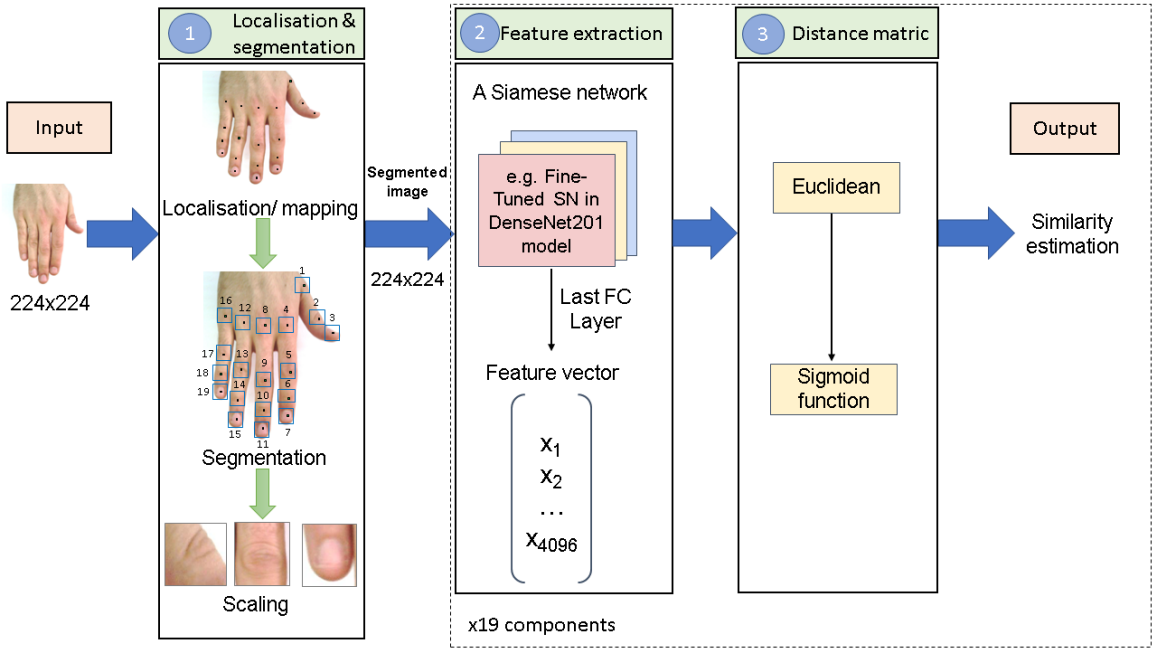

Figure 5.1: A diagram of the proposed Siamese network approach for identifying people using dorsal fingernails and knuckle patterns.

5.2.1 Pre-processing of the hand images

In this stage, hand images passed through the pre-processing, in which the keypoints of the hand were segmented and scaling was conducted as illustrated in the previous subsection 3.2.1. Therefore, the outcomes of this stage are the segmentation of the dorsal fingernails, base, major, and minor knuckles.

5.2.2 The Siamese networks for extracting features

The Siamese networks use pairs of segmented images as inputs to determine whether or not the images belong to the same class, and the same network is utilised for feature extractions. Two base models were investigated of DLNNs, including the DenseNet201, and ResNet152. Based on the results of evaluations, the best base model was DenseNet201, which then fine-tuned.

5.2.2.1 Dividing the dataset into positive and negative pairs

The aim of this stage is to construct the dataset by dividing it into positive and negative pairs. Positive pairs are sub-images of the same class that are expected to have a similarity value greater than 0.5 (a positive match and the two sub-images are similar). Thus, negative pairs are the inverse. At random, pairs of a sub-image from one subject and a sub-image from another subject were chosen. The advantage of this is that the same sub-image can be used twice. As a result, the size of the training data is doubled for both positive and negative matches. Negative pairs are annotated with a 0 and positive pairs with a 1 and both used to train the model.

5.2.2.2 Constructing the Siamese network

The structure of the Siamese network consists of two inputs for the segmented images as shown in the figure 5.2. Various base models of DLNN were utilised to compare between their performance. These base models are DenseNet201 (G. Huang et al., 2017), VGG16 Simonyan and Zisserman, 2014, and ResNet152 (He et al., 2016). The DLNN model is utilised to extract abstract high-level features and the last FC layer was used to construct the feature vector. A lambda layer was added that acquired features from pairs of sub-images as inputs and calculated the Euclidean distance between them. The distance value was then passed to a dense layer with a Sigmoid activation function for a gradient descent combined with the Nesterov momentum at a score of 0.9. The function was used to predict the similarity based on the distance. The following is the equation for the Euclidean distance:

$$E(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + \ (y_n - x_n)^2}, \tag{5.1}$$

where n in the formula is equal to 4096, the dimension of the segmented image's feature vector. Also, x and y are two different sub-image vectors. Finally, the Sigmoid will convert the obtained distance value into a value between 0 and 1. If that value is greater than 0.5, the matching is positive; otherwise, the matching is negative. In training of the base model, the built model uses early stopping and the loss function is binary cross-entropy because there are only two labels (match or no match). In the second step, the DenseNet152 was fine-tuned, because its performance is the highest

among others as illustrated in the subsection 5.3.1. The fine-tuning model was built as demonstrated in the previous work (Mona Alghamdi, P. Angelov, and Alvaro, 2022). Many layers were added to the original model of DenseNet201 (G. Huang et al., 2017). These layers were: 2D global average pooling; batch normalisation with a 0.9 momentum; dropout at 0.5; dense layer with 4096 dimensions; Relu activation function; dropout at 0.6; batch normalisation with a 0.9 momentum; and a softmax layer with the dimension numbers of subjects. Epochs of training and augmentations of the training and validation data were used the same as this previous work (Mona Alghamdi, P. Angelov, and Alvaro, 2022). These augmentations were applied to the training data, which improved model learning and addressed the issue of small training data. The augmentation comprises pixel rescaling from 0 to 1, horizontal flipping, a randomised rotation range of 30, a randomised width and height shift range of 30, and a randomised zoom range of 0.9 to 1.1. The DLNN was trained on 150 epochs. The built model used an early stopping, the learning optimization was stochastic gradient descent, with a learning rate of 0.001, a Nesterov momentum of 0.9, and a loss function of binary cross-entropy. Later, the same network was again retrained with all of its layers and parameters with 100 epochs. Regarding the time complexity, the proposed approach of hand region identification using the Siamese network was processed in different stages. The first stage was the pre-processing, including resizing, localisation, mapping, segmentation, and scaling, which took around 5 to 7 minutes per hand region. The second stage was the feature extraction using two models: the base and fine-tuning model (including retraining and validation), which required about 10 to 20 and 30 to 40 minutes, respectively. The last stage was testing, which required around 3 to 7 minutes. Also, Tensorflow and Keras platform with Tesla V100 GPU was used to run this experiment.

5.3 The results and discussion

The results of the evaluation of the proposed approach on the datasets '11k Hands' and 'PolyU' are presented in this section.

5.3.1 Datasets description

Two well-known datasets were used: the '11k Hands' and 'PolyU'. These two datasets described previously in the subsection 3.3.1.

This chapter focused on the identifiability of individual fingers in a uni-modal approach using the Siamese network, incorporating the performance of all fingernails and knuckles belonging to a particular finger. The performance metric of this study and comparisons with different studies were assessed using the rank-1 recognition rate as presented in the table 5.1. Pre-processing was performed by hand, including

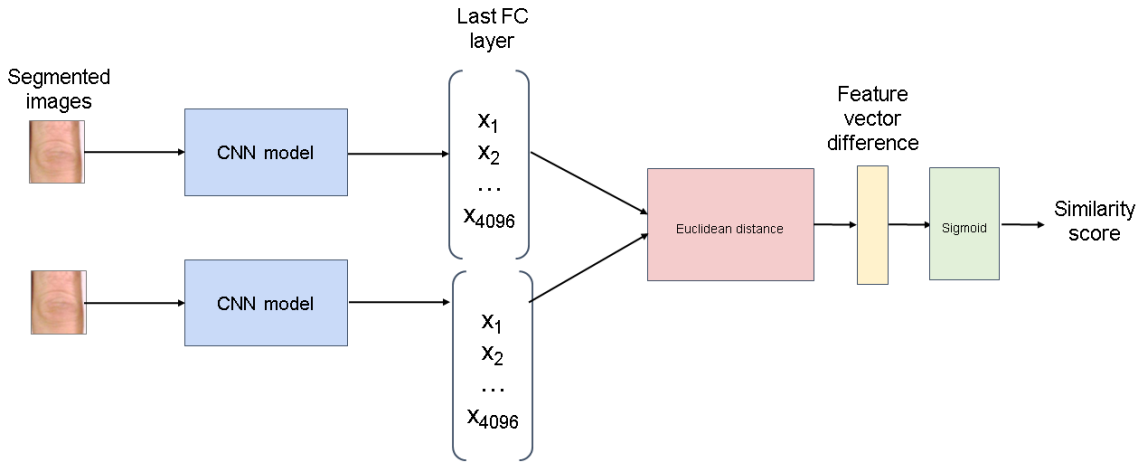

Figure 5.2: A schematic illustration of the Siamese Network

downsizing to 224×224 to be compatible with the localization of the ROI keypoints method in (Simon et al., 2017), mapping the keypoints to the original size of the image, segmentation, and scaling, as described in subsection 5.2.1. The features were retrieved in the second step by utilising the base model of DenseNet201 and ResNet152 that pre-trained on the famous 'ImageNet' dataset. In this experiment, retraining the CNN model of DenseNet201 on parts of the '11K Hands' and 'PolyU' datasets was also used because this model showed a good performance and to compare the performance of various feature extractor methods, as detailed in the subsection 5.2.2.

The datasets, including the '11k Hands' and 'PolyU', were separated into two parts: The first part was 90%, which was divided again into around 77% and 33% for training and validation, respectively, and the second part was 10% for testing. In the '11k Hands' dataset, $47,717$ and $49,513$ segmented images were used from all hand components including the left and right '11k Hands' datasets, respectively, to train and validate the fine-tuned model. Also, $4,650$, and $4,908$ images from the left and right '11 k Hands' dataset were used for testing. For the 'PolyU' dataset, $75,297$ sub-images from all components, were used for training and validation, respectively. The testing comprised of $7,312$ segmented images.

