A Multimodal Hand Vein Database and Recognition System

Felix Olanrewaju Babalola

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Engineering

Eastern Mediterranean University
July 2022
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy in Computer Engineering.

Prof. Dr. H. Işık Aybay
Chair, Department of Computer Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Computer Engineering.

Assoc. Prof. Dr. Yıltan Bitirim
Co-Supervisor

Prof. Dr. Önsen Toygar
Supervisor

Examining Committee

1. Prof. Dr. Hasan Demirel

2. Prof. Dr. Bilge Günsel

3. Prof. Dr. Fikret S. Gürgen

4. Prof. Dr. Önsen Toygar

5. Assoc. Prof. Dr. Duygu Çelik Ertuğrul

ABSTRACT

Biometric studies over the years have made the usage of physiological traits in human authentication technologies popular around the world. More recent studies in this field have given rise to more reliable, faster and user friendly security systems. One of the relatively new area of this field is hand vein biometrics where vascular patterns of hands are used for human recognition. This has some advantages over other physiological traits such as inherent spoof-proof attribute, lack of occlusion and noninvasiveness. Additionally, vein patterns can be captured from different parts of the hand, which could in turn be used in a multimodal system. Multimodal systems are generally preferred because they ensure a more robust and secure system compared to unimodal frameworks.

In general, this study introduced a hand vein database named FYO with multiple hand vein datasets for palm, dorsal and wrist vein for the purpose of implementing hand vein multimodal biometric systems. Subsequently, feature descriptors such as Histogram of Oriented Gradients, Gabor filter and Binarized Statistical Image Features, and Convolutional Neural Network models such as AlexNet, VGG-16, VGG-19 and ResNet-50 are applied to show the efficiency of the proposed methodologies. Varieties of architectures for improving the robustness of hand vein recognition systems in both unimodal and multimodal forms are proposed in this study.

Additionally, all experiments performed with the datasets acquired are similarly carried out on datasets from publicly available databases such as Badawi, Bosphorus, PUT, Tongji Contactless Palm Vein database and VERA, while the performances of the proposed systems are effectively compared to similar studies in the field.

Keywords: Multimodal biometrics, Feature fusion, Hand vein recognition, Dorsal vein, Palm vein, Wrist vein, CNN models, BSIF.

ÖZ

Yıllar boyunca yapılan biyometrik çalışmalar, insan kimlik doğrulama teknolojilerinde fizyolojik özelliklerin kullanımını dünya çapında popüler hale getirmiştir. Bu alandaki daha yakın tarihli çalışmalar, daha güvenilir, daha hızlı ve kullanıcı dostu güvenlik sistemlerine yol açmıştır. Bu alanın nispeten yeni olan alanlarından biri, ellerin vasküler örüntülerinin insan tanımasında kullanıldığı el damar biyometrisidir. Bunun, doğal sahtekarlığa karşı koruma özelliği, tıkanıklık olmaması ve invaziv olmama gibi diğer fizyolojik özelliklere göre bazı avantajları vardır. Ek olarak, elin farklı bölgelerinden damar desenleri yakalanabilir ve bu da çoklu bir sistemde kullanılabilir. Çok modlu sistemler, tek modlu çerçevelere göre daha sağlam ve güvenli bir sistem sağladıkları için genellikle tercih edilir.

Genel olarak, bu çalışma, el damarı için çoklu biyometrik sistemlerini uygulamak amacıyla avuç içi, el sırtı ve bilek damarı için çoklu el damarı veri kümeleriyle FYO adlı bir el damarı veritabanını sunmuştur. Daha sonra, Yönlendirilmiş Gradyanların Histogramı (HOG), Gabor filtresi ve İkili İstatistiksel Görüntü Öznitelikleri (BSIF) gibi öznitelik tanımlayıcıları ve AlexNet, VGG-16, VGG-19 ve ResNet-50 gibi Evrişimli Sinir Ağı (CNN) modelleri uygulanmıştır. Bu çalışmada hem tek modlu hem de çok modlu formlarda el damarı tanıma sistemlerinin sağlamlığını geliştirmek için çeşitli mimariler önerilmiştir.

Ek olarak, elde edilen veri setleri ile yapılan tüm deneyler, Badawi, Bosphorus, PUT, Tongji Temassız Avuçiçi Damar veri tabanı ve VERA gibi kamuya açık veri tabanlarından alınan veri kümeleri üzerinde benzer şekilde gerçekleştirilirken,

sistemlerin performansı literatürdeki benzer çalışmalarla etkili bir şekilde karşılaştırılmıştır.

Anahtar Kelimeler: Çok modlu biyometri, Öznitelik kaynaşımı, El damarı tanıma, El sırt damarı, Avuç içi damarı, Bilek damarı, CNN modelleri, BSIF.

DEDICATION

Dedicated to the Babalola family

ACKNOWLEDGMENT

Special recognition to my supervisors and motivators, Prof. Dr. Önsen Toygar and Assoc. Prof. Dr. Yıltan Bitirim. The outstanding encouragement and kindness they showed me during the course of this study cannot be overemphasized. It's been my profound honor, working with them.

I will also like to articulate my utmost gratitude to the dynamic head of the Department of Computer Engineering, Prof. Dr.H. Işık Aybay, and all the academic and non-academic staff of the University especially of the Computer Engineering department. My time as a Research and Teaching Assistant at the department was of immense advantage both to my growth as person and as a researcher.

My family has been of invaluable gift to me since the beginning of my academic pursuits especially my wonderful brother, Engr. Stephen Babalola and my parents Mr. and Mrs. F.N Babalola; I could not have done it without them. I thank them for believing in me all through this journey.

I also thank the amazing Research Assistants I worked with at the department of Computer Engineering and members of St Cyril's Catholic Community, Famagusta, the entire Catholic community in North Cyprus and the Joint Christian Council, Famagusta.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Background

Biometric systems and technologies have become very popular and successful in recent times for person recognition and authentication and there have been several types of physiological modalities/traits proposed and implemented with ranging level of efficiency, ease of use, user's willingness to use them and robustness in the face of alterations such as rotation, color changes due to cosmetics, and impediments [1]. Multimodal biometric systems have become the favored type of these systems because of its capability to deal with most of the issues around biometrics and physiological traits. For example, they are more likely to be spoof-proof because of the high level of effort it will take to acquire and present multiple fake parts of the body of the same person. They are also more robust and effective in comparison to unimodal systems as they produce better results in the face of anomalies such as orientation, rotation, contrast and color [2], this has consequently led to significant attention in the hand vein field of biometrics.

Figure 1.1 shows the network of veins on human hands; it can be seen that patterns can be obtained from different parts of the hand with proper devices. Hence, palmar vein patterns which consists of common palmar digital arteries and superficial palmar arch, dorsal vein pattern made up of dorsal metacarpal veins, dorsal venous arch and intercapitular veins and wrist vein pattern composed of intermediate antebrachial vein

and perforating veins can be identified and cropped for human identification. Due to the location of the vein, under the skin, it has an innate feature which makes it relatively immune to spoof attack [3]. Additionally, the evolution of low-cost devices that have the ability to effectively capture vein patterns from different area of the body in a fast manner has also made hand vein favored in authentication systems requiring high security.

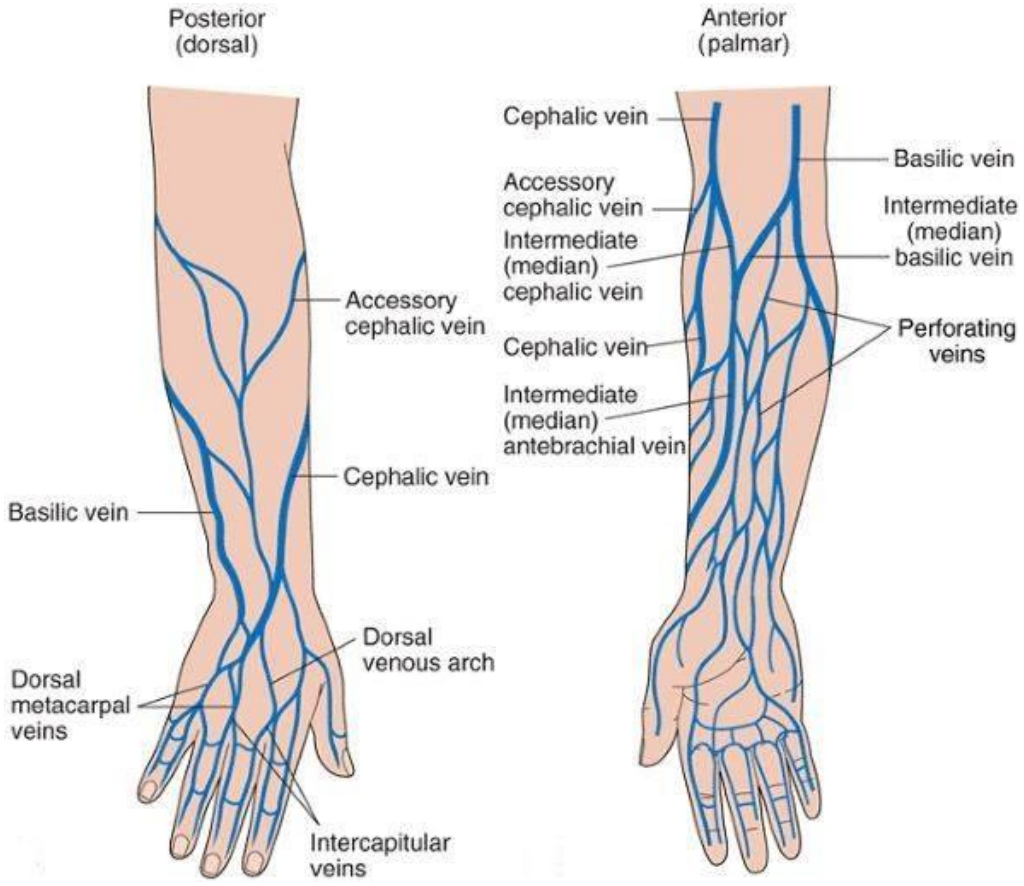


Figure 1.1: Human hand venous network. Adapted from [4].

A hand vein database composed of three datasets; dorsal, palm and wrist vein, was introduced in this study. The data were acquired with the aid of a relatively cheap medical device for vein detection, equipped with near infrared-light source camera. The images were acquired from willing volunteers who were mainly staff and students of Eastern Mediterranean University. In literature, there is no multimodal vein

database including the aforementioned three traits available and a few unimodal vein databases exist. After the construction of the multimodal vein database, person identification by employing robust feature extractors, such as Binarized Statistical Image Features (BSIF), Gabor Filters, and Histogram of Oriented Gradients (HOG), is carried out using the datasets as well as other available datasets for proper comparison.

Similarly, individual hand vein traits were used for unimodal biometric systems both with handcrafted feature extraction methods and deep learning models. The experiments include, a palm vein recognition system where a texture-based descriptor (BSIF) was combined with a deep learning based model, a wrist vein based biometric system where texture-based feature descriptors (BSIF, HOG and 2D Gabor filter) were combined at the decision level, dorsal hand vein based recognition system using score-level combination of overlapping regions as well as comparison with other fusion algorithms. Another experiment is the analysis of color models being utilized in hand vein recognition systems, and a proposal to fuse multiple color space channels for a more robust palm vein recognition system, among others.

1.2 Thesis Contribution

This study established FYO database, a publicly available multimodal vein database named after the main contributors of the project. The database consists of three datasets; palm, dorsal and wrist vein, thereby providing valuable data to explore and aid research work in the area of hand vein patterns as traits in multimodal authentication systems. Currently available hand vein databases were also reviewed in this study.

Consequently, the study proposed and implemented different unimodal and multimodal hand vein biometric systems, combining the strength of multiple traits as well as multiple algorithms (handcrafted methods and deep learning models) in different studies to enhance and improve hand vein biometric systems.

1.3 Thesis Layout

The first chapter presents what encouraged this study and an overview of experiments in this thesis. Chapter 2 gives a review of related research works especially in the field of hand vein biometrics as well as datasets that have been established for the purpose of biometric research. Chapter 3 provides a review of hand vein databases available in literature and compared them to FYO database established in the course of this study, while Chapter 4 gives proposed methods for human recognition using hand vein patterns. Chapter 5 shows results of performed experiments in this study and related discussions while Chapter 6 discusses the conclusions drawn from our studies and proposed future direction of our work.

Chapter 2

LITERATURE REVIEW

Several research works in biometrics have been carried out over the years, resulting in numerous algorithms for biometrics, from face recognition and identification studies [5] to palm print and finger print recognition systems [6]. There are also iris recognition systems [7], ear pattern recognition systems [8] and hand shape and geometry recognition systems [9] among numerous studies that exist in literature.

2.1 Hand Vein Biometrics

This study focused on traits that can be found in different areas of the hand in the form of vein patterns. We identified that vein patterns can be captured from four parts of the hand; palm, finger, wrist vein and back of the hand (dorsal). These characteristics represent four different hand vein biometric traits that can be employed for human recognition. Three of these traits were captured and used in this study.

2.1.1 Dorsal Vein

Although hand vein biometrics is relatively new, there have been numerous research work in this area. Some studies in dorsal vein biometrics include; texture features and shape clue combination for hand dorsal vein pattern recognition [10], four alternative feature sets fusion on hand vein images for person identification [11], hand vein authentication using line segments extracted using Hough transform matched by modified Hausdorff distance [12], and improvement of smartphone recognition technology with dorsal vein pattern extracted using oriented FAST and rotated BRIEF (ORB) algorithm [13].

Other studies include a multimodal system which combines finger vein and dorsal vein features using monogenic local binary pattern and Improved Gaussian Matched Filter (IGMF) for human recognition [14], dorsal vein recognition through fractal technique computed by box counting method for tissue properties identification and SUAS database acquisition [15], and three trait multimodal system which fuses dorsal, palm and wrist vein patterns combining both Binarized Statistical Image Features and a CNN-based model [3]. Samples of hand dorsal images acquired from different datasets that have been used in related studies are shown in Figure 2.1.

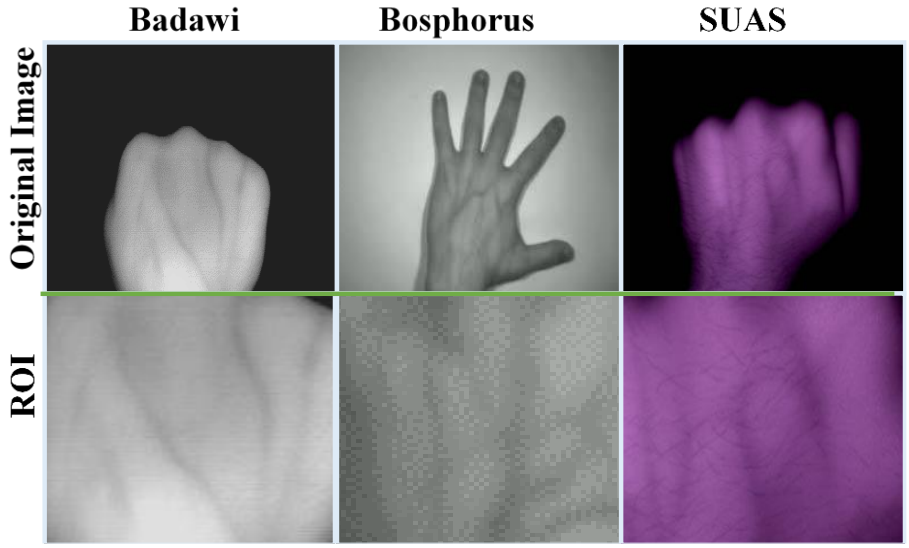


Figure 2.1: Samples of acquired hand dorsal vein and their corresponding ROIs

2.1.2 Palm Vein

Palm vein as a physiological modality is the most common in hand vein biometrics. Studies in this field include; non-vein pixels determination using a bank of directional filters to obtain line-based features from palm images [16], graphic depiction of palmar vein patterns where vein patterns are seen as blocks and then distance between blocks are viewed as edges, where the blocks are segmented with the aid of a randomly generated value which employees a key given to ensure a secure system [17]. Another

related study is a palm vein recognition system which applies competitive coding algorithm to multi-scale LBP, and optimized by ant colony optimization allowing the system work perfectly even with noisy images or low quality images [18].

There has also been multiple review of the processes used in palm vein authentication and recognition systems especially methods for classification and feature identification and extraction [19] and compared to similar modalities such as finger vein and finger print [20]. Samples of palm images available different datasets that have been used in related studies are shown in Figure 2.2.

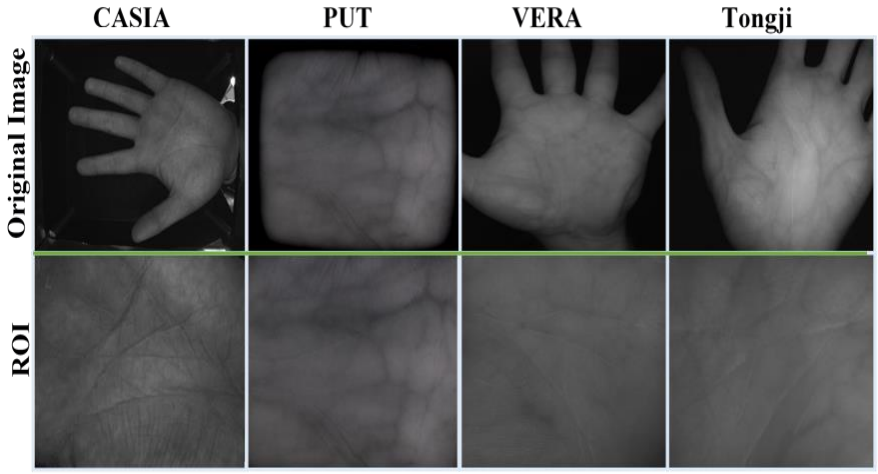


Figure 2.2: Samples of acquired palm vein images and corresponding ROIs

2.1.3 Wrist Veins

Although wrist vein modality is least used vascular trait for human recognition, numerous research works in this area exist as well, including; wrist vein feature extraction aided by spatial and orientation attributes of vein networks [21], using scale invariant algorithms such as SIFT, SURF and ORB for wrist vein recognition [22]. This is because the wrist also has an easily identifiable neurovascular structure, providing sufficiently distinct anatomical features that can be used for identification

and authentication. The location of the wrist in the body and since it is relatively usually free from makeup and cosmetics, it is an easily accessible part of the body [23].

