WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

2020

Palmprint Gender Classification Using Deep Learning Methods

Minou Khayami
WVU, mk0057@mix.wvu.edu

Palmprint Gender Classification Using Deep Learning Methods

Minou Khayami

Thesis submitted to the

Benjamin M. Statler College of Engineering and Mineral Resources

at West Virginia University

in partial fulfillment of the requirements

for the degree of

Master of Science

in

Electrical Engineering

Jeremy Dawson, Ph.D., Chair

Nasser Nasrabadi, Ph.D.

Matthew Valenti, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2020

Keywords: Palmprint, Gender Classification, Convolutional Neural Network, Deep Learning

Abstract

Palmprint Gender Classification Using Deep Learning Methods

Minou Khayami

Gender identification is an important technique that can improve the performance of authentication systems by reducing searching space and speeding up the matching process. Several biometric traits have been used to ascertain human gender. Among them, the human palmprint possesses several discriminating features such as principal-lines, wrinkles, ridges, and minutiae features and that offer cues for gender identification. The goal of this work is to develop novel deep-learning techniques to determine gender from palmprint images. PolyU and CASIA palmprint databases with 90,000 and 5502 images respectively were used for training and testing purposes in this research. After ROI extraction and data augmentation were performed, various convolutional and deep learning-based classification approaches were empirically designed, optimized, and tested. Results of gender classification as high as 94.87% were achieved on the PolyU palmprint database and 90.70% accuracy on the CASIA palmprint database. Optimal performance was achieved by combining two different pre-trained and fine-tuned deep CNNs (VGGNet and DenseNet) through score level average fusion. In addition, Gradient-weighted Class Activation Mapping (Grad-CAM) was also implemented to ascertain which specific regions of the palmprint are most discriminative for gender classification.

I dedicate this thesis to my parents and friends.

What is essential is invisible to the eye[1] …

---

[1] The Little Prince by Antoine de Saint Exupéry

Acknowledgments

I would like to express my deepest appreciation and gratitude to my research advisor, Dr. Jeremy Dawson, for giving me the opportunity and guidance through this journey. His knowledge, dedication, encouragement, and trust have inspired me in many ways and has allowed me to learn and discover exciting and novel machine learning methods. I would also like to express my sincere thanks to Dr. Nasser Nasrabadi for sharing his expertise in the field of neural networks, and for the beneficial guidance and discussions on the topic. I also give my appreciation to Dr. Matthew Valenti for his continued support throughout my study. I am indebted to all the three for their constant encouragement and wise suggestions which helped carve not only my thesis but my approach towards quality research. I have learned a lot from my committee members, sometimes even without their knowledge.

I extend my special gratitude to my parents, for their love, patience, and continued support in making this research effort a reality. Without their support, I would have not come this far.

Lastly, I would like to express my gratitude to all my friends who have given me guidance, encouragement, and inspiration during my work.

Contents

List of Figures

List of Tables

Chapter 1:

Introduction

1.1.   Executive summary

Gender classification plays a very important role in many applications such as security, surveillance, context-based indexing, human-computer interaction, demographic research, and biometrics. The problem of automated gender assessment is typically treated as a two-class classification problem in which features extracted from a set of images corresponding to male and female subjects are used to train two-class classifiers. In earlier research, it was observed that there exist measurable differences between male and female biometric characteristics which can be exploited to improve the performance of recognition applications in surveillance and computer vision [1][2]. Besides, automatic human gender classification using palmprint will be among the next most popular biometric technology, especially in forensic applications, because of its uniqueness and strength[3]. It has been indicated by law enforcement agencies that at least 30 percent of the prints lifted from crime scenes are of palms and not fingers [4], achieving higher accuracy on palmprint gender classification which leads to better palmprint identification is important.

Although many achievements have been obtained in the face [5][6][7], iris [8][9], and fingerprint-based [10][11][12] gender classification, there are a very few efforts focused on addressing gender classification by palmprint images. Palmprint images have a large surface area containing rich information and can be easily collected. Furthermore, its texture features are much more stable [13]. Moreover, in [14] it was reported that biological characteristics of the hands are the least infringed when they are collected.

Few references have reported the identification of human attributes such as gender with the help of palmprint images. Wu and Yuan [15] developed a gender classification system that relied on the length, width, and aspect ratio of the palm. Polynomial Smooth Support Vector Machine (PSSVM) was employed for gender classification based on geometry features of the palm. Geometric properties like boundary and Fourier descriptors on Zernike moments were reported by Amayeh et al. [16] Advantages of the fingers and palm region were considered separately by dividing them into six different parts, the distance of a given part from two different eigenspaces was computed, then score level fusion and LDA was used to identify human gender. Kanchan et al. [17] reported that a measurable difference in palmprint ridge

density between women and men exists. It has been revealed that the palmprint ridge density is significantly higher among women than men. Another study was reported by Amayesh et al. [18] based on the north Indian community to infer gender in forensic investigation using palmprints. But recently, Convolutional Neural Networks have achieved great performance in image classification [19][20] after the pioneering work by Krizhevsky et al.[21] . some studies [22][23] used CNN for face and fingerprint gender classification and achieved promising results. Thus, this was the motivation to use CNN and design and develop a deep learning methodology for palmprint gender classification.

Although a few studies have been performed on palmprint gender classification, more research is required to improve the precision (accuracy rate) and speed of the process. Some of the unsolved challenges associated with palmprint gender classification are listed below:

1. Dataset availability – Most deep learning architectures require thousands of sample images for training and testing the performance of the network. To date, there are only a few publicly available palmprint image datasets with accompanying gender information that can be used to train a deep convolutional neural network model.

2. Time-consuming process – Computation cost for prior methods in gender classification based on palmprint is high, and they cannot be used in real-time applications.

3. Not generalizable – All of the conventional methods in deploying palmprint gender classification have used the handcrafted and limited dataset to test their system and analyze their result, which is not quite generalizable to other datasets.

4. Image composition – In the available datasets, some of the images are from the whole hand and not only the palmprint region of interest, requiring automated segmentation of the regions of interest. However, these larger images also contain features such as, creases, wrinkles, and principal lines that could be leveraged for identification purposes.

5. Data stratification - Most of the publicly available palmprint data is unbalanced between the female and male subjects.

From the literature, it is understood that previous studies were accomplished on limited datasets with conventional hand-crafted features and associated machine learning techniques. The first aim of this research is to improve the available datasets in terms of overcoming the

problem of image composition and data stratification by performing ROI extraction and data augmentation. Referring to the previous studies, one main drawback was using low-quality images which resulted in low accuracy. However, in this research, low quality contactless palmprint images have been used to not only speed up the time of the process and achieve a real-time system but also enhance the accuracy rate. Hence this work uses large publicly available datasets rather than creating a small, limited dataset of palmprints, it can be generalized to every other palmprint dataset available or created in the future.

The goal of this research is to perform gender classification on the publicly available palmprint image datasets and achieve high accuracy by using deep learning and CNNs, which has not been addressed in previous studies. The architecture of this robust method benefits from transfer learning, so it works with any size of datasets and gives the highest accuracy rate to the best of found knowledge. A Grad-CAM (Gradient-weighted Class Activation Mapping) is also implemented to ascertain which specific regions of the palmprint are most discriminative for gender classification.

Gender classification can assist in individual identification in many ways as mentioned above. This work will help individual identification based on palmprint, by reducing the search space of biometric databases, as a preprocessing step, which will greatly improve the performance of identification tasks. In addition, this work is an effective and real-time method for palmprint gender classification due to the proposed architecture and its features. This method can also be used for palmprint latent matching to specify the gender of the prints found in crime scenes prior to the identification process. Lastly, gender classification on palmprint provides spoof detection ability in such cases that, it can verify if a claimed gender for a particular palmprint matches its ground truth or not.

1.2. Context of the research and motivation

Automatic gender classification has received increasing attention since gender carries rich and distinguished information concerning male and female social activities [24]. Gender classification is used to recognize the gender of a person based on the characteristics that differentiate between masculinity and femininity. In the area of artificial intelligence, gender classification is one of the most important applications of pattern recognition methods [25]. The progress of gender classification research has driven many potential applications. For instance, a computer system with gender recognition functions has a wide range of applications in fundamental and applied research areas that includes: human-computer interaction (HCI) [26], the security and surveillance industry [27], demographic research [28], commercial development [29], and mobile application and video games [30]. Furthermore, multi-mechanisms are proposed to enhance the performance of gender recognition in terms of both accuracy and efficiency [31].

Although many achievements have been obtained in the face, iris, and fingerprint-based gender classification, there are very few researches addressing gender classification by palmprint images. Since all of the conventional methods in deploying palmprint gender classification have used the handcrafted and limited dataset to test their system and analyze their result, it is important to develop a method that can employ much larger and publicly available datasets so that the method would be generalizable to any other datasets. and recently, Convolutional Neural Networks have achieved great performance in gender classification on other biometric traits such as face [6][7], iris, and fingerprint [22][23], thus these were the motivations to use CNN and design and develop a robust deep learning methodology for palmprint gender classification to improve the precision and the speed of the process so it can be used in real-time application.

1.3.   Literature review

Identification of gender using palmprint does not have much literature, however, a few studies are reported on gender identification using hand geometry, hand shape, and palmprint ridge density, which are discussed in the next section.

1.3.1.   Background and related works of palmprint gender classification

As mentioned, few references have reported the identification of human attributes such as gender with the help of palmprint images. Wu and Yuan [15] developed a gender classification system that relied on the length, width, and aspect ratio of the palm. This gender classification method was comprised of two main attributes, features were extracted by image processing, and Polynomial Smooth Support Vector Machine (PSSVM) was employed for gender classification based on geometry features of the palm. Their database contained a total of 180 palm images that were collected from 30 persons to verify the validity of their proposed gender classification approach, and the classification rate of 85% was achieved.

Figure 1. 1. Wu and Yuan Database in thumbnails [15]

Amayeh et al. [16] system segmented the hand silhouette into six different parts corresponding to the palm and fingers. The geometry of each part was represented by region and boundary features based on Zernike moments and Fourier descriptors. For classification, the distance of a given part from two different eigenspaces was computed, one corresponding to

6

male class and the other corresponding to the female class. Using each part of the hand separately as well as fusing information from different parts of the hand has experimented. They used a small database of 20 males and 20 females, and the classification result of 98% was reported using score-level fusion and LDA.
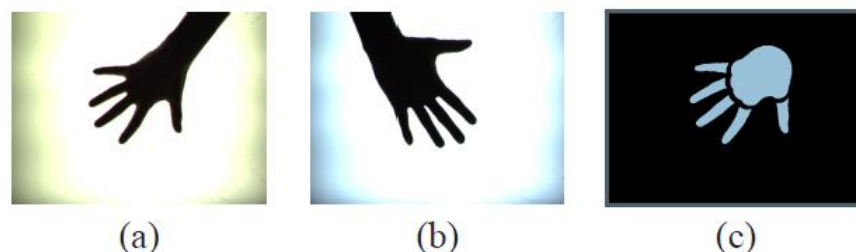
Figure 1. 2. (a) Female hand image, (b) male hand image, (c) segmented female hand image [16].

Kanchan et al. [17] studied the variation in ridge density in different areas of palmprints among men and women. The four prominent areas were analyzed on the palm prints that included the central prominent part of the thenar eminence (P1), hypothenar region, inner to the proximal axial triradius (P2), medial mount, proximal to the triradius of the second digit (P3) and lateral mount, proximal to the triradius of the fifth digit (P4). The mean palmprint ridge density among women was reported significantly higher than men in all the designated areas in both hands except for the P3 area in the right hand. In their research, significant differences were observed in the palmprint ridge density between the different palm areas in men and women in right and left hands. A total of 131 individual, 73 men, and 58 women, participated in their study
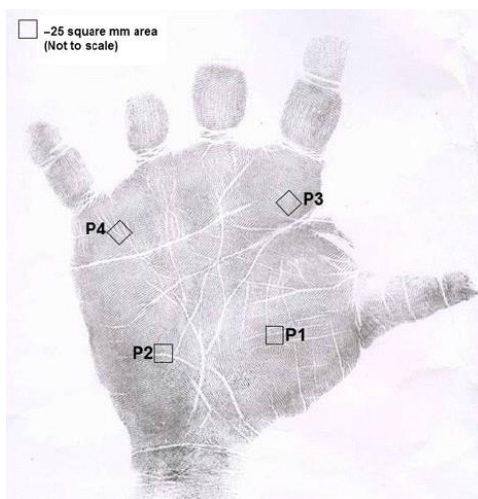
Figure 1. 3. Four designated areas on the palmprints that were analyzed for the palmprint ridge density [17].

1.4. Problem statement and objectives

Gender classification is still an important problem that exists in biometric recognition. Although a few studies have been performed on palmprint gender classification, still some challenging tasks need more enhancement and continuous research. As mentioned before, the main problems in palmprint gender classification are **dataset availability**, to date, there are only a few publicly available palmprint image datasets with accompanying gender information that can be used to train a deep convolutional neural network model. This challenge requires the transfer learning method to take advantage of the pre-trained convolutional neural network models on larger datasets and fine-tune it on available palmprint datasets to have a more accurate result despite insufficient data. The next challenge in palmprint gender classification is the **time-consuming process**, and cannot be used in a real-time application, deep learning methodology would accelerate the process and makes it more agile. The next challenge this thesis tries to overcome is the **generalizability** of the method to any other dataset. As mentioned before, it is understood from the literature that all of the conventional methods in palmprint gender classification have used the handcrafted and fairly small datasets to test their system, which is not quite generalizable to other datasets. With the use of larger and publicly available datasets to train and evaluate this thesis' method generalizability of it on many applications can be ensured. **Image composition** and **data stratification** of the image data are the next challenges, and this research aims to improve the available datasets in terms of overcoming these problems by performing ROI extraction and data augmentation, which are explained in detail in their respective sections.

The goal of this research is to perform gender classification on the publicly available palmprint image datasets and achieve high accuracy by using deep learning methods and CNNs, which has not been addressed in previous studies. The architecture of this robust method benefits from transfer learning, so it works with any size of datasets and gives the highest accuracy rate to the best of found knowledge. A Grad-CAM (Gradient-weighted Class Activation Mapping) is also implemented to ascertain which specific regions of the palmprint are most discriminative for gender classification.

1.5.   Contributions of the thesis

In this research, an experimental study is conducted on palmprint gender classification which is established that using convolutional neural networks with transfer learning method, is superior to other tested models. In order to achieve the objectives of this work, some contributions are suggested for gender classification approaches based on palmprint. The contributions of this thesis are summarized as follows:

First, data augmentation and ROI (region of interest) extraction have been performed on 4 different publicly available palmprint image datasets (KTU, IITD, CASIA, PolyU) for the purpose of training deep neural networks.

Second, as a solution, a new and time-efficient network structure is proposed, the proposed work combines and customizes two deep CNN model (VGGNet and DenseNet) pre-trained on ImageNet [32]. This final model is fine-tuned to learn complex patterns in ridges and minutiae of the palmprint images. To the best of found knowledge, this model has the highest accuracy rate on the CASIA and PolyU palmprint image datasets.

Third, performed Gradient-weighted Class Activation Mapping (Grad-CAM), this method uses the class-specific gradient information flowing into the final convolutional layer of a CNN to produce a coarse localization map of the important regions in the palmprint image and specified the specific area of the palmprint that is more important to distinguish between female and male class.

1.6.   Organization of the thesis

The thesis is structured as follows:

- Chapter 2 reports the general context of biometry, neural network, and related works.
- Chapter 3 describes the overall experimental setup, datasets, data augmentation, and ROI extraction. It presents the design and development of the networks used in this research from the initial baseline network to the final version of the proposed network design.
- Chapter 4 presents the experimental evaluations and results of the methods described in chapter 3.
- Finally, chapter 5 concludes the thesis and discusses its most important results, contributions, and future works.