After splitting the data, a pre-processing procedure was performed: scaling that increased the performance of the proposed approach. The framework used the final FC layer with 4096 dimensions to extract abstract high-level characteristics from the

data. The euclidean distance was then calculated from the feature vector and passed to the Sigmoid function to generate the similarity score as displayed in the figure 5.2. Thus, random pairs were generated and run through the model, which gave a similarity score. If the score was more than 0.5, it was labelled as the same class. In the other case, it was labelled as not the same. Later, it was compared with an actual label to validate a prediction, so the number of correct predictions was counted from pairs at the end. Thus, the accuracy was calculated using the correct predictions over the total number of predictions. The testing was repeated three times, which was then averaged out, as indicated in the table 5.1.

The proposed approach obtained outstanding results for the identification rate of individual modalities, as shown in the table 5.1. Therefore, in the first step, the proposed method was evaluated using the base models of two Siamese networks: the DenseNet201 and ResNet152. The DenseNet201 performed better than ResNet152, with a rank-1 accuracy score of over 90%. However, the fine-tuning of the DenseNet201 enhanced the performance of the identification rate. Also, in general, the left hands of the '11k Hands' dataset scored higher identification accuracy than the right hands. In addition, in this experiment the fingernails achieved the best, then the major, minor and base knuckle patterns are from the highest to the lowest. The following subsections will explain the results per regions using the Siamese network approach and the comparisons with existing studies, such as (Mona Alghamdi, P. Angelov, and B. Williams, 2021; Mona Alghamdi, P. Angelov, and Alvaro, 2022; Vyas, Rahmani, et al., 2021), as shown in the table 5.1.

5.3.1.1 Fingernail region

For the fingernail region, the base models gave good results. For example, in the '11k Hands' dataset the rank-1 score gained an accuracy of 95.82% for the index finger of the left hand using the base model of DenseNet201, 95.41% for the ring finger of the left hand using the base model of ResNet152 and 95.33% for the ring finger in the 'PolyU' dataset using the base model of ResNet152, as displyed in the table 5.1.

In addition, the fine-tuning of the DenseNet201 achieved better results. For instance, using the '11k Hands' dataset, the middle finger gained a rank-1 of 98.47% for the left hands, 97.62% for the ring finger of the right hands, 97.51% for the ring finger of the left hands and 96.85% for the ring finger of the right hands.

Unfortunately, there needs to be more studies on the Siamese networks to compare with the current work. However, it can be compared with the (Thapar, Jaswal, and Nigam, 2019) study, in which the fingernail gained the best performance with an estimated rate of 94.83%. However, the study did not define the fingers. Moreover, the suggested approach outperformed the performance of previous studies in (Mona Alghamdi, P. Angelov, and B. Williams, 2021; Mona Alghamdi, P. Angelov, and

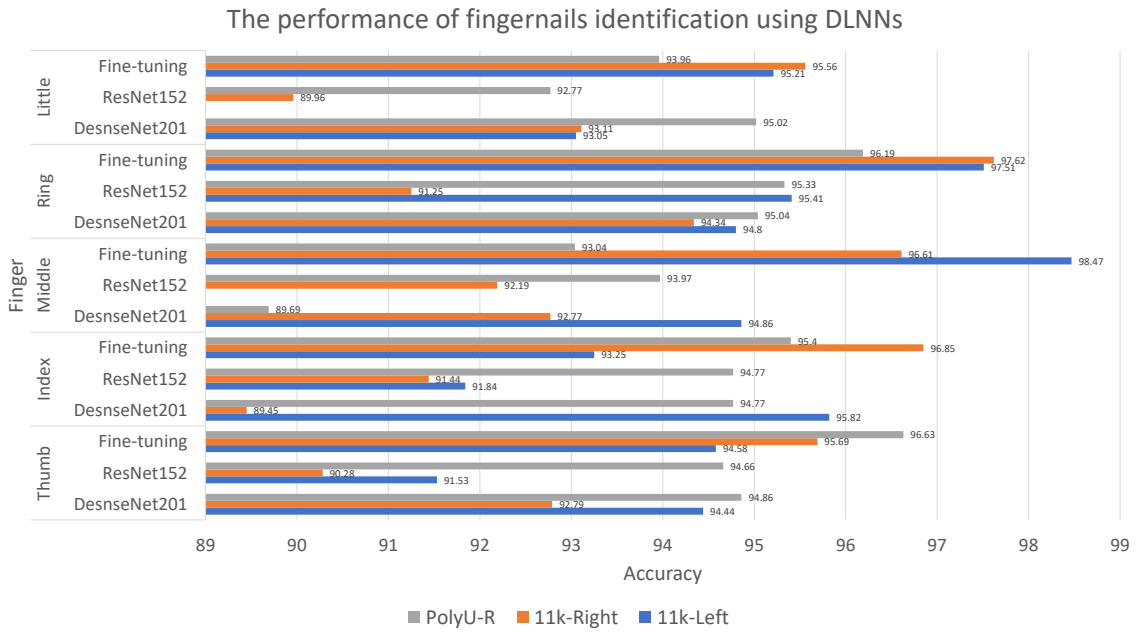

Figure 5.3: The identification results for fingernails using DLNNs and Euclidean distance employed in the Siamese network.

Alvaro, 2022; Vyas, Rahmani, et al., 2021), as illustrated in the table 5.1. For example, the thumb finger of the right hand in the '11k Hands' dataset achieved 95.69%. In addition, the middle and ring fingers gained 93.04% and 96.19%, respectively, in the 'PolyU' dataset.

5.3.1.2 Minor knuckles

Regarding the base models, as displayed in the figure 5.4, the minor knuckles of the ring and little fingers performed 95.05%, and 95.04% for the right hands of the '11k Hands' and 'PolyU' datasets, respectively, using the DenseNet201 model. The little finger of the left hand in the '11k Hands' scored 94.9% using the DenseNet201 model. Also, the rank-1 rate was 94.26% for the left hand's ring finger in the '11k Hands' dataset.

Concerning the fine-tuning model of the DenseNet201, the estimation rate was 97.98% for the middle finger, 97.62% for the little finger, and 97.03% for the index finger of the left hand in the '11k Hands' as shown in the figure 5.4. In comparison, the minor knuckle achieved only 88.73% accuracy in (Thapar, Jaswal, and Nigam, 2019). Moreover, the performance of the proposed approach exceeded the identification rate in the previous studies (Vyas, Rahmani, et al., 2021; Mona Alghamdi, P. Angelov,

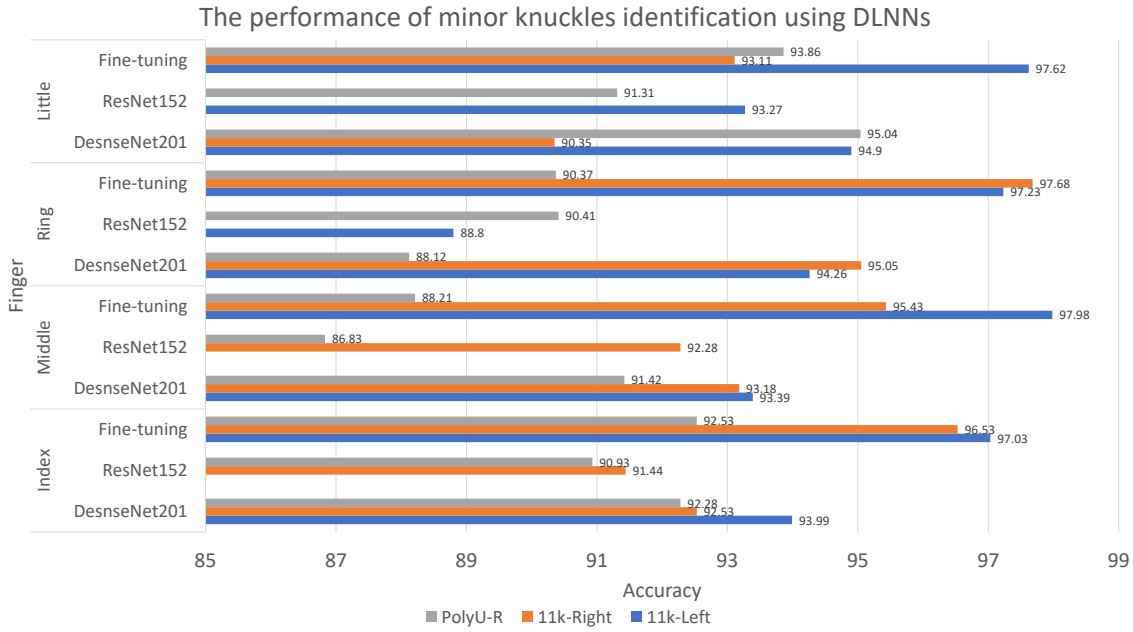

Figure 5.4: The identification results for minor knuckles using DLNNs and Euclidean distance employed in the Siamese network.

and B. Williams, 2021; Mona Alghamdi, P. Angelov, and Alvaro, 2022), in which the index and middle finger performed identification score of 96.53%, 92.53%, 95.43 and 93.21% for the right hand of the '11k Hands', and 'PolyU' datasets, respectively.

5.3.1.3 Major knuckles

The major knuckle achieved the highest results compared to the other knuckles, with estimation rates of 97.33%, 96.40%, and 95.41% for the ring finger of the left, middle, and index finger of the right hands of the '11k Hands' dataset, respectively, as shown in the figure 5.5.