Wrist vein has also been fused with other traits to obtain more robust identification systems, such as combining wrist vein with palm vein pattern and using 2D Gabor as feature extraction algorithm [24], and combining wrist vein features from both right and left hands of volunteers through a combinatory algorithm which obtains two binarized images one based on local thresholding and another by global thresholding, and then multiply them [25] among others. Furthermore, with the abundant use of smartphones today, the availability of ever improving imaging features, and biometric authentication being incorporated, research works that employ wrist vein pattern in smartphones for person identification with the aid of infrared camera has been proposed [26]. Samples of wrist images acquired from different datasets that have been used in related studies are shown in Figure 2.3.

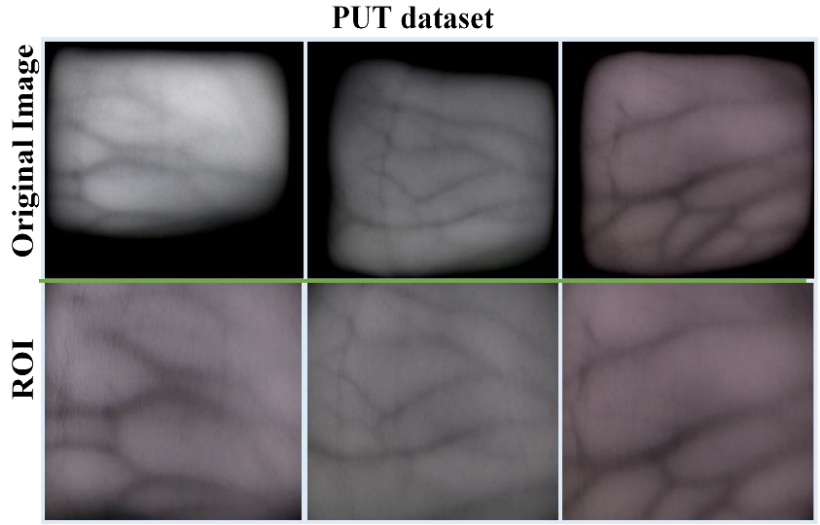


Figure 2.3: Samples of acquired wrist vein images and their ROIs

Chapter 3

HAND VEIN DATABASES

Datasets are an essential part of biometric research, such that their availability or lack of it determines how far a research study could go. We therefore established a hand vein database in this study comprising of three distinct datasets (hand dorsal, palmar and wrist); the details are given in this chapter. We also examined publicly available hand vein databases in this chapter, especially those that are relevant to this study.

3.1 Database Acquisition

The images of the database were obtained using a cheap medical device for vein detection. The device has a 1/3 inch infrared complementary metal-oxide semiconductor camera (CMOS) bounded by twelve infrared LED as light sources. It was designed with the ability to detect veins in any place on the human body, weight, age and skin color notwithstanding. The acquisition process was carried out using a laptop connected to the vein finder with a USB cord. Therefore, captured images can be viewed and saved in PNG format of size 800 x 600 x 24.

The device comes with an adjustable holder capable of adjusting up to 360 degree angles clamed unto a stand that equally adjustable in length. However, there is a need for some uniformity in the acquisition process for consistency, hence a fixed position was adopted throughout the acquisition process as shown in Figure 3.1. The acquisition process lasted for about a month. The height of the camera from the ground was maintained at 35cm, while three hand guides (7cm high) were constructed, one for

each hand vein modality. The guides are necessary in order to reduce rotation of the hand and guide users more easily. Consequently, volunteers' hands were about 26-28 cm from the sensor, depending trait and hand size.

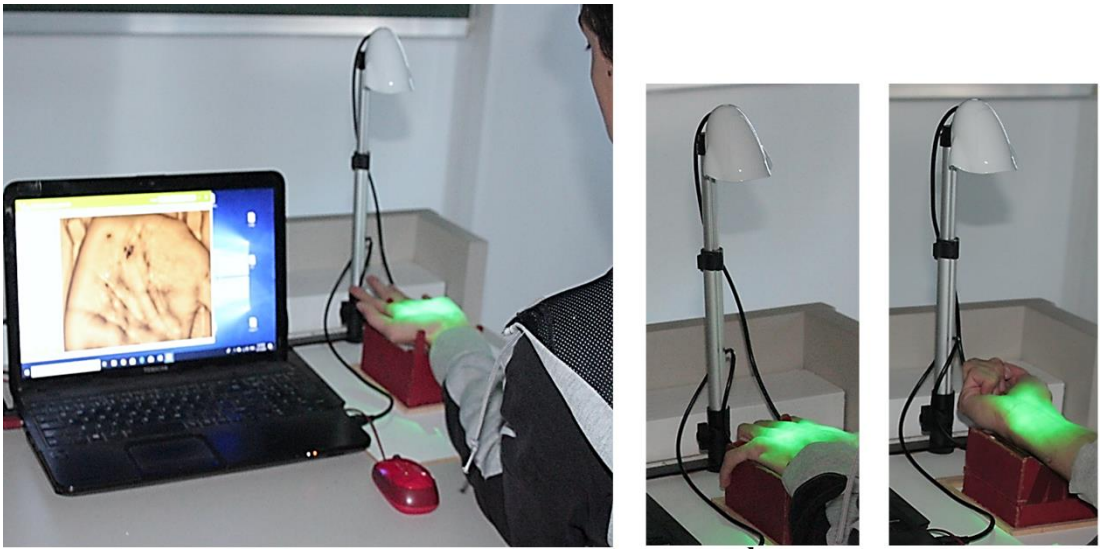
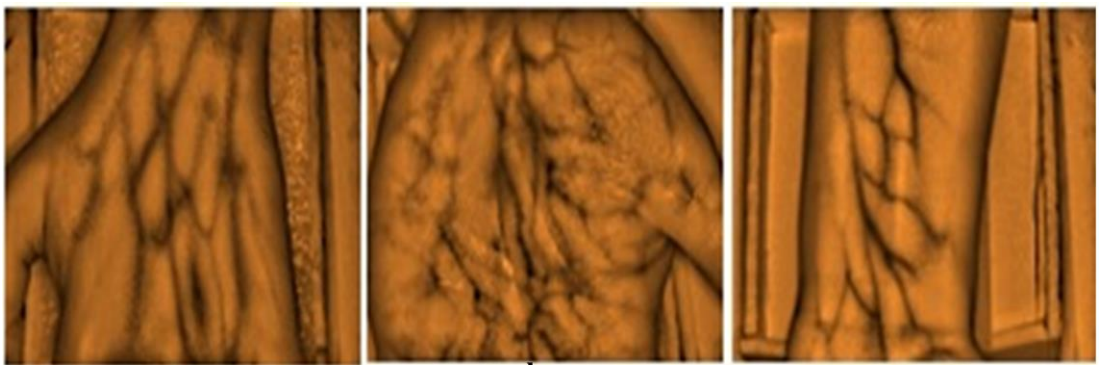


Figure 3.1: Data acquisition process

The acquisition of the three modalities was done successively in less than one minute. There are approximately 10 seconds apart from each acquisition, while a second session starts right after the first. Figure 3.2 shows samples of images captured using the described setup.



(a) (b) (c)

Figure 3.2: Samples of acquired images (a) dorsal, (b) palmar, (c) wrist

A total of 160 volunteers mainly from Turkey, North Cyprus, Nigeria, and from other neighboring Middle Eastern and African countries participated in the acquisition process. We acquired sample images from both hands of every subject in two independent sessions with one image per session from each hand and trait. This sums up to 4 images per trait and 12 images per individual. The total images in each dataset is 640 giving a total of 1940 samples in the database. The age range of the volunteers is shown in Figure 3.3, where it can be seen that it ranges from 17 to 63 and male to female ratio is 69.27% to 30.73% representing 111 to 49 individuals since there are 160 subjects [3]. The database is publicly available to researchers only upon request at a webpage provided by the university: https://fyo.emu.edu.tr/, a screenshot of the website is provided in Appendix A. A sample of the application form for the database, as available on the website, is shown in Appendix B.

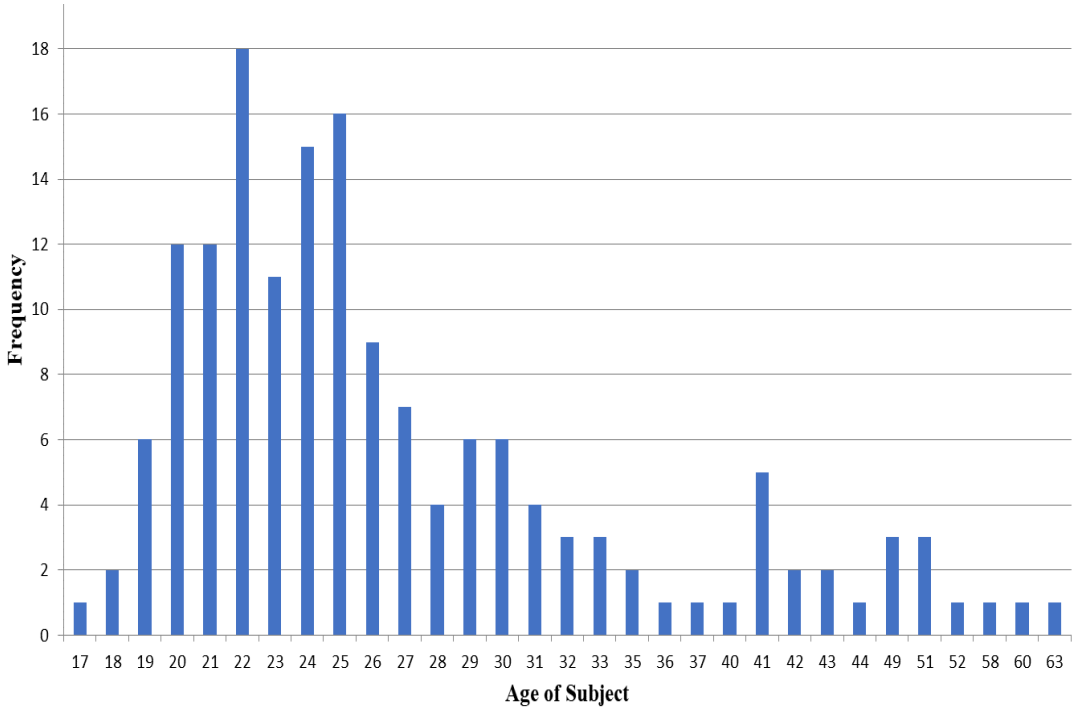


Figure 3.3: Volunteers' age distribution

3.2 Related Databases

Hand vein databases with similar features as our database were reviewed in this subsection. The aim is to examine how they can be used in our multimodal biometric experiments while also comparing them to our database.

- **PUT** hand vein pattern database was introduced in 2011; the first publicly available dataset for vein recognition. It consists of palm and wrist vein datasets with a total of 2400 samples. The data were obtained from both left and right hands of 50 individuals, adding up to 100 unique hand vein patterns for both palm and wrist vein [27].
- **Tongji** contactless database is a relatively sizeable database composed of palm vein and palm print images obtained from 300 subjects which include 108 females and 192 males. Their ages ranged from 20 to 50 years. The acquisitions were carried out in two sessions where 10 images per trait were captured from both left and right hands, respectively. Consequently, the database is composed of a total of 12,000 images captured from 300 people [28].
- **VERA** palm vein database was established at Idiap Research Institute in Martigny and Haute Ecole Spécialisée, Switzerland. It has a total of 2200 samples taken from 110 different volunteers. The data were taken in 2 sessions and in 2 different locations, where 5 samples were acquired from each hand of participants. The average age of volunteers in the database was 33 years, ranging between 18 and 60 years [29].
- **Bosphorus** hand vein database contains 1200 left hand images from 100 volunteers taken in 4 sessions to represent different physical conditions that

human hands could be subjected to before presenting the hand for recognition purpose in real-life. The activities include:

i. carrying a 3 kg bag for about a minute before image acquisition,
ii. squeezing an elastic ball by opening and closing repeatedly for about a minute
iii. placing an ice pack on the hand to cool it
iv. Normal condition.

Three sample images were taken per subject in each session [11].

- **Badawi** database was taken from 50 volunteers but from both hands of each person, coming to 100 hand vein classes. It has a total of 500 sample images, acquired by taken 5 images from both left and right hands of indiviuals at varying intervals [30].
- **SUAS** hand vein database was collected at Sakarya University of Applied Sciences, Turkey. The volunteers included 80 male and 75 females and the images were taken using infrared camera. There are a total of 919 sample images comprising of 3 samples per hand [31].
- **CASIA** Multi-Spectral Palmprint Image Database has a total of 7200 palm vein sample images obtained from 100 volunteers. The samples were captured in two sessions where six palm images were captured simultaneously with six varying electromagnetic spectrums (460 nm, 630 nm, 700 nm, 850 nm, 940nm and white light) [32].

Table 3.1 shows a comparison of other publicly available hand vein databases with our acquired database. It shows that majority of hand vein databases have only one

vein trait except PUT and our database (FYO) which are composed of two and three traits, respectively.

Table 3.1: Hand vein databases comparison

Database	**Badawi**	**Bosphorus**	**PUT**	**SUAS**	**Tongji**	**VERA**	**FYO**
Subjects	50	100	50	155	300	110	160
Male/Female	--	58/42	--	80/75	192/108	70/40	111/49
Left/Right hands	2	2	2	2	2	2	2
Sessions	1	4	3	1	2	2	2
Samples per session	5	3	4	3	10	5	1
Total number of samples	500	1575	2400	919	12000	2200	1920
Dorsal	√	√	--	√	--	--	√
Palm	--	--	√	--	√	√	√
Wrist	--	--	√	--	--	--	√

Chapter 4

METHODOLOGY

The objective of this thesis is to propose multimodal biometric methods where hand vein patterns are used as biometric traits. The study explored different ways of combining physiological traits and feature extraction methods in a view to propose and implement more robust biometric authentication systems. Hence, various multimodal set-ups were examined in the study, including multimodal authentication system which combines features extracted from palm vein, dorsal vein and wrist vein. Hand dorsal vein authentication system proposed segments images into five overlapping regions and combines CNN models' prediction from each of the segments. Other proposed methods use only one biometric trait but combine multiple algorithms. This include palm vein recognition system which combines a texture-based descriptor method with a CNN model while another palm vein biometric system combines three CNN models at the decision level. A third system based on palm vein explored different color models by combining the most contributing channel from each of them for a more reliable authentication system. The wrist vein system combines decisions from three texture-based descriptors. The feature extraction methods and the several proposed systems are explained in the following subsections.

4.1 Methods Used

Feature extraction methods used in this study can be categorized into two types: handcrafted methods (mainly texture based feature extraction methods including HOG, Gabor and BSIF) and deep learning methods (where CNN models such

AlexNet, ResNet-50, VGG-16 and VGG-19 were modified to reduce training time). The algorithms are explained in the following subsections.

4.1.1 Handcrafted Descriptors

Handcrafted descriptors are feature descriptors where the set of filters have been hand-picked by the designers; they rely on the knowledge of experts to define a common way of extracting features from images. They can be grouped into different types based on their construction and usage, ranging from point-wise-based, graph-based, and texture-based, among others. A few of texture-based descriptors were used in this study which are explained below.

4.1.1.1 BSIF

Binarized Statistical Image Features (BSIF) is a texture-based feature descriptor originally proposed for face recognition and texture classification. Local Binary Patterns (LBP) and Local Phase Quantization (LPQ) descriptors are the foundation/basis on which this method was built. LBP and LPQ model the neighborhood of every pixel as a binary code which are gotten filtering the image with a set of manually constructed linear filters. The feedbacks of the filters are subsequently binarized by quantization algorithm. Hence, each bit of the code string corresponds to responses of different filters after it has been binarized. These methods have shown to be effective, recording outstanding performance in several pattern recognition studies.

On the other hand, a bank of filters in BSIF were automatically obtained from a set of natural images as oppose to the manually crafted filters used in LPQ and LBP [33].

4.1.1.2 M-BSIF

The filter bank in BSIF is made up of filters of different odd square sizes, ranging from 5×5 to 17×17, meaning that there are seven different sizes for each filter. Additionally, there are also varying length of the filters, ranging from 5 to 12. These different sizes and depth gives a total of 57 filters in the BSIF bank. However, in this study, we proposed choosing five filter sizes that performed best; this would be determined by a preliminary studies to identify the best performing filters. Subsequently, filters with 12 depth was chosen, and the five highest sizes. M-BSIF is therefore described as a BSIF feature description/extraction technique that uses five kernels.

4.1.1.3 Gabor Filter

Gabor filter is a linear filter which got its name from Dennis Gabor, its originator. It has been shown to effectively analyze texture of images and favored for detecting edges and shapes. A band-pass spatial filter with the ability to choose both spatial frequency and orientation is known as a 2-D Gabor filter. Its sinusoid wave impulse response function is surrounded by a Gaussian function with oscillation orientation and frequency.