Chapter 2:

Background

2.1.   Introduction

In this section, an introduction of the biometrics system, different biometric modalities, advantages, and disadvantages of different modalities are explained. Features of palmprint trait, its advantages, and challenges are introduced. Background information of the convolutional neural network, its architectural components, and the deployed design components are discussed before moving on to the details explained in the following chapter.

2.2.   Biometrics system

Automatic personal authentication systems that identify or verify the identity of an individual using biometric information such as the face, iris, fingerprint, and palmprint play an important role in the application of security, access control, forensic, and banking. Unlike conventional methods for personal identification, such as personal identification number, password, and key, these features cannot be duplicated, lost, or stolen. biometrics are typically time-invariant, easy to acquire, and unique in every individual. A biometric system is essentially a pattern recognition system that operates by acquiring biometric data from an individual, extracting a feature set from the acquired data, and comparing this feature set against the template set in the database. The benefit of these types of systems is that they make a great solution for identification or verification and they are also simple and easy to use.

Depending on the application context, a biometric system may operate either in verification mode or identification mode. In verification mode, the system validates a person's identity by comparing the captured biometric data with her own biometric templates stored in the system database. In such a system, an individual who desires to be recognized claims an identity, usually via a personal identification number (PIN), a username, or a smart card, and the system conducts a one-to-one comparison to determine whether the claim is true or not. Identity verification is typically used for positive recognition, where the aim is to prevent multiple people from using the same identity [33]. In the identification mode, the system recognizes an individual by searching the templates of all the users in the database for a match. Therefore, the system conducts a one-to-many comparison to establish an individual's identity, without the subject

having to claim an identity. Identification is a critical component in negative recognition applications where the system establishes whether the person is who she denies being. The purpose of negative recognition is to prevent a single person from using multiple identities[33].

2.2.1. Biometric modalities

There are different kind of biometric modalities which can be classified into two categories: Physical biometric and behavioral biometric. Physical biometrics are like the image of a part of the human body such as the face, iris, fingerprint, or palmprint. As for behavioral biometric, it uses a person's behavioral traits such as signature, the way each person types, or walks. There are also morpho-behavioral biometrics namely voice, which relies at the same time on the morphology of vocal cords and the behavior of it, since voice may change easily depending on the emotional state of the person. Figure 2.1. demonstrates some examples of different biometric modalities.
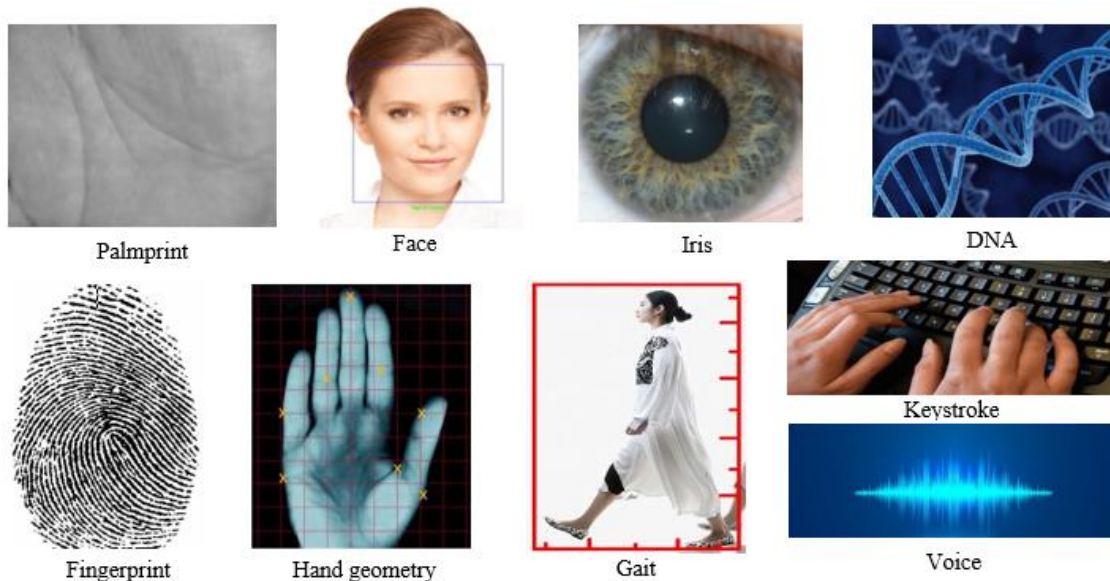
Figure 2. 1: Different biometric modalities.

2.2.1.1. Palmprint features

Several biometric characteristics exist and are in use in various applications. Each biometric has its strengths and weaknesses, and the choice depends on the application. Voice is very less accurate, keystroke needs a long observation time, the face is affected due to pose, illumination, and aging factors, iris sensor is very expensive, hand geometry varies as children grow, wearing of rings and rapid growth of pregnant ladies in a short time, DNA is not user-friendly, the fingerprint is unclear for elderly persons and manual labors and even missing, gait is influenced by medical conditions, clothing, surface, and footwear, Signature is easy to forge, palm print images are large in size and thus the sensor is bulkier. No single biometric is expected to effectively meet the requirements of all the applications. In other words, no biometric is "optimal." The match between a specific biometric and an application is determined depending upon the operational mode of the application and the properties of the biometric characteristic.

As the search for a more accurate biometrics system continues, hand biometrics and in particular palmprints, are becoming the biometric of choice, especially for deployment for access control at the point of entrance such as airports, federal buildings, and other highly sensitive places. Palmprint is one of the important biometrics traits and has several advantages over other characteristics, such as low-resolution imaging, low-intrusiveness, stable line features, and low-cost capturing device. Indeed, since the principal lines and wrinkles can be observed under low-resolution images (e.g., 100 dpi or lower), palmprint systems do not require high resolution capturing devices. The area of the palm is much larger than the area of a finger and as a result, palmprint contains more information than fingerprint beside ridges, such as principal lines and wrinkles so it is more distinctive [34]. Patterns in palmprint such as Principal lines, wrinkles,
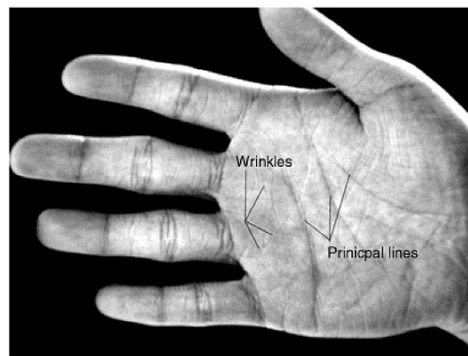
Figure 2. 2: Figure Palmprint feature definitions with principal lines and wrinkles [3].

ridges, minutiae points, singular points, and texture are a very reliable biometric trait and to extract these features we need minimum cooperation from the user.

Various features can be extracted at different image resolutions. For features such as minutiae points, ridges, and singular points, a high-resolution image, with at least 400 dpi (dots per inch), is required for feature extraction. However, features like principal lines and wrinkles, which are defined in figure 2.2., can be obtained from a low-resolution palmprint image with less than 100 dpi. In general, high-resolution images are essential for some applications such as law enforcement, where ridges, singular points, and minutiae points are extracted and matched in latent prints for identification and verification. For civil and commercial applications, low-resolution palmprint images are more suitable than high-resolution images because of their smaller file sizes, which results in shorter computation times during preprocessing and feature extraction. Therefore, they are useful for many real-time palmprint applications [35]. The details of these features hardly change throughout the life of a person. And, these distinctive features such as principal lines and wrinkles can be captured even with a lower resolution scanner, therefore in comparison with iris, the cost of capture devices for palmprint is much less. When using a high-resolution palmprint scanner, all the features of the palm such as hand geometry, ridge and valley features (e.g., minutiae and singular points such as deltas), principal lines, and wrinkles may be combined to build a highly accurate biometric system. Compared with ridges and wrinkles, principal lines are usually the consequence of genetic effects. therefore, they are the most significant features in palmprint images and have good permanence.

2.2.2. Properties of biometric modalities

Biometric authentication presents several advantages compared to traditional identification ways like passwords or cards because it establishes a strong link between an individual and their identity. The principal properties of a biometric modality are the following:

- **Universality**: The whole population should possess this modality (physical or behavioral characteristic).

- **Distinctiveness**: Two different individuals must have different biometrics representations.

- **Stability**: To ensure individual authentication success, biometric modality should be relatively stable over time and it also must be stable regardless conditions of acquisition (external conditions, emotional conditions of the person, etc.)

- **Collectability:** The biometric modality must be acquired.

- **Acceptance**: The acceptance and the facility of usage are related to the acquisition constraints of a biometric modality.

- **Circumvention**: The biometric modality must not be easily falsified.

- **Performance**: Biometric recognition should be accurate, fast, and robust with regards to operational and environmental changes.

All modalities do not possess all these properties or may possess them with different degrees. Hence, there is no ideal or perfect modality. The trade-off between the presence and absence of some of these properties is required according to each system's needs, regarding the choice of biometric modality.

2.2.3. Advantage and disadvantage of each modality

A comparison between different biometric modalities according to the seven properties mentioned previously (universality, distinctiveness, stability, collectability, acceptance, Circumvention, performance) is presented in table 2.1. The quality degree accorded to each modality is classified into three classes: high, medium, and low referred to the following notations • • •, • • and •, respectively. This table originating from [33] indicates that behavioral biometric modalities (keystroke, odor, or signature) present low recognition performances unlike intrusive data such as methods based on DNA or retina. However, data are given from palmprint or hand geometry presents the advantage that they do not possess any low criterion compared to

other modalities. Moreover, they are acquired from a single biometric modality which is hand trait. These advantages justify the choice of palmprint modality which is treated in this thesis.

Table 2. 1: comparison of biometric modalities according to seven properties: Universality (U), Distinctiveness (D), Stability (S), Collectability (Co), Acceptance (A), Circumvention (Ci), Performance (P). ●●●: high degree, ●●: medium degree, ●: low degree [1].

Biometric modality	U	D	S	Co	A	Ci	P
Face	●●●	●	●●	●●●	●●●	●	●
Iris	●●●	●●●	●●●	●●	●	●	●●●
Fingerprint	●●	●●●	●●●	●●	●●	●●	●●●
Hand geometry	●●	●●	●●	●●●	●●	●●	●●
Palmprint	●●	●●●	●●●	●●	●●	●●	●●●
Keystroke	●	●	●	●●	●●	●●	●
Odor	●●●	●●●	●●●	●	●●	●	●
Retina	●●●	●●●	●●	●	●	●	●●●
Signature	●	●	●	●●●	●●●	●●●	●
Voice	●●	●	●	●●	●●●	●●●	●
Hand vein	●●	●●	●●	●●	●●	●	●●
DNA	●●●	●●●	●●●	●	●	●	●●●

2.2.4. Structure of a biometric system

The structure of a biometric system consists of two different phases: enrollment and authentication, as shown in figure 2.3. Enrollment is common for both verification and identification modes. It is the preliminary phase where the biometric data of a user is registered for the first time in the system. During this phase, one or more biometric modalities are captured and stored as templates in the database. This phase is very crucial since it influences, later, the whole recognition process. In fact, the quality of enrolled data is essential for ulterior
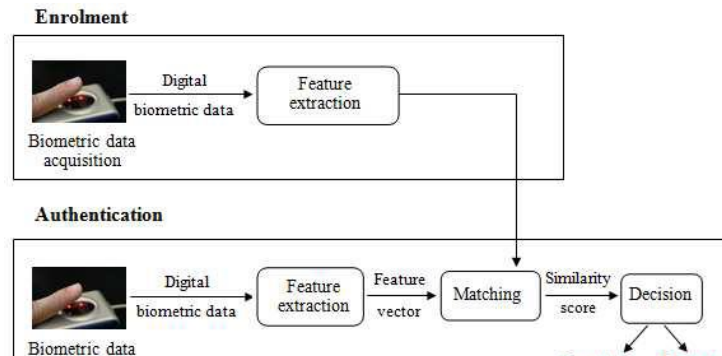
Figure 2. 3: Structure of biometric systems.

identification phases because acquired data are considered as references for the person. A set of samples should be captured to consider the variability of the biometric modality of a person.

A biometric system is composed of 4 modules, some of them are common for enrollment and authentication phases namely acquisition, features extraction, matching, and making a decision. Acquisition and feature extraction are two modules presented in the enrollment and authentication phases. Features extraction is a data representation (e.g. image or signal) as a vector that should be representative for data and discriminant versus other data of other individuals. During the enrollment phase, the feature vector extracted from the biometric sample is called reference and stored in a database. During the authentication phase, acquisition and features extraction modules allow achieving a representation of biometric data to be tested later in features space. The matching module is used during the authentication phase to compare the feature vector extracted for a test with the reference feature vector. The decision module consists of deciding the output of a matching module which generates a similarity score between two feature vectors. For verification applications, the matching is executed only once, between reference data and test data, and a decision of" true" or" false" is taken. For identification applications, the matching is carried out between all references stored in the database and the decision is the answer to the following question:" Does this person exist in the database, and if so, who is he/she?"

2.2.5. Evaluation of a biometric system

To evaluate the performance of a biometric system, three principal criteria have to be already defined clearly:

1. **F**alse **R**ejection **R**ate or **FRR**: This rate represents the percentage of individuals expected to be recognized but they are nevertheless rejected by the system.

2. **F**alse **A**cceptance **R**ate or **FAR**: This rate represents the percentage of individuals expected to be not recognized but they are nevertheless accepted by the system.

3. **E**qual **E**rror **R**ate or **EER**: This rate represents the optimal performance measure and is computed depending on the first two criteria. It is achieved when FAR=FRR, i.e. the best trade-off between false rejections and false acceptances.

Figure 2.4 shows the FAR and FRR diagram according to distributions of genuine and imposter scores. The EER is represented in Figure 2.5.
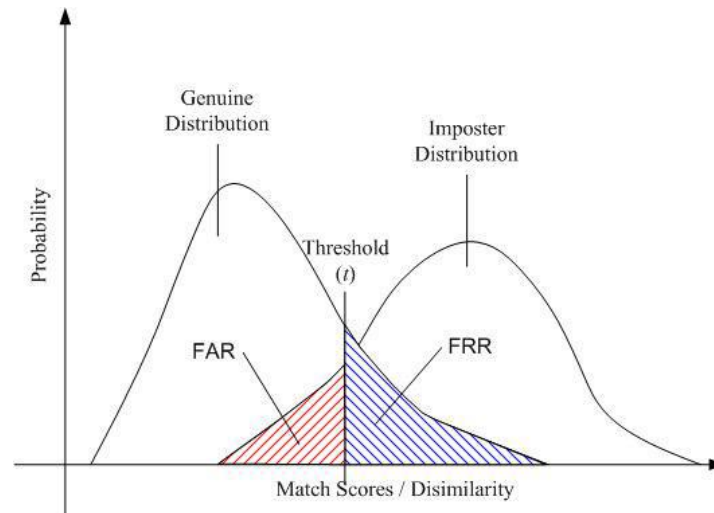
Figure 2. 4: FAR and FRR diagram
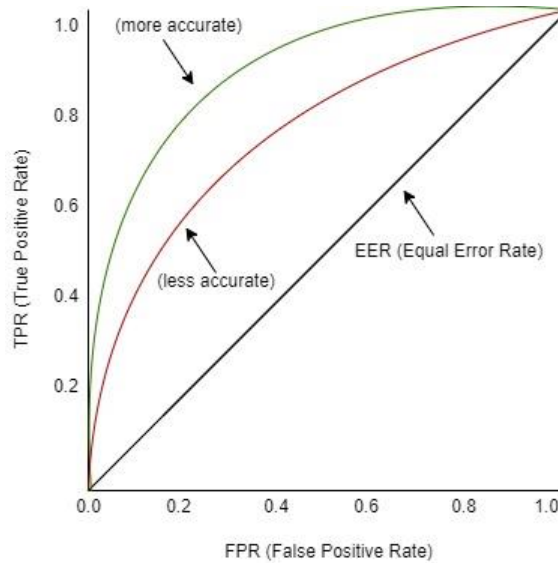
Figure 2. 5: ROC curve

There are two ways to measure the biometric system performance, according to the mode (authentication or identification):

- If the system works in authentication mode, then the ROC (Receiver Operating Characteristic) curve is used. This curve draws the false rejection rate depending on the false acceptance rate. The more this curve fits the mark shape the more the system is efficient with a high Recognition Rate (RR). The results of this research are measured based on this evaluation method (The area under the ROC curve).