For the fine-tuning model, overall, compared to other ROIs, the major knuckles gained the highest score of rank-1 with an accuracy of 99.06%, 98.17%, and 99.18% for the index, middle, and ring fingers of the left '11k Hands' dataset, as displayed in the table 5.1. In (Thapar, Jaswal, and Nigam, 2019) study, the major knuckle achieved only 90.52% accuracy without determining the fingers or side of the hand. In addition, the offered approach performed the best with a rank-1 score of 91.40% for the major knuckle of the thumb finger from the 'PolyU' database, 96.71% for the index finger of the right hand from the '11k Hands' dataset comparing to other studies in (Vyas, Rahmani, et al., 2021; Mona Alghamdi, P. Angelov, and B. Williams, 2021;

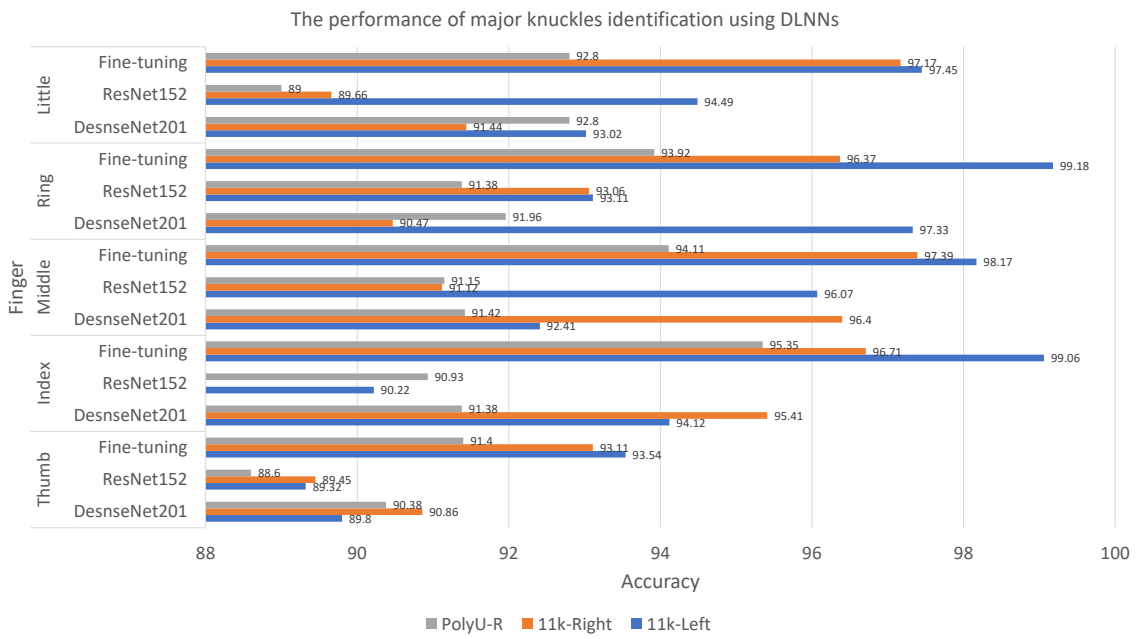

Figure 5.5: The identification results for major knuckles using DLNNs and Euclidean distance employed in the Siamese network.

Mona Alghamdi, P. Angelov, and Alvaro, 2022). The middle and ring fingers also gained recognition rates of 97.39% and 94.11% for the right hand in both datasets, and 99.18%, 96.37%, and 93.92% for the left and right hands of both datasets, which is the highest compared to other methods of other studies.

5.3.1.4 Base knuckles

Figure 5.6 displays the performance of the identification of the base knuckle using DLNNs and Euclidean distance employed in the '11k Hands' and 'PolyU' datasets. The DenseNet201 gained high results in the base knuckle region. The model scored 95.48%, and 95.43% for the index finger of the left and right hand, respectively, and 92.77% for the right thumb of the '11k Hands' dataset, as shown in table 5.1.

Regarding the fine-tuning model, the identification performance improved. For the '11k Hands' dataset, the index and little finger achieved 96.76% and 96.63% and the middle finger scored a rank-1 of 96.40% for the right hand, as illustrated in the figure 5.6.

The suggested approach performed the best in the base knuckle of fingers among previous studies (Vyas, Rahmani, et al., 2021; Mona Alghamdi, P. Angelov, and B. Williams, 2021; Mona Alghamdi, P. Angelov, and Alvaro, 2022). For example, the

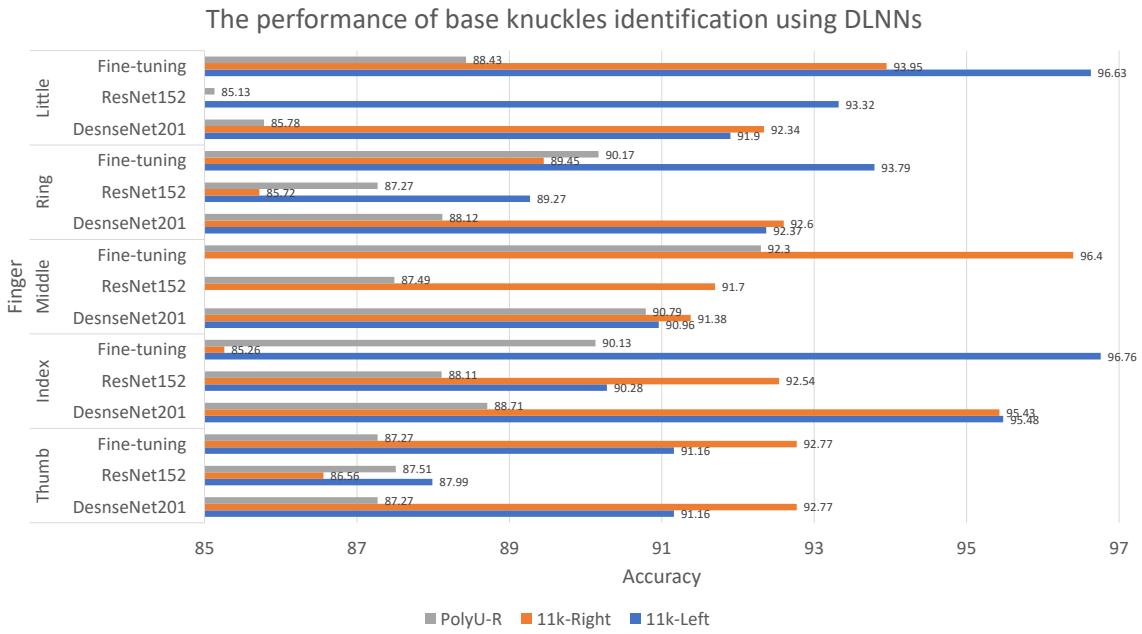

Figure 5.6: The identification results for base knuckles using DLNNs and Euclidean distance employed in the Siamese network.

approach scored an identification rate of 96.26%, 96.40%, 94.45%, and 87.27% for the index, middle, and ring fingers of the right hand of the '11k Hands' dataset, and the thumb finger of the 'PolyU' dataset.

In the following subsections, the results of (dis-)similarity prediction will be shown in terms of the confusion matrix and matching samples. A paired result below 0.5 indicates a dis-similarity, and the sub-images do not belong to the same person. The figure 5.7 shows an example of two pairs of sub-images and a confusion matrix from the major knuckle of the middle finger for the left hands of the '11k Hands' dataset. The left sample is a correct prediction, which is dis-similar because the actual label is 0 and the prediction is less than 0.5. In contrast, the sample on the right side is a wrong prediction or match due to estimating an inaccurate score, which is 0.86.

Another example is shown in the figure 5.8 for a confusion matrix from the major knuckle of the ring finger of the left hands of the '11k Hands' dataset. In this matrix, 200 samples predicted dis-similarity between the pairs, and 210 samples correctly predicted similarity. Also, in this figure, two pairs of sub-images are shown. The left example is a wrong match; because the actual label is 0, which indicates no match. However, the framework estimation matches a score of 0.77. On the other side, the prediction is 0 for the pairs, and the true label is 0; therefore, the right pair is correct.

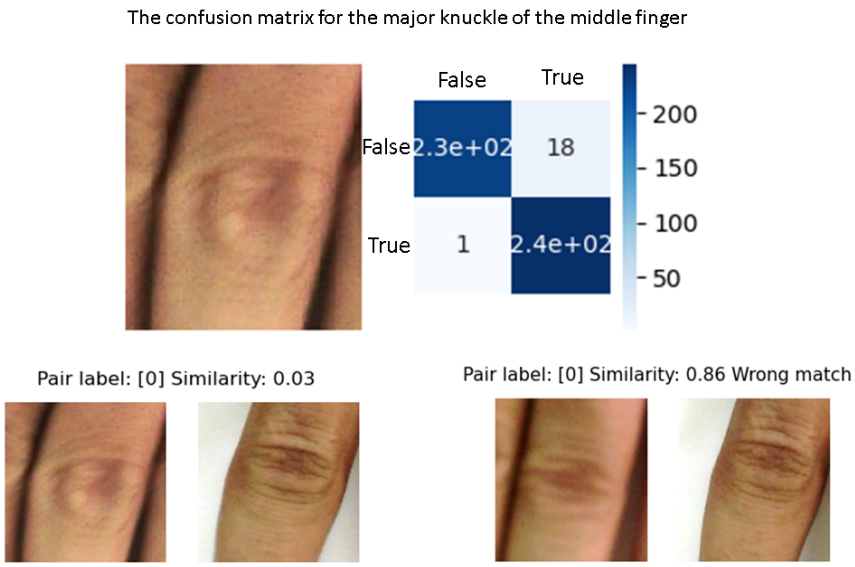

Figure 5.7: A sample of a confusion matrix, correct, and wrong similarity predictions applied in the major knuckle of the middle finger of the '11k Hands'.

Table 5.1: The rank-1 recognition rate (shown in %) for the base and fine-tuning models of the Siamese network and comparisons with different studies using the '11k Hands' and 'PolyU' datasets.

Region	Finger	Method	11k-Left	11k-Right	PolyU
Fingernail	Thumb	DesnseNet201	94.44	92.79	94.86
		ResNet152	91.53	90.28	94.66
		Fine-tuning	94.58	**95.69**	96.63
		PIFK^{+} (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	95.00	**100**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	87.83	84.21	93.81
		(Vyas, Rahmani, et al., 2021)	-	-	-

Table 5.1: The rank-1 recognition rate (shown in %) for the base and fine-tuning models of the Siamese network and comparisons with different studies using the '11k Hands' and 'PolyU' datasets.

Region	Finger	Method	11k-Left	11k-Right	PolyU
	Index	denseNet201	95.82	89.45	94.77
		ResNet152	91.84	91.44	94.77
		Fine-tuning	93.25	96.85	95.40
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	**100**	**95.83**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	89.42	88.95	90.40
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Middle	DenseNet201	94.86	92.77	89.69
		ResNet152	50.00	92.19	93.97
		Fine-tuning	98.47	96.61	**93.04**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	**100**	91.83
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	90.48	89.47	87.65
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Ring	DenseNet201	94.8	94.34	95.04
		ResNet152	95.41	91.25	95.33
		Fine-tuning	97.51	97.62	**96.19**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	**100**	93.88
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	93.65	91.58	85.10
		(Vyas, Rahmani, et al., 2021)	-	-	-

Table 5.1: The rank-1 recognition rate (shown in %) for the base and fine-tuning models of the Siamese network and comparisons with different studies using the '11k Hands' and 'PolyU' datasets.

Region	Finger	Method	11k-Left	11k-Right	PolyU
	Little	DenseNet201	93.05	93.11	95.02
		ResNet152	48.25	89.96	92.77
		Fine-tuning	95.21	95.56	93.96
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	**100**	**95.83**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	84.13	80.00	87.30
		(Vyas, Rahmani, et al., 2021)	-	-	-
Minor Knuckle	Index	DenseNet201	93.99	92.53	92.28
		ResNet152	50.00	91.44	90.93
		Fine-tuning	97.03	**96.53**	**92.53**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	95.00	91.67
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	84.66	76.84	72.47
		(Vyas, Rahmani, et al., 2021)	86.35	89.56	74.81
	Middle	DenseNet201	93.39	93.18	91.42
		ResNet152	50.00	92.28	86.83
		Fine-tuning	97.98	**95.43**	**93.21**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	95.00	85.71
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	85.19	82.11	68.92

Table 5.1: The rank-1 recognition rate (shown in %) for the base and fine-tuning models of the Siamese network and comparisons with different studies using the '11k Hands' and 'PolyU' datasets.