The result of filtering an image with 2D Gabor filters is a complex value of imaginary and real parts. The process is carried out by convolving through sample image using both the imaginary components responses to kernels and the real components [34, 35].

4.1.1.4 HOG

Histogram of Oriented Gradients (HOG) is another common texture based descriptor favored for objects identification. The algorithm counts the occurrences of gradient orientation within a portion of an image, this is done throughout the image. The values of gradients are first generated using horizontal and vertical one dimensional masks. The second step is the grouping the gradients into units/cells and obtaining the

histograms of all units. Then determine the orientation of each unit by rated polls drawn from every pixel in the unit according to the gradient values obtained earlier. Finally, the strength of gradients are normalized locally, this is carried by segmenting/categorizing units in the same neighborhood together into larger blocks. With the help of this normalization step, changes illumination, contrast and others are resolved. The descriptor is developed by joining the elements of the histograms of normalized cell from each of the identified blocks [36, 37].

4.1.2 Deep Learning Methods

A number of popular and successful CNN models were used in this study, namely AlexNet, VGG-16, VGG-19 and ResNet-50. They are explained below along with the changes made.

4.1.2.1 AlexNet

AlexNet is a CNN model made up of five convolution layers and each of these layers are followed by an activation function called ReLU, and Batch Normalization and Maxpool layers. The original structure is composed of different filter sizes in different convolution layer, between 3 x 3 and 11 x 11 kernels. The number of filters in the layers range from 96 to 384. A Dropout layer, a Fully-connected layer and Softmax function ends the structure after the five layers [38].

In this study, however, the number of kernels in each convolution layer has been reduced drastically to greatly decrease computation time involved in the training process while still maintaining the efficiency of the model. The resultant architecture is presented in Figure 4.1. In addition, 3 x 3 filter size was upheld for each convolutional layer all through the model in this experiment as oppose to different

kernel sizes used in the original AlexNet model. Table 4.1 shows the level of modification in the modified version in comparison with the original model.

Table 4.1: AlexNet CNN model filter reduction

Layer	Original	Modified	Reduction percentage
First layer	11 x 11 x 96	3 x 3 x 32	66.67%
Second layer	5 x 5 x 256	3 x 3 x 64	75.00%
Third layer	3 x 3 x 384	3 x 3 x 128	66.67%
Fourth layer	3 x 3 x 384	3 x 3 x 128	66.67%
Fifth layer	3 x 3 x 256	3 x 3 x 64	75.00%

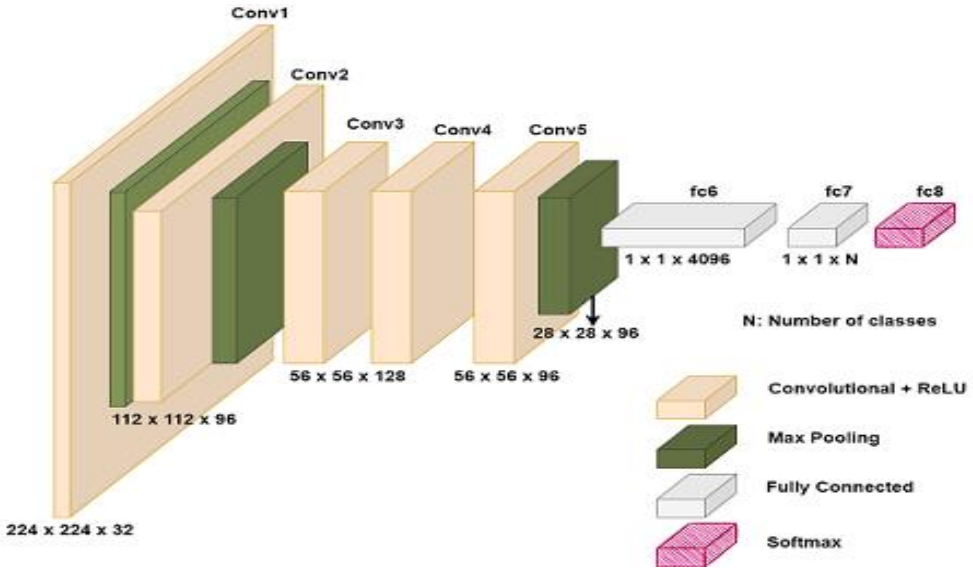


Figure 4.1: Modified AlexNet structure

4.1.2.2 VGG-16

VGG-16 is another CNN model which was introduced in 2014 at the Imagenet competition. It has sixteen convolution layers corresponding to its name. Therefore, the structure has a large network, regarded as one of the high achieving CNN models. All its convolutional layers has 3 x 3 kernels and ReLU activation. However, convolutional layers, are grouped into five layers as shown in Figure 4.2 where the

first layer is made up of 2 convolution layers, the second layer is also made up of 2 convolution layers, while the third, fourth and last layers have 3 convolution layers each). Each layer is followed by Batch Normalization and Maxpool layers of size 2 x 2 and two strides. At the end, there are a couple of Fully-connected layers, as well as a Softmax function which is used to normalize the output; setting them to between 0 and 1. The number of filters in convolution layer range from 64 to 512 [39].

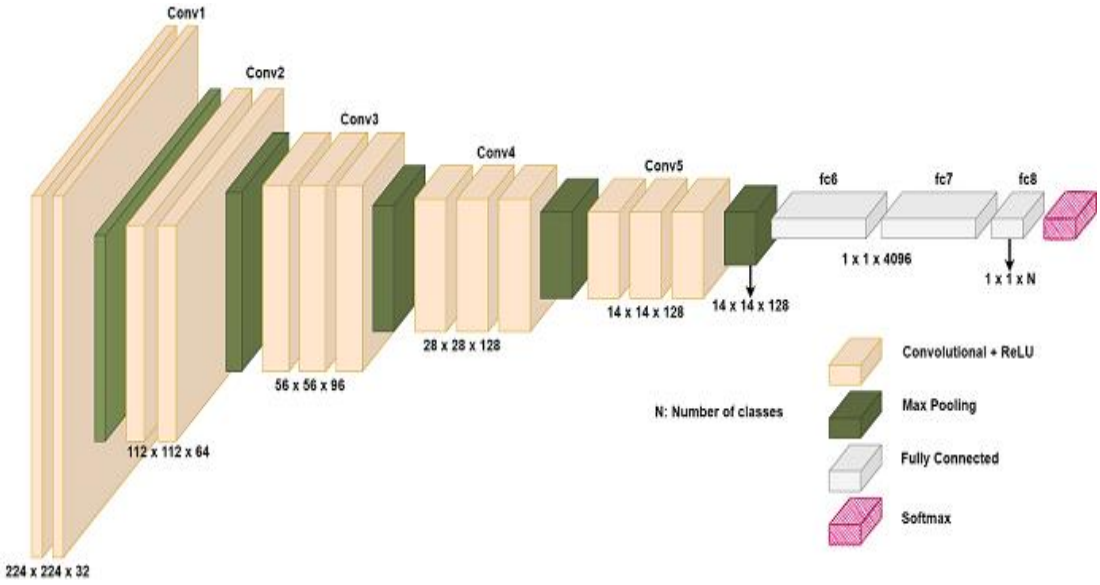


Figure 4.2: Modified VGG-16 architecture

4.1.2.3 VGG-19

Similar to VGG-16, the number 19 in VGG-19 corresponds to the nineteen convolutional layers that exist in the model. It has all the attributes of VGG-16 but one additional convolutional layer has been added to third, fourth and fifth layers of the model as shown in Figure 4.3. The same filter reduction as in the aforementioned VGG-16 model is applied to this model as well. Table 4.2 presents the changes in the

number of kernels both in the modified version of VGG-16 and VGG-19 models proposed in this study as well as the original [40].

Similar to AlexNet, however, the original VGG-16 and VGG-19 were modified in this study by reducing the amount of filters in every layer as shown in Figure 4.2 and Figure 4.3, respectively, in order to decrease training time while still maintaining the efficiency of the CNN model. Table 4.2 shows the percentage reduction in filter per layer.

Table 4.2: VGG-16 and VGG-19 filter reduction

Layer	**Original**	**Modified**	**Reduction percentage**
First layer	64	32	50.00%
Second layer	128	64	50.00%
Third layer	256	96	62.50%
Fourth layer	512	128	75.00%
Fifth layer	512	128	75.00%

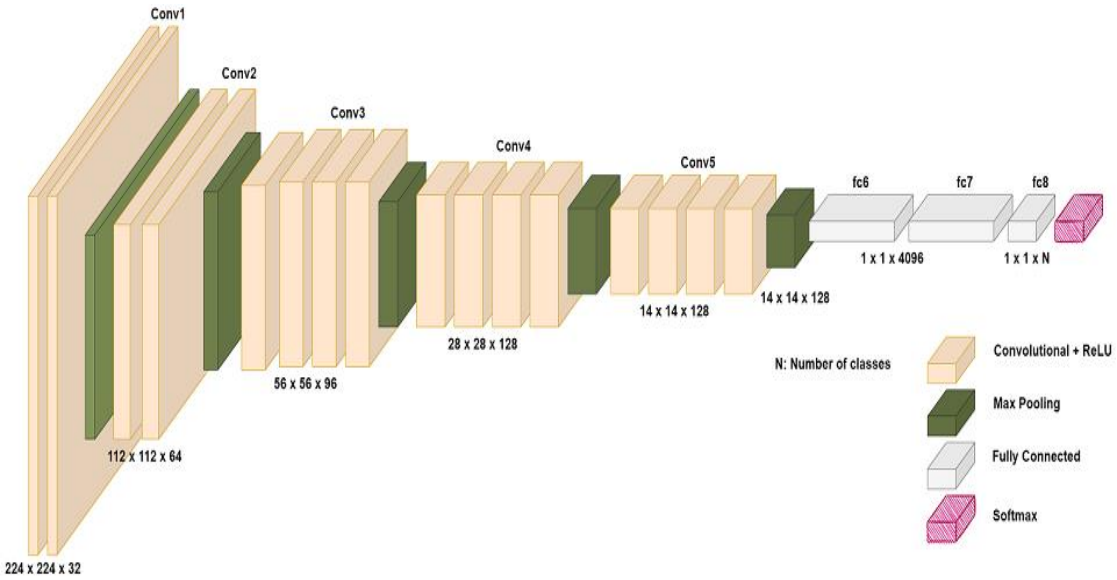


Figure 4.3: Modified VGG-19 architecture

4.1.2.4 ResNet-50

ResNet-50 is another CNN model, a variant of the original model called Residual networks (ResNet). It is made up of blocks (Convolution and Identity blocks) built to address the gradient vanishing problem of deep learning models, giving it a rather different structure compared to AlexNet and VGG. The structure is characterized by shortcuts to jump over some layers which give them the ability to train really deep networks without caring about vanishing gradient. The structure of the model is represented in Figure 4.4.

Each Convolution block is made up of two parallel lines of convolution layers concatenated at the end. The first line is composed of three sequential convolution layers followed by ReLU activation and batch normalization. On the other hand, the second line is made up of one convolution layer as shown in Figure 4.5. Similarly, the identity blocks are made up of a line of three sequential convolution layers with ReLU activation and batch normalization in each layer. However, the second line in this case only concatenates the initial input with the output of the first line as shown in Figure 4.5 [41, 42].

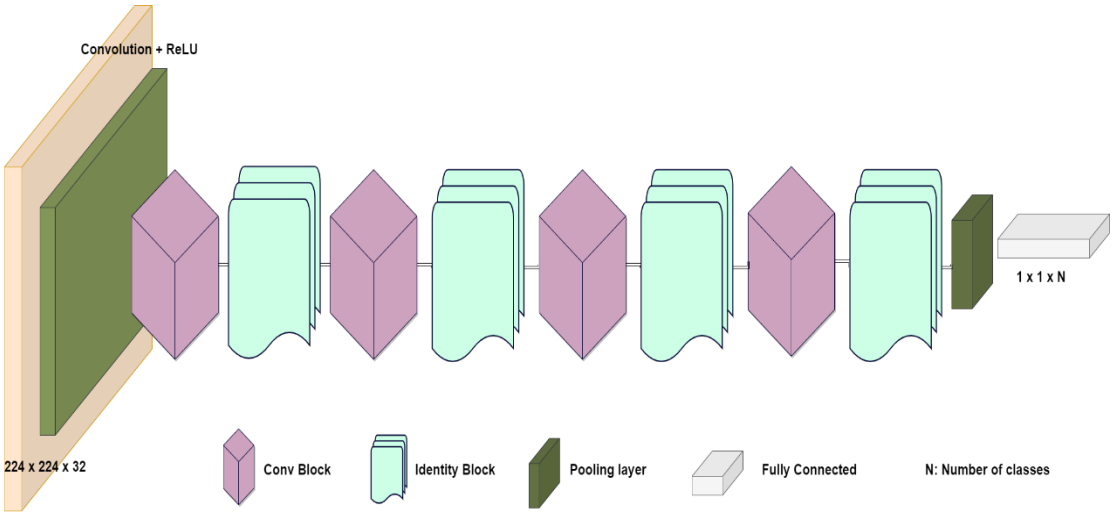


Figure 4.4: ResNet-50 architecture

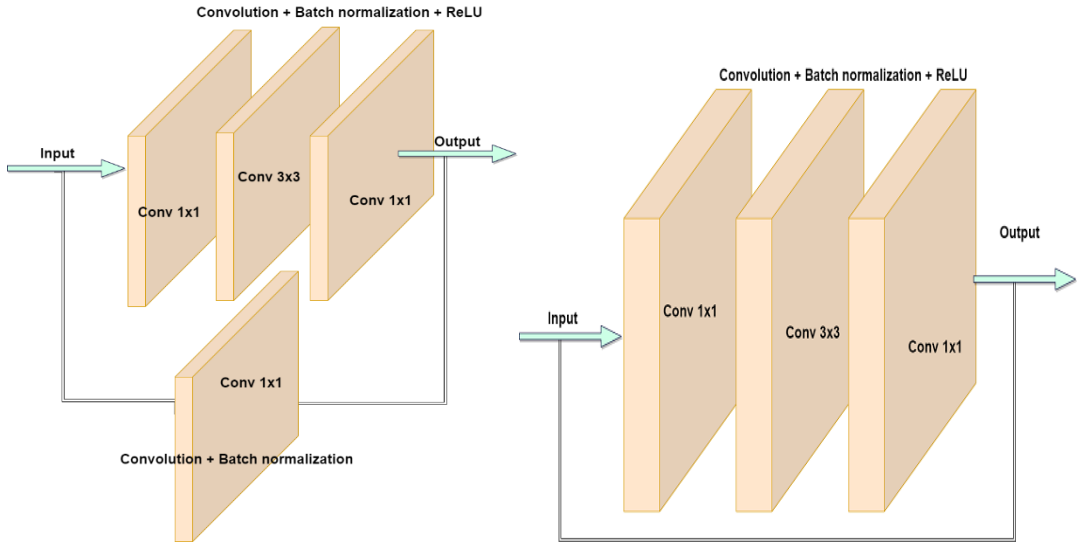


Figure 4.5: Convolution block and identity block of ResNet-50

Similar to AlexNet and VGG, the number of kernels per convolution layer was also decreased in this model. The changes made are shown in Table 4.3 where it can be seen that at least 50% reduction has been made at each stage of the model from stage 2 to stage 5.

Table 4.3: Filter reduction per stage in ResNet-50

Stage	Original Filters 1,2,3	Modified Filters 1,2,3	Percentage Reduction
2	64, 64, 256	32, 32, 128	50%
3	128, 128, 512	32, 32, 128	75%
4	256, 256, 1024	64, 64, 256	75%
5	512, 512, 2048	128, 128, 512	75%

4.2 Proposed Multimodal Biometric Systems

Multimodal biometric systems combines the evidence or information presented by more than one sources of information in order to deal with shortcomings of unimodal systems. Unimodal biometric systems are often confronted by varieties of issue such

as noisy data, intra-class variations, non-universality, spoof attacks, and high error rates. These can be addressed by employing a multimodal system in place of unimodal system [42].

One common pre-processing step is the region of interest (ROI) specification and trimming. This is because hand vein sample images in all modalities are usually taken along with some parts of the background/surrounding/place-holder. Identifying and cropping out the ROI in other to eliminate the background and other areas of the hand that are not required in this study, such as fingers, upper parts of the hand. A semi-automatic process is used for this process. It involves identifying the best fitting dimension that best fits the ROI for all samples and employing it to crop every image. Furthermore, there may be samples that are badly cropped, the samples are checked to remove those that are badly cropped. The process of identifying the best fit dimension and cropping is repeated for this set. This continues till all images are trimmed as desired.

The following sub-sections gives three types of multimodal systems proposed in this study in details. Firstly, combination of multiple traits (that is, vein patterns from palmar, dorsal and wrist areas of the hand) and combining features from different segments of an image are explained in section 4.2.1. Secondly, section 4.2.2 explains methods of combining feature descriptors, such as combining CNN models and handcrafted descriptors for palm vein recognition, fusion of handcrafted methods in wrist vein biometrics and combination of scores from three CNN models for palm vein biometrics. Lastly, another architecture which combines channels of different color models for palm vein biometrics is explained in section 4.2.3.

4.2.1 Multi-Traits

Most multimodal biometrics combine features of multiple traits to obtain a more robust system, such as combining face features and iris features, or fusing palm print and finger print features etc. either at the feature level, score level or at the decision level. Combination of feature from the palm vein, wrist vein and dorsal vein was carried out in this study in two separate experiments. A third experiment in this category fuses features from defined segments of an image; only one biometric is used in a multi-modal structure.