- In the case of the identification mode, the CMC (Cumulative Match Characteristic) curve is used. The CMC curve provides the percentage of recognized individuals according to a variable called a rank. A system is said to recognize at the rank 1 when the nearest image is selected as the recognition result, and a system is said to recognize at the rank 2 when it selects, among two images, the one that best matches the input image. Subsequently, the more the rank is high the more the correspondent recognition rate is related to a low-security level.

2.3.   Soft biometrics

Recently for improving the performance of traditional biometric systems, the use of soft biometrics has grown. Soft biometrics are characteristics that can be used to describe an individual, but they are not fully distinctive by themselves in recognition tasks. Soft biometric traits include characteristics such as gender, height, weight, hair type, scars, marks, etc. The beginnings of soft biometrics science were laid by Alphonse Bertillon in the nineteenth century, who first introduced the idea for a personal identification system based on biometric, morphological, and anthropometric determinations. He used traits like colors of eye, hair, beard, and skin; shape and size of the head; general discriminators like height or weight and also a description of indelible marks such as birthmarks, scars, or tattoos. A great majority of those descriptors fall at present into the category of soft biometrics. Soft biometrics is defined to be a set of characteristics that provide some information about the individual but are not able to individually authenticate the person, mainly due to lack of distinctiveness and permanence.

soft biometrics are not expensive to compute, can be sensed at a distance, do not require the cooperation of the surveillance subjects, and have the aim to narrow down the search from a group of candidate individuals. Moreover, we here note that the human compliance of soft biometrics is the main factor, which differentiates soft biometrics from classical biometrics offering new application fields. Researchers have a new definition of soft biometrics which is as follows, Soft biometric traits are physical, behavioral, or adhered human characteristics, classifiable in pre-defined human compliant categories. These categories are, unlike in the classical biometric case, established and time-proven by humans to differentiate individuals. In other words, the soft biometric traits instances are created in a natural way, used by humans to distinguish their peers.

Recently, soft biometric traits have been employed to preliminary narrow down the search of a database, in order to decrease the computational time for the classical biometric trait. Another application approach is to fuse soft biometrics and classical biometric traits to increase system reliability. Soft biometrics impart systems substantial advantages: they can be partly derived from the main detected classical biometric identifier, their acquisition is nonobtrusive and does not require enrollment, training can be performed in advance on individuals out of the specific identification group.

2.3.1. Gender features of palmprint

In the recent past, gender dimorphism of the fingerprint ridge density has been explored by researchers in different populations worldwide. In [36] Jantz observed differences in gender and ethnicity for finger ridge count correlations, and several other kinds of research pointed out that females have finer epidermal ridge details than males [37][38]. Palmprint and fingerprint ridge density is determined by two characters: ridge width and distance between ridges. The thickness of the epidermal ridges varies between individuals, females having finer ridges, and therefore, greater ridge density than males [39]. The differences in the ridge density between males and females may also be the result of overall size dimorphism. The larger the size of the palm, the greater is the distance between the ridges and lesser will be the ridge density [40][41].

Palmprints share most of the features with fingerprints and also possess much larger skin area and as explained before, other discriminative features such as principal lines and creases. Palmprint recognition has considerable potential as a personal identification technique.

According to a study by Jain and Feng [42], about 30% of the latent prints recovered from crime scenes are from palms. Hence, there is a growing need for palmprint databases worldwide. When a person touches a surface or grabs an object, an impression of the friction ridges may be left behind. Detailed analysis of fingerprints and palmprints at the crime scene, becomes useful to identify the suspect in a particular case. Often, the prints collected from crime scenes, weapons of an offense such as knife hilts, gun grips, steering wheels, and windowpanes belong to palms and fingers [43]. The sex differences in palm print ridge density can even be valuable in the identification of a dismembered hand during medicolegal investigations to establish the identity of an individual in cases of mass disasters/mass homicides [17]. Though analysis of ridge density has been used for the distinction of sex from fingerprints [41][44][45][46], studies on its application in palmprint ridge density are limited. If the gender of the suspect could be inferred from the prints available at the crime scene application of interest, then the burden of the investigation is reduced to half and the investigations may be directed towards either a male or a female suspect accordingly.

Several studies show that females have finer ridge patterns over the palmar surface than males and show significantly greater palmprint ridge density than males. The distal region of the palm is shown to have finer ridges than the proximal region among males and females. Therefore, apart from the matters relating to sex inference at the crime scene, palmprint ridge density can be considered as a trait for individual and population variation in forensic anthropology [47].

Gender classification of palmprint images is an interesting yet challenging problem, it is a rather novel research topic in computer vision. Automatic gender classification could be of important value in human-computer interaction, such as personal identification. Also, it is a useful preprocessing step for palm recognition. A computer system with the capability of gender classification has a wide range of applications in basic and applied research areas, including man-machine communication, security, law enforcement, demographics studies, psychiatry, education, and telecommunication. Gender classification on palmprint reduces the searching

space and speeds up the identification process. It also gives us spoof detection ability in such case that a fake palmprint data may contain different gender information than the actual gender of the subject.

2.4.   Convolutional Neural Networks

Convolutional Neural Network is a type of artificial neural network whose layers resemble the simple and complex cells in the primary visual cortex [48]. The connectivity is inspired by the biological brain neurons, where each neuron receives input signals (x_0) from its dendrites and produces output signals along its axon, which in turn connects to different other neurons. The signals are multiplied with the dendrites ($x_0.w_0$) of other neurons while passing along the axons. The signals are only sent (fired) through the axon when the summation is above a certain threshold, which can be modeled as activation function f [49]. This phenomenon can be shown as a mathematical model as in Figure 2.6.
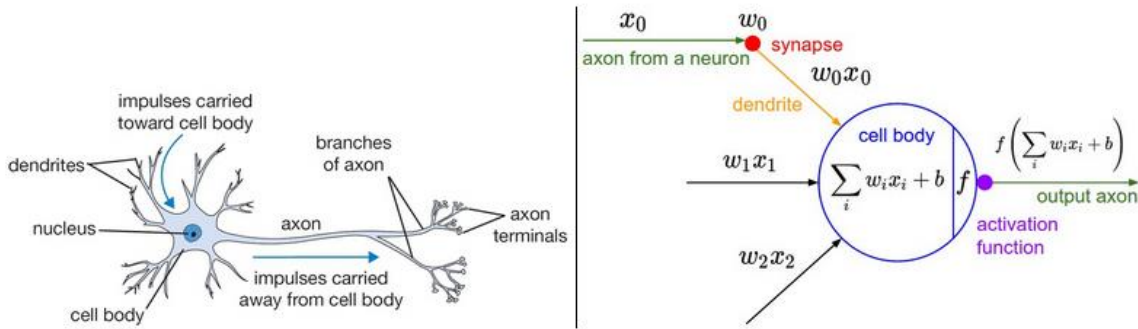
Figure 2. 6: Biological neuron and its mathematical model [18].

The CNNs have local connectivity property, where each neuron is only connected to a small region (receptive field) of adjacent layers. The neurons in the layers of a CNN are three dimensional (height, width, and depth) and are connected only to the receptive field of the previous layer. CNNs have weight sharing property and form a feature map with replicated units sharing the same weight and bias, which makes all the neurons in the layer detect the same features. This feature decreases the parameters and so helps in reducing the memory cost.

The architectural overview of a convolutional neural network is composed of distinct layers, which transforms the input to output through certain functions. The layers and activation functions are described below.

2.4.1. Architectural components

i. Convolutional layer:

Convolutional Layer is the core building block of the network. It consists of a set of learnable filters with small receptive fields but extends through the full depth of the input volume. When the input passes through the convolution layer, each filter is slid across the width and height of the input volume computing the dot product between the entries of the filter and the input at that position. This produces a two-dimensional activation map and these activation maps are stacked along the depth dimension for all filters, producing the output of the convolutional layer.
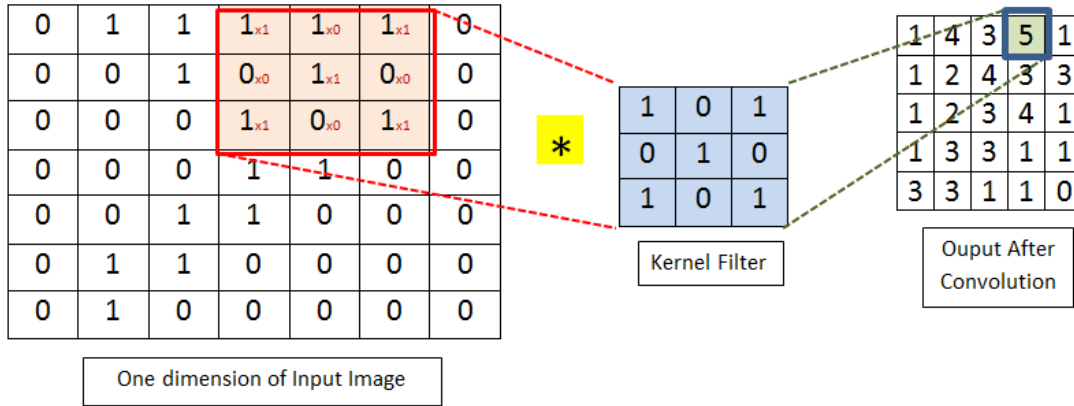
Figure 2. 7: Illustration of convolutional layer output calculation.

ii. Max pulling layer:

Pooling is a form of the non-linear down-sampling method used to achieve spatial invariance by progressively reducing the spatial size of feature maps and the number of parameters. The pooling layer is also called a subsampling layer. There are many variants of pooling, such as average pooling, L2-norm pooling, etc., in which Max pooling is a popular one, outperforming the other subsampling operations [50].

In the construction of neural networks in this thesis, the max pool layer is used as a sub-sampling layer. Generally, the pooling layer comes after the convolutional layer but is not necessary. The most common pooling layer window size is 2×2, with a stride of 2 and is applied independently along with the depth of the input, by which the width and height are reduced by 2 while the depth remains the same. The max-pooling function applies the window to the input patch and computes the maximum in the neighborhood as shown in Figure 2.8.

0.54	0.7	0.11	0.66
0.1	0.8	0.23	0.45
2	0.11	0.44	7
0.45	0.89	5	0.12

0.8	0.66
2	7

Applying Max pooling of 2x2 window with stride 2 on One dimension of the input patch
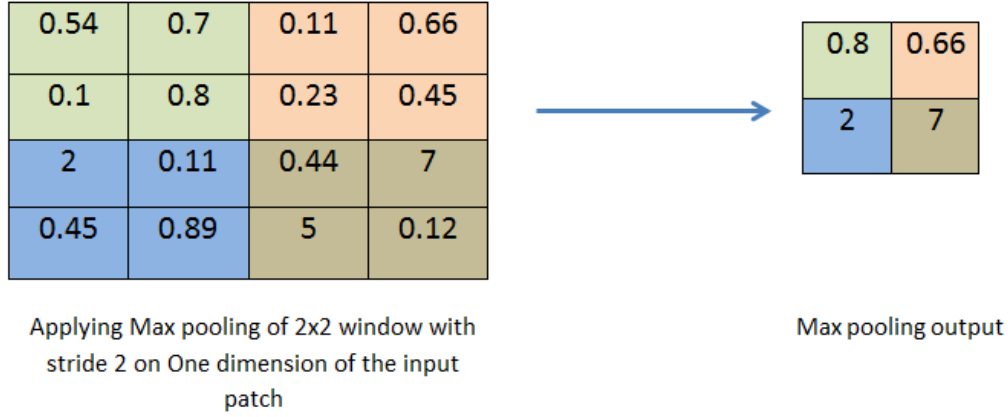
Max pooling output

Figure 2. 8: Illustration of Max pooling layer calculation.

iii.   Rectified Linear Unit layer:

This layer applies a non-saturating nonlinearity activation function to model the neuron's output [51]. There are different kinds of activation functions such as, tanh, sigmoid, softplus, etc. Rectified Linear Unit (ReLU) is used in this thesis, as ReLUs are said to train several times faster than their equivalents with tanh units and also reduce over-fitting [21].

The ReLU activation function is defined as:

$$f(x) = max\ (0, x) \tag{2. 1}$$

The plot of the Rectified Linear Unit function is shown in Figure 2.9. The function thresholds the activations at zero. This creates a disadvantage that if at all any neuron is updated to zero, then the gradient passing through it will always be zero since the gradient of zero is a zero. Generally, ReLU is added after each pair of convolution and pooling layer.
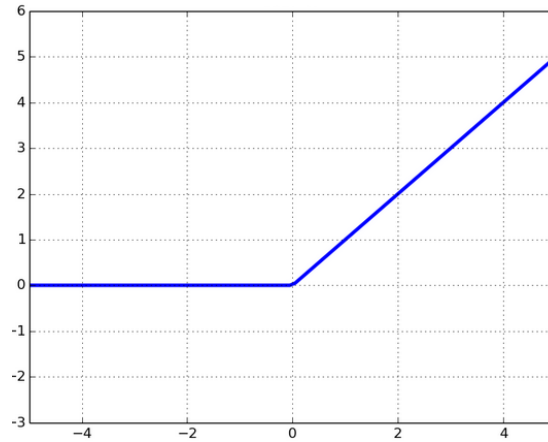
Figure 2. 9: Plot of ReLU function

iv.   **Fully Connected Layer:**

In the fully connected layer, neurons have connections to all activations in the previous layer. It is generally used as the last layer which holds the output scores and is also called the "output layer". Its activations are computed with a matrix multiplication and bias offset. The output of the fully connected layer is 1×1×C, where C is the number of class labels.

2.4.2. Employed design components

In this section, preprocessing steps and the design method chosen for training all of the CNNs model used in this research are explained. While the individual architectures are explained in their respective sections, the design steps taken for modeling the data and the network are explained as the following.

i.   **Data preprocessing:**

Data preprocessing is a method used to transform the raw input data into an evenly distributed, scaled data within a range, understandable to the network. Data preprocessing is proven to improve the training process of the neural networks. Data can be preprocessed by mean subtraction or normalization. In this thesis, all the training data is normalized by centering the data to have a zero mean and scaling to [-1 1] range.

ii.   Weight initialization:

The network weights are initialized before training in order to avoid the diminishing of the input signal variance as it passes through the layers. In CNN-a, which is not a very deep network, the weights are initialized randomly with Gaussian random numbers. This scaling of inputs ensures that all the neurons in the network initially have approximately the same output distribution and empirically improves the rate of convergence [49]. Very low weights shrink the signal, while with very large weights the signal grows at each layer and becomes useless.

iii.   Batch normalization:

Ioffee and Szegedy [52] address the internal covariate shift, a phenomenon named as the change in the distribution of network activations due to the change in parameters during training, which needs a careful parameter initialization and lower learning rates. They have introduced the Batch Normalization method to avoid saturating nonlinearities by applying normalization of the layer inputs for each training mini-batch.

iv.   Regularization Method: Drop out:

The neural network is prone to overfitting, a phenomenon where the network has a very low training loss but could not generalize to novel data. Dropout is a regularization method to address the overfitting issue to some extent and has shown to improve performance [53]. The dropout technique drops out the units in a neural network below a certain threshold p, where p is between [0, 1]. A dropout of 0.5 is used in the CNN-a network and 0.7 in all of the other networks.