Region	Finger	Method	11k-Left	11k-Right	PolyU
		(Vyas, Rahmani, et al., 2021)	93.69	93.17	84.57
	Ring	DenseNet201	94.26	95.05	88.12
		ResNet152	88.8	49.74	90.41
		Fine-tuning	97.23	**97.68**	90.37
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	95.00	**91.84**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	84.13	77.89	71.46
		(Vyas, Rahmani, et al., 2021)	91.45	89.56	80.80
	Little	DenseNet201	94.9	90.35	95.04
		ResNet152	93.27	50.00	91.31
		Fine-tuning	97.62	93.11	**95.86**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	**95.00**	91.67
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	81.48	76.84	78.43
		(Vyas, Rahmani, et al., 2021)	83.91	80.83	73.25
Major Knuckle	Thumb	DesnseNet201	89.80	90.86	90.38
		ResNet152	89.32	89.45	88.60
		Fine-tuning	93.54	93.11	**91.40**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**95.00**	**95.00**	91.11

Table 5.1: The rank-1 recognition rate (shown in %) for the base and fine-tuning models of the Siamese network and comparisons with different studies using the '11k Hands' and 'PolyU' datasets.

Region	Finger	Method	11k-Left	11k-Right	PolyU
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	85.71	84.21	76.83
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Index	DenseNet201	94.12	95.41	91.38
		ResNet152	90.22	48.19	90.93
		Fine-tuning	99.06	**96.71**	95.35
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	95.00	**97.87**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	82.45	83.16	79.71
		(Vyas, Rahmani, et al., 2021)	93.28	93.93	91.23
	Middle	DenseNet201	92.41	96.4	91.42
		ResNet152	96.07	91.12	91.15
		Fine-tuning	98.17	**97.39**	**94.11**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	95.00	93.75
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	85.11	82.54	76.64
		(Vyas, Rahmani, et al., 2021)	94.70	95.26	88.90
	Ring	DenseNet201	97.33	90.47	91.96
		ResNet152	93.11	93.06	91.38
		Fine-tuning	**99.18**	**96.37**	**93.92**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	90.00	95.00	91.49

Table 5.1: The rank-1 recognition rate (shown in %) for the base and fine-tuning models of the Siamese network and comparisons with different studies using the '11k Hands' and 'PolyU' datasets.

Region	Finger	Method	11k-Left	11k-Right	PolyU
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	80.95	83.68	83.23
		(Vyas, Rahmani, et al., 2021)	94.30	92.03	90.34
	Little	DenseNet201	93.02	91.44	92.80
		ResNet152	94.49	89.66	89.00
		Fine-tuning	**97.45**	97.17	97.60
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	95.00	**100**	**97.87**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	83.07	75.79	82.06
		(Vyas, Rahmani, et al., 2021)	88.59	86.34	88.57
Base Knuckle	Thumb	DesnseNet201	90.42	89.99	87.19
		ResNet152	87.99	86.56	87.12
		Fine-tuning	91.16	92.77	**87.27**
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	**100**	79.59
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	87.30	85.26	58.35
		(Vyas, Rahmani, et al., 2021)	-	-	-
	Index	denseNet201	95.48	95.43	88.71
		ResNet152	90.28	92.54	88.11
		Fine-tuning	96.76	**96.26**	90.13

Table 5.1: The rank-1 recognition rate (shown in %) for the base and fine-tuning models of the Siamese network and comparisons with different studies using the '11k Hands' and 'PolyU' datasets.

Region	Finger	Method	11k-Left	11k-Right	PolyU
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	90.00	**95.92**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	78.84	81.58	64.34
		(Vyas, Rahmani, et al., 2021)	78.82	84.82	62.38
	Middle	DenseNet201	90.96	91.38	90.79
		ResNet152	84.82	91.70	87.49
		Fine-tuning	81.31	**96.40**	92.30
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	95.00	**97.96**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	80.95	77.37	67.27
		(Vyas, Rahmani, et al., 2021)	84.32	85.96	61.49
	Ring	DenseNet201	92.37	92.60	88.12
		ResNet152	89.27	85.72	87.27
		Fine-tuning	93.79	**94.45**	90.17
		PIFK$^+$ (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	100	80.00	**97.96**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	80.42	77.37	66.47
		(Vyas, Rahmani, et al., 2021)	76.99	73.62	57.16
	Little	DenseNet201	91.90	92.34	85.78
		ResNet152	93.32	50.00	85.13
		Fine-tuning	96.63	93.95	88.43

Table 5.1: The rank-1 recognition rate (shown in %) for the base and fine-tuning models of the Siamese network and comparisons with different studies using the '11k Hands' and 'PolyU' datasets.

Region	Finger	Method	11k-Left	11k-Right	PolyU
		PIFK^{+} (Mona Alghamdi, P. Angelov, and Alvaro, 2022)	**100**	**100**	**97.96**
		PIFK (Mona Alghamdi, P. Angelov, and B. Williams, 2021)	84.13	80.53	67.94
		(Vyas, Rahmani, et al., 2021)	83.71	81.78	59.16

5.4 Conclusion

This chapter presented a method for identifying individuals from knuckle crease and fingernail sub-images using the Siamese networks approach. It introduces a framework for automatic person identification that includes hand image localization of many components' ROI, recognition and segmentation of the detected components using bounding boxes, and similarity matching between two sets of segmented images. The proposed framework relies heavily on feature extraction methods. For extracting discriminative abstract features, this experiment employs several DLNNs, including DenseNet201, ResNet152, and fine-tuning of DenseNet201. In addition, the Euclidean distance was used in the matching process. The proposed method was validated against well-known benchmarks such as the 11k Hands dataset and the Hong Kong Polytechnic University dataset. The proposed approach exceeded the state-of-the-art methods in some regions of the hand. In future studies, the recommendation is to apply fusion techniques to the Siamese network.

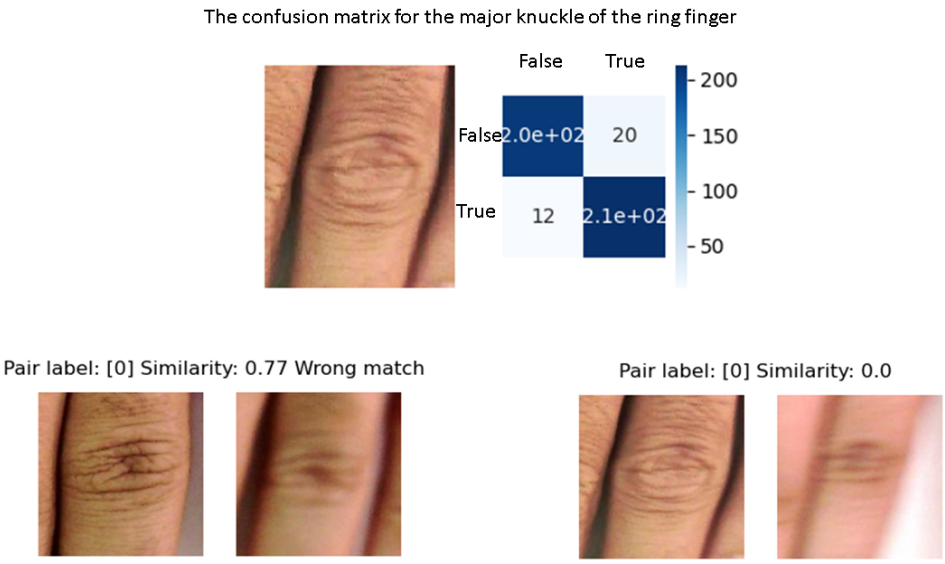

Figure 5.8: A sample of a confusion matrix, correct, and wrong similarity predictions applied in the major knuckle of the ring finger of the '11k Hands'.

Chapter 6

Conclusion

6.1 Introduction

This chapter will sum up the work by illustrating the key research findings that related to the research aims, objectives, and research questions. It will highlight the research contributions and define the limitations of the investigation, as well as suggest plans for future research.

6.2 Conclusion of the thesis

The aim of this study was to investigate human identification from RGB dorsal hand images. The results indicated the flat posture of dorsal hand images could be used for this purpose. The hypothesis was to calculate the similarity of two different sub-images, taken from the knuckle pattern and fingernails after segmentation from hand images. When the result of the similarity is high, this indicates that the two sub-images belong to the same subject or person. Otherwise, they belong to different persons.

The ability to recognise images and differentiate between distinct subjects and types of images has been revolutionised by advancements in ML, CV, and DLNN technologies. This can significantly advance and facilitate biometrics, forensic, and biological application research. The human hand, in particular, offers a substantial quantity of information that may be unique to each individual and can therefore aid in identifying people or specific medical issues. For instance, identifying people from videos and images has become more crucial than ever in criminal investigations.

The human hand is a highly complicated part of the anatomy. It has yet to be extensively studied, especially in light of today's advanced level of computational intelligence and, more specifically, the developments in AI, ML, and CV.

The objectives of this thesis were to develop a new framework for person identification, which would include joint or keypoint detection, segmentation, feature extraction, and finding the best match using similarity metrics; to use the base model and fine-tune DLNN feature extraction and various similarity matching to improve the framework; to use score-level fusion to improve the identification framework; and to use a cutting-edge method, like the Siamese network, to enhance the framework.