4.2.1.1 Fusion of Dorsal, Palm and Wrist Vein Biometrics

Dorsal vein, palmar vein and wrist vein features are combined in two separate experiments. The first is illustrated in Figure 4.6, where hand-crafted feature descriptors, namely BSIF, Gabor and HOG, were used to obtain features from hand vein images from three parts described in this study: dorsal vein, palm vein and wrist vein. The features of these separate areas are combined before being used to recognition, depicting a Feature-Level Fusion process.

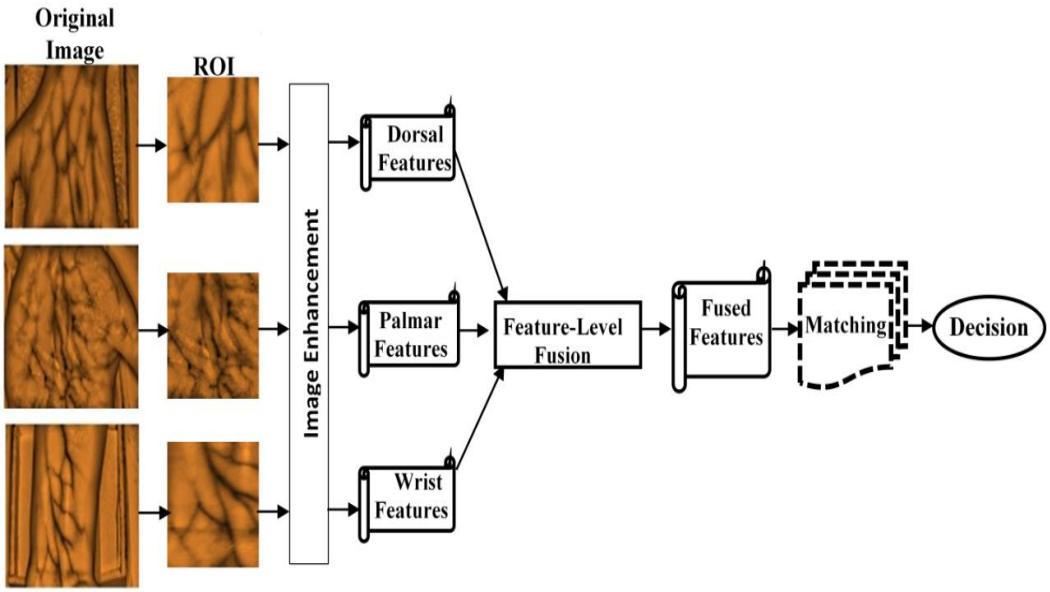


Figure 4.6: Multimodal structure for texture-based feature descriptors

The second method in this category is depicted in Figure 4.7 where same hand vein images as the first method are used (dorsal, palm and wrist). However, deep learning methods; AlexNet, VGG-16 and VGG-19, were used for feature extraction in this case. Furthermore, decision level fusion was used in this case for combining the traits together.

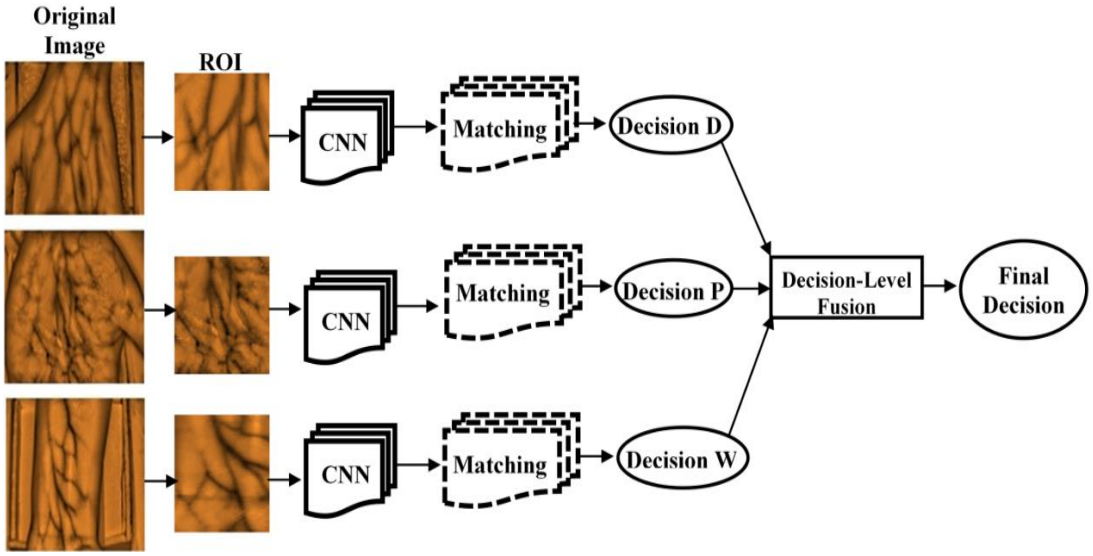


Figure 4.7: Multimodal CNN model structure

4.2.1.2 Hand Dorsal Vein Recognition System

The proposed system uses five overlapping vein image regions as used in the palm vein recognition experiment. It trains these regions separately by a CNN architecture and then combines the obtained CNN predictions using score-level fusion. The steps are illustrated in Figure 4.8. Firstly, it involves pre-processing stage where input images are converted into required dimension, and then, the images are enhanced by removing noise after cropping out background areas. The overlapping areas of the images are identified as in the previous experiment and each of them is used in its corresponding CNN training module, followed by classification module which generates five different sets of scores for each test sample. These scores are added

together and they are employed in the determination of the class a test sample belongs to.

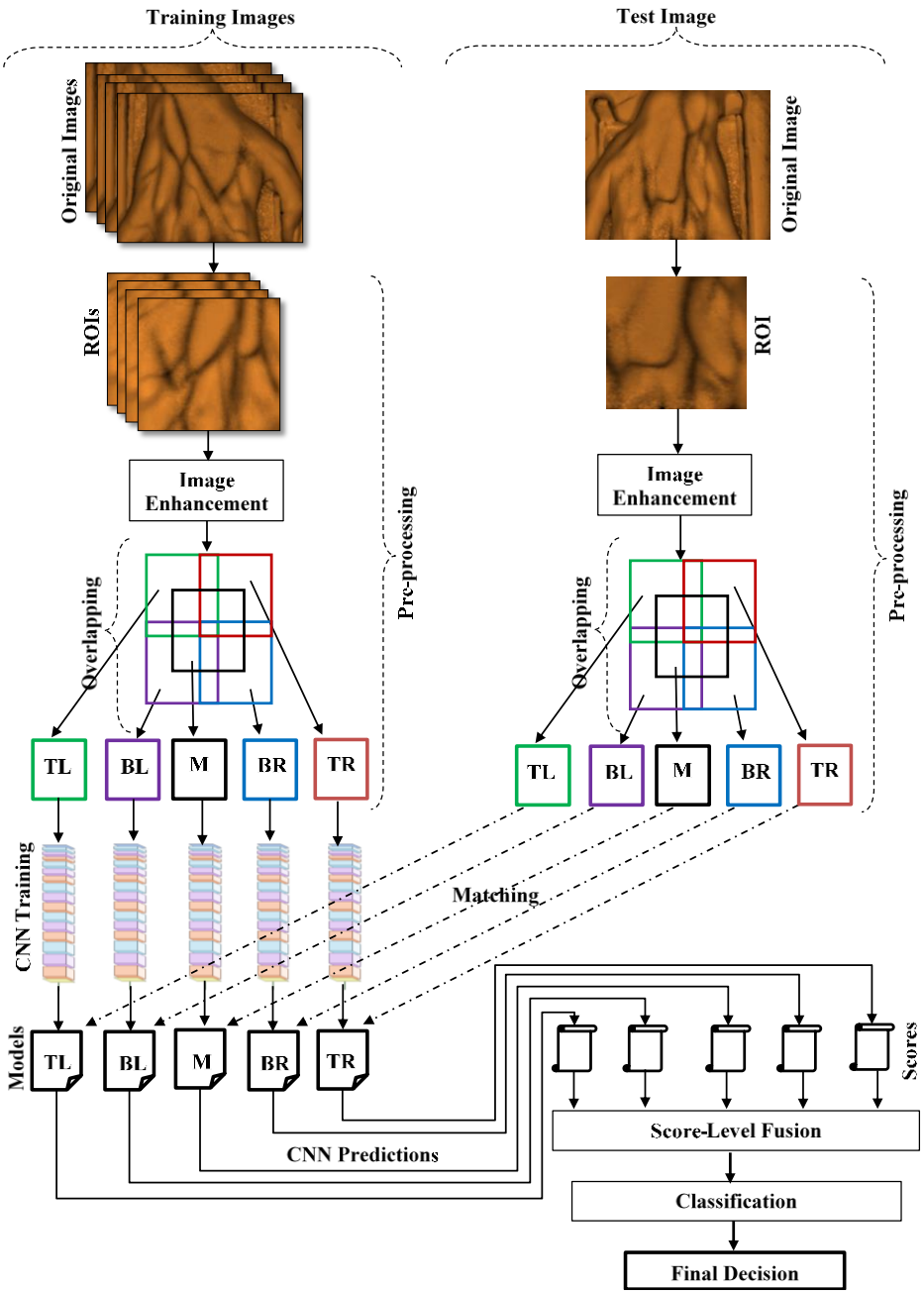


Figure 4.8: Dorsal vein recognition using separately trained five overlapping regions

As part of pre-processing, ROIs are re-sized to 214 x 214 pixels, while a region is defined as 128x128 pixels. Therefore, the overlapping regions are 128x128 kernels shifted to the corners of the image, that is, top-left, top-right, bottom-left and bottom-

right corners of the ROI images. However, the middle overlapping region is 128x128 pixels centered at the center of the ROI image. Image enhancement using histogram equalization is also carried. Histogram equalization is a well-known image enhancement technique that can be used to improve contrast in images. This can potentially remove noise from images, and make vein pattern more distinctive.

In the last phase of pre-processing stage, ROI images are categorized as training and test sets. Then from training images, for each overlapping region type, namely; Top-Left (TL), Top-Right (TR), Bottom-Left (BL), Bottom-Right (BR) and Middle (M), a training set is formed. Hence, there are five training sets which are Training-set TL (for TL region), Training-set TR (for TR region), Training-set BL (for BL region), Training-set BR (for BR region) and Training-set M (for M region). Class labels are assigned to every region in the training set.

Training of the system is done with the AlexNet based CNN model described in the previous session. Datasets used were from Badawi, Bosphorus and FYO databases. However, the sample size of each of the above named datasets are too small for training in deep learning. Training samples generally perform better when they are about 5000 samples or more. Therefore data reinforcement was performed using Keras data generator which creates new images using parameters such as varying zoom, height and width shift, brightness adjustment, and applying slight rotation on the original images. Consequently, the datasets used for testing the CNN models were organized as below:

- **Badawi**: 11 new images were generated for each sample in the original dataset (500 images, 5 samples per person), summing up to an aggregate of 5500

images, 5000 of which were for training (about 91 percent), and 500 were used as test samples.

- **Bosphorus**: 5 new images were generated for each sample in the original dataset (1200 images, 12 samples per person), summing up to a total of 6000 images with 60 samples per subject. 90 percent of the augmented dataset was
- **FYO**: The generated dorsal vein dataset already available in database was used. It has 6400 images (20 samples per subject). Similarly, 90 percent of the dataset was for training, while remaining was used to test the system.

4.2.2 Multi-Descriptors

Another type of multimodal system structure is where feature extraction algorithms or methods are combined instead of features. The following subsections present proposed multimodal biometric systems of this type, including two palm vein systems where one combines a CNN model with features extracted from defined overlapping segments of the image by a handcrafted descriptor (BSIF) while the second one combines three CNN models (AlexNet, VGG-16 and VGG-19) at the decision level. Other systems include a dorsal vein system which combines five CNN models obtained from five overlapping regions at the score level and a wrist vein system which combines decisions of three texture-based descriptors (M-BSIF, HOG and 2D Gabor Filter).

4.2.2.1 Palm Vein Recognition System

A palm vein biometric recognition method that combines a texture-based method with a CNN-based method was proposed in this study. The structure is presented in Figure 4.9. The texture-based system obtains features from five overlapping regions of ROI of palm vein sample images using BSIF described in the previous subsection. The

scores are obtained by matching the training sets with test sets which are then fused in this study to obtain one of the decisions of the system.

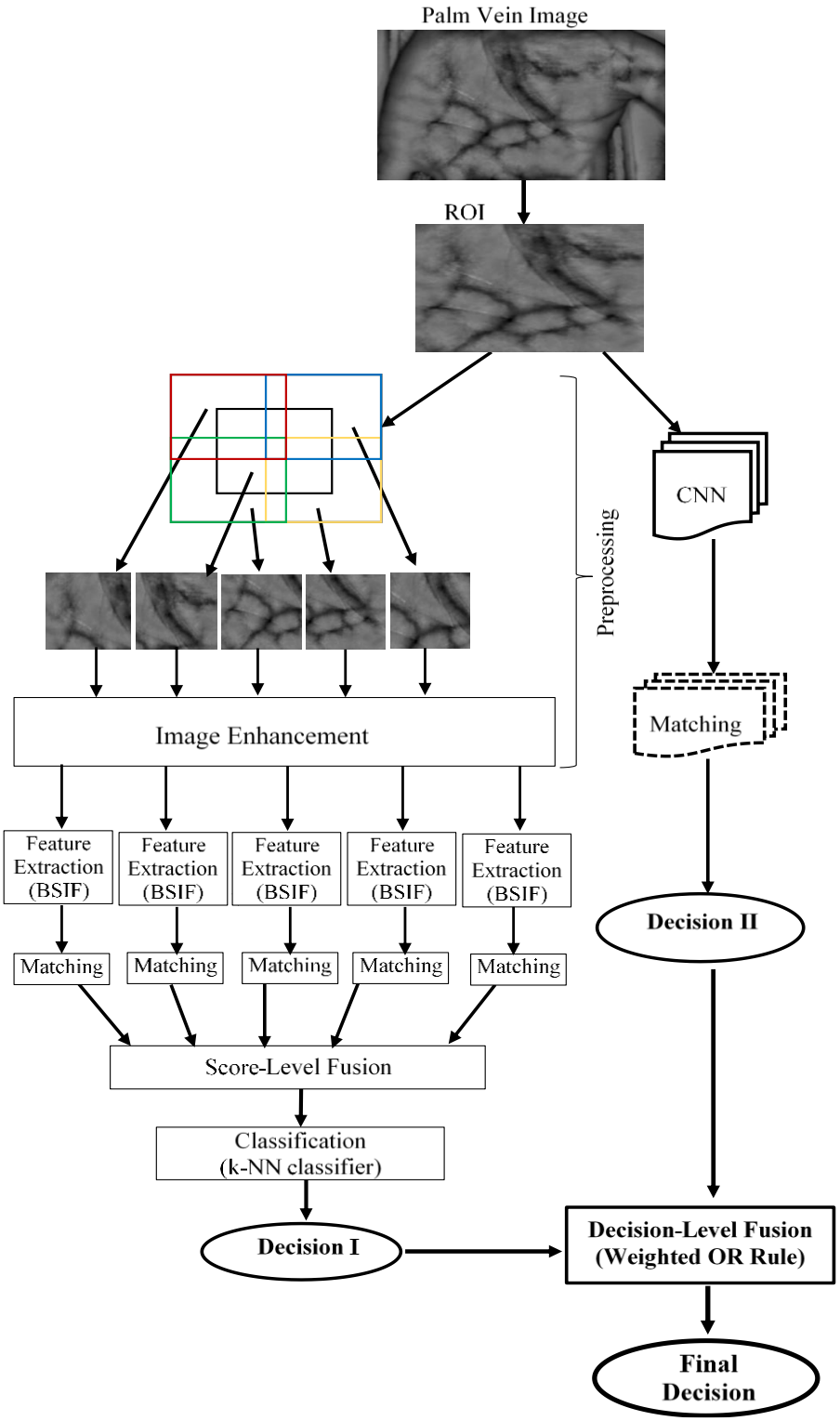


Figure 4.9: Palm vein recognition which combines BSIF in five regions with CNN models

The sub-regions are defined as middle region, top-left corner region, top-right corner region, bottom-left corner region, bottom-right corner region similar to the system in section 4.2.1.2. The ROI image is first resized to 500 × 450 pixels while each region is defined as 300 × 250 pixels. The middle region is therefore a 300 × 250 pixels box centered at the middle of the image. The other four regions are windows of 300 × 250 pixels pushed to the top-right, top-left, bottom-right and bottom-left corners of the ROI. Match scores of the regions are fused using score-level fusion to obtain a single match score vector which is employed in classification process to classify test samples.

The second method is a CNN model: AlexNet. The model is trained to obtain the second decision. The final decision of the system is achieved by combining both decisions gotten above by a Weighted OR Rule. Every decision outputs either true recognition or false recognition. A true decision is assigned 1, while false decisions get 0 weight. Subsequently, the weights are added and measured a threshold of 0.9 to arrive at the final decision of the system.

Figure 4.10 shows sample palm vein images from different datasets used in the experiment, namely; CASIA, FYO, PUT, VERA and Tongji palm vein datasets.