(a) Standard Neural Net   (b) After applying dropout.
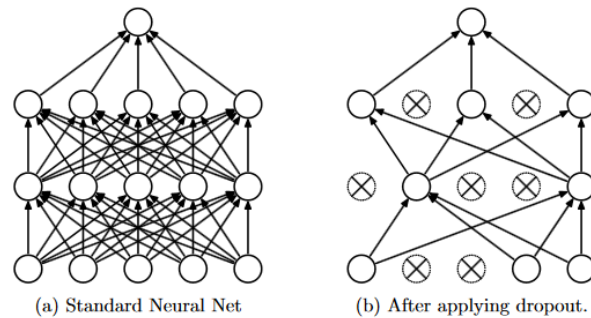
Figure 2. 10: Schematic diagram showing the thinning of neural network after dropout [24].

v.   Softmax cross-entropy Loss:

The softmax layer is used as the classifier layer in the neural network and the network is trained under a loss function. In the networks in this thesis, the cross-entropy loss is used as a cost function to train the network. The cost function is used to calculate the loss between the predicted and true labels and this layer will be right after the fully connected layer. The softmax function scales the scores to be in [0,1] range and all scores sum to one. To reduce the chance of over-fitting the ℓ_2 norm can be added to the loss function, resulting in an overall loss function as:

$$\mathcal{L}_{final} = \mathcal{L}_{class} + \lambda_1 \|W_{fc}\|_F^2 \tag{2.2}$$

Where $\mathcal{L}_{class} = -\sum_i p_i \log(q_i)$ is the cross-entropy loss, and $\|W_{fc}\|_F^2$ denotes the Frobenius norm of the weight matrix in the last layer. This loss function can be minimized using a stochastic gradient descent algorithm.

vi.   Optimization Method: Stochastic Gradient Descent:

An optimization method is used to minimize the loss function. Gradient descent is an iterative method of minimizing a differentiable loss function $J(\theta)$ by updating the parameters in the negative direction of the gradient of the loss function. Stochastic gradient descent is a variant of gradient descent, which performs parameter update in the negative direction for each training iteration to minimize the gradient. The gradient is computed, and the weights are updated over small batches of training data for every iteration. The general form of the stochastic gradient can be given as:

$$\theta = \theta - \eta . \nabla_\theta J(\theta; x^{(i)}; y^{(i)}) \tag{2.3}$$

where η is the learning rate or step size in this equation. The learning rate can be kept constant or changed for every or after certain iterations. There are other methods of gradient descent algorithms: Adagrad, Adam, RMSprop, and Adadelta, which after empirical evaluation of these methods, stochastic gradient descent (SGD) is employed in all the trained neural networks in this thesis.

Chapter 3:

Methodology

3.1.   Introduction

In this section, an overview of the detailed approaches of the experimental setup is described. The main contributions of the thesis are ROI extraction of the palmprint image datasets and developing a deep neural network model to achieve high accuracy in classifying the gender of palmprint images.

Four different publicly available palmprint image datasets were divided into the train, validation, and test sets to evaluate the performance of four deep neural network models that are presented in this thesis. for simplicity, each network is given the following abbreviations.

1. CNN-a: Initial baseline CNN based on VGGNet.
2. CNN-VGG: Pretrained VGGNet-E on ImageNet, fine-tuned, and modified based on the palmprint image datasets.
3. CNN-Dense: Pretrained DenseNet121 on ImageNet, modified to train on smaller available datasets.
4. CNN-b: The proposed deep learning neural network which is a combination of CNN-VGG and CNN-Dense.
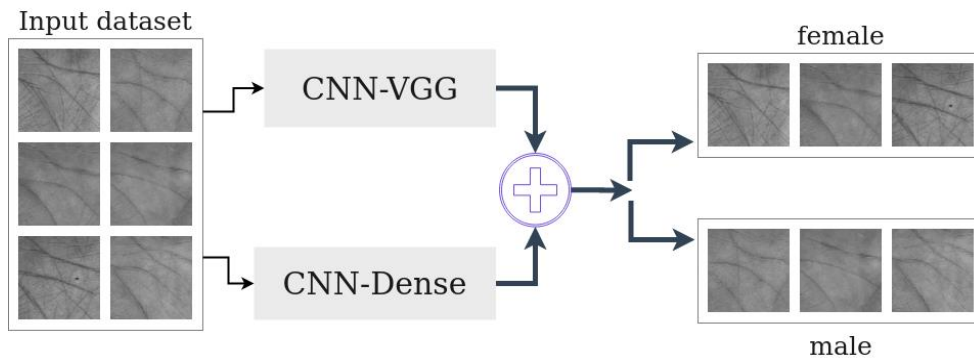
Figure 3. 1: Overview of the experimental setup.

3.2. Datasets

This section represents all of the publicly available palmprint datasets that are used in this research.

i. KTU CVPR Lab. Palmprint Database-2

This database contains 15440 color images corresponding to 534 different palms in BMP image format. Each image in this database was taken using a low-cost CCD camera with a resolution of 768×576 pixels at 75 dpi. The number of samples per individual in the database was restricted between 13 and 40. Since the images were taken in different sessions some variations were appeared in the images, like focus, illuminations, and pose of the hand [54].

Figure 3. 2: Sample of images in KTU database.

ii. CASIA Palmprint Image Database

CASIA Palmprint Image Database contains 5502 palmprint images captured from 312 individuals (74 females and 238 males). For each subject, palmprint images from both left and right hands were collected. All palmprint images are 8 bit gray-level JPEG files [55].

Figure 3. 3: Sample of images in CASIA database.

iii.   PolyU Multispectral Palmprint Database

Multispectral palmprint images were collected from 250 volunteers, including 206 males and 44 females. The PolyU Multispectral Palmprint database contains 15 kinds of palmprint images, which were taken from 15 different spectra using a piece of equipment developed by the PolyU Biometric Research Center. The spectra included Blue, Blue_Infrared, Green, Green_Blue, Green_Blue_Infrared, Green_Infrared, Infrared, Red, Red_Blue, Red_Blue_Infrared, Red_Green, Red_Green_Blue, Red_Green_Blue_Infrared, Red_Green_Infrared, and Red_Infrared. Each hand has 12 images under each spectrum. This database includes 250 people and 90,000 images [56].
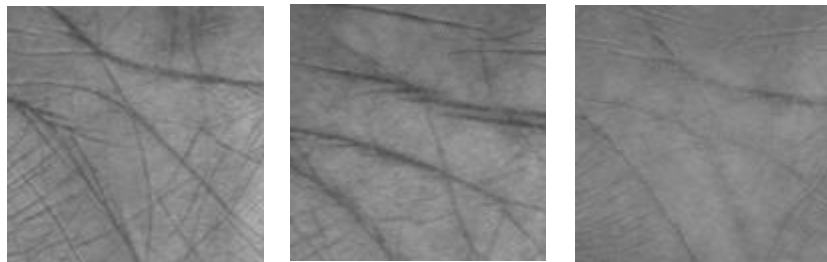
Figure 3. 4: Sample of images in PolyU database.

iv.   IIT Delhi Touchless Palmprint Database

The IIT Delhi palmprint image database consists of the hand images collected from 230 users, all the images are in bitmap (*.bmp) format. Seven images from each subject, from each of the left and right hand, are acquired in varying hand pose variations. The resolution of the images in this dataset is 800×600 pixels [57].

Figure 3. 5: Sample of images in IIT Delhi Palmprint database.

Each dataset was divided into three different subsets for training, validating, and testing purposes. Table 3.1. shows detailed information on the palmprint datasets used in this project.

Table 3. 1: Detail information of the palmprint datasets

Dataset	Number of images	Number of female's hand	Number of male's hand	Number of palmprints images for each hand	Number of spectra
PolyU Multispectral Palmprint [56]	90,000	88	412	24	15
CASIA Palmprint Image Database [55]	5,502	74	238	1~16	1
KTU Palmprint Database [54]	15,440	-	-	13~40	1
IIT Delhi Palmprint Database [57]	3,220	-	-	14	1

3.2.1. Palmprint ROI extraction

ROI extraction is a crucial step and it greatly affects the final results. Most of the existing methods extract ROI according to some key points between fingers or in palm boundary, or some external factors. And these methods assume that the palmprint images have been well aligned before performing feature extraction and matching. It is important to define a coordinate system that is used to align different palmprint images for matching. To extract the central part of a palmprint, for reliable feature measurements, the gaps between the fingers are used as reference points to determine a coordinate system [1]. The five major steps in processing the image are as follows:

1. Applied a lowpass filter, $L(u, v)$, such as Gaussian smoothing, to the original image, $O(x, y)$. A threshold, T_p, is used to convert the convolved image to a binary image, $B(x, y)$, as shown in Figure 3. 5(b).

2. Obtained the boundaries of the gaps, $(F_i x_j, F_i y_j)(i = 1,2)$, between the fingers, using a boundary tracking algorithm, Figure 3. 6(c). The boundary of the gap between the ring and middle fingers is not extracted since it is not useful for the following processing.

3. The tangent of the two gaps is computed. (x_1, y_1) and (x_2, y_2) are assumed to be any points on $(F_1 x_j, F_1 y_j)$ and $(F_2 x_j, F_2 y_j)$, respectively. If the line $(y = mx + c)$ passing through these two points satisfies the inequality, $F_i y_j \leq m F_i x_j + c$, for all i and j (in Figure 3. 6(d)), then the line $(y = mx + c)$ is considered to be the tangent of the two gaps.

4. Lined up (x_1, y_1) and (x_2, y_2) to get the Y-axis of the palmprint coordinate system, and used a line passing through the midpoint of these two points, which is perpendicular to the Y-axis, to determine the origin of the coordinate system as shown in Figure 3. 6(d).

5. Extracted a sub-image of a fixed size based on the coordinate system. The sub-image is located at a certain area of the palmprint image for feature extraction (Figure 3. 6(e) and 3. 6(f)).
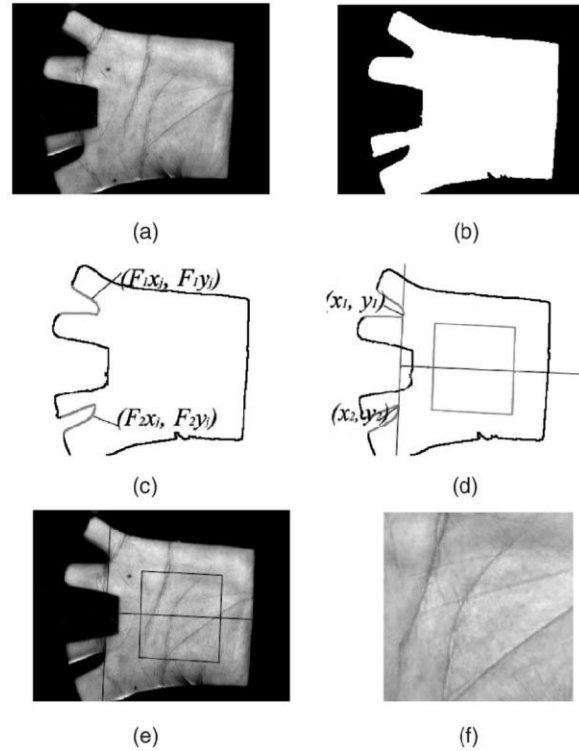
Figure 3. 6: The main steps of ROI extraction. (a) Original image, (b) binary image, (c) boundary tracking, (d) building a coordinate system, , (e) extracting the central part as a sub image, and (f) preprocessed result [1].

3.2.2. Data augmentation

The data is augmented by techniques such as flipping, adding a small amount of distortion to increase the number of samples. Data augmentation resulted in creating a more robust model that learns better characteristics for making distinctions between data images. The generalization of the model as well as training the network with a small amount of data achieved with data augmentation. The images are randomly rotated between [6°, 24°], horizontally shifted between x and y values of [15, 20], and randomly noised with Gaussian and salt & pepper noise at different random noise levels between [0.05, 0.3]. Sometimes, one image may also be combinedly shifted and added noise randomly.

3.3. Baseline CNN-a network

CNN-a is the abbreviation given to the baseline convolutional neural network based on the VGGNet model in this research.

3.3.1. Architecture

The VGGNet achieves state-of-the-art image classification performance, which also took second place in the ILSVRC-2014 (ImageNet Large Scale Visual Recognition Challenge) classification challenge [58]. There are five configurations of VGGNet, which are denoted as VGGNet-A, B, C, D, E. Here, we choose VGGNet-E (widely known as VGG-19), which gives the best result in this experiment. This network with some modifications is considered a baseline network in this work. It takes grayscale images for training and testing. The CNN-a network architecture built for palmprint gender classification consists of 19 layers, where each convolutional and max pool layer pair are followed by a ReLU layer. The last fully connected layer is followed by a softmax layer, which calculates the probabilities and the label with the highest probability will be assigned as the final output.

The detailed network architecture is presented in Figure 3. 6. Each input palmprint image, after ROI extraction and resizing, is normalized to a 128×128 grayscale images. Each pair of

Conv + Maxpool is followed by a rectified linear layer, which allows the network to train several times faster [59]. A dropout layer is added as a regularization layer to prevent overfitting [53]. The final layer, after the fully connected layer, is a softmax layer that classifies the input palmprint image into the respective gender class.
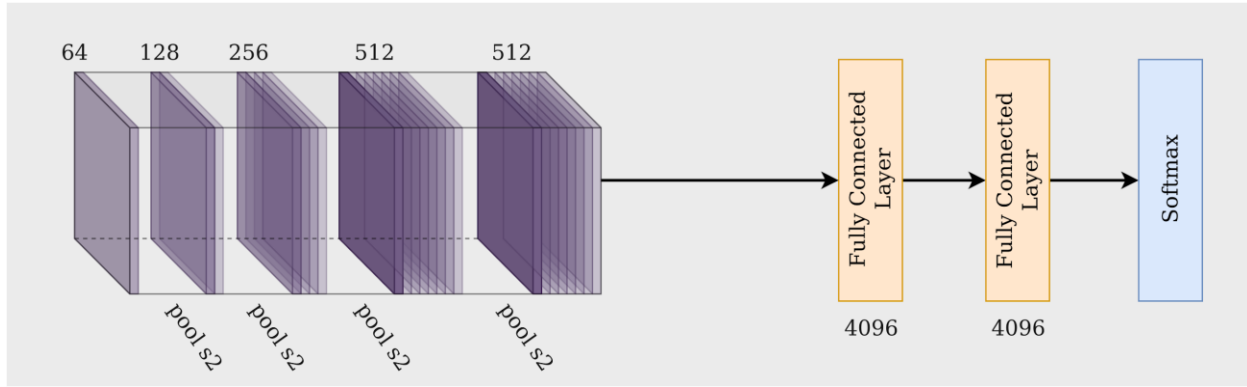
Figure 3. 7: CNN-a architectural elements

All of the layers initialized by Gaussian random numbers and in FC7, the biases were initialized with 1, the value of the standard deviation of the Gaussian distribution was set to 0.005 and the activation function *f* as said before was Rectified Linear Units (ReLU):

$$x^\ell = f(u^\ell), with\ u^\ell = W^\ell x^{\ell-1} + b^\ell \tag{3. 1}$$

$$ReLU(u^\ell) = max(u^\ell, 0) \tag{3. 2}$$

Where ℓ represents the current layer, W^l is the weight of ℓ layer, x^{l-1} represents the output of ℓ-1 layer (input of current layer), b^l is the bias. u^l is the input of ReLU. In FC8 (which is the last fully connected layer), the biases were initialized with 0, the value of the standard deviation of the Gaussian distribution was 0.01. The above-mentioned values for biases and standard deviations were chosen empirically through the experiment. There were three fully connected layers in the configuration of network CNN-a, the first two had 4,096 neurons, and the third one performed 2-class classification and thus contained 2 neurons (one for each class). The final layer was the soft-max transform layer.