This thesis included many contributions as followings:

- A novel approach to segmenting fingernails and knuckle creases for individual identification: The study enhanced the multi-view bootstrapping, a well-known hand detection technique, to automatically label and segment hand components.
- Dividing the left and right human dorsal hands into 19 separate halves, considering the fingernails and the knuckle creases. The MCP, PIP, DIP, and the fingernails of the five fingers on the hand's left and right dorsal surfaces are among the hands' areas that can be sensed.
- Examining a wide range of DLNN base models, such as VGG16, ResNet, MobileNetV2, and others, as well as similarity metrics, such as Cosine, City-block, and Bray-Curtis, for feature extraction and similarity matching in the PIFK framework for suggested person identification. According to the analysis, the DenseNet201 and Bray-Curtis metric base models performed the best.
- By developing the DenseNet201 DLNN model with subsets from segmented hands in the "11k Hands" and "PolyU," the author increased the person identification framework's PIFK+ performance.
- Developing PIFK+'s performance using fusion methods. The fusions were carried out based on each finger's component or region, including the fingernails, major, minor, base knuckle creases, and the components of the entire hand. Additionally, researching the fusion using Bray-Curtis, Cosine, and Euclidean similarity metrics. Finally, the majority voting (MV) and weighted average (WA) fusion rules were applied.
- Based on the Siamese network, a cutting-edge deep learning method, a novel framework for identifying individuals was introduced. It involved recognizing persons based on several sub-images of the human dorsal hand. DLNN models, including DenseNet201, ResNet152, and VGG16, were also investigated for feature extraction. In order to enhance the identification outcome, DenseNet201, the best model of DLNN, received fine-tuning.
- The fingernail components outperformed the other components, which is an exciting finding. For example, the prominent knuckle creases often yielded better

identification results than other features. Additionally, compared to the right segmented sub-images of the hands, the left segmented sub-images showed more remarkable recognition ability.

The first chapter included identifying the main problem of the thesis, which was identifying persons. Then, the chapter introduced human identification, biometric recognition, its techniques, characteristics, challenges, applications, aims, and the research structure. The second chapter explored the literature on different aspects of human identification from a hand. These aspects were image capturing devices, image types, detecting, localisation of the hand's joints, pose estimations, feature extraction of various hand components, feature matching in a uni-modal recognition system, and fusion methods.

In addition, to fulfil the aim of the study, different ML approaches including DLNN, and image processing were applied to build person identification frameworks (PIFK), enhanced PIFK^{+}, and extensions of the later framework. Many methods and approaches such as image prepossessing, including resizing, scaling, segmentations and augmentations, were used to solve the issue of identification. A critical element of this research is the feature extraction techniques. Many abstract high-level feature extraction methods, the base models of DLNN were employed, such as VGG16, MobileNet, (fine-tuned) DenseNet201, ResNet50V2 and others. In addition, the study investigated Cityblock, Bray-Curtis, Cosine, and Euclidean distance metrics to match pairs of sub-images to the correct subjects.
A type of DLNN structure called the Siamese network was utilised to find the similarity between sub-images of the human hands. This network included different base models of DLNN, such as VGG16, ResNet, and DenseNet201, for better feature extraction. The DenseNet201 was also fine-tuned and used in extracting the features of fingernails and knuckle creases.

In addition, to the best of our knowledge, this is the first research to use a holistic matching approach in sub-images of various knuckles and fingernails from all five fingers of both hands, resulting in enhanced recognition. The approach also used ensemble learning or fusion to accomplish this aim. In addition, the fusion of Bray-Curtis, Cosine, and Euclidean similarity metrics was researched, which has not before been addressed in the literature. When different score-level fusion approaches, such as MV and WA, were used, the identification became more robust and accurate. The proposed approaches were evaluated using two popular datasets, the '11k Hands', and 'PolyU'.

Interesting findings were discovered. The first finding was that the left hands are more person-identified than the right hand. The second finding was that although the knuckle creases are a more permanent biometric trait, the fingernails have a rigid

shape that leads to easier individual identification. Finally, the major and minor knuckle patterns are relatively more distinguishable than the base knuckle patterns. In addition, the Bray-Curtis similarity metric is the best compared to other metrics, which was found when implementing the $PIFK$ and $PIFK^{+}$ frameworks.

Unlike other studies, e.g,(Afifi, 2019) that investigated the identification of individuals based on the whole hand, this research studies all the elements of the hand including fingernails and knuckle patterns to build a uni-model PIFK which, to the best of the author's knowledge, have not been studied before.

This study can be a good starting point for person identification from the hand image using RGB images of the hand in a flat dorsal position. However, various challenges can be encountered, such as the quality of the acquired hand images. They can be of poor quality, with missing details due to occlusions, and poor lighting. Furthermore, the hand image can be acquired with different poses. For example, dorsal angled, palm angled position, occluded hand, or any other uncontrolled position in an uncontrolled environment and background. These conditions can affect the identification system's performance and decrease overall performance. Thus, it is helpful to construct an extensive hand image database combining all these elements and focus on constructing a more sophisticated identification framework.

6.3 Future work

New directions of research in person identification problems can be drawn. Many methods implemented in computer vision tasks can be applied in person identification based on the hand. First, the localisation and detection of the dorsal hand components, including knuckle patterns and fingernails, can be further investigated and improved. This thesis mainly used a popular hand pose estimation method called the multi-view bootstrapping method to identify the location of hand joints using a pretrained detector. However, the detector can sometimes identify the joint incorrectly. Thus, the detection of hand components needs to be improved. In addition, we can compare images of hand components with poorly detected joints with those of good-quality detected joints.

Also, many supervised detection methods based on DLNN are available in the literature for different computer vision tasks. For example, these methods can be investigated in the problem of knuckle pattern and fingernail regions' localisation and detection tasks. These methods include YOLO, Regions with CNN features (R-CNN), and Faster R-CNN. The latter method can employ VGG19, InceptionV2, or others pretrained CNN models as a backbone. Also, combining different methods, such as YOLO and R-CNN, can improve the localisation task.

The U-net method can be used to predict the pattern of the minor, major, and

base knuckles or the vein of the dorsal hand when an image passes to the method and can generate the pattern of the knuckle. Different similarity distances can then be used to predict the person's identity.

The process of giving each pixel a label is known as semantic segmentation. Semantic segmentation aims to divide an image into pairs of mutually exclusive subsets, each corresponding to a significant portion of the original image. Since semantic segmentation seeks to separate each object region from the background region precisely, it has a more significant application in hand-based person identification. It can be used to segment each pixel of a dorsal hand image into different components, such as a minor, major, base knuckle, and fingernail. Diverse techniques have used deep neural networks to produce encouraging results in semantic segmentation. In general, a deep neural network is trained to learn a mapping between a semantic label and its varied visible representations by providing enough images and associated pixel-wise labelling maps as training data (Hao, Yuan Zhou, and Guo, 2020). Many DLNN models are frequently employed in semantic segmentation. For example, VGG, with its different versions like VGG-13, VGG-16, and VGG-19, can serve as the framework for several semantic segmentation models.

A recurrent neural network (RNN)-based alternative to convolutional networks (ReNet) offers an alternate method of building the network architecture. The network substitutes multi-direction RNNs for convolutional layers. An RNN-based model for semantic segmentation (ReSeg) is a ReNet-based representative semantic segmentation technique. Many semantic segmentation techniques, like the DeepLab family and PSPNet, use ResNet as their backbone. Also, there are three examples of DenseNet-based semantic segmentation techniques: Densely connected Atrous Spatial Pyramid Pooling (DenseASPP) (M. Yang et al., 2018), Fully Convolutional DenseNet (FC-DenseNet) (Jégou et al., 2017), and Stacked Deconvolutional Network (SDNet) (Fu et al., 2019). DenseNet's advantage connects every layer, in contrast to the conventional approach that makes a network broader (Hao, Yuan Zhou, and Guo, 2020).

In addition, instance segmentation can be employed to segment the hand into various components. Semantic segmentation provides fine inference by anticipating labels for each pixel in the incoming image. Each pixel is assigned a label based on the object class it is included in. Regions with CNN features (RCNN), Fast RCNN, Faster RCNN, YOLO, and Detectron2 can be used as DLNN-based detectors, for instance segmentation (Hafiz and Bhat, 2020).

DLNN, or representation learning, is a popular machine-learning technique that has shown tremendous success recently. It can be used for high-level feature extraction in a person identification framework. There are many structures of DLNN. For example, recursive neural networks (RvNN), recurrent neural networks (RNN), convolutional neural networks (CNN), and Deep Generative Networks. These

networks could be investigated and optimised using many procedures, such as identifying the network parameters that reduce the loss function.

Then, these optimised models can be used for extracting the features of hand joints. These features can be clustered based on different clustering methods, for example, hierarchical, k-mean, density-based clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Ordering Points to Identify Clustering Structure (OPTICS), hierarchical density-based spatial clustering of applications with noise (HDBSCAN), fuzzy clustering, grid-based clustering, and many others. Also, comparisons between these methods can be conducted.

In addition, an image can be segmented based on different colours using a fuzzy-based method, and an interpretable model can be built using semantic segmentation. The hand image can be semantically segmented using a prototype-based technique. The most representative local focal points in terms of the features are prototypes. It is possible to utilise linguistic IF...THEN rules to describe the results obtained while choosing the winning class label by working in a latent feature space and projecting to the raw feature space. Scientists and specialists make these rules comprehensible and interpretable and they can enable users to examine and audit the decision-making process.

To extend the framework, ensemble learning or fusion techniques can be investigated in the issue of person identification from fingernail and knuckle patterns. Different types of fusion techniques can be implemented, such as feature-level fusion. Features extracted from the knuckles, fingernails, dorsal veins, finger veins, and palm prints can be combined. Also, these features can be integrated into the score or algorithm level. One feature can be obtained at the algorithm level using different methods or algorithms. Then, the obtained features can be fused at the sensor, feature, score, or decision level.

Also, fusion can be employed with the Siamese network. Different images of hand elements, for example dorsal vein, and dorsal knuckles, can be passed to the Siamese network, which can obtain the elements' features using the different base models of finetuning model of DLNN such as DenseNet201, VGG19, ResNetV2, or ResNet152. Also, various distances can be utilised to accomplish the best result. Then, the fusion can be implemented at the feature or score level.

Another study direction is the multi-modalities recognition system of different sub-image inputs of different hand features combined with different algorithms. Various features can be obtained from handcrafted methods that support interpretability and DLNN-based methods that support high performance. Thus, the advantages of both approaches can be achieved.

References

Afifi, Mahmoud (2019). "11K Hands: Gender recognition and biometric identification using a large dataset of hand images". In: *Multimedia Tools and Applications* 78, pp. 20835–20854. DOI: 10.1007/s11042-019-7424-8. URL: https://doi.org/10.1007/s11042-019-7424-8.

Agnihotri, Manish et al. (2019). "Learning Domain Specific Features using Convolutional Autoencoder: A Vein Authentication Case Study using Siamese Triplet Loss Network." In: *ICPRAM*, pp. 778–785.

Ahmed, Ejaz, Michael Jones, and Tim K Marks (2015). "An improved deep learning architecture for person re-identification". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3908–3916.

Alghamdi, M. et al. (2019). "Self-Organising and Self-Learning Model for Soybean Yield Prediction". In: *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS 2019*. ISBN: 9781728129464. DOI: 10.1109/SNAMS.2019.8931888.