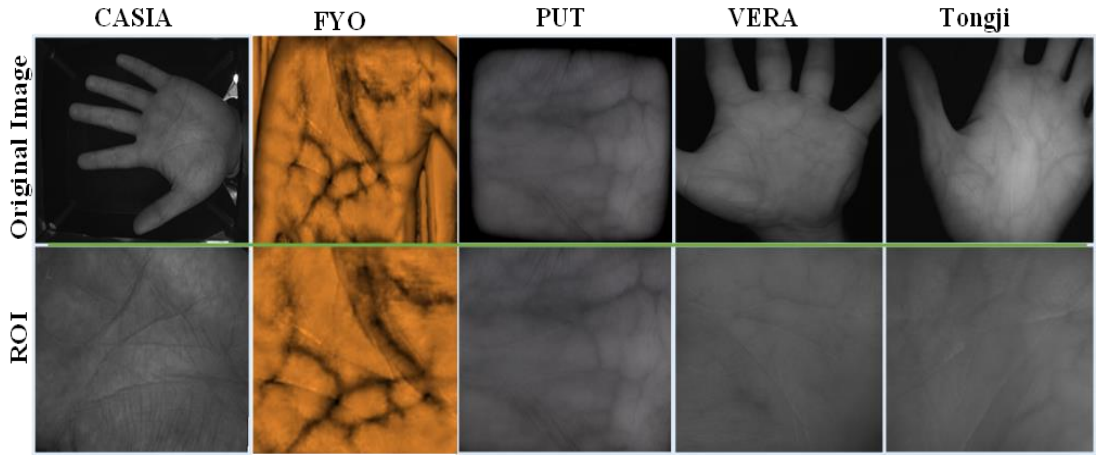


Figure 4.10: Sample palm vein images from different database and corresponding ROI

4.2.2.2 Wrist Vein Recognition System

This system integrates the fortitude of three different common methods, M-BSIF, 2D Gabor and HOG that are all texture-based feature descriptors. This is because texture descriptors are relatively prominent in biometric studies with numerous studies attesting to their efficiency. The implementation and validation of this system was performed on two datasets from FYO and PUT databases. The system is presented in Figure 4.11 where three methods were combined. The structure is made up of three phases; pre-processing, feature extraction, and algorithm combination.

In the preprocessing phase, the ROI is obtained. The second phase is the feature extraction processes. It include separate feature extractions by M-BSIF, that is BSIF with multiple filters described in subsection 4.1.2, 2D Gabor and HOG. However, in the case of 2D Gabor, histogram is performed pro to feature extraction in order to increase the image contrast, thereby making features more easily observable.

Third phase of the system is the Decision-Level Fusion and classification of sample tests. The three decisions are obtained at the phase, Decision I from M-BSIF, Decision II from 2D-Gabor and Decision III from HOG. Weight of 1 is given to correct recognition and incorrect recognition gets 0 in each of the methods. This brings the weights range for each sample to between 0 and 3. Therefore a threshold of 1.9 is fixed to reach a final decision for the system. FYO and PUT wrist vein datasets were used in this study.

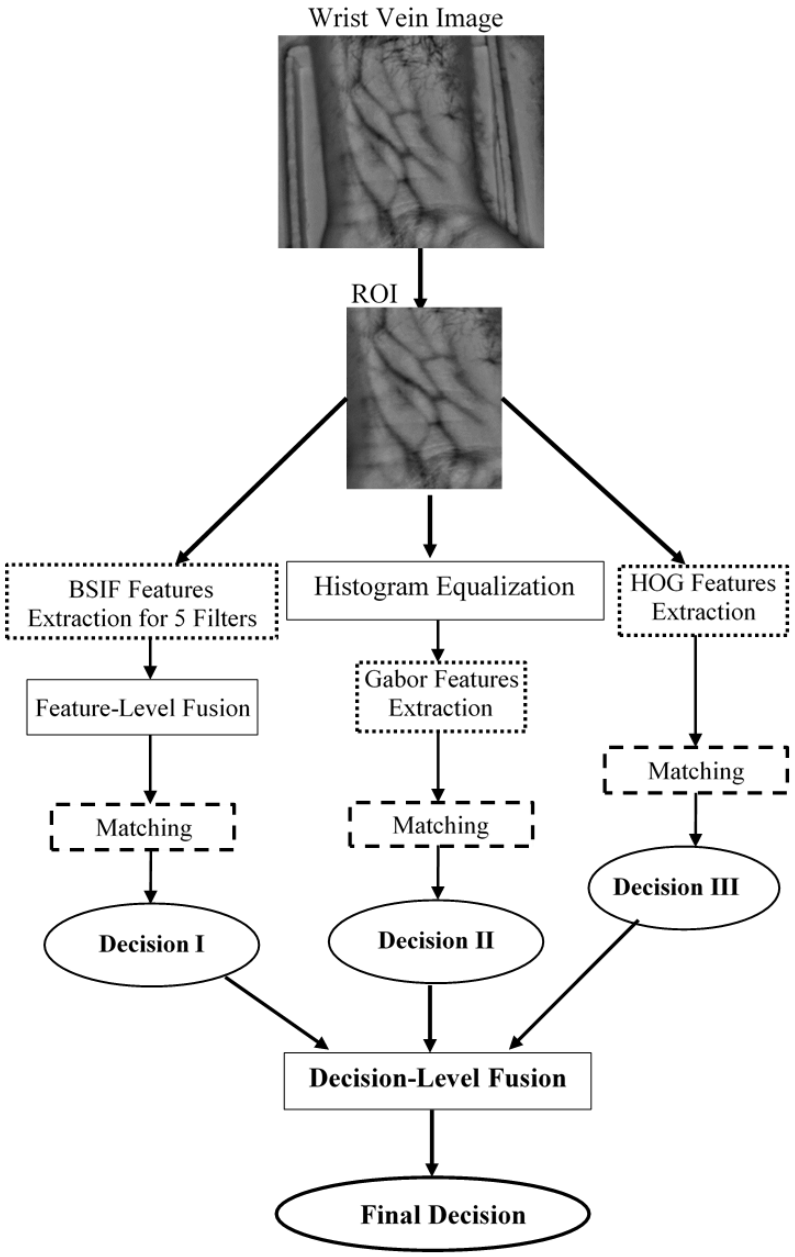


Figure 4.11: Wrist vein authentication system using multiple texture-based feature descriptors

4.2.2.3 Palm Vein Recognition System with CNN Models

This study combined three CNN models, namely, VGG-16, VGG-19 and AlexNet, using the modified versions presented in the previous section. This palm vein recognition system was experimented on three palm vein datasets obtained from VERA, PUT and FYO databases. The structure of the system is shown in Figure 4.12.

The system involves acquiring three separate decisions, one for each CNN models used; Decision I comes from VGG-16, Decision II from VGG-19 while Decision III is from AlexNet. A binary results of either True or False is expected for each of the models. Hence, weight of 1 is given to correct recognition and incorrect recognition gets 0. All weights are summed up per test sample. Therefore, in the final decision, weights 2 and 3 are given True as the overall outcome of the system, while False is for weights 0 and 1.

Similar to the previous experiment, however, there is a need for data augmentation using Keras data generator in order to increase sample images. The sample images were increased to 6400 samples in FYO, 6000 samples in PUT and 5500 samples in VERA. Subsequently, we organized test-set and training datasets as below:

- FYO: 5760 images in the training set (18 images per individual), Test-set: 640 (2 images per individual)
- PUT: 5400 images in the training set (54 images per individual), Test-set: 600 (6 images per individual)
- VERA: 4950 images in the training set (45 images per individual), Test-set: 550 (5 images per individual).

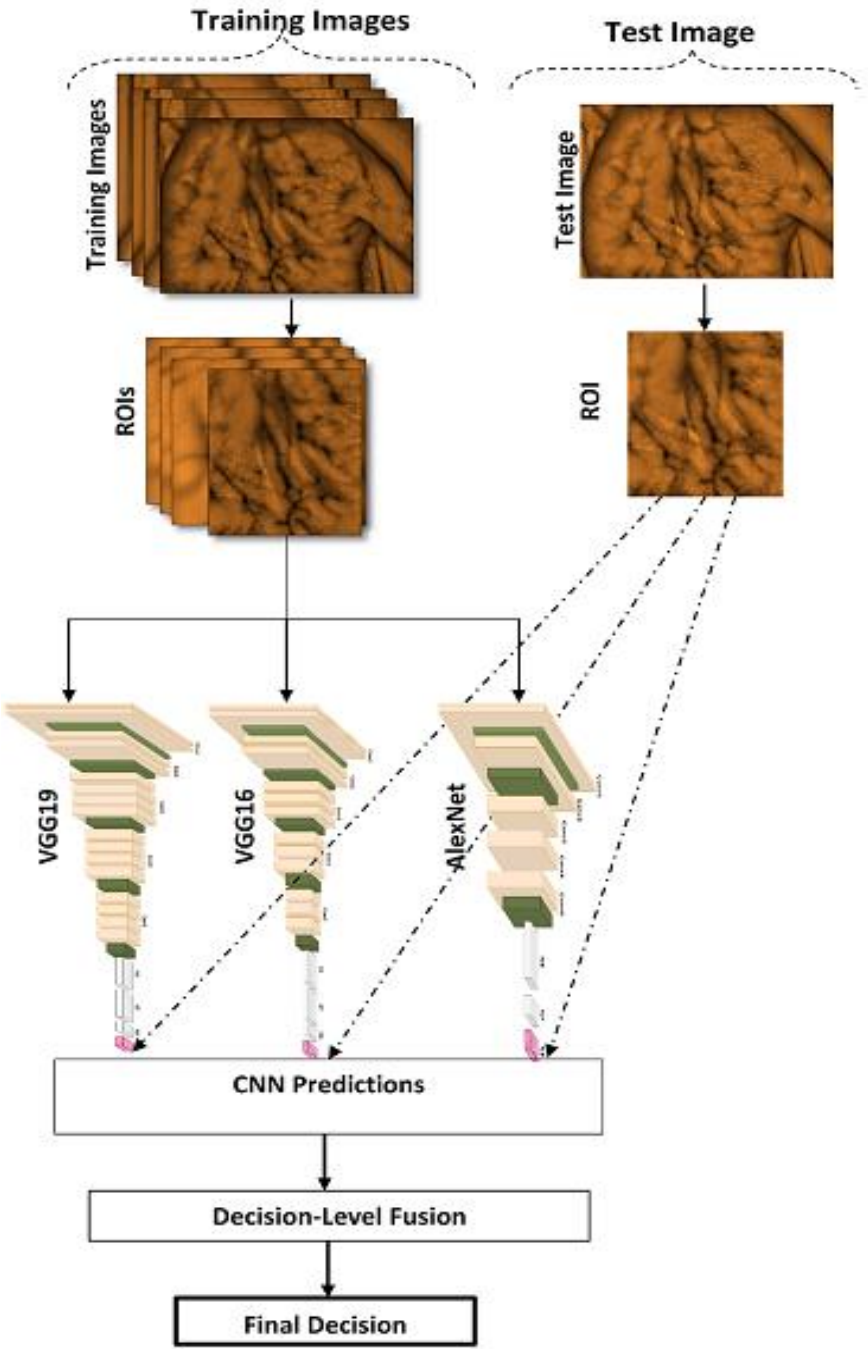


Figure 4.12: Multi-CNN models for palm vein recognition

4.2.3 Multi-color spaces

Another palm vein system which explored different color models by combining the most contributing channel from each of them for a more reliable authentication system is proposed here. The method identifies that the best channel from each of five chosen color spaces (R from RGB, X from XYZ, Y from YUV, L from LAB and V from

HSV). These channels are used to a train a CNN model separately, consequently generating five different predictions which are combined at the decision level.

There are different color models in use in different areas of our life. Color models are abstract computative representation of color according to how they are perceived. The representation is in form of tuples of numbers, typically as three components. On the other hand, color spaces are specific organization of colors which supports reproducible representations of color either in analog or a digital representation and this representations are supported by various technological devices [43].

Application of color space information has been applied pattern recognition research works as well as in computer vision related works, such as analyzing RGB, YCbCr and L$\alpha\beta$ to examined the pros of applying color information on a region-based face authentication system [44], and combining RGB and HSV color spaces for recognizing Iraqi currencies by average color estimation from database of currencies for stronger recognition [45].

There has also been research about human skin detection method by combining RGB, HSV and YCbCr color models in images [46], comparative study of color segmentation by a vector quantization algorithm called Kekre median codebook generation in YCbCr, RGB, and HSV color spaces [47], and comparative study of HSV and YCbCr color spaces in nucleus of white cells detection [48].

In this study, different color spaces were examined, they are grouped into five types according to their origin and their structure. They are categories are explained below.

- **Category 1: RGB** is the standard color representation in digital imaging known as an additive model, it is the only one in this category. RGB stands for the colors red, green, and blue. Colors are added together in this model to make up what is perceived on the digital screens. A digital screen creates tiny pixels that, when observed under a magnifying glass, one of red, green and blue colors can be seen [43]. A typical illustration of this color model is shown in Figure 4:13.

Figure 4.13: Additive RGB color channel representation. Adapted from [49]

- **Category 2: XYZ** was introduced in 1931 by the Commission internationale de l'éclairage (CIE) which translates to International Commission on Illumination in English. It is a gamma correction (linear transformation) of RGB. It is also the only color space in this category. The X-channel corresponds approximately to the red/green part of a color, the Y-channel

corresponds approximately to the lightness, while the Z-value corresponds to the blue/yellow part of a color. The values in X ranges from 0 to 95, it ranges from 0 to 100 in the Y-channel while the Z-value ranges between 0 and 109 [50]. Figure 4.14 shows a typical illustration XYZ from RGB color model.

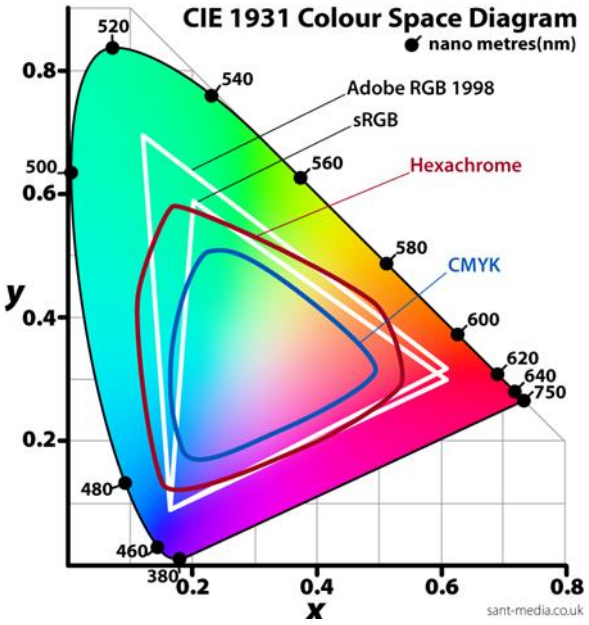


Figure 4.14: XYZ representation from RGB. Adapted from [50]

- **Category 3: LAB and LUV** are non-linear transformation of XYZ. They were developed by CIE in 1976. In LAB, The L component represents the lightness in the color, that is, a white object may have an L value of about 100 while a black object will have close to 0. Hue is the color tone of a color while chroma is the level of saturation of a color represented by channels A and B of the color model as shown in Figure 4.15. Colors with high chroma will be bright or clear, while colors with low chroma will be dull [51, 52].

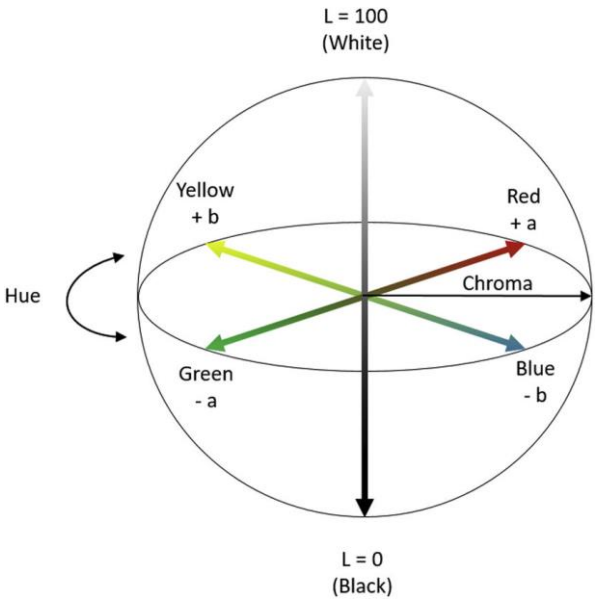


Figure 4.15: LAB color model formation structure. Adapted from [52]

- **Category 4: HSV and HSL** are cylindrical transformation of RGB as shown Figure 4.16. The channels of these models represents saturation, hue and value/lightness [53].

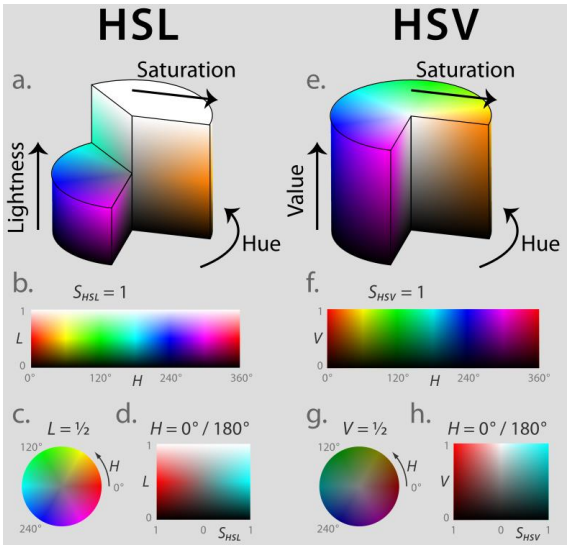


Figure 4.16: Color structure of HSV and HSL [53]

- **Category 5: YCbCr and YUV**: These models are formed by linear transformation of RGB based on the formulas shown in Figure 4.17. Simple

mathematical calculation performed on each pixel by the given formulas, can be used to transform from RGB to YUV or YCbCr and back to RGB.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad \begin{matrix} Y \in [0, 255] \\ C_b \in [0, 255] \\ C_r \in [0, 255] \end{matrix}$$

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \begin{matrix} Y \in [0, 255] \\ U \in [-111, 111] \\ V \in [-157, 157] \end{matrix}$$

Figure 4.17: YCbCr and YUV formation from RGB.