Score normalization:

Normalization is used to improve the speed by reducing the difference in the data [60]. The scores obtained from the CNN-a are preferably normalized to get the best results. Two types of score normalizations are evaluated, Soft normalization and Hard normalization, computed using the following formulae (3.3)(3.4) [61].

    i.   **Soft Normalization:**

$$Norm = (X - Mean)/(StdDev) \qquad\qquad (3.3)$$

X: input score vector

Mean: mean of the input vector

StdDev: standard deviation of the input vector

    ii.   **Hard Normalization:**

$$Norm = (X - Min)/(Max - Min) \qquad\qquad (3.4)$$

X: input score vector

Min: minimum value of the vector

Max: maximum value of the vector

3.3.2. CNN-a + SVM classifier

Linear Support Vector Machines (SVM) are widely used for binary classification, which is one of the main points of this research and is based on statistical learning theories [62]. SVM is relatively insensitive to the number of data points and the classification complexity does not depend on the dimensionality of the feature space [63], which helps the method to learn a larger set of patterns. SVM classifier used for classification in this research mainly because as said before is a binary classifier that can optimally separate two classes. It exhibits many advantages in solving the small sample size problem, and it has been successfully applied in gender classification [64][65][66]. The features from the last fully connected layer were used to train

SVM directly. To find out the best model 10-fold cross-validation was carried out on the CASIA Palmprint Image database and PolyU Multispectral Palmprint Database.

3.3.3. CNN-a + PCA + SVM

Principal Component Analysis or PCA is a technique used to emphasize variation and bring out strong patterns in a dataset and is widely used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. Reducing the number of variables or features comes at the expense of accuracy and on the other hand, it makes the large data set simpler, easy to explore, and visualize. Also, it reduces the computational complexity of the model which makes analyzing data much easier and faster for machine learning algorithms without extraneous variables to process. After training the SVM with features from the last fully connected layer of the CNN-a network, PCA was applied to reduce dimension, and SVM has trained again.

3.4. VGGNet transfer learning

In transfer learning, a model trained on one task is repurposed on another related task, usually by some adaptation toward the new task. For example, one can imagine using an image classification model trained on ImageNet [32] to perform texture classification. It would be plausible to use the representation learned by a model, trained for general-purpose classification, for a different image processing task. There have been many works based on pre-trained deep learning models to perform a different task in the past few years.

There are two main ways in which the pre-trained model is used for a different task. In one approach, the pre-trained model, a language model as an example, is treated as a feature extractor, and a classifier is trained on top of it to perform classification. In the other approach, the whole network, or a subset of it, is fine-tuned on the new task. Therefore, the pre-trained model weights are treated as the initial values for the new task and are updated during the training stage.

Transfer learning is used to train a model on a small or of an insufficiently sized dataset to achieve good performance results, where a network trained on a larger dataset is used to

initialize the weights. Transfer learning is an approach applied for leveraging the knowledge of already learned networks, whose extracted features are found to be very powerful in object detection and classification problems [67].

The VGGNet-E (up to 19 layers) is fine-tuned to classify the gender of palmprint images for the following two reasons. First, it is a good model that has been trained by a large-scale dataset, whereas there are limited palmprint images to train a net directly. Second, VGGNet is a deep network that can extract many abstract or high-level palmprint features. Thus, it is conducive to find out the difference between different gender palmprint images.

CNN-VGG is a pre-trained VGGNet-E model on the ImageNet dataset which was used and then fine-tuned using the train set which is a subset of the palmprint datasets presented before. One more fully connected layer is also added on top of the CNN-a network to form CNN-VGG, because in the CNN-a model the neurons in the last fully connected layer changed from 4096 to 2 and such a sudden change may influence performance, thus in an effort to transition smoothly the number of neurons in the additional fully connected layer was set to 1,000. The entire model design of CNN-VGG ends up with a total of 20 layers including fully connected layers. Unlike the CNN-a network, all the layers of the CNN-VGG model up to the last fully connected layer were initialized by the weights from the pre-trained model of the VGGNet-E model on the ImageNet dataset.
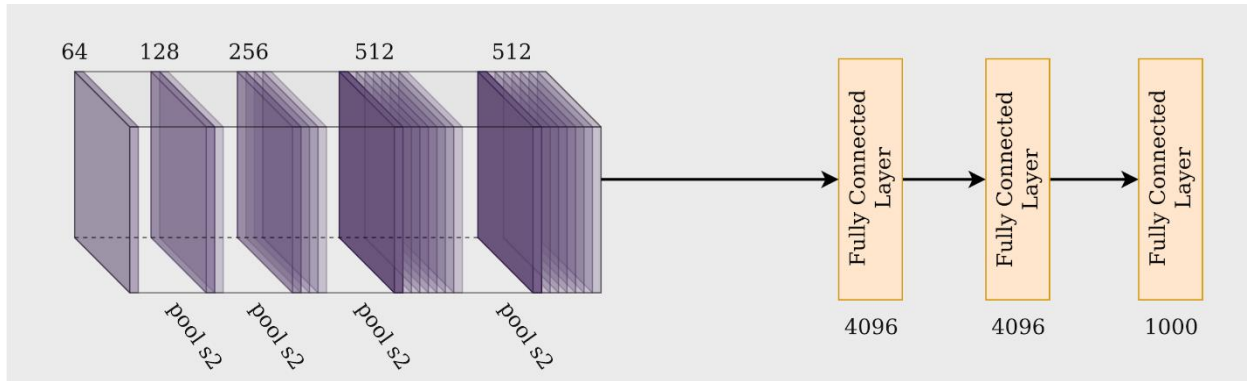

Figure 3. 8: CNN-VGG architectural elements

3.5. DenseNet transfer learning

DenseNet is a network architecture where each layer is directly connected to every other layer in a feed-forward fashion (within each *dense block*). For each layer, the feature maps of all preceding layers are treated as separate inputs whereas its own feature maps are passed on as inputs to all subsequent layers. This connectivity pattern yields state-of-the-art accuracies on CIFAR 10/100 (with or without data augmentation) and SVHN. On the large scale ILSVRC 2012 (ImageNet) dataset, DenseNet achieves similar accuracy as ResNet but using less than half the number of parameters.

In this thesis, CNN-Dense has been introduced for gender classification on palmprint images. The underlying research work is two-fold:

a. The proposed work customizes DenseNet121, a preeminent CNN model pre-trained on ImageNet(animal dataset). As palmprint features substantially differ from animals, this model is fine-tuned to learn complex patterns in ridges and minutiae of the palmprint images.

b. In order to ascertain the aid of specific layers of the proposed framework, the layer-specific feature analysis is carried out through the visualization of the palmprint features learned by arbitrary layers. It helps in selecting the layers of a pre-trained model to fine-tune for palmprint gender classification.

The proposed model incorporates DenseNet121 [68] as a basic building block with major customization. DenseNet is best suitable for the aforesaid problem due to some important observations as it facilitates feature reuse at each subsequent layer inside a dense block and constitutes more feature maps. It also resolves the vanishing gradient problem using a direct path to all preceding layers to route residuals during backpropagation. Furthermore, it outperformed another state-of-the-art CNN architectures in the ImageNet challenge [69]. DenseNet was induced from ResNet [70], where a layer receives outputs from the previously second or third layer through residual connections and the outputs are added on the same depth as shown in Figure 3.8(a). During backpropagation, these residual connections route residuals (errors) directly to the previous layers to let the network learn faster. DenseNet uses a series of dense blocks with transition layers between them. Figure 3.8(b) shows a dense block where a separate connection exists between each layer to all subsequent layers inside a block. Hence, a dense block with L layers has a total $L(L+1)/2$ connections among them [68].
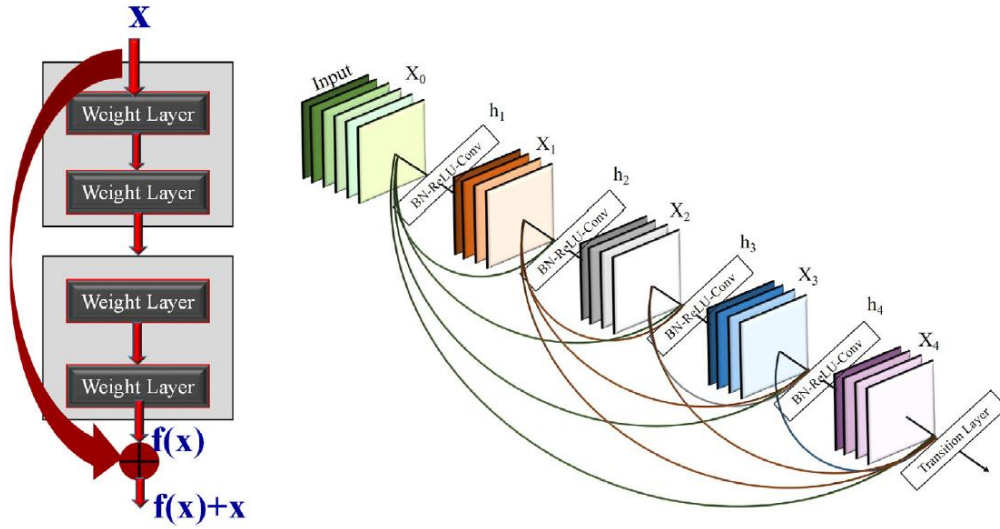
Figure 3. 9: (a) Residual connection, and (b) Dense block [63].

In ResNet, the output of layer-l is

$$x[l] = f(W \times x[l-1] + x[l-2]) \tag{3.5}$$

whereas, in DenseNet

$$x[l] = f(W \times h(x[l-1], x[l-2], \dots, x[l])) \tag{3.6}$$

Here, h denotes the stacking of layers. DenseNet is considered better then ResNet in the manner that; it concatenates features as opposed to ResNet which adds them up. Moreover, DenseNet constitutes immense features and has fewer learning parameters, in fact, less than half as compared to ResNet. These observations raise motivation to apply DenseNet for palmprint gender classification. In a dense block, the nth layer gets feature-maps of all previous layers, $X_0, X_1, \dots, X_{n-1}$ as input. Therefore, the feature map of the nth layer represented by:

$$X_n = F_n([X_0, X_1, \dots, X_{n-1}]) \tag{3.7}$$

Here $[X_0, X_1, \dots, X_{n-1}]$ constitutes the concatenation of all previous feature-maps output in layer $0, \dots, n-1$. F_n is a function performing three subsequent operations, i.e. $BN \rightarrow ReLU \rightarrow Conv_{(3 \times 3)}$. These represent Batch normalization (BN), Rectified Linear Unit function (ReLU), and Convolution respectively [68].

In this research, the Schematic design of the CNN-Dense is illustrated in Figure 3.10 where DenseNet acts as a feature extractor, and SVM is a classifier. DenseNet121 is a deep CNN model pre-trained on the ImageNet dataset, which contains 4 dense blocks containing 6, 12, 24, and 16 dense layers, respectively. Each dense layer constitutes k feature maps as output yet receives comparatively more inputs. Besides, a bottleneck layer is introduced between each dense layer which performs $1 * 1$ convolution. Therefore, a dense block has the following arrangement of layers, $BN \rightarrow ReLU \rightarrow Conv_{(1 \times 1)} \rightarrow BN \rightarrow ReLU \rightarrow Conv_{(3 \times 3)}$. At the end of the second dense block, a transition layer is used that performs convolution with the kernel size of $1 * 1$ and average pooling [68]. Therefore, this transition layer is defined by $BN \rightarrow Conv_{(1 \times 1)} \rightarrow \rightarrow Avgpooling_{(2 \times 2)}$. Collectively, these constitute over 400 layers (including BN, ReLU, dropouts, etc.).

Since this model is huge and could not be trained with palmprint images in the available datasets, the CNN-Dense selects only up to the first two dense blocks (with approx. 50 layers) for feature extraction. The output of the pooling layer (pool_2) after the second dense block represents the gender features of palmprint learned by the CNN-Dense model. Further, a flatten layer is added to form a 401 408-dimensional feature vector. For further down-sampling, two fully connected (fc) layers are added with 512 and 128 neurons, respectively. Besides, two dropout layers (with probability 0.4) are also used to remove overfitting in fc layers.

The FC layers (first sub-block of classification block) with 3 units corresponded to fc1, fc2, and fc3. Indeed, the three FC layers (fc1, fc2, and fc3) are used during model training while the softmax classifier is employed. In such a case, the last FC layer i.e. fc3 contains two neurons corresponding to two gender categories, and the softmax classifier is applied to this layer. However, model training requires the softmax classifier, due to which an additional fc layer consisting of two neurons, is added after the second fc layer. Once the training is completed, this softmax layer (fc3) is replaced with the SVM classifier. Then, the SVM is trained on the features constituted by the second fc layer (fc2) and further used to map the given palmprint image to the correct gender category. Here, it is worth noting that in the CNN-Dense model, the feature extractor part and SVM classifier is trained separately. Though in our experiments SVM classifier exhibits better performance compared to Softmax, therefore, it is added as a classifier after the second fc layer.

The entire model design ends up with a total of 56 layers including Batch Normalization, ReLU, and Dropouts. Moreover, since the initial layers are aimed to constitute basic features, such as points, edges, blobs, etc., they are not involved in training, while the remaining layers are fine-tuned. In our experiment, the initial 27 layers are kept frozen and the rest 29 layers are retrained as shown in Figure 3.10 on palmprint images of the available palmprint datasets.
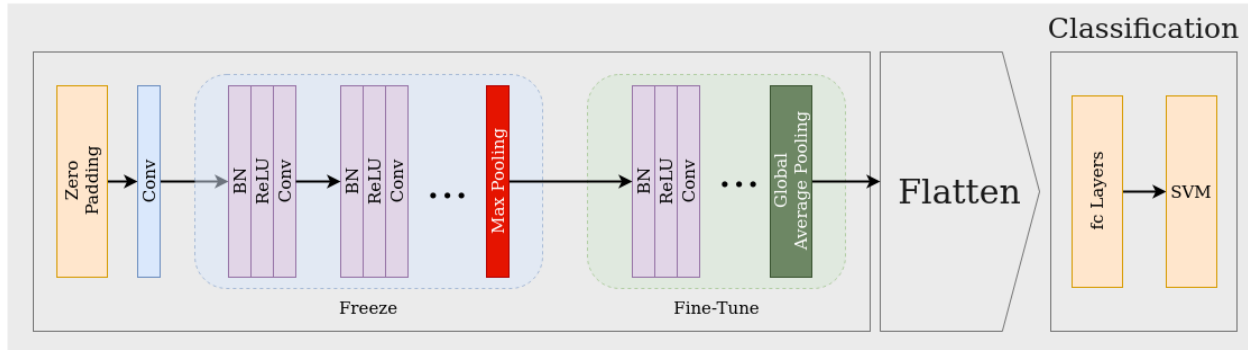
Figure 3. 10: CNN-Dense architectural elements

3.6.   Proposed CNN-b network

CNN-b is the proposed network for performing palmprint gender classification on the PolyU Multispectral Palmprint database and CASIA Palmprint Image Database. After investigating the performance of deep convolutional neural networks on classification tasks, an idea to combine two networks at the training phase was implemented. Many attempts were made in order to determine at what layer or after how many layers the features had to be combined in a vector. Here, the network which has achieved the highest test accuracy of 90.70% on the CASIA palmprint dataset and 94.87% on the PolyU palmprint image database is discussed.