Alghamdi, Mona (2022). "A Multi-modal Biometric Approach Based on Score-level Fusion and Fine-tuning Deep Learning Features". In: *IEEE IS'22*.

Alghamdi, Mona, Plamen Angelov, and Lopez Pellicer Alvaro (2022). "Person identification from fingernails and knuckles images using deep learning features and the Bray-Curtis similarity measure". In: *Neurocomputing* 513, pp. 83–93. ISSN: 18728286. DOI: 10.1016/j.neucom.2022.09.123. URL: https://doi.org/10.1016/j.neucom.2022.09.123.

Alghamdi, Mona, Plamen Angelov, and Bryan Williams (2021). "Automated Person Identification Framework Based on Fingernails and Dorsal Knuckle Patterns". In: *2021 IEEE Symposium Series on Computational Intelligence, SSCI 2021 - Proceedings*. DOI: 10.1109/SSCI50451.2021.9659850.

Alshehri, Helala et al. (2018). "Cross-Sensor Fingerprint Matching Method Based on Orientation, Gradient, and Gabor-HoG Descriptors With Score Level Fusion". In: *IEEE Access* 6, pp. 28951–28968. DOI: 10.1109/ACCESS.2018.2840330.

Amraoui, Amine, Youssef Fakhri, and Mounir Ait Kerroum (2017). "Finger knuckle print recognition system using compound local binary pattern". eng. In: *2017 International Conference on Electrical and Information Technologies (ICEIT)*.

Vol. 2018-. IEEE, pp. 1–5. ISBN: 1538615169. DOI: 10.1109/EITech.2017.8255216.

Anbari, Mohammad and Ali M Fotouhi (2021). "Finger knuckle print recognition for personal authentication based on relaxed local ternary pattern in an effective learning framework". In: 32, p. 55. DOI: 10.1007/s00138-021-01178-6. URL: https://doi.org/10.1007/s00138-021-01178-6.

Andriluka, Mykhaylo et al. (2014). *2D Human Pose Estimation: New Benchmark and State of the Art Analysis*. Tech. rep.

Angelov, Plamen P and Xiaowei Gu (2018). "Deep rule-based classifier with human-level performance and characteristics". In: *Information Sciences* 463, pp. 196–213.

Attia, Abdelouahab (2022). "Deep learning-driven palmprint and finger knuckle pattern-based multimodal Person recognition system". In: pp. 10961–10980.

Attia, Abdelouahab, Zahid Akhtar, and Youssef Chahir (2021). "Feature-level fusion of major and minor dorsal finger knuckle patterns for person authentication". In: *Signal, Image and Video Processing* 15.4, pp. 851–859.

Attia, Abdelouahab, Zahid Akhtar, Nour Elhouda Chalabi, et al. (2020). "Deep rule-based classifier for finger knuckle pattern recognition system". In: *Evolving Systems* 0123456789. ISSN: 18686486. DOI: 10.1007/s12530-020-09359-w. URL: https://doi.org/10.1007/s12530-020-09359-w.

Attia, Mohamed et al. (2019). "Fingerprint synthesis via latent space representation". In: *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, pp. 1855–1861.

Babenko, Artem et al. (2014). "Neural codes for image retrieval". In: *European conference on computer vision*. Springer, pp. 584–599.

Bolle, Ruud M et al. (2005). "The relation between the ROC curve and the CMC". In: *Fourth IEEE workshop on automatic identification advanced technologies (AutoID'05)*. IEEE, pp. 15–20.

Bromley, Jane et al. (1994). "Signature Verification using a "Siamese" Time Delay Neural Network". In.

Cao, Zhe et al. (2017). "Realtime multi-person 2D pose estimation using part affinity fields". In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-January.Xxx, pp. 1302–1310. DOI: 10.1109/CVPR.2017.143. arXiv: arXiv:1812.08008v2.

Chlaoua, Rachid et al. (2019). "Deep learning for finger-knuckle-print identification system based on PCANet and SVM classifier". In: *Evolving Systems* 10.2, pp. 261–272.

Chopra, Sumit, Raia Hadsell, and Yann LeCun (2005). "Learning a similarity metric discriminatively, with application to face verification". In: *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005* I, pp. 539–546. DOI: 10.1109/CVPR.2005.202.

Choudhury, Surabhi Hom, Amioy Kumar, and Shahedul Haque Laskar (2019). "Biometric authentication through unification of finger dorsal biometric traits". In: *Information sciences* 497, pp. 202–218.

Connor, Patrick and Arun Ross (Feb. 2018). "Biometric recognition by gait: A survey of modalities and features". In: *Computer Vision and Image Understanding* 167, pp. 1–27. ISSN: 1090235X. DOI: 10.1016/j.cviu.2018.01.007.

Csurka, Gabriella et al. (2004). "Visual categorization with bags of keypoints". In: *Workshop on statistical learning in computer vision, ECCV*. Vol. 1. 1-22. Prague, pp. 1–2.

Daas, Sara et al. (Dec. 2020). "Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion". In: *IET Image Processing* 14 (15), pp. 3859–3868. ISSN: 17519659. DOI: 10.1049/iet-ipr.2020.0491.

Daniel III, C Ralph, Bianca Maria Piraccini, and Antonella Tosti (2004). "The nail and hair in forensic science". In: *Journal of the American Academy of Dermatology* 50.2, pp. 258–261.

Erol, Ali et al. (2007). "Vision-based hand pose estimation: A review". In: *Computer Vision and Image Understanding* 108.1-2, pp. 52–73.

Fei, Lunke et al. (2019). "LEARNING DISCRIMINATIVE FINGER-KNUCKLE-PRINT DESCRIPTOR Department of Computer and Information Science , University of Macau , Taipa , Macau , China School of Computer Science and Technology , Guangdong University of Technology , Guangzhou , China". In: pp. 2137–2141.

Ferrer, Miguel A et al. (2007). "Low cost multimodal biometric identification system based on hand geometry, palm and finger print texture". In: *2007 41st Annual IEEE International Carnahan Conference on Security Technology*. IEEE, pp. 52–58.

Font-Aragones, Xavier, Marcos Faundez-Zanuy, and Jiri Mekyska (2013). "Thermal hand image segmentation for biometric recognition". In: *IEEE aerospace and electronic systems magazine* 28.6, pp. 4–14.

Fu, Jun et al. (2019). "Stacked deconvolutional network for semantic segmentation". In: *IEEE Transactions on Image Processing*.

Gao, Guangwei et al. (2019). "Reconstruction in gabor response domain for efficient finger-knuckle-print verification". In: *ANZCC 2018 - 2018 Australian and New Zealand Control Conference*, pp. 110–114. DOI: 10.1109/ANZCC.2018.8606590.

Ge, Liuhao et al. (2019). "3D Hand Shape and Pose Estimation from a Single RGB Image". In.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep learning*. MIT press.

Green, Manfred. (2009). *Identity, security and democracy the wider social and ethical implications of automated systems for human identification*. eng. Ed. by Manfred Green et al. NATO science for peace and security. Series E, Human and societal dynamics, vol. 49. Amsterdam ; Oxford: Ios Press. ISBN: 6612070552.

Grover, Jyotsana and Madasu Hanmandlu (2015). "Hybrid fusion of score level and adaptive fuzzy decision level fusions for the finger-knuckle-print based authentication". In: *Applied Soft Computing Journal* 31, pp. 1–13. ISSN: 15684946. DOI: 10.1016/j.asoc.2015.02.001. URL: http://dx.doi.org/10.1016/j.asoc.2015.02.001.

Grzejszczak, Tomasz, Michal Kawulok, and Adam Galuszka (2016). "Hand landmarks detection and localization in color images". In: *Multimedia Tools and Applications* 75.23, pp. 16363–16387.

Guru, D. S., K. B. Nagasundara, and S. Manjunath (2010). "Feature level fusion of multi-instance finger knuckle print for person identification". In: *ACM International Conference Proceeding Series*, pp. 186–190. DOI: 10.1145/1963564.1963595.

H-unique (2019). https://www.lancaster.ac.uk/security-lancaster/research/h-unique/. [Online; accessed 19-October-2022].

Hafiz, Abdul Mueed and Ghulam Mohiuddin Bhat (2020). "A survey on instance segmentation: state of the art". In: *International journal of multimedia information retrieval* 9.3, pp. 171–189.

Hammouche, Rabah, Abdelouahab Attia, and Samir Akhrouf (2022). "Score level fusion of major and minor finger knuckle patterns based symmetric sum-based rules for person authentication". In: *Evolving Systems* 13.3, pp. 469–483. ISSN: 18686486. DOI: 10.1007/s12530-022-09430-8. URL: https://doi.org/10.1007/s12530-022-09430-8.

Han, Jiawei, Micheline Kamber, and Jian Pei (2012). *Data mining concepts and techniques*. eng. Ed. by Micheline Kamber and Jian Pei. 3rd ed. Morgan Kaufmann series in data management systems. Burlington, Mass.: Elsevier. ISBN: 1-283-17117-1.

Han, Wei-Yu and Jen-Chun Lee (2012). "Palm vein recognition using adaptive Gabor filter". In: *Expert Systems with Applications* 39.18, pp. 13225–13234.

Hanif, Muhammad Shehzad (2019). "Patch match networks: Improved two-channel and Siamese networks for image patch matching". In: *Pattern Recognition Letters* 120, pp. 54–61.

Hao, Shijie, Yuan Zhou, and Yanrong Guo (2020). "A brief survey on semantic segmentation with deep learning". In: *Neurocomputing* 406, pp. 302–321.

He, Kaiming et al. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Heidari, Hadis and Abdolah Chalechale (2022). "Biometric authentication using a deep learning approach based on different level fusion of finger knuckle print and fingernail". In: *Expert Systems with Applications* 191, p. 116278.

Hom Choudhury, Surabhi, Amioy Kumar, and Shahedul Haque Laskar (2019). "Biometric Authentication through Unification of Finger Dorsal Biometric Traits". In: *Information Sciences* 497, pp. 202–218. ISSN: 00200255. DOI: 10.1016/j.ins.2019.05.045. URL: https://doi.org/10.1016/j.ins.2019.05.045.

Hong, Lin, Anil K Jain, and Sharath Pankanti (1999). "Can multibiometrics improve performance?" In: *Proceedings AutoID*. Vol. 99. Citeseer, pp. 59–64.

Howard, Andrew G. et al. (2017). "MobileNets: Efficient convolutional neural networks for mobile vision applications". In: *arXiv*. ISSN: 23318422. arXiv: 1704.04861.