Sample palm vein images in the different color spaces used are shown in Figure 4.18, where it can be seen that all the different color spaces represents images in different ways or formats.

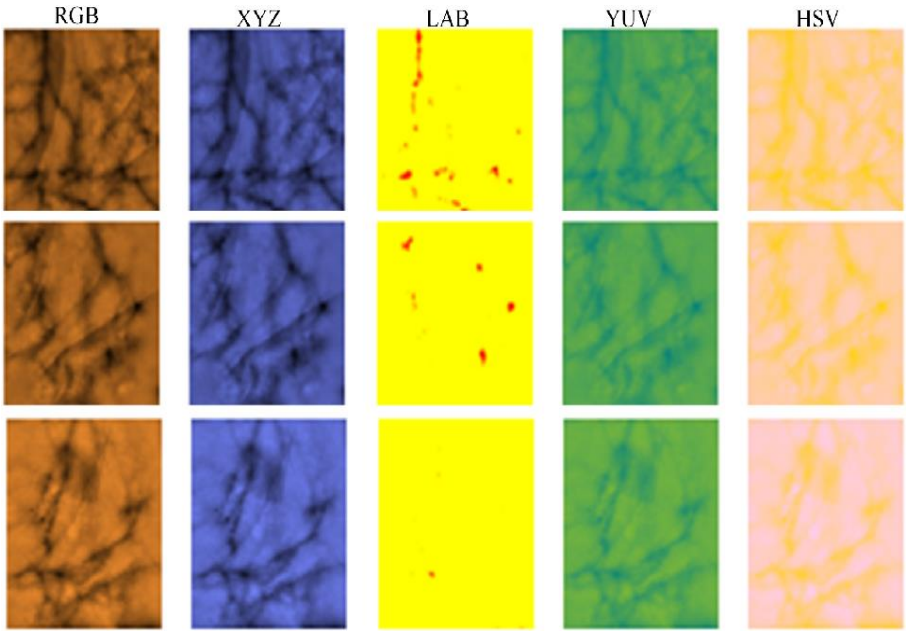


Figure 4.18: Sample hand vein images in different color spaces

The proposed architecture for person identification in this study is combination of decisions from different color categories mentioned above on palm vein images. This

architecture is shown in Figure 4.19, where the decisions from five channels are combined for one final decision of the system. The system is based on modified CNN models discussed in the previous sessions.

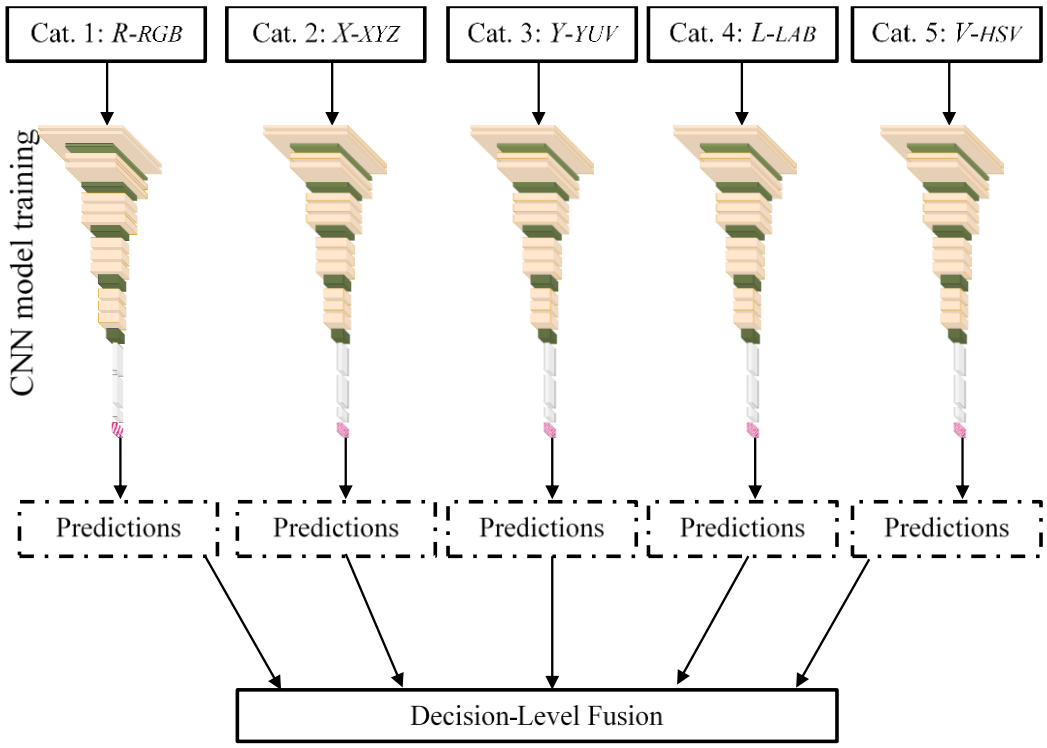


Figure 4.19: Five channels combination for palm vein recognition system

Chapter 5

RESULTS AND DISCUSSION

Experimental results from the various hand vein experiments carried out in this study are presented in this chapter and analyzed. Six different experiments as highlighted in Chapter 4 are conducted. They include multimodal authentication systems where features extracted from palm vein, hand dorsal vein and wrist vein which are combined using different fusion technique; these experiments were carried out using both handcrafted descriptors and CNN models, and as much as 100 percent accuracy was recorded across the datasets and algorithms used [3]. The proposed hand dorsal vein authentication system segments images into five overlapping regions and combines CNN models' prediction from each of the segments using score-level fusion.

Single biometric trait was used in other experiments while combining multiple algorithms. These include a palm vein recognition system which combines a texture-based descriptor method (BSIF) with a CNN model (AlexNet) described in Chapter 4. The results showed approximately 100 percent accuracy across all palm vein datasets used [54]. Similarly, the presented wrist vein system combines three texture-based descriptors (M-BSIF, HOG and 2D Gabor) at the decision level to obtain relatively comparable accuracy of 95.63 percent [55].

Furthermore, another palm vein biometric system combines three CNN models (AlexNet, VGG-16 and VGG-19) at the decision level; the experimental results show

that this system is at least 99 percent accurate across all three datasets used [56]. Lastly, a third authentication system based on palm vein explored different color models by combining the most contributing channel from each of them for a more reliable authentication system. The technique combines decisions from models trained with channels R of RGB, X of XYZ, Y of YUV, L of LAB and V of HSV and obtained approximately 100 percent accuracy across three CNN models (ResNet-50, VGG-19 and AlexNet).

5.1 Result of Multi-Trait Systems

The preliminary stage of this system involve experiments on individual vein biometric trait to determine their individual efficiency before combining them with other traits. This was implemented using three handcrafted feature extractors and the AlexNet CNN model discussed section 4.1.2.1. The corresponding results in Table 5.1 show good performances in general for the CNN model as well as BSIF, which performed best among the handcrafted methods.

Table 5.1: Unimodal system results

Feature Extraction Approach	PUT (Palm)	PUT (Wrist)	VERA (Palm)	Tongji (Palm)	Badawi (Dorsal)	Bosphorus (Dorsal)	FYO (Palm)	FYO (Dorsal)	FYO (Wrist)
BSIF	91.00	82.25	91.45	99.67	95.50	80.17	95.94	95.31	92.50
Gabor	90.75	80.00	90.91	99.93	89.50	59.00	80.00	82.50	83.44
HOG	80.00	65.25	82.55	99.90	79.00	39.00	48.44	52.81	50.94
AlexNet	100.00	98.33	99.20	94.67	98.67	100.00	99.53	99.38	98.13

The main goal of this proposed system, however, is a stronger biometric systems which combines hand vein traits like the three traits available in our database. This was done by applying feature-level fusion; for the hand-crafted feature descriptors, and decision-

level fusion; for the CNN architecture, to raise the robustness of the biometric authentication system. We also carried out the same experiment using PUT datasets since the database is compose of two vein traits; palm and wrist. The results in Table 5.2 show that this system improves the accuracy of the authentication system to 100% in BSIF and the CNN model.

Table 5.2: Multi-trait system results

Feature Extraction Approach	**PUT DB**	**FYO DB**			
	Palm + Wrist	Dorsal + Palm	Dorsal + Wrist	Palm + Wrist	Dorsal + Palmar+ Wrist
BSIF	94.00	99.06	98.13	100.00	100.00
Gabor	96.50	93.13	94.06	94.06	97.50
HOG	84.00	70.31	71.25	70.31	82.50
AlexNet	100.00	100	99.27	99.83	100.00

5.2 Hand Dorsal Vein Recognition System

A preliminary experiment was carried out in order to show that each region is different in composition by checking the entropy of each overlapping region. Entropy is a measure of the degree of randomness in an image given by the following equation:

$$H = -\sum_{j=1}^{k} p_j \, log_2(p_j) \tag{5.1}$$

where *k* is the number of grey levels/ normalized histogram in the image, and p_j is the probability or frequency of occurrence of each grey level [57]. The entropies are shown in Table 5.3, where the lowest total entropy value is shown for the middle region which implies that the middle part is the most stable segment generally.

Table 5.3: Sum of entropies of each region based on FYO dataset

Region	Sum of entropies in the dataset
Top-Left	9128.50
Top-Right	9265.13
Middle	8564.34
Bottom-Left	9147.63
Bottom-Left	9307.43

Subsequently, experiments were performed to ascertain the performance of the proposed system as well as experiments on other algorithm and systems in order to compare the results with similar studies. BSIF and M-BSIF which have been used in similar experiments were also implemented using the same datasets to compare corresponding results to the proposed method. These methods are some of the most efficient texture-based methods in the literature. For further comparison, we implemented the CNN model which takes the ROI as a whole, to compare our method where separate training of five regions were carried out. For all these implementations, the aforementioned training and test datasets were used. The results (accuracies) shown in Table 5.4 showed that our proposed system earned the best accuracy for each of the datasets.

Table 5.4: Hand dorsal vein recognition results

Method	Database	Badawi	Bosphorus	FYO	Average
Handcrafted methods	BSIF	98.20	92.42	95.00	95.21
	M-BSIF	98.20	94.41	96.88	96.50
CNN models	AlexNet	99.00	99.67	98.90	99.19
	VGG16	97.60	98.50	94.37	96.82
CNN (5 regions) SLF Fusion	AlexNet	99.70	99.83	99.84	99.79
	VGG16	99.60	99.83	99.38	99.60

Additionally, we explored other fusion techniques that have been used in literature for score level fusion rather than the traditional techniques where values are only normalized and appended. For example, symmetric addition of scores has been proposed [58] while Weighted Quasi-Arithmetic Mean (WQAM) for fusion at the score stage in multi-biometric systems was proposed in [59]. WQAM involve the mathematical functions such as trigonometry functions, cosine, sine and tangent along with some defined weight for score level fusion. General formula of M for this method is given as follows:

$$M(s) = f^{-1}\left(\sum_{j=1}^{n} w_j f\left(s_j\right)\right), \tag{5.2}$$

where *s* is the input sample, *f* is the mathematical function and f^{-1} is its corresponding inverse, while w_j is weight assigned to each score. The weights add up to 1.

Therefore, WQAM functions adapted from [59] were used to compare the original SLF in this study. The results given both in Tables 5.5 and 5.6 for experiments carried

out on AlexNet and VGG-16, respectively, indicate that the system works effectively across different score level fusion techniques.

Table 5.5: Accuracy of score level fusion techniques in AlexNet

Fusion Method			**Badawi**	**Bosphorus**	**FYO**
Normal SLF			99.70	99.83	99.84
	f	**M**			
WQAM	$\sin(\frac{\pi}{2}s)$	$\frac{2}{\boldsymbol{\pi}}\arcsin(s)\left(\sum_{j=1}^{n} w_j \sin(\frac{\boldsymbol{\pi}}{\mathbf{2}} s_j)\right)$	99.80	99.92	100.00
	$\cos(\frac{\pi}{2}s)$	$\frac{2}{\boldsymbol{\pi}}\arccos(s)\left(\sum_{j=1}^{n} w_j \cos(\frac{\boldsymbol{\pi}}{\mathbf{2}} s_j)\right)$	99.40	99.42	99.06
	$\tan(\frac{\pi}{2}s)$	$\frac{2}{\boldsymbol{\pi}}\arctan(s)\left(\sum_{j=1}^{n} w_j \tan(\frac{\boldsymbol{\pi}}{\mathbf{2}} s_j)\right)$	99.30	99.08	99.06
	s^r	$\left(\sum_{j=1}^{n} w_j (s_j)^r\right)^{1/r}$	99.40	99.33	98.91
	r^s	$log_r\left(\sum_{j=1}^{n} w_j r^{s_j}\right)$	99.70	100.00	100.00

Table 5.6: Accuracy of score level fusion techniques in VGG16

Fusion Method		**Badawi**	**Bosphorus**	**FYO**
Normal SLF		99.60	99.83	99.38
WQAM	f=$\sin(\frac{\pi}{2}s)$	100.00	100.00	99.53
	f=$\cos(\frac{\pi}{2}s)$	98.90	98.70	99.06
	f=$\tan(\frac{\pi}{2}s)$	98.20	98.20	98.44
	f= s^r	98.90	98.70	98.59
	f= r^s	100.00	100.00	99.69

5.3 Palm Vein Recognition System

Table 5.7 presents different experimental results carried out on the proposed palm vein recognition system. It shows that several other experiments were also carried out along with the proposed system for the purpose of comparison. The table show that the system proposed here attained the best results across all datasets, with accuracies as much as 100% in some of the datasets used. Furthermore, different data fusion algorithms were examined for the sake of comparison (Feature-level fusion, Score-level fusion, Decision-level fusion). The results also showed that score-level fusion out-performed the other methods and therefore was the preferred fusion method for fusing the five overlapping regions proposed.

Table 5.7: Palm vein authentication results

Database / **Method**	**CASIA**	**FYO**	**PUT**	**Tongji**	**VERA**
BSIF	96.25	95.00	98.00	98.33	99.77
BSIF (Middle region)	92.17	88.13	97.50	98.33	98.86
BSIF (FLF 5 regions)	97.33	95.63	98.00	99.00	99.77
BSIF (DLF 5 regions)	95.25	87.19	97.50	98.33	99.32
BSIF (SLF 5 regions) **(Decision I)**	97.33	95.94	98.00	98.83	99.55
AlexNet **(Decision II)**	95.17	98.91	91.50	98.00	94.55
Proposed method	99.83	100.00	99.00	100.00	100.00

5.4 Wrist Vein Authentication System

Over-fitting is often a concern at the training phase, therefore, to prevent over fitting in the system, we carried out cross validation in each of the wrist vein experiment as follows:

- FYO wrist vein dataset is made up of samples taken in sessions, and there is only one image per session. Therefore, the first session was employed for training and while the second one was employed for testing. And for cross-examination, these sessions were swapped to carry out the experiment second time, that is; the second session was used to train the system while the first was to test the system. The average outcome of the two attempts was taken as the mean accuracy of the system.
- PUT dataset was captured in three sessions with 4 images in each session. In this experiment, two of these sessions were employed for training while the other one was used for testing. Therefore, for cross validation, datasets for testing and training were swapped three times in other to have three separate implementations where different sets of sessions are used for testing and training. Subsequently, the mean of the three results was considered as the system's overall accuracy.

Table 5.8 shows the accuracy of the individual methods. It can be inferred from the results that BSIF shows the best performance and even better performance when used as M-BSIF. At this stage, three separate decisions are obtained, and they are used to decide overall decision of the system using decision-level fusion with varying weights. BSIF and M-BSIF perform substantially better than 2D Gabor and HOG, hence M-BSIF was assigned 2 as the weight while 2D Gabor and HOG decisions weighed 1

each. The overall system decision is made by setting 2 as the lower bound of true detections while others outside the category fall under false positives.

Table 5.8: Wrist vein authentication system results

Experiment	**FYO DB**	**PUT DB**
HOG	69.38	83.58
Gabor	85.78	86.83
BSIF	93.28	92.42
M-BSIF	94.69	93.83
BSIF+Gabor (FLF)	87.81	87.58
BSIF+Gabor+HOG (DLF)	94.38	93.33
M-BSIF+Gabor+HOG (DLF)	95.63	93.92

The results shown in Table 5.8 show that the system proposed here is superior to the individual methods across the datasets used. For the sake of comparison, we tested a similar method by combining only Gabor and BSIF using at the feature stage and combining BSIF, HOG and 2D Gabor at the decision stage. Table 5.8 shows that the method proposed here out-performs both methods.

5.5 Palm Vein Authentication System Using Multiple CNN Models

This study proposes a palm vein authentication system where three CNN model results are fused together. In addition, the study also compared the efficiency of three CNN models used (VGG-16, VGG-19 and AlexNet) in palm vein biometrics. Datasets used for the experiment include datasets from VERA, PUT and FYO databases.