3.6.1. Architecture

CNN-b network is a fusion of two networks, of which one is CNN-VGG and the other CNN-Dense. Both the networks take in the image input and the features extracted from them are combined at feature level and the loss is calculated accordingly. The schematic diagram of CNN-b architecture is shown in figure 3.11. Both the networks take the same batch of input images.

both networks as said before are pre-trained on ImageNet datasets and also they are trained on all of the images from the KTU Palmprint Database and IIT Delhi Palmprint Database which do not have a gender label but they help to fine-tune the network for gender classification. With this design obtaining different features from both networks increases the image information and as a result, the network achieves better performance.
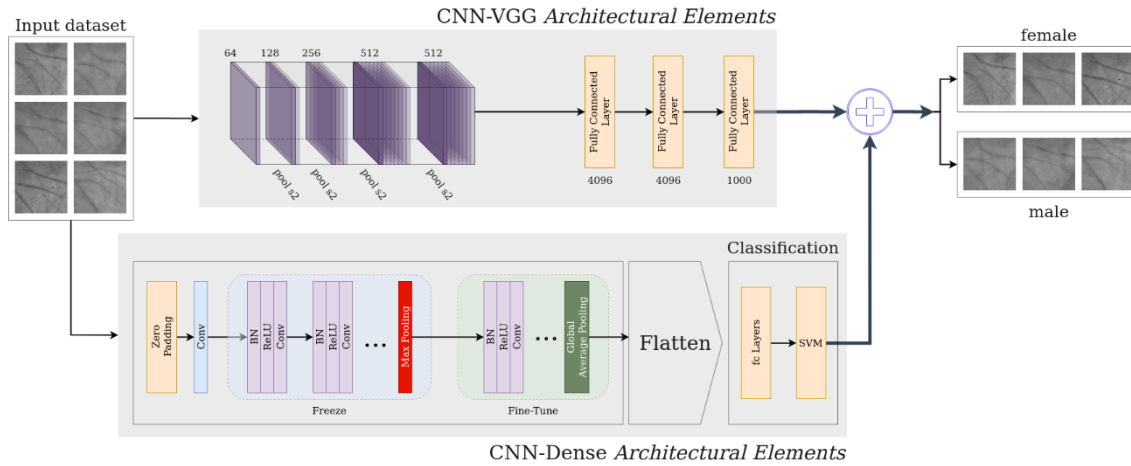
Figure 3. 11: Architectural elements of the proposed CNN-b network

i. Score Level Fusion

Fusion is a method of combining the scores from different classifiers to improve classification performance. Even though there are many types of classifier fusion schemes [71], the most used are the following - feature level fusion and score level fusion.

1. **Feature level fusion:** In this method, the features extracted from the trained models are fused into a vector.

2. **Score level fusion:** In this method, the features are extracted from different classifiers, and their soft output scores (between [0,1]) are fused by minimum, maximum, product, and average fusion rules. These techniques are class-conscious fusion techniques for soft labels [72].

In this thesis, score level fusion with an average rule is used. The fusion of classifiers does not produce any better results if all the classifiers produce the same errors or are correlated. After

doing an empirical study on the different fusion methods, it is established that the average score level fusion of CNN-VGG + CNN-Dense has shown improved accuracy.

ii.   Score level Average Fusion

Following [72], a simple aggregation rule for average is used. Let for a test input and Class labels $= (1, 2, \dots, C)$, the scores for each classifier be:

$s1 = \{ \alpha_1, \alpha_2, \dots, \alpha_C \}, where\ s1 \in [0,1]$

$s2 = \{ \beta_1, \beta_2, \dots, \beta_C \}, where\ s2 \in [0,1]$ (3. 8)

$s3 = \{ \gamma_1, \gamma_2, \dots, \gamma_C \}, where\ s3 \in [0,1]$

Then the average fusion of the score be,

$S[i] = (s1[i] + s2[i] + s3[i])/3 ; for\ i \in C$ (3. 9)

The output label is the max(S).

3.7.   Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM), uses the class-specific gradient information flowing into the final convolutional layer of a CNN to produce a coarse localization map of the important regions in the image.

convolutional layers naturally retain spatial information, which is lost in fully connected layers, it is expected that the last convolutional layers to have the best compromise between high-level semantics and detailed spatial information. The neurons in these layers look for semantic class-specific information in the image. Grad-CAM uses the gradient information flowing into the last convolutional layer of CNN to assign importance values to each neuron for a particular decision of interest. To obtain the class-discriminative localization map $L^c_{Grad-CAM} \in \mathbb{R}^{u \times v}$ of width u and height v for any class c, first, the gradient of the score for class c, yc (before the softmax) is computed, with respect to feature map activations Ak of a convolutional layer, i.e. $\frac{\partial y^c}{\partial A^k}$. These gradients flowing back are global-average-pooled over the

width and height dimensions (indexed by i and j respectively) to obtain the neuron importance weights α_k^c:

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i \sum_j}^{\text{global average pooling}} \quad \underbrace{\frac{\partial \mathcal{Y}^c}{\partial A_{ji}^k}}_{\text{gradients via backprop}} \tag{3.10}$$

During the computation of α_k^c while backpropagating gradients concerning activations, the exact computation amounts to successive matrix products of the weight matrices and the gradient with respect to activation functions till the final convolution layer that the gradients are being propagated to. Hence, this weight α_k^c represents a partial linearization of the deep network downstream from A, and captures the 'importance' of feature map k for a target class c.

Then a weighted combination of forwarding activation maps was performed and followed by ReLU to obtain,

$$L_{\text{Grad-CAM}}^c = ReLU \underbrace{\left(\sum_k \alpha_k^c A^k \right)}_{\text{linear combination}} \tag{3.11}$$

This results in a coarse heatmap of the same size as the convolutional feature maps. ReLU was applied to the linear combination of maps to only focus on the features that have a positive influence on the class of interest. As expected, without this ReLU, localization maps sometimes highlight more than just the desired class and perform worse at localization.

Chapter 4:

Experimental results and analysis

4.1. Baseline CNN-a: Training and results

CNN-a is the abbreviation given to the baseline convolutional neural network in this research. The following explains the training of the network and its results. the architecture of this network is explained in section 3.3.1.

The CNN-s is trained for roughly 3000 epochs on normalized grayscale images, with a learning rate fixed at 0.001 throughout the training with a batch-size of 64. Within 3,000 iterations the verification accuracy was seen to be improved, and the training loss was declining. However, after 3,000 iterations, training loss stopped declining and further training could not improve the model. Thus, we trained the network for 3,000 iterations, also the softmax layer is used as the classifier at the end of the network. It is a logistic function that computes the probabilities. Finally, the CNN-a model was obtained. The results of this network on both dataset CASIA palmprint image database and PolyU Multispectral Palmprint Database are shown in Table 4.1 and Table 4.2.

4.1.1. CNN-a + SVM: Training and results

The hyperparameter selection is done by grid search using n-Fold cross-validation. The grid search is done through an automatic selection of a range of C and g values. This search range is changed automatically until the best cross-validation accuracy is found. Through this study, different kernel functions were tested, and their results are reported in Table 4.1 and Table 4.2. The average accuracy was obtained by the following equations:

$$X_i = \frac{C}{N} \times 100\% \tag{4.1}$$

$$X = \frac{\sum_{i=1}^{n} X_i}{n} \tag{4.2}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n(n-1)}} \tag{4.3}$$

Where X_i denotes the i-th correct classification rate, C is the number of correct trials, N is the number of testing samples. And $\bar{X}$ is the average of X_i , because we carry out 10-fold cross-validation, so $n = 10$. S denotes the standard deviation.

4.1.2. CNN-a + PCA + SVM: Training and results

Principal Component Analysis or PCA is a technique used to emphasize variation and bring out strong patterns in a dataset and is widely used to reduce the dimensions. After applying PCA, SVM has trained again and the whole network was tested on the test set for both CASIA and PolyU palmprint image databases. The results of this experiment are shown in table 4.1 and Table 4.2.

Table 4. 1: Test result of CNN-a, CNN-a + SVM, CNN-a + PCA + SVM on CASIA Palmprint Image Database.

Method	Kernel function	Average Accuracy (%)	Standard Deviation (%)	Average Accuracy of Female (%)	Average Accuracy of male (%)
CNN-a	Linear	58.12	9.56	26.06	92.28
	Polynomial	48.32	3.31	0.12	98.67
	Radial Basis	42.65	1.87	87.67	10.08
CNN-a + SVM	Linear	62.23	10.08	27.87	93.67
	Polynomial	54.46	2.56	0.09	99.56
	Radial Basis	48.12	1.67	85.78	12.34
CNN-a + PCA + SVM	Linear	69.43	4.65	50.45	87.34
	Polynomial	70.02	6.94	69.43	69.92
	Radial Basis	73.88	7.08	69.51	77.50

Table 4. 2: Test result of CNN-a, CNN-a + SVM, CNN-a + PCA + SVM on PolyU Multispectral Palmprint Database.

Method	Kernel function	Average Accuracy (%)	Standard Deviation (%)	Average Accuracy of Female (%)	Average Accuracy of male (%)
CNN-a	Linear	68.12	8.32	58.19	78.25
	Polynomial	52.11	2.01	20.42	89.03
	Radial Basis	78.23	0.27	87.15	60.11
CNN-a + SVM	Linear	68.09	8.98	38.87	80.34
	Polynomial	70.43	1.30	20.62	92.0
	Radial Basis	72.87	1.04	90.32	60.55
CNN-a + PCA + SVM	Linear	78.23	2.94	87.45	73.12
	Polynomial	80.22	2.51	69.43	72.86
	Radial Basis	90.56	1.38	89.59	99.00

4.2. CNN-VGG: Retraining and results

Here, in this work, the VGGNet-19 pre-trained network model with one additional fully connected layer is used as a feature extractor. The available VGGNet-19 trained model on the ILSVRC-2014 dataset is used with the tutorial code from TensorFlow[2] to retrain the network's final softmax layer for gender classification on palmprint. Keeping all the layer weights of the model, only the final softmax layer is retrained. In short, the layers acted as a feature extractor, and the final layer is trained on these extracted features to produce probabilities of 2 classes. The following describes the retraining of CNN-VGG on the train set.

---

[2] https://github.com/machrisaa/tensorflow-vgg

First, the network is trained on all of the data from KTU and IITD palmprint datasets, then for the CASIA dataset the training set contained 4,032 images (126 hands from female and male respectively), and 448 images (14 hands from female and male respectively) selected as a test set. For the PolyU palmprint dataset all training images of 15 spectra combined as a whole training set to train CNN-VGG. Each spectrum has 88 hands for training (44 hands from female and male respectively, totally 2,112 images and half of them are original images), 32 hands for validation (16 hands from female and male respectively, totally 768 images and half of them are original images), and 56 hands for testing (28 hands from female and male respectively, including 672 original and 672 augmented data). There is no imbalance between males and females in these three datasets. This network achieves a 78.31% accuracy rate on CASIA and 92.26% on PolyU test sets. feature reduction by PCA could not improve performance. This is probably because with enough training data, VGGNet is not overfitting and the extracted features could represent the image well, thus linear feature reduction brings a negative effect.

4.3. CNN-Dense: Retraining and results

For training this model, Stochastic Gradient Descent (SGD) optimizer with momentum is employed, as plain SGD can make erratic updates on non-smooth functions. SGD with momentum updates the weights with the moving average of the changes in individual weights for a single training sample. If everything is labeled with t then the moving average is given as:

$$\Delta W_{hl}^{(n)}(t) = \eta \sum_p delta_l^{(n)}(t) + \alpha . \Delta W_{hl}^{(n)}(t-1) \qquad (4.4)$$

where the second term $\alpha . \Delta W_{hl}^{(n)}(t-1)$ is the momentum, which increases for dimensions where gradients point in the same direction and decrease for those with changing direction. Therefore, leads to fast convergence. Here, α is the momentum parameter, when α is zero, it works as simple SGD, whereas small values of α fluctuate updating function to a great extent. If α increases, it adds a contribution from previous training samples [73]. In our experiment, α is set to 0.9, as a large value turns out to have a smoother curve.

In order to train the feature extractor, the softmax loss function is used to calculate residuals which are backpropagation for weight adjustment in the early layers. After training the feature extractor, the further task is to train the SVM classifier. To achieve this, the softmax classifier is removed and training features are extracted from the second fc layer. This network is also retrained on the same train set as CNN-VGG and achieves an accuracy rate of 85.64% for CASIA and 92.69% for the PolyU palmprint image database.

4.4.   Proposed CNN-b: Training and results

The CNN-b network (architecture explained in section 3.6.1.) is trained on the same train sets as CNN-VGG and CNN-Dense networks. After an empirical study, the initial learning rate of 0.001 is used with stochastic gradient descent and softmax cross-entropy loss function. a dropout of 0.7 is used. This network is trained for 10k steps with a batch size of 64. The result of this network on both CASIA and PolyU palmprint database along with CNN-a, CNN-VGG, CNN-Dense is shown on the ROC curve. To the best-known CNN-b network has the highest accuracy rate among the literature and the other networks implemented in this research. it achieves a 94.87% accuracy rate on the PolyU multispectral palmprint database and 90.70% on the CASIA palmprint image database.

4.4.1. K-Fold cross-validation

Cross-validation or rotation estimation is a validation technique used to do a statistical analysis of the data and to avoid overfitting the model on train data [74]. K-Fold cross-validation is a non-exhaustive cross-validation technique, where the original data are divided into K equal sized subsamples. In each rotation, each subsample is retained as a validation set and the remaining K-1 subsamples are used for training the K models.

A common choice for K-Fold cross-validation is K=10. The augmented data is randomly split into 10 sets and the training follows the training parameters of the CNN-b and is trained for 10k steps.
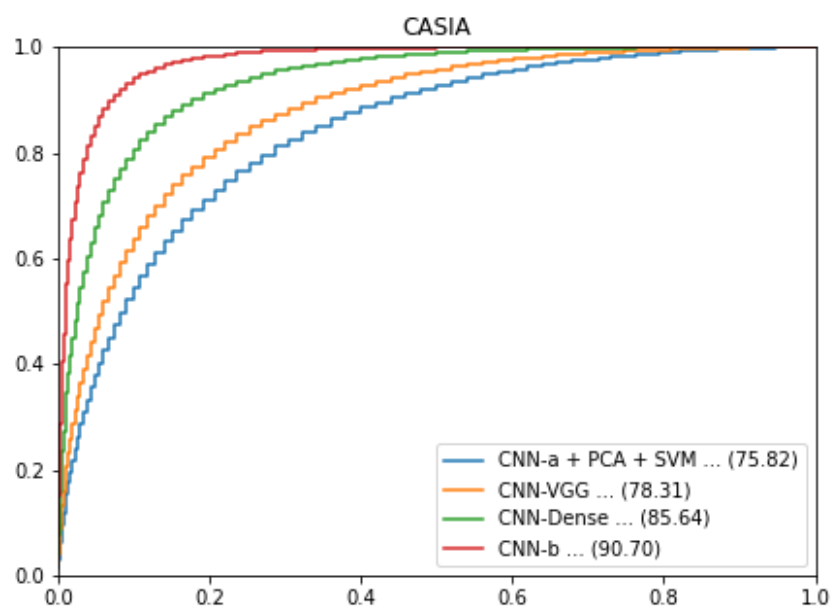
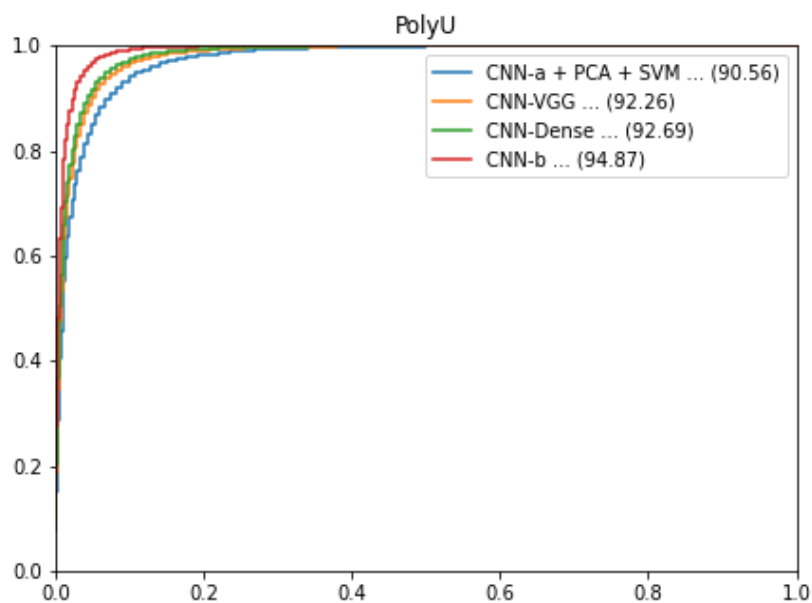Figure 4. 2: ROC curve of all the CNN models tested on CASIA palmprint dataset.