Huang, Gao et al. (2017). "Densely connected convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

Jain, A K, A Ross, and S Prabhakar (2004). "An introduction to biometric recognition". In: *IEEE Transactions on Circuits and Systems for Video Technology* 14.1, pp. 4–20. ISSN: 1051-8215. DOI: 10.1109/TCSVT.2003.818349.

Jain, Anil, Lin Hong, and Ruud Bolle (1997). "On-line fingerprint verification". In: *IEEE transactions on pattern analysis and machine intelligence* 19.4, pp. 302–314.

Jain, Anil K et al. (2008). *Handbook of biometrics*. New York: Springer.

Jaswal, Gaurav, Amit Kaul, and Ravinder Nath (2016). "Knuckle print biometrics and fusion schemes - Overview, challenges, and solutions". In: *ACM Computing Surveys* 49.2. ISSN: 15577341. DOI: 10.1145/2938727.

Jaswal, Gaurav, Aditya Nigam, and Ravinder Nath (2017a). "DeepKnuckle: revealing the human identity". In: *Multimedia Tools and Applications* 76.18, pp. 18955–18984. ISSN: 15737721. DOI: 10.1007/s11042-017-4475-6.

— (2017b). "Finger knuckle image based personal authentication using DeepMatching". In: *2017 IEEE International Conference on Identity, Security and Behavior Analysis, ISBA 2017*. DOI: 10.1109/ISBA.2017.7947706.

Jégou, Simon et al. (2017). "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 11–19.

Jha, Ranjeet Ranjan et al. (2020). "PixISegNet: pixel-level iris segmentation network using convolutional encoder–decoder with stacked hourglass bottleneck". In: *IET biometrics* 9 (1), pp. 11–24. ISSN: 2047-4938. DOI: 10.1049/iet-bmt.2019.0025.

Jia Deng et al. (2009). "ImageNet: A large-scale hierarchical image database". eng. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 248–255. ISBN: 9781424439928. DOI: 10.1109/CVPR.2009.5206848.

Joshi, JC et al. (2019). "Finger Knuckleprint based personal authentication using siamese network". In: *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, pp. 282–286.

Juefei-Xu, Felix, Eshan Verma, and Marios Savvides (2017). "DeepGender2: A generative approach toward occlusion and low-resolution robust facial gender classification via progressively trained attention shift convolutional neural networks (PTAS-CNN) and deep convolutional generative adversarial networks (DCGAN)". In: *Deep Learning for Biometrics*. Springer, pp. 183–218.

Jun, Wang and Wang Guoqing (2017). "Quality-Specific Hand Vein Recognition System". In: *IEEE Transactions on Information Forensics and Security* 12.11, pp. 2599–2610. ISSN: 1556-6013. DOI: 10.1109/TIFS.2017.2713340.

Kamboj, Aman et al. (2021). "CED-Net: context-aware ear detection network for unconstrained images". In: *Pattern Analysis and Applications* 24, pp. 779–800. DOI: 10.1007/s10044-020-00914-4. URL: https://doi.org/10.1007/s10044-020-00914-4.

Khan, Salman (2018). *A Guide to Convolutional Neural Networks for Computer Vision*. San Rafael.

Khellat-Kihel, S et al. (2016). "Multimodal fusion of the finger vein, fingerprint and the finger-knuckle-print using Kernel Fisher analysis". In: *Applied Soft Computing* 42, pp. 439–447.

Kim, Chan Sik, Nam Sun Cho, and Kang Ryoung Park (2019). "Deep Residual Network-Based Recognition of Finger Wrinkles Using Smartphone Camera". In: *IEEE Access* 7, pp. 71270–71285. ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2920391.

Koch, Gregory et al. (2015). "Siamese Neural Networks for One-Shot Image Recognition". In: *Cs.Toronto.Edu* 2.14, pp. 900–916. ISSN: 00278424. arXiv: 1906.05586. URL: http://www.cs.toronto.edu/%7B~%7Dgkoch/files/msc-thesis.pdf.

Kulkarni, Sujata, R.D Raut, and P.K Dakhole (2015). "Wavelet Based Modern Finger Knuckle Authentication". eng. In: *Procedia computer science* 70, pp. 649–657. ISSN: 1877-0509. DOI: 10.1016/j.procs.2015.10.101.

Kumar, A (2012). "Can we use minor finger knuckle images to identify humans?" eng. In: *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, pp. 55–60. ISBN: 9781467313841. DOI: 10.1109/BTAS.2012.6374558.

Kumar, Ajay (2014). "Importance of Being Unique From Finger Dorsal Patterns: Exploring Minor Finger Knuckle Patterns in Verifying Human Identities". eng. In: *IEEE transactions on information forensics and security* 9.8, pp. 1288–1298. ISSN: 1556-6013. DOI: 10.1109/TIFS.2014.2328869.

Kumar, Ajay and Ch Ravikanth (2009). *Personal Authentication using Finger Knuckle Surface*. Tech. rep. 1, pp. 98–110.

Kumar, Ajay and Bichai Wang (2015). "Recovering and matching minutiae patterns from finger knuckle images". In: *Pattern Recognition Letters* 68, pp. 361–367. ISSN: 01678655. DOI: 10.1016/j.patrec.2015.08.013.

Kumar, Ajay and Xu Zhihuan (2016). "Personal Identification Using Minor Knuckle Patterns From Palm Dorsal Surface". In: *IEEE Transactions on Information Forensics and Security* 11 (10), pp. 2338–2348. ISSN: 1556-6013. DOI: 10.1109/TIFS.2016.2574309.

Kumar, Ajay and Yingbo Zhou (2009). "Human identification using knucklecodes". In: *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*. IEEE, pp. 1–6.

Kumar, Amioy, Shruti Garg, and Madasu Hanmandlu (2014). "Biometric authentication using finger nail plates". In: *Expert systems with applications* 41.2, pp. 373–386.

Kurosaki, K, T Matsushita, and S Ueda (1993). "Individual DNA identification from ancient human remains". In: *American journal of human genetics* 53.3, p. 638. ISSN: 0002-9297.

Kuzu, Ridvan Salih et al. (2020). "On-the-Fly Finger-Vein-Based Biometric Recognition Using Deep Neural Networks". In: *IEEE transactions on information forensics and security* 15, pp. 2641–2654. ISSN: 1556-6013. DOI: 10.1109/TIFS.2020.2971144.

Lahat, Dana, Tülay Adali, and Christian Jutten (2015). "Multimodal data fusion: an overview of methods, challenges, and prospects". In: *Proceedings of the IEEE* 103.9, pp. 1449–1477.

Lecun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep learning". In: *Nature* 521.7553, pp. 436–444. ISSN: 14764687. DOI: 10.1038/nature14539.

Lee, J-C et al. (2014). "Dorsal hand vein recognition based on 2D Gabor filters". In: *The Imaging Science Journal* 62.3, pp. 127–138.

Li, Xiaoxia, Di Huang, and Yunhong Wang (2016). "Comparative study of deep learning methods on dorsal hand vein recognition". In: *Chinese Conference on Biometric Recognition*. Springer, pp. 296–306.

Lin, Chenhao and Ajay Kumar (2017). "Multi-siamese networks to accurately match contactless to contact-based fingerprint images". In: *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, pp. 277–285.

Ma, Chao et al. (2015). "Hierarchical convolutional features for visual tracking". In: *Proceedings of the IEEE international conference on computer vision*, pp. 3074–3082.

Majtner, Tomáš, Sule Yildirim-Yayilgan, and Jon Yngve Hardeberg (2017). "Combining deep learning and hand-crafted features for skin lesion classification". In: *2016 6th International Conference on Image Processing Theory, Tools and Applications, IPTA 2016*. DOI: 10.1109/IPTA.2016.7821017.

Melekhov, Iaroslav, Juho Kannala, and Esa Rahtu (2016). *Siamese Network Features for Image Matching*. Tech. rep.

Minaee, Shervin et al. (2019). "Biometrics recognition using deep learning: A survey". In: *arXiv preprint arXiv:1912.00271*.

Mitra, Sushmita and Tinku Acharya (2007). "Gesture recognition: A survey". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37.3, pp. 311–324.

Muthukumar, A. and A. Kavipriya (2019). "A biometric system based on Gabor feature extraction with SVM classifier for Finger-Knuckle-Print". In: *Pattern Recognition Letters* 125, pp. 150–156. ISSN: 01678655. DOI: 10.1016/j.patrec.2019.04.007. URL: https://doi.org/10.1016/j.patrec.2019.04.007.

Nanni, Loris, Stefano Ghidoni, and Sheryl Brahnam (2017). "Handcrafted vs. non-handcrafted features for computer vision classification". In: *Pattern recognition* 71, pp. 158–172. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2017.05.025.

Neverova, Natalia et al. (2015). "Hand segmentation with structured convolutional learning". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9005, pp. 687–702. ISSN: 16113349. DOI: 10.1007/978-3-319-16811-1_45.

Nigam, Aditya, Kamlesh Tiwari, and Phalguni Gupta (May 2016). "Multiple texture information fusion for finger-knuckle-print authentication system". In: *Neurocomputing* 188, pp. 190–205. ISSN: 18728286. DOI: 10.1016/j.neucom.2015.04.126. URL: http://dx.doi.org/10.1016/j.neucom.2015.04.126.

Poppe, Ronald (2010). "A survey on vision-based human action recognition". In: *Image and vision computing* 28.6, pp. 976–990.

Qian, Jianjun, Jian Yang, and Guangwei Gao (2013). "Discriminative histograms of local dominant orientation (D-HLDO) for biometric image feature extraction". In: *Pattern Recognition* 46.10, pp. 2724–2739. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2013.03.005.

Qian, Jianjun, Jian Yang, Ying Tai, et al. (2016). "Exploring deep gradient information for biometric image feature representation". In: *Neurocomputing* 213, pp. 162–171. ISSN: 18728286. DOI: 10.1016/j.neucom.2015.11.135. URL: http://dx.doi.org/10.1016/j.neucom.2015.11.135.

Le-Qing, Zhu and Zhang San-Yuan (2010). "Multimodal biometric identification system based on finger geometry, knuckle print and palm print". In: DOI: 10.1016/j.patrec.2010.05.010.

Ranjan Jha, Ranjeet et al. (2017). *UBSegNet: Unified Biometric Region of Interest Segmentation Network*. Tech. rep. arXiv: 1709.08924v1.

Rawat, Waseem and Zenghui Wang (2017). "Deep convolutional neural networks for image classification: A comprehensive review". In: *Neural computation* 29.9, pp. 2352–2449.

Reid, Paul (2004). *Biometrics for network security*. Prentice Hall Professional.

Ren, Shaoqing et al. (2017). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE transactions on pattern analysis and machine intelligence* 39.6, pp. 1137–1149. ISSN: 0162-8828.