Furthermore, the effect of shrinking the amount of kernels/filters used in the convolutional layer of these CNN models was checked in terms of computation time. Table 5.9 presents the training time or the modified versions versus the original

versions of the three CNN models. It shows a reduction of approximately 450% across the models when the number of training epoch is fixed at 30.

Table 5.9: Computation time (sec) for original and modified CNN models

Experiment / Database	VGG-16	Modified VGG-16	VGG-19	Modified VGG-19	AlexNet	Modified AlexNet
FYO	131325.41	27843.18	136282.18	29402.61	109850.63	18446.27
PUT	107353.45	24838.32	123842.00	27353.93	141014.50	17300.00
VERA	97840.38	22762.88	115968.66	24901.59	132259.94	15728.60

Significantly sufficient results were obtained from three models employed in this study. Additionally, the result of the proposed study which fuses them at the decision level showed significant improvement in accuracy of the recognition system. The system attained close to 100% across the datasets: 99.83%, 99.26% and 99.06% accuracy on PUT, VERA and FYO datasets, respectively as shown in Table 5.10.

Table 5.10: Accuracy (%) of palm vein authentication methods

Database / Experiment	FYO	PUT	VERA
AlexNet	90.53	94.33	98.53
VGG-16	94.50	97.67	98.72
VGG-19	96.09	97.67	94.50
All three combined	**99.06**	**99.83**	**99.26**

5.6 Multi-Color Spaces Authentication System

This study proposes the use of different color models for human recognition by fusing the results of most reliable or contributing channel to hand vein pattern. VERA, PUT

and FYO palm vein datasets were used for this study. Each of the datasets were augmented to ensure that there are enough samples for the CNN models. Selecting the best channels from the color spaces is an important aspect of this study. Each CNN model was trained using one channel as shown in Figure 5.1.

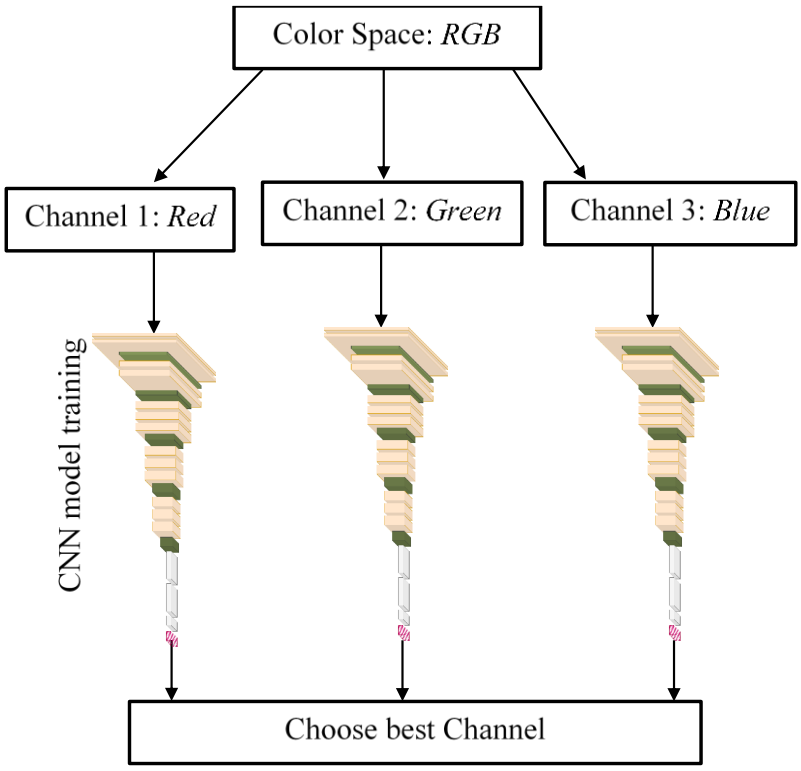


Figure 5.1: Preliminary experiment to determine best channel

The best channels generally have high accuracy across CNN models and palm vein dataset as shown in Table 5.11. The outcomes of the preliminary experiments are given in Table 5.11. These results were used to identify the best channel per color space as follow:

- **RGB, XYZ**: Due to the construction of the color spaces; red-green-blue colors in RGB, and their linear transformation in XYZ; all the channels performed favorably. However, R from RGB, and X from XYZ were chosen by popular experimental votes.

- **YCbCr, YUV**: Luminance component Y performed well in both cases, others are chroma components with very low performance.
- **LAB, LUV**: L component which is closely related to human perception of light performed favorably. Similarly, other channels which are chroma components performed poorly.
- **HSV, HSL**: H (hue) and S (saturation) in both color spaces performed poorly. V value (HSV) and L lightness (HSL) are therefore preferred.

Table 5.11: Preliminary color channel efficiency results

CNN model		**AlexNet**			**ResNet-50**			**VGG-19**		
DB / **Color space**		FYO	PUT	VERA	FYO	PUT	VERA	FYO	PUT	VERA
RGB	R	96.72	98.67	96.70	93.59	96.50	99.82	91.41	97.50	91.38
	G	93.28	98.50	98.90	95.63	98.83	59.45	95.31	87.83	35.41
	B	91.09	98.33	73.57	88.28	95.33	38.71	95.63	92.83	98.17
XYZ	X	96.09	93.33	99.63	98.44	98.50	98.72	94.22	89.00	99.08
	Y	96.25	98.83	28.07	93.78	27.33	33.21	91.87	82.00	99.08
	Z	93.91	94.50	18.35	17.81	69.00	76.33	90.78	94.83	89.36
YCrCb	Y	94.69	59.00	99.63	99.75	85.50	99.75	91.25	95.83	97.80
	Cr	86.09	24.83	0.92	12.19	1.50	0.92	62.97	4.83	0.92
	Cb	92.50	27.00	0.91	81.25	20.67	0.92	94.84	1.17	0.92
YUV	Y	96.09	77.17	98.53	99.75	98.67	99.75	93.12	97.17	99.27
	U	52.97	1.67	0.92	1.09	2.67	0.92	92.97	32.00	0.92
	V	80.00	2.33	2.92	7.50	2.17	0.92	92.19	10.67	0.92
LAB	L	97.50	77.50	96.70	99.84	91.50	99.88	93.28	79.50	99.08
	A	13.59	1.50	0.92	10.31	6.33	0.92	6.25	1.83	0.92
	B	63.91	33.67	0.92	0.94	1.83	0.92	82.03	1.83	0.92
HSV	H	2.81	75.67	0.92	0.78	10.67	0.92	41.56	67.17	0.92
	S	21.56	89.17	0.91	0.97	71.00	0.92	85.78	76.00	0.92
	V	95.31	96.50	99.45	98.46	99.00	90.46	94.38	98.90	97.80
HLS	H	8.59	58.00	0.91	0.92	18.00	0.92	22.66	48.67	0.92
	L	97.03	95.33	96.88	99.53	63.17	99.63	92.19	88.83	98.35
	S	76.25	98.00	0.91	0.63	95.83	0.92	0.92	95.50	0.92

Additionally, after identifying the best channels in all the color spaces used in the experiment, further confirmation that the channels would give the best results in all cases, were needed. This was carried out by adding precision and recall experiments as shown by the results in Table 5.12 and Table 5.13, respectively. **Recall,** also known as Sensitivity, is the percentage of Real Positive samples that are accurately predicted. It examines the scope of the Real Positive samples by the Predicted Positive rule while **Precision,** also known as Confidence, refers to the percentage of positive samples that are really positive samples [60].

Table 5.12: Precision results of chosen color channels

	CNN Model	AlexNet			ResNet-50			VGG-19		
	DB	FYO	PUT	VERA	FYO	PUT	VERA	FYO	PUT	VERA
Chosen Channels	**R:RGB**	98	87	97	100	97	100	98	98	94
	X:XYZ	98	95	97	99	99	99	97	91	99
	L:LAB	99	97	99	100	94	100	97	85	99
	Y:YUV	98	97	100	100	99	97	96	97	99
	V:HSV	98	95	100	99	99	95	97	98	100
	YCbCr	98	86	91	100	89	100	97	96	98
	L:HLS	98	86	98	100	86	100	96	92	99

The results of the system that combines five color channels from five different color models and categories are given in Table 5.14. Four possible combinations were examined as seen in the table, although the proposed combination is Red channel from RGB, X channel from XYZ, V channel from HSV, channel L from LAB and channel Y from YUV. The results show that the proposed method has the potential to

significantly increase the accuracy of a recognition system, with close to 100 percent accuracy across all three datasets used.

Table 5.13: Recall results of chosen color channels

CNN Model		AlexNet			ResNet-50			VGG-19		
DB		FYO	PUT	VERA	FYO	PUT	VERA	FYO	PUT	VERA
Chosen Channels	**R:RGB**	97	79	97	100	96	100	96	97	91
	X:XYZ	96	94	96	98	98	99	94	89	99
	L:LAB	97	96	98	100	92	100	93	80	99
	Y:YUV	97	96	100	100	99	96	93	97	99
	V:HSV	95	94	99	98	99	90	94	97	99
	YCbCr	95	76	68	100	85	100	95	96	98
	L:HLS	96	66	97	100	63	100	92	89	98

Table 5.14: Five color channels combination results

	CNN Model	AlexNet			ResNet-50			VGG-19		
	DB	FYO	PUT	VERA	FYO	PUT	VERA	FYO	PUT	VERA
DLF: 5 channels	R:RGB,X:XYZ, L:HLS,L:LAB, Y:YCbCr	100.00	95.50	99.63	100.00	99.67	100.00	99.38	98.67	100.00
	R:RGB,X:XYZ, L:HLS,L:LAB, Y:YUV	99.84	98.33	100.00	100.00	99.50	100.00	99.69	98.33	100.00
	R:RGB,X:XYZ, V:HSV,L:LAB, Y:YCbCr	100.00	98.83	100.00	100.00	99.67	100.00	99.69	99.00	100.00
	R:RGB,X:XYZ, V:HSV,L:LAB, Y:YUV	**99.69**	**99.33**	**100.00**	**100.00**	**99.67**	**100.00**	**99.84**	**99.33**	**100.00**

For the sake of comparison, three combinations instead of five were also examined as shown in Table 5.15. The results are comparable to the proposed method, therefore

this method could also be considered, although we recommend the five channels combination for a more robust system.

Table 5.15: Three channels combination results

	CNN Model	AlexNet			ResNet-50			VGG-19		
	DB	FYO	PUT	VERA	FYO	PUT	VERA	FYO	PUT	VERA
DLF: 3 channels	R:RGB, X:XYZ, L:LAB	99.84	96.50	100.00	99.84	99.67	100.00	99.06	96.17	100.00
	R:RGB, X:XYZ, V:HSV	99.84	96.17	100.00	99.69	99.50	99.63	98.91	99.00	100.00
	R:RGB, X:XYZ, Y:YUV	99.06	96.83	100.00	99.84	99.50	100.00	98.91	98.83	100.00
	R:RGB, L:LAB, V:HSV	99.38	97.67	100.00	99.84	99.67	100.00	99.06	98.50	100.00
	R:RGB, L:LAB, Y:YUV	99.69	98.00	100.00	100.00	99.50	100.00	98.59	98.33	100.00
	L:LAB, V:HSV, Y:YUV	98.91	98.83	100.00	100.00	99.50	99.63	98.28	98.17	100.00
	X:XYZ, V:HSV, Y:YUV	99.06	98.83	100.00	100.00	99.33	99.27	98.44	98.83	100.00

5.7 Overall Comparison

Finally, we compared all the methods presented in this thesis with the state-of-the-art methods that are available in literature. These methods include a system which fuses finger vein and dorsal vein features by employing monogenic local binary pattern and Improved Gaussian Matched Filter for human recognition [14], CNN feature learning and transfer method on hand dorsal vein pattern [61], a dorsal vein recognition system where fractal technique computed by box counting method for tissue properties identification is used [15] and combination of vein features from both right and left

wrists using a combinatory algorithm where two binarized images one based on local thresholding are multiplied by global thresholding [25].

Mirmohamadsadeghi and Drygajlo compared operators and histograms of multi-scale Local Binary Patterns (LBPs) as a descriptors for palm vein patterns and higher-order local pattern descriptors based on Local Derivative Pattern (LDP) histograms, using them for palm vein and palmprint verification and identification [62] while Lee constructed a palm vein image database in 2012 and extracted features from them using 2D Gabor filter [63]. A deep CNN method was used for palm vein and palm print recognition using a large scale database constructed, named Tongji in [64].

The methods compared to the proposed methods range from unimodal to multimodal systems as shown in Table 5.16, where it can be seen that our proposed systems compared favorably against other methods, especially the proposed multi-color spaces authentication system and multi-modal palm, dorsal and wrist system where 100 percent accuracy was recorded in experiments performed with FYO datasets and other datasets used for comparison.

Table 5.16: Comparison of proposed methods the with state-of-the-art methods

	Method	Year	Trait (Vein)	Database	Accuracy (%)
State-of-the-art methods	Zhang et al. [65]	2007	Palm	Own Database	98.80
	Mirmohamadsadeghi & Drygajlo [61]	2011	Palm	CASIA	97.20
	Lee [62]	2012	Palm	Own Database	99.18
	Zhang et al. [63]	2018	Palm	Tongji	100.00
	Trabelsi et al. [14]	2013	Dorsal & Finger	Badawi	98.80
	Al-johania and Elrefaei, [64]	2019	Dorsal	Badawi & Bosphorus	97.00
	Çimen et al. [15]	2021	Dorsal	Bosphorus & SUAS	99.00
	Abed [25]	2017	Palm & Wrist	PUT	94.49
Methods in this study	Multi-modal System (Palm + Wrist + Dorsal)				100.00
	Hand dorsal vein recognition system (5 segments)				99.38
	Palm vein recognition system (AlexNet + BSIF)				100.00
	Wrist vein authentication system (M-BSIF + HOG + 2D Gabor)				95.63
	Palm vein authentication system using multiple CNN models				99.06
	Multi-color spaces authentication system				100.00

Chapter 6

CONCLUSION

6.1 Main Findings

This study establishes a multimodal vein database named FYO composed of three biometric modalities; dorsal vein, palmar vein and wrist vein with the aim of proposing a robust biometric system that is highly efficient and equipped with innate defense against spoof attacks. Furthermore, the availability of these datasets from the same individuals/subjects in one database embroiders multimodal vein biometric recognition research works.

Consequently, two multi-trait recognition systems which combine features of three hand vein traits are proposed: The first one is a feature-level fusion of texture-based feature descriptor BSIF which is a highly effective algorithm based on LBP and LPQ. The second is a decision-level fusion of features extracted using a modified version of AlexNet CNN model. The resultant outputs of the proposed methodologies show significant improvement in system accuracy in comparison with when individual traits are used, as 100 percent accuracy was obtained in both methods.

A third method exploited the benefit of combining multiple traits but from only one trait by using overlapping segments of an image. This architecture is a deep learning structure of the five overlapping regions proposed for dorsal vein recognition system. The system extracts features from each region using variants of CNN models; AlexNet

and VGG-16, and combines them using score-level fusion. The accuracy of this architecture is recorded as above 99 percent in both CNN models. The method is also implemented with handcrafted methods for the sake of comparison, while also comparing with the state-of-the-art methods where it shows a considerable superiority or similar efficiency in all cases.

Furthermore, a number of methodologies which combine more than one algorithm in order to improve hand vein biometric system efficiency are proposed. Palm vein patterns are employed in two different methods. The first one divides images into five overlapping regions and extracts BSIF features from each before applying score-level fusion of the five regions, and a CNN model (AlexNet or VGG-16) is trained in parallel using the whole ROI images. The two parallel structures are then combined using decision-level fusion. The results show improvement in recognition system when compared with other algorithms as 99 – 100 percent accuracy is achieved across all five datasets used in the experiment.

On the other hand, a method that combines three CNN models, namely, VGG-16, VGG-19 and AlexNet as modified in this study is also proposed as a palm vein authentication system. The system trains the datasets three times using a separate CNN model for each training, then the subsequent predictions from each trained model are fused at the decision level. This system achieved at least 99.06 percent accuracy across all three datasets used in the experiment.

Similarly, a wrist vein recognition system which combines handcrafted methods as oppose to deep learning models is proposed. The system fuses three texture-based feature descriptors, that is; M-BSIF (BSIF of multiple filters), 2D Gabor, and HOG

using decision-level fusion. Significant improvement is achieved using this method as oppose to using individual algorithms; 95.63 percent and 93.92 percent are obtained in FYO and PUT datasets respectively.

Finally, the contribution of color in pattern recognition in hand vein biometrics is examined. Different color models are considered and grouped into five based on their origin and structure. The aim of the experiment is to find the best channel in each of the color models for hand vein recognition and then combine these channels for a more robust authentication system. The channels identified are; Red channel from RGB, X channel from XYZ, V channel from HSV, channel L from LAB and channel Y from YUV. These also show that color channels representing luminance contribute most in vein recognition. The accuracy of the overall system shows that combining the best channels improves recognition to approximately 100 percent in all datasets used.

In general, our studies indicate the importance of having multiple traits in human biometric recognition systems, as well as the possibility of improvement when more algorithms are employed. This is evident both in handcrafted methods and deep learning models.

6.2 Future Work

The fallout of this study and research works we came across in the duration of this study have opened many prospective areas of research interest which will be pursued in the nearest future. These will include methods to improve the efficiency of previous studies as well as new areas in relation to this study, especially exploring more deep learning methods and different fusion methods and algorithms in order to obtain an even more efficient system with minimal additional computation time.