Figure 4. 1: ROC curve of all the CNN models tested on PolyU palmprint dataset.

4.5. Comparison of performances between different methods of palmprint and fingerprint gender classification

In this section, the performance of the proposed network is compared with similar methods found in the literature. In [15] Wu and Yuan used very basic geometric properties such as length, aspect ratio, and width with polynomial smooth- SVM and got 85% accuracy in gender prediction. Whereas in [16] the authors used Geometric features, boundary, and Fourier descriptors on Zernike moments of hand geometry and achieved a 98% result on a dataset of 40 Palmprints. In another study by Xie et al. [75] fine-tuned a VGGNet to achieve a 20-layer network and experimental results show that they achieved a 92.1% accuracy rate on PolyU multispectral palmprint database and 80.09% accuracy on the CASIA palmprint image database. All the accuracy results in Table 4. 3 are reported in the papers mentioned in this section.

The drawback of the reported method in [16] is that it is not suitable for low resolution and far distance images captured using the touch-free method. As they need touch-based palm acquisition. On the contrary, the proposed method in this study is suitable for both the approaches: touch-based and touch-free. The performance of the proposed method in this research outperformed the other methods with an accuracy rate of 94.87% on the PolyU dataset and 90.70% accuracy on the CASIA dataset.

Table 4. 3: Comparative analysis of palmprint-based gender classification methods

Authors	Database	Classification	Accuracy
Amayeh et al. [16]	20 Male, 20 Female	Score Level Fusion	98%
Wu & Yuan [15]	180 Palmprint images of 30 voluntaries	PSSVM	85%
Xie et al. [75]	PolyU & CASIA datasets	CNN	92.1% 80.09%
Proposed method	PolyU & CASIA datasets	CNN	94.87% 90.70%

In the following, a review of research papers on fingerprint-based gender classification is represented. Based on the results reported in Table 4. 3 the performance of the proposed method for palmprint-based gender classification is comparable to the performance of fingerprint-based gender classification methods.

Table 4. 4: Analysis of the literature for fingerprint-based gender classification [76]

Reference Publication	Feature Extraction	Database Description	Algorithms	Accuracy
Ashish Mishra et al [77]	orientation field & singular points, minutiae, incipient ridges	NIST	SVM, NN	SVM 76.06% NN 83.7%
Ronny Merkel et al [78]	Binary Pixels, Mean print gray value, mean variance, Local image regions, Gradients, Tamura contrast, coherence, Benford's Law	CWL & CLSM	WEKA Machine learning toolbox, LMT	-
Prabha et al [79]	multi-resolution statistical	Own dataset	Back propagation neural network	96.60%
S.S. Gornale et al [80]	DWT and Gabor based	Own dataset	LDA & QDA	97%
Dr. Ashish Mishra et al [81]	Ridge count, Ridge density, Ridge thickness to valley thickness ratio (RTVTR), Ridge width and Fingerprint patterns and Pattern types	Own dataset	SVM, NN	SVM 76.06% NN 83.7%
Sri Suwarno et al [82]	RIDGE DENSITY DWT	-	-	-
A. S. Falohun et al [83]	RTVTR DWT/PCA	own dataset	ANN	Female 80% Males 72.86%
Shivanand Gornale [84]	DWT and Gabor based	Own dataset	LDA & QDA	97%
Mangesh K. Shinde et al [85]	DWT and SVD	Own dataset	KNN	Male 82.90% Female 82.60%

4.6.   GRAD-CAM: Results

This is the result of applying GRAD-CAM. GRAD-CAM uses the gradients of any target concept, flowing into the final convolutional layer to produce a coarse localization map highlighting important regions in the image for predicting the concept which is gender in this research. To explain it in more detail, the final convolutional feature map is taken and the weight of every channel in that feature with the gradient of the class concerning the channel is calculated. It indicates how intensely the input image activates different channels by how important each channel is with regards to the class, and it does not require any re-training or change in the existing architecture.

As explained in Section 3.7, the gradient of the class output value with respect to the feature map is computed. Then the gradients are pooled over all the axes leaving out the channel dimension. Finally, the output feature map with the computed gradient values is weighted. Then the average of the weighted feature map along the channel dimension is calculated and resulted in a heat map. Then the heat map is normalized to make the values between 0 and 1. Then the existing heat map is resized to match the palmprint image size. The original palmprint image and the heat map are blended to superimpose the heat map on to the image.

As shown in Figures 4.3 and 4.4, the GRAD-CAM has specified the specific areas of the palmprint are considered by CNN to distinguish between female and male classes. These heat maps show that the network is looking at the parts that are highlighted in red for making the classification decision, which are different areas for male and female classes. To recognize the male class, it can be observed that the network focuses on the hypothenar part of the palmprint. For the female class, it looks between the principal lines and the interdigital part of the palmprint.
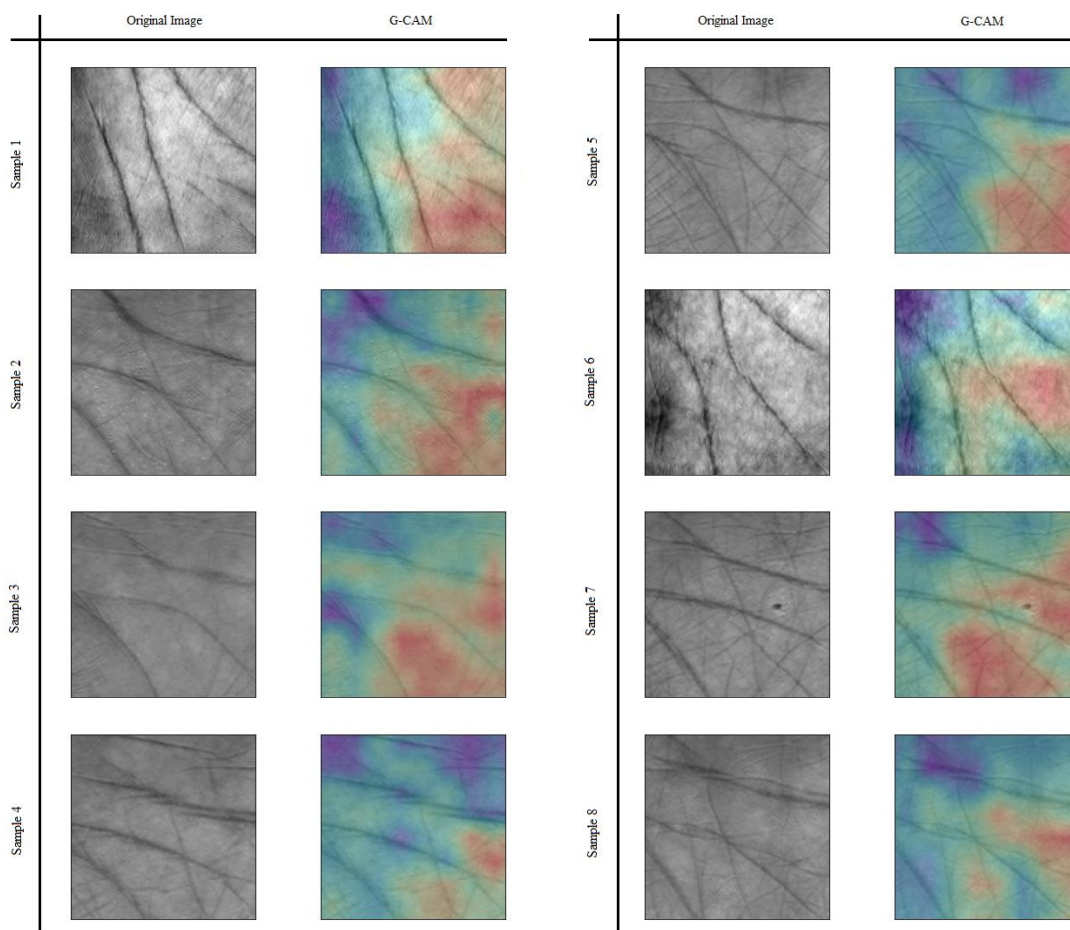
Figure 4. 3: Original image and GRAD-CAM visualization of 8 different samples for the "male" class.
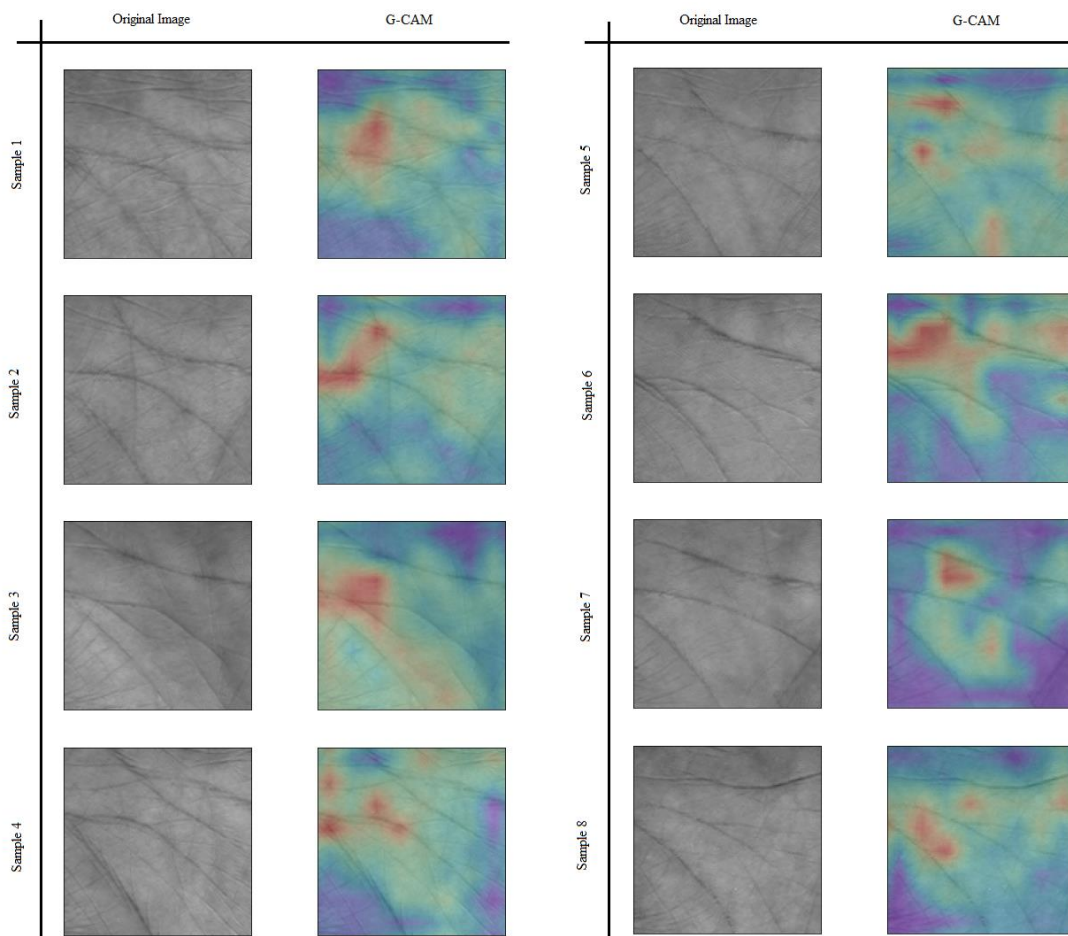
Figure 4. 4: Original image and GRAD-CAM visualization of 8 different samples for the "female" class.

Chapter 5:

Conclusion and future work

5.1. Conclusion

In this work, an effective and feasible method for gender classification by palmprint images is proposed which is a supervised deep learning approach for the problem of palmprint gender classification that was developed and investigated. Four publicly available palmprint image datasets were used for training and testing the deployed methods. Preprocessing steps such as ROI extraction and data augmentation was performed on the data. Four deep convolutional neural networks were trained and tested on the image dataset (CNN-a, CNN-VGG, CNN-Dense, CNN-b). CNN-VGG and CNN-Dense were already pre-trained on the ImageNet dataset for improving the performance of the network as well as reducing the time of computation greatly.

The proposed CNN-b network which is made by combining two deep convolutional neural networks (CNN-VGG and CNN-Dense) achieved the best result to the best known, the accuracy rate for the CASIA palmprint dataset was 90.70% and on PolyU multispectral palmprint image datasets was 94.87%. In this work, since the model was trained on publicly available datasets and because of the techniques used for designing the final model, it can be easily generalizable for any other dataset with any size of data.

The GRAD-CAM method is also performed and the resulted heat maps showed the specific area of the palmprint that is more important to distinguish between female and male classes (hypothenar for male class and interdigital/between principal lines for female class). These heat maps indicate where is the network looking when making the classification decision.

5.2. Future work

Since gender classification can assist in individual identification in many ways this work can be considered as a helpful preprocessing step on the task of individual identification based on palmprint, by reducing the search space of biometric databases, the identification task could achieve a higher accuracy rate with the less computational process.

Multimodal Biometrics is an emerging subject that integrates two or more biometric features to overcome certain limitations of using them individually. For future research, a robust suggestion of extracting gender information from multiple source samples can be discussed to minimize the search time for the identification and verification process. Since various researches have been carried out on the determination of gender from fingerprints it can be a good suggestion to combine it with palmprint to enhance the accuracy rate of the system. The fusion of information for this kind of system can be done at four different levels such as sensor level, match score level, feature level, and decision level fusion [86].

This method can also be used for palmprint latent matching to specify the gender of the prints found in crime scenes before the identification process. It can also be performed on partial palmprint latent found on crime scenes as GRAD-CAM results presented which part of the palmprint is more important in distinguishing between female and male. So, as researched, it can be investigated for that matter as well.

Also, more image data can be acquired for training the deep models, since there are proven results that the increased data improve the classification performance and widens the features available for the system to learn.

References

[1]   X. Wu, D. Zhang, and K. Wang, "Fisherpalms based palmprint recognition," *Pattern Recognit. Lett.*, vol. 24, no. 15, pp. 2829–2838, 2003, doi: 10.1016/S0167-8655(03)00141-7.

[2]   X.-Y. Jing, Y.-Y. Tang, and D. Zhang, "A Fourier–LDA approach for image recognition," *Pattern Recognit.*, vol. 38, no. 3, pp. 453–457, Mar. 2005, doi: 10.1016/J.PATCOG.2003.09.020.

[3]   S. S. Gornale, A. Patil, M. Hangarge, K. A. Science, R. A. Pardeshi, and K. Arts, "Smart Computational Strategies: Theoretical and Practical Aspects," *Smart Comput. Strateg. Theor. Pract. Asp.*, no. January, 2019, doi: 10.1007/978-981-13-6295-8.

[4]   S. K. Dewan, "Elementary, Watson: Scan a palm, find a clue," *New York Times*, vol. 21, p. 3, 2003.

[5]   G. Azzopardi, A. Greco, and M. Vento, "Gender Recognition from Face Images Using a Fusion of SVM Classifiers," in *Image Analysis and Recognition*, 2016, pp. 533–538.

[6]   K. Zhang, L. Tan, Z. Li, and Y. Qiao, "Gender and Smile Classification Using Deep Convolutional Neural Networks," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 739–743, 2016, doi: 10.1109/CVPRW.2016.97.

[7]   S. S. Liew, M. Khalil-Hani, S. Ahmad Radzi, and R. Bakhteri, "Gender classification: A convolutional neural network approach," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 24, no. 3, pp. 1248–1264, 2016, doi: 10.3906/elk-1311-58.

[8]   V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, "Learning to predict gender from iris images," *IEEE Conf. Biometrics Theory, Appl. Syst. BTAS'07*, pp. 0–4, 2007, doi: 10.1109/BTAS.2007.4401911.

[9]   S. Lagree and K. W. Bowyer, "Predicting ethnicity and gender from iris texture," *2011 IEEE Int. Conf. Technol. Homel. Secur. HST 2011*, pp. 440–445, 2011, doi: 10.1109/THS.2011.6107909.

[10]   R. Kaur, S. G. Mazumdar, and D. Bhonsle, "A Study on Various Methods of Gender Identification Based on Fingerprints," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 4, pp. 532–537, 2012.

[11]   R. Kaur and S. G. Mazumdar, "Fingerprint Based Gender Identification Using Frequency Domain Analysis," *Int. J. Adv. Eng. Technol.*, vol. 3, no. 1, pp. 295–299, 2012.

[12]   D. Gnana Rajesh and M. Punithavalli, "An efficient fingerprint based gender classification system using dominant un-decimated wavelet coefficients," *Res. J. Appl. Sci. Eng. Technol.*, vol. 8, no. 10, pp. 1259–1265, 2014, doi: 10.19026/rjaset.8.1093.

[13]   M. Watanabe *et al.*, "Palm Dorsal Vein," *Encycl. Biometrics*, pp. 1027–1027, 2009, doi: 10.1007/978-0-387-73003-5_2361.

[14]   Wei Shu and D. Zhang, "Palmprint verification: an implementation of biometric technology," no. 1, pp. 219–221, 2002, doi: 10.1109/icpr.1998.711120.

[15]   M. Wu and Y. Yuan, "Gender classification based on geometry features of palm image," *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/734564.

[16]   G. Amayeh and R. U. S. A. Data, "Hand-based gender classification," vol. 2, no. 12, 2014.

[17]   T. Kanchan, K. Krishan, K. R. Aparna, and S. Shyamsundar, "Is there a sex difference in palm print ridge density?," *Med. Sci. Law*, vol. 53, no. 1, pp. 33–39, 2013, doi: 10.1258/msl.2012.011092.

[18]   G. Amayeh, G. Bebis, and M. Nicolescu, "Gender classification from hand shape," *2008 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work.*, no. June 2008, 2008, doi: 10.1109/CVPRW.2008.4563122.

[19]   M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," *Eur. Conf. Comput. Vis.*, vol. 8689, pp. 818–833, 2014.

[20]   C. Szegedy, W. Liu, and Y. Jia, "Going deeper with convolutions," *EEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1–9, 2015.

[21]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[22]   F. H. . Tivive and A. Bouzerdoum, "A gender recognition system using shunting inhibitory convolutional neural networks," *IEEE Int. Jt. Conf. Neural Netw. Proc.*, pp. 5336–5341, 2006.

[23]   H. Zhang, Q. Zhu, and X. Jia, "An Effective Method for Gender Classification with Convolutional Neural Networks," *nternational Conf. Algorithms Archit. Parallel Process.*,

pp. 78–91, 2015.

[24] J. R. Udry, "The nature of gender," *Demography*, vol. 31, no. 4, pp. 561–573, 1994, doi: 10.2307/2061790.

[25] O. Singh, G. Bommagani, S. R. Ravula, and V. K. Gunjan, "Pattern Based Gender Classification," vol. 3, no. 10, pp. 888–895, 2013.

[26] L. Beckwith, M. Burnett, S. Wiedenbeck, and V. Grigoreanu, "Gender HCI: Results To Date Regarding Issues in Problem-Solving Software," *AVI 2006 Gend. Interact. Real Virtual Women a Male World Work. Pap.*, no. May, pp. 1–4, 2006.

[27] M. Demirkus, K. Garg, and S. Guler, "Automated person categorization for video surveillance using soft biometrics," in *Proc.SPIE*, 2010, vol. 7667, doi: 10.1117/12.851424.

[28] J. H. P. Hoffmeyer-Zlotnik, J. H. P. Hoffmeyer-Zlotnik, and C. Wolf, *Advances in cross-national comparison: A European working book for demographic and socio-economic variables*. Springer Science & Business Media, 2003.

[29] R. Maldonado, P. Tansuhaj, and D. Muehling, "The impact of gender on ad processing: A social identity perspective," *Acad. Mark. Sci. Rev.*, vol. 2003, no. 3, p. 1, 2003.

[30] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, "A Study on Gait-Based Gender Classification," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1905–1910, 2009, doi: 10.1109/TIP.2009.2020535.

[31] S. A. Khan, M. Ahmad, M. Nazir, and N. Riaz, "A comparative analysis of gender

classification techniques," *Middle - East J. Sci. Res.*, vol. 20, no. 1, pp. 1–13, 2014, doi: 10.5829/idosi.mejsr.2014.20.01.11434.

[32]   J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[33]   JAMES L. WAYMAN, "Fundamentals of biometric authentication technologies," *Int. J. Image Graph.*, vol. 01, no. 01, pp. 93–113, 2001.

[34]   A.K. Jain ; A. Ross ; S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, 2004, doi: 10.1109/TCSVT.2003.818349.

[35]   D. Zhang ; Wai-Kin Kong ; J. You ; M. Wong, "Online palmprint identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1041–1050, 2003, doi: 10.1109/TPAMI.2003.1227981.

[36]   A. Leguebe, S. Vrydagh, and J. Ducros, "Finger ridge-count correlations: Race, sex and hand differences," *J. Hum. Evol.*, vol. 10, no. 6, pp. 453–466, Sep. 1981, doi: 10.1016/S0047-2484(81)80092-9.

[37]   E. A. Ohler and H. Cummins, "Sexual differences in breadths of epidermal ridges on finger tips and palms," *Am. J. Phys. Anthropol.*, vol. 29, no. 3, pp. 341–362, 1942, doi: 10.1002/ajpa.1330290302.

[38]   R. T. Moore, "Automatic fingerprint identification systems," *Adv. Fingerpr. Technol.*, pp. 163–191, 1994.

[39] E. Gutiérrez-Redomero, M. C. Alonso, and J. E. Dipierri, "Sex differences in fingerprint ridge density in the Mataco-Mataguayo population," *HOMO*, vol. 62, no. 6, pp. 487–499, Dec. 2011, doi: 10.1016/J.JCHB.2011.05.001.

[40] D. Z. Loesch and M. Lafranchi, "Relationships of epidermal ridge patterns with body measurements and their possible evolutionary significance," *Am. J. Phys. Anthropol.*, vol. 82, no. 2, pp. 183–189, 1990, doi: 10.1002/ajpa.1330820207.

[41]   and C. Y. Wang, Jen-Feng, Chen-Liang Lin, Yung-Hsien Chang, Mark L. Nagurka, Chen-Wen Yen, "Gender Determination Using Fingertip Features," *Internet J. Med. Updat.*, vol. 3, pp. 22–29, 2088.

[42] Anil K. Jain ; Jianjiang Feng, "Latent Palmprint Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1032–1047, 2009.

[43] T. Kanchan, A. Gupta, and K. Krishan, "Estimation of sex from mastoid triangle – A craniometric analysis," *J. Forensic Leg. Med.*, vol. 20, no. 7, pp. 855–860, Oct. 2013, doi: 10.1016/J.JFLM.2013.06.016.

[44] M. A. Acree, "Is there a gender difference in fingerprint ridge density?," *Forensic Sci. Int.*, vol. 102, no. 1, pp. 35–44, May 1999, doi: 10.1016/S0379-0738(99)00037-7.

[45] Sudesh Gungadin, "Sex determination from fingerprint ridge density," *Internet J. Med. Updat.*, vol. 2, no. 2, 2007.

[46] E. Gutiérrez-Redomero, C. Alonso, E. Romero, and V. Galera, "Variability of fingerprint ridge density in a sample of Spanish Caucasians and its application to sex determination,"

Forensic Sci. Int., vol. 180, no. 1, pp. 17–22, Aug. 2008, doi: 10.1016/J.FORSCIINT.2008.06.014.

[47] K. Krishan, T. Kanchan, R. Sharma, and A. Pathania, "Variability of palmprint ridge density in a North Indian population and its use in inference of sex in forensic examinations," *HOMO*, vol. 65, no. 6, pp. 476–488, Dec. 2014, doi: 10.1016/J.JCHB.2014.08.003.

[48] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex.," *J. Physiol.*, vol. 148, no. 3, pp. 574–591, 1959.

[49] Andrej Karpathy, "neural networks part 1: Setting up the architecture," 2015.

[50] V. Suárez-Paniagua and I. Segura-Bedmar, "Evaluation of pooling operations in convolutional architectures for drug-drug interaction extraction," *BMC Bioinformatics*, vol. 19, pp. 92–101, 2018, doi: 10.1186/s12859-018-2195-1.

[51] M. J. Brown, L. A. Hutchinson, M. J. Rainbow, K. J. Deluzio, and A. R. De Asha, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proc. 27th Int. Conf. Mach. Learn.*, pp. 807–814, 2010, doi: 10.1123/jab.2016-0355.

[52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, 2015.

[53] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15,

pp. 1929–1958, 2014.

[54] "KTU CVPR Lab. Palmprint Database-2." [Online]. Available: http: //ceng2.ktu.edu.tr/~cvpr/palmDB2.htm.

[55] "CASIA Palmprint Database." [Online]. Available: http://biometrics.idealtest.org.

[56] "PolyU multispectral palmprint Database." .

[57] "IIT Delhi Palmprint Image Database version 1.0." [Online]. Available: https://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Palm.htm.

[58] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[59] A. Krizhevsky, I. Sutskever, and G. . Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, pp. 1097–1105, 2012.

[60] A. B. A. Graf and S. Borer, "Normalization in support vector machines," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2191, pp. 277–282, 2001, doi: 10.1007/3-540-45404-7_37.

[61] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, 2005, doi: 10.1016/j.patcog.2005.01.012.

[62] H. Byun and S. W. Lee, "Applications of support vector machines for pattern recognition: A survey," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect.*

Notes Bioinformatics), vol. 2388, pp. 213–236, 2002, doi: 10.1007/3-540-45665-1_17.

[63]   W. Li and Z. Liu, "A method of SVM with normalization in intrusion detection,"

Procedia Environ. Sci., vol. 11, no. PART A, pp. 256–262, 2011, doi:

10.1016/j.proenv.2011.12.040.

[64]   *Computational Intelligence and Communication Networks (CICN)*. 2012.

[65]   S. S. Gornale, M. Basavanna, and R. Kruthi, "Gender Classification Using Fingerprints

Based On Support Vector Machines," vol. 6, no. 7, pp. 588–593, 2015.

[66]   B. Moghaddam and Ming-Hsuan Yang, "Gender classification with support vector

machines," in *Proceedings Fourth IEEE International Conference on Automatic Face and

Gesture Recognition (Cat. No. PR00580)*, 2000, pp. 306–311, doi:

10.1109/AFGR.2000.840651.

[67]   S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*,

vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[68]   G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected

convolutional networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition,

CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[69]   O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J.

Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[70]   D. Menotti *et al.*, "Deep Representations for Iris, Face, and Fingerprint Spoofing

Detection," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 4, pp. 864–879, Apr. 2015, doi:

10.1109/TIFS.2015.2398817.

[71]   D. Ruta and B. Gabrys, "An Overview of Classifier Fusion Methods," *Comput. Inf. Syst.*, vol. 7, no. 1, pp. 1–10, 2000.

[72]   L. I. Kuncheva and E. Alpaydin, *Combining Pattern Classifiers: Methods and Algorithms*, vol. 18, no. 3. 2007.

[73]   L. Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent," *Lechevallier Y., Saporta G. Proc. COMPSTAT'2010*, pp. 3–4, 2010, doi: 10.1007/978-3-7908-2604-3.

[74]   R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, 1995, pp. 1137–1143.

[75]   Z. Xie, Z. Guo, and C. Qian, "Palmprint gender classification by convolutional neural network," *IET Comput. Vis.*, vol. 12, no. 4, pp. 476–483, 2018, doi: 10.1049/iet-cvi.2017.0475.

[76]   C. P. Divate and S. Z. Ali, "Study of Different Bio-Metric Based Gender Classification Systems," *Proc. Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2018*, no. Icirca, pp. 347–353, 2018, doi: 10.1109/ICIRCA.2018.8597340.

[77]   G. Vitello, F. Sorbello, G. I. M. Migliore, V. Conti, and S. Vitabile, "A novel technique for fingerprint classification based on fuzzy C-means and naive bayes classifier," *Proc. - 2014 8th Int. Conf. Complex, Intell. Softw. Intensive Syst. CISIS 2014*, vol. 169, no. 7, pp.

155–161, 2014, doi: 10.1109/CISIS.2014.23.

[78] R. Merkel, J. Dittmann, and C. Vielhauer, "A First Public Research Collection of High-Resolution Latent Fingerprint Time Series for Short- and Long-Term Print Age Estimation," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 10, pp. 2276–2291, 2017, doi: 10.1109/TIFS.2017.2705622.

[79] P. P., J. Sheetlani, and R. Pardeshi, "Fingerprint based Automatic Human Gender Identification," *Int. J. Comput. Appl.*, vol. 170, no. 7, pp. 1–4, 2017, doi: 10.5120/ijca2017914910.

[80] S. S. Gornale, M. Basavanna, and R. Kruti, "Fingerprint based gender classification using local binary pattern," *Int. J. Comput. Intell. Res. ISSN*, pp. 973–1873, 2017.

[81] N. B. Hills, R. Sahu, and N. B. Hills, "A Survey: Gender Classification Based on Fingerprint," vol. 117, no. 20, pp. 985–992, 2017.

[82] S. Suwarno and P. I. Santosa, "Short Review of Gender Classification based on Fingerprint using Wavelet Transform," *IJACSA*, vol. 8, no. 11, 2017.

[83] A. S. Falohun, O. . Fenwa, and F. . Ajala, "A Fingerprint-based Age and Gender Detector System using Fingerprint Pattern Analysis," *Int. J. Comput. Appl.*, vol. 136, no. 4, pp. 43–48, 2016.

[84] S. Gornale, A. Patil, and V. C., "Fingerprint based Gender Identification using Discrete Wavelet Transform and Gabor Filters," *Int. J. Comput. Appl.*, vol. 152, no. 4, pp. 34–37, 2016, doi: 10.5120/ijca2016911794.

[85] M. K. Shinde and S. A. Annadate, "Analysis of Fingerprint Image for Gender Classification or Identification: Using Wavelet Transform and Singular Value Decomposition," in *2015 International Conference on Computing Communication