Romero Lopez, Adria et al. (2017). "Skin lesion classification from dermoscopic images using deep learning techniques". In: *Proceedings of the 13th IASTED International Conference on Biomedical Engineering, BioMed 2017*, pp. 49–54. DOI: 10.2316/P.2017.852-053.

Schaefer, Gerald et al. (2011). "Colour and contrast enhancement for improved skin lesion segmentation". In: *Computerized Medical Imaging and Graphics* 35, pp. 99–104. DOI: 10.1016/j.compmedimag.2010.08.004.

Schmidhuber, Jürgen (2015). "Deep learning in neural networks: An overview". In: *Neural Networks* 61, pp. 85–117. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2014.09.003.

Shariatmadar, Zahra S. and Karim Faez (2011). "A novel approach for Finger-Knuckle-Print recognition based on Gabor feature fusion". In: *Proceedings - 4th International Congress on Image and Signal Processing, CISP 2011* 3, pp. 1480–1484. DOI: 10.1109/CISP.2011.6100450.

Shyam, Radhey and Yogendra Narain Singh (2015). "Face recognition using augmented local binary pattern and Bray Curtis dissimilarity metric". In: *2nd International Conference on Signal Processing and Integrated Networks, SPIN 2015*, pp. 779–784. DOI: 10.1109/SPIN.2015.7095267.

Sid, K. et al. (2017). "Finger Knuckle print features extraction using simple deep learning method". In: *Int. J. Comput. Sci. Commun. Inf. Technol. (CSCIT)* 5, pp. 12–18.

Sid, Khaled Ben et al. (2017). "Finger Knuckle Print Features Extraction using Simple Deep Learning Method". In: 5, pp. 12–18.

Sim, Hiew Moi et al. (2014). "Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images". eng. In: *Expert systems with applications* 41.11, pp. 5390–5404. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2014.02.051.

Simon, Tomas et al. (2017). *Hand Keypoint Detection in Single Images using Multiview Bootstrapping*. Tech. rep. arXiv: 1704.07809v1.

Simonyan, Karen and Andrew Zisserman (2014). "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556*.

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

Spurr, Adrian et al. (2018). "Cross-modal deep variational hand pose estimation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 89–98.

Stratton, H (2015). "Dorsal hand feature analysis an aid to forensic human identification". Thesis.

Su, Hong-Ren et al. (2017). "A deep learning approach towards pore extraction for high-resolution fingerprint recognition". In: IEEE, pp. 2057–2061. ISBN: 1509041176. DOI: 10.1109/ICASSP.2017.7952518.

Sun, Yi, Michael Reale, and Lijun Yin (2008). "Recognizing partial facial action units based on 3D dynamic range data for facial expression recognition". In: *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, pp. 1–8.

Sun, Zhenan et al. (2005). "Ordinal palmprint represention for personal identification [represention read representation]". In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. IEEE, pp. 279–284.

Sundararajan, Kalaivani and Damon L. Woodard (Apr. 2018). *Deep learning for biometrics: A survey*. DOI: 10.1145/3190618.

Taigman, Yaniv et al. (2014). "Deepface: Closing the gap to human-level performance in face verification". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1708.

Tan, Xiaoyang and Bill Triggs (2010). "Enhanced local texture feature sets for face recognition under difficult lighting conditions". In: *IEEE transactions on image processing* 19.6, pp. 1635–1650.

Tandel, Gopal S., Ashish Tiwari, and O. G. Kakde (2021). "Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification". In: *Computers in Biology and Medicine* 135.June, p. 104564. ISSN: 18790534. DOI: 10.1016/j.compbiomed.2021.104564. URL: https://doi.org/10.1016/j.compbiomed.2021.104564.

Tang, Su et al. (2019). "Finger vein verification using a Siamese CNN". In: *IET Biometrics* 8.5, pp. 306–315. ISSN: 20474946. DOI: 10.1049/iet-bmt.2018.5245.

Tarawneh, Ahmad S et al. (2022). "DeepKnuckle: Deep Learning for Finger Knuckle Print Recognition". In: *Electronics* 11.4, p. 513.

Thapar, Daksh, Gaurav Jaswal, and Aditya Nigam (2019). "FKIMNet: A Finger Dorsal Image Matching Network Comparing Component (Major, Minor and Nail) Matching with Holistic (Finger Dorsal) Matching". In.

Uhl, Andreas (2019). "State of the Art in Vascular Biometrics". In: DOI: 10.1007/978-3-030-27731-4_1.

Usha, K and M Ezhilarasan (2016a). "Fusion of geometric and texture features for finger knuckle surface recognition". eng. In: *Alexandria engineering journal* 55.1, pp. 683–697. ISSN: 1110-0168. DOI: 10.1016/j.aej.2015.10.003.

— (2016b). "Personal recognition using finger knuckle shape oriented features and texture analysis". eng. In: *Journal of King Saud University. Computer and information sciences* 28.4, pp. 416–431. ISSN: 1319-1578. DOI: 10.1016/j.jksuci.2015.02.004.

Usha, K. and M. Ezhilarasan (2018). "Robust personal authentication using finger knuckle geometric and texture features". In: *Ain Shams Engineering Journal* 9.4, pp. 549–565. ISSN: 20904479. DOI: 10.1016/j.asej.2016.04.006. URL: https://doi.org/10.1016/j.asej.2016.04.006.

Vančo, Marek, Ivan Minarik, and Gregor Rozinaj (2014). "Evaluation of static hand gesture algorithms". In: *IWSSIP 2014 Proceedings*. IEEE, pp. 83–86.

Vizilter, Yuri Valentinovich et al. (2017). "Real-time face identification via cnn and boosted hashing forest". In: *Computer Optics* 41 (2), pp. 254–265. ISSN: 0134-2452. DOI: 10.18287/2412-6179-2017-41-2-254-265.

Vyas, Ritesh, Hossein Rahmani, et al. (2021). "Robust End-To-End Hand Identification via Holistic Multi-Unit Knuckle Recognition". In: *2021 IEEE International Joint Conference on Biometrics, IJCB 2021*. DOI: 10.1109/IJCB52358.2021.9484356.

Vyas, Ritesh, Bryan M Williams, et al. (2022). "Ensemble-Based Bounding Box Regression for Enhanced Knuckle Localization". In: *Sensors* 22.4, p. 1569.

Waghode, Arti B and CA Manjare (2017). "Biometric Authentication of Person using finger knuckle". In: *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. IEEE, pp. 1–6.

Wang, Yangang, Baowen Zhang, and Cong Peng (2020). "SRHandNet: Real-Time 2D Hand Pose Estimation With Simultaneous Region Localization". eng. In: *IEEE transactions on image processing* 29, pp. 2977–2986. ISSN: 1941-0042. DOI: 10.1109/tip.2019.2955280.

Wang, Yiding et al. (2010). "Study of hand-dorsa vein recognition". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6215 LNCS, pp. 490–498. ISSN: 03029743. DOI: 10.1007/978-3-642-14922-1_61.

Wildes, R.P (1997). "Iris recognition: an emerging biometric technology". eng. In: *Proceedings of the IEEE* 85.9, pp. 1348–1363. ISSN: 0018-9219. DOI: 10.1109/5.628669.

Wong, Tzu-Tsung (2015). "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation". In: *Pattern recognition* 48 (9), pp. 2839–2846. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2015.03.009.

Wu, Jonathan, Prakash Ishwar, and Janusz Konrad (2016). "Two-stream CNNs for gesture-based verification and identification: Learning user style". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 42–50.

Yang, Maoke et al. (2018). "Denseaspp for semantic segmentation in street scenes". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3684–3692.

Yann, Lecun, Bengio Yoshua, and Hinton Geoffrey (2015). "Deep learning". In: *Nature* 521.7553, p. 436. ISSN: 0028-0836. DOI: 10.1038/nature14539.

Yoruk, Erdem et al. (2006). "Shape-based hand recognition". In: *IEEE transactions on image processing* 15.7, pp. 1803–1815.

Yu, Hongyang et al. (2015). "A new finger-knuckle-print ROI extraction method based on Two-stage center point detection". In: *International Journal of Signal Processing, Image Processing and Pattern Recognition* 8.2, pp. 185–200.

Yuksel, Aycan, Lale Akarun, and Bulent Sankur (2011). "Hand vein biometry based on geometry and appearance methods". In: *IET computer vision* 5.6, pp. 398–406.

Zhai, Yikui et al. (2018). "A novel finger-knuckle-print recognition based on batch-normalized CNN". In: *Chinese conference on biometric recognition*. Springer, pp. 11–21.

Zhang, Chao et al. (2020). "Multimodal intelligence: Representation learning, information fusion, and applications". In: *IEEE Journal of Selected Topics in Signal Processing* 14.3, pp. 478–493.

Zhang, D et al. (2003). "Online palmprint identification". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.9, pp. 1041–1050. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2003.1227981.

Zhang, David et al. (2003). "Online palmprint identification". In: *IEEE Transactions on pattern analysis and machine intelligence* 25.9, pp. 1041–1050.

Zhang, Fan et al. (2020). "MediaPipe Hands: On-device Real-time Hand Tracking". In: *arXiv preprint arXiv:2006.10214*.

Zhang, Lei et al. (2011). "Ensemble of local and global information for fingerknuckle-print recognition". In: *Pattern Recognition* 44.9, pp. 1990–1998. ISSN: 00313203. DOI: 10.1016/j.patcog.2010.06.007. URL: http://dx.doi.org/10.1016/j.patcog.2010.06.007.

Zhang, Lin and Hongyu Li (Dec. 2012). "Encoding local image patterns using Riesz transforms: With applications to palmprint and finger-knuckle-print recognition". In: *Image and Vision Computing* 30 (12), pp. 1043–1051. ISSN: 0262-8856. DOI: 10.1016/J.IMAVIS.2012.09.003.

Zhang, Lin, Lei Zhang, and David Zhang (2009). *Finger-Knuckle-Print Verification Based on Band-Limited Phase-Only Correlation*. Tech. rep., pp. 141–148.

Zhang, Lin, Lei Zhang, David Zhang, and Hailong Zhu (July 2010). "Online finger-knuckle-print verification for personal authentication". In: *Pattern Recognition* 43.7, pp. 2560–2571. ISSN: 00313203. DOI: 10.1016/j.patcog.2010.01.020.

Zhao, Shi, Yi-Ding Wang, and Yun-Hong Wang (2008). "Biometric identification based on low-quality hand vein pattern images". In: *2008 International Conference on Machine Learning and Cybernetics*. Vol. 2. IEEE, pp. 1172–1177.