REFERENCES

[1] Gayathri M., Malathy C., and Prabhakaran M. (2020). A Review on Various Biometric Techniques, Its Features, Methods, Security Issues and Application Areas. In: Smys S., Tavares J., Balas V., Iliyasu A. (eds) *Computational Vision and Bio-Inspired Computing. ICCVBIC* 2019. Advances in Intelligent Systems and Computing 1108, Springer, Cham. doi: 10.1007/978-3-030-37218-7_99.

[2] Dahea W., and Fadewar H. S. (2018). Multimodal biometric system: A review. *International Journal of Engineering and Technology*. 4. 25-31. doi: 10.13140/RG.2.2.34056.65287.

[3] Toygar Ö., Babalola F.O., and Bitirim Y. (2020). FYO: A Novel Multimodal Vein Database With Palmar, Dorsal and Wrist Biometrics. *IEEE Access* 8, 82461-82470. doi: 10.1109/ACCESS.2020.2991475.

[4] iStock. Retrieved on June 2022 from https://www.istockphoto.com/tr/vekt%C3%B6r/kol-ill%C3%BCstrasyon-vekt%C3%B6r-beyaz-zemin-%C3%BCzerine-komplo-damar-t%C4%B1bbi-kavram%C4%B1-gm695674742-128773341.

[5] Subban R., and Mankame D.P. (2014). Human Face Recognition Biometric Techniques: Analysis and Review. In: Thampi S., Abraham A., Pal S., Rodriguez J. (eds) Recent Advances in Intelligent Informatics. *Advances in Intelligent Systems and Computing*. 235. Springer, Cham. doi: 10.1007/978-3-319-01778-5_47.

[6] Mouad M.H.A., Ashok G., and Pravin Y. (2018). A Review: Palmprint Recognition Process and Techniques. *International Journal of Applied Engineering Research.* 13. 7499-7507.

[7] Karhe R., and Patil M. (2016). Iris Recognition: Review and Open Issues. *International Journal of Latest Trends in Engineering & Technology,* 7(1). 168-175. doi: 10.21172/1.71.023.

[8] Saranya S., Jothi R., and Palanisamy V. (2019). A Review on Ear Recognition Techniques Based On Local Texture Descriptors. *International Journal of Computer Sciences and Engineering*, 7(5), 1583-1587. doi: 10.26438/ijcse/v7i5.15831587.

[9] Bakshe R.C., and Patil A. M. (2014). Hand Geometry Techniques: A Review. *International Journal of Modern Communication Technologies & Research (IJMCTR),* 2(11).

[10] Huang D., Zhu X., Wang Y., and Zhang D. (2016). Dorsal hand vein recognition via hierarchical combination of texture and shape clues. *Neurocomputing*, 214, 815-828. doi: 10.1016/j.neucom.2016.06.057.

[11] Yüksel A., Akarun L., and Sankur B. (2010). Biometric Identification through Hand Vein Patterns, *International Workshop on Emerging Techniques and Challenges for Hand-Based Biometrics,* pp. 1-6, doi: 10.1109/ETCHB.2010.5559295.

[12] Hamid H., Narang V.K., and Singh D. (2017). Dorsal Hand Vein Analysis for Security Systems. *International Journal of Engineering and Technology*, 9(4), 3075-3080. doi: 10.21817/ijet/2017/v9i4/170904044.

[13] Sayed M., Taha M., and Zayed H. (2021). Real-Time Dorsal Hand Recognition Based on Smartphone. *IEEE Access*, 9, 151118-151128. doi: 10.1109/access.2021.3126709.

[14] Trabelsi R., Alima D., and Sellami D. (2013). A New Multimodal Biometric System Based on Finger Vein and Hand Vein Recognition. *International Journal of Engineering and Technology*. 5. 3175-3183.

[15] Çimen M., Boyraz O., Yildiz M., and Boz, A. (2021). A new dorsal hand vein authentication system based on fractal dimension box counting method. *Optik*, 226, 165438. doi: 10.1016/j.ijleo.2020.165438.

[16] Wu K., Lee J., Lo T., Chang K. and Chang C. (2013). A secure palm vein recognition system. *Journal of Systems and Software*, 86(11), pp.2870-2876. doi: 10.1016/j.jss.2013.06.065.

[17] Nayar G., Thomas T. and Emmanuel, S. (2021). Graph based secure cancelable palm vein biometrics. *Journal of Information Security and Applications*, 62, 102991. doi: 10.1016/j.jisa.2021.102991.

[18] Aberni Y., Boubchir L., and Daachi B. (2020). Palm vein recognition based on competitive coding scheme using multi-scale local binary pattern with ant colony

optimization. *Pattern Recognition Letters*. 136, 101-110. doi: 10.1016/j.patrec.2020.05.030.

[19] Savitha A., and Ramegowda. (2017). A Comparative Study of Palm Vein Feature Extraction and Classification. Materials Today: Proceedings, 4(11), 11882-11887. doi: 10.1016/j.matpr.2017.09.107.

[20] Adiraju R., Masanipalli K., Reddy T., Pedapalli R., Chundru S., and Panigrahy A. (2021). An extensive survey on finger and palm vein recognition system. *Materials Today: Proceedings*, 45, 1804-1808. doi: 10.1016/j.matpr.2020.08.742.

[21] Pflug A., Hartung D., and Busch C. (2012). Feature extraction from vein images using spatial information and chain codes. *Information Security Technical Report*. 17, pp. 26-35. doi: 10.1016/j.istr.2012.02.003.

[22] Garcia-Martin R., and Sanchez-Reillo R. (2020). Wrist Vascular Biometric Recognition Using a Portable Contactless System. *Sensors* (Basel, Switzerland), 20(5), 1469. doi: 10.3390/s20051469.

[23] Why are the blood veins in your wrist blue? Retrieved on June 2022 from https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/why-are-the-blood-veins-in-your-wrist-blue/articleshow/65994347.cms.

[24] Akhloufi M., and Bendada A. (2008). Hand and wrist physiological features extraction for near infrared biometrics, *Computer and Robot Vision CRV'08*. Canadian Conference on, pp. 341-344.

[25] Mohammed H. A. (2017). Wrist and palm vein pattern recognition using gabor filter. *Journal of Qadisiyah Computer Science and Mathematics*. 9(1), pp. 49-60.

[26] Garcia-Martin R., and Sanchez-Reillo R. (2020). Vein Biometric Recognition on a Smartphone. *IEEE Access*. 8, pp. 104801-104813, doi: 10.1109/ACCESS.2020.3000044.

[27] Kabaciński R., and Kowalski, M. (2011). Vein pattern database and benchmark results. *Electronics Letters*, 47(20), 1127 – 1128. doi: 10.1049/el.2011.1441.

[28] Zhang L., Cheng Z., Shen Y., and Wang D. (2018). Palmprint and Palmvein Recognition Based on DCNN and a New Large-Scale Contactless Palmvein Dataset. *Symmetry*, 10, 78. doi: 10.3390/sym10040078.

[29] Tome P., and Marcel S. (2015). On the vulnerability of palm vein recognition to spoofing attacks, pp. 319–325. In: 8th IAPR *International Conference on Biometrics (ICB)*. doi: 10.1109/ICB.2015.7139056. Retrieved November 2020 from http://publications.idiap.ch/index.php/publications/show/3096.

[30] Shahin M., Badawi A., and Kamel M. (2007). Biometric authentication using fast correlation of near infrared hand vein patterns, *Int. J. Biol. Med. Sci*., vol. 2, no. 3, pp. 141-148.

[31] Yildiz M. Z., Boyraz O. F., Guleryuz E., Akgul A., and Hussain, I. (2019). A Novel Encryption Method for Dorsal Hand Vein Images on a Microcomputer. *IEEE Access*, 7, 60850-60867.

[32] Hao Y., Sun Z., Tan T., and Ren C. (2008). Multi-spectral palm image fusion for accurate contact-free palmprint recognition. Proceedings of *IEEE International Conference on Image Processing*, pp.281-284, USA.

[33] Kannala J., and Rahtu E. (2012). BSIF: Binarized statistical image features. *21st International Conference on Pattern Recognition (ICPR2012)*, Tsukuba, pp. 1363-1366.

[34] Liu C., Koga M., and Fujisawa H. (2005). Gabor feature extraction for character recognition: comparison with gradient feature. *8th International Conference on Document Analysis and Recognition (ICDAR'05)*, pp. 121-125 Vol. 1, doi: 10.1109/ICDAR.2005.119.

[35] Anuj S. (2018). Through The Eyes of Gabor Filter. June 2022 from https://medium.com/@anuj_shah/through-the-eyes-of-gabor-filter-17d1fdb3ac97.

[36] Sharma M. and Ghosh H. (2015). Histogram of gradient magnitudes: A rotation invariant texture-descriptor. *IEEE International Conference on Image Processing (ICIP)*. pp. 4614-4618, doi: 10.1109/ICIP.2015.7351681.

[37] Tyagi M. (2021). HOG (Histogram of Oriented Gradients): An Overview. Retrieved on June 2022 from https://towardsdatascience.com/hog-histogram-of-oriented-gradients.

[38] Krizhevsky A., Sutskever I., and Hinton G. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90. doi: 10.1145/3065386.

[39] Ha I., Kim H., Park S. and Kim H. (2018). Image retrieval using BIM and features from pretrained VGG network for indoor localization. *Building and Environment*, 140, pp.23-31. doi: 10.1016/j.buildenv.2018.05.026.

[40] Kaiming H., Xiangyu Z., Shaoqing R. and Jian, S. (2016). Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770-778. doi: 10.1109/CVPR.2016.90.

[41] Brital A. (2021). Residual Networks with Examples. Retrieved on April 2022 from https://medium.com/@AnasBrital98/residual-networks-with-examples-80b47cacecf4.

[42] Ross A. and Jain A. K. (2004). Multimodal biometrics: An overview. *12th European Signal Processing Conference*, pp. 1221-1224.

[43] Hasting G. and Rubin, Alan. (2012). Colour spaces - a review of historic and modern colour models. *African Vision and Eye Health*. doi: 71. 10.4102/aveh.v71i3.76.

[44] Tome P., Vera-Rodriguez R., Fierrez J., and Ortega-Garcia J. (2013). Fusion of Facial Regions Using Color Information in a Forensic Scenario. 8259. 399-406. doi: 10.1007/978-3-642-41827-3_50.

[45] Aziz M. M. (2016). Iraqi currency recognition system using RGB and HSV color average. *International Journal of Business and Administrative Studies*, 2(1), 9-15. doi: 10.20469/ijbas.2.10003-1.

[46] Kolkur S., Kalbande D., Shimpi P., Bapat C., and Jatakia, J. (2017). Human Skin Detection Using RGB, HSV and YCbCr Color Models, in *International Conference on Communication and Signal Processing 2016 (ICCASP 2016)*, Atlantis Press. doi: 10.2991/iccasp-16.2017.51.

[47] Alkinani, F. and Rahma A. M. (2019). A Comparative Study of KMCG Segmentation Based on YCbCr, RGB, and HSV Color Spaces. doi: 10.29304/jqcm.2019.11.1.468.

[48] Vaghela H., Modi H., Pandya M., and Potdar B. M. (2016). Comparative Study of HSV Color Model and Ycbcr Color Model to Detect Nucleus of White Cells. *International Journal of Computer Applications*. 150. 38-42. doi: 10.5120/ijca2016911614.

[49] The Difference between RGB and CMYK Retrieved on May 2022 from https://blog.printaura.com/blog/art-resources/the-difference-between-rgb-and-cmyk.

[50] XYZ / CIE Color Spaces. Pipette. Retrieved on June 2022 from https://www.sttmedia.com/colormodel-xyz.

[51] Beetsma J. The CIELAB L*a*b* System – the Method to Quantify Colors of Coatings. September 11, 2020. Retrieved on May 2022 from https://knowledge.ulprospector.com/10780/pc-the-cielab-lab-system-the-method-to-quantify-colors-of-coatings/.

[52] Ly, B. Dyer, E., Feig, J., Chien, A., and Bino, S. (2020). Research Techniques Made Simple: Cutaneous Colorimetry: A Reliable Technique for Objective Skin Color Measurement. *The Journal of investigative dermatology*. 140. 3-12.e1. doi: 10.1016/j.jid.2019.11.003.

[53] HSL and HSV. Retrieved on June 2022 from https://en.wikipedia.org/wiki/HSL_and_HSV.

[54] Babalola F. O., Bitirim, Y., and Toygar, Ö. (2020). Palm vein recognition through fusion of texture-based and CNN-based methods. *Signal, Image and Video Processing*. 15(3), 459-466. doi: 10.1007/s11760-020-01765-6.

[55] Babalola F. O., Toygar Ö., and Bitirim Y. (2021). Wrist Vein Recognition by Fusion of Multiple Handcrafted Methods. *3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA),* 2021. pp. 1-5. doi: 10.1109/HORA52670.2021.9461367.

[56] Babalola F. O., Toygar Ö., and Bitirim Y. (2021). A Palm Vein Recognition Approach by Multiple Convolutional Neural Network Models. *European Journal of Science and Technology*. 29, pp. 237-242. doi:10.31590/ejosat.1016532.

[57] Yurtkan, K., Demirel, H. (2013). Entropy based feature selection for improved 3D facial expression recognition. *Signal Image Video Process*. 8(2), 267–277. doi: 10.1007/s11760-013-0543-1.

[58] Cheniti M., Boukezzoula N., and Akhtar Z. (2017). Symmetric sum-based biometric score fusion. *IET Biometrics*. 7(5), 391-395.

[59] Abderrahmane H., Noubeil G., Lahcene Z., Akhtar Z., and Dasgupta D., 2020. Weighted quasi-arithmetic mean based score level fusion for multi-biometric systems. *IET Biometrics*, 9(3), 91-99.

[60] Powers D. M. W. (2011). Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*. 2. 37-63. doi: 10.9735/2229-3981.

[61] Al-johania N., and Elrefaei L. (2019). Dorsal hand vein recognition by convolutional neural networks: Feature learning and transfer learning approaches. *International Journal of Intelligent Engineering and Systems*, 12(3), 178-191.

[62] Mirmohamadsadeghi L., and Drygajlo A. (2011). Palm vein recognition with local binary patterns and local derivative patterns. *2011 International Joint Conference on Biometrics (IJCB)*. 1-6. doi: 10.1109/IJCB.2011.6117804.

[63] Lee, J. (2012). A novel biometric system based on palm vein image. *Pattern Recognit. Lett*. 33, 1520–1528.

[64] Zhang, L., Cheng, Z., Shen, Y., et al. (2018). Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset. *Symmetry*. 10, 78.

[65] Zhang Y., Li Q., and You J. (2007). Palm vein extraction and matching for personal authentication. In:Qiu,G., Leung, C.,Xue,X., Laurini, R. (eds.) *Advances in Visual Information Systems (VISUAL).* Lecture Notes in Computer Science, vol. 4781. Springer, Berlin, Heidelberg.

APPENDICES

Appendix A: FYO Hand Vein Webpage

FYO: Multimodal Vein Database with Palmar, Dorsal and Wrist Biometrics

Eastern Mediterranean University

Download

The FYO Multimodal Vein Database is intended for research on multimodal biometrics based on dorsal, palmar and wrist vein patterns. The database is open access to researchers upon request as described below. The vein patterns were acquired in Computer Engineering Department laboratories of Eastern Mediterranean University from volunteers by the researchers PhD Reasearch Assistant Felix Babalola, Assoc.Prof.Dr. Yıltan Bitirim and Prof.Dr.Önsen Toygar.

The images of the database are acquired using a low cost medical vein finder device equipped with 1/3 inch infrared CMOS camera rounded with 12 pieces of infrared LED light sources. The hardware setup for vein acquisition system is shown in the following figure. Note that three separate hand guides were used to capture images from the dorsal, palmar and wrist parts of the body.

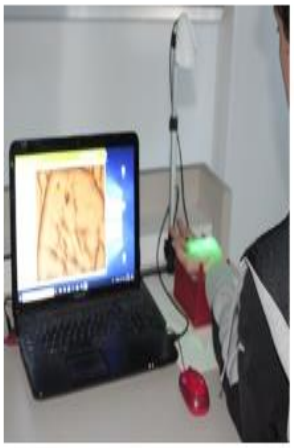


Figure. Hardware Setup for Vein Acquisition

There are overall 1920 images from 160 subjects. From each subject, the following images are acquired:

Appendix B: Agreement for Downloading FYO Database

FYO Multimodal Vein Database License Agreement

The goal of our research is to explore new methods for multimodal biometrics based on dorsal, palmar and wrist vein patterns, and to serve a source based on dorsal, palmar and wrist vein patterns to biometrics research community.

Release of the database:

To advance the research efforts in the field, the database will be made available to researchers. The researcher must agree to the conditions below, sign the document and send its scanned copy with her/his request via e-mail. Then, the link to download the database will be sent.

Conditions:

1. Redistribution: I agree that the copy I receive is to be used for research purposes by myself/my team/my lab only. I will not redistribute a part or the whole database.

2. Modification: I agree not to modify the database in any form.

3. Commercial use: The database is released for research purposes only. I will not use the original data or modified copies for any commercial purpose.

4. Warranty: The database comes without any warranty. The authors of the FYO database are not responsible for any damage caused by the use of the database.

5. Citations: All documents and papers that report any research regarding the FYO database must cite the following article:

   Ö. TOYGAR, F.O. BABALOLA, Y. BİTİRİM, "FYO: A Novel Multimodal Vein Database with Palmar, Dorsal and Wrist Biometrics", IEEE Access, 2020.

Full name and surname with title: _____

E-mail: _____

Telephone: _____

Institution address: _____

Signature: _____

Date: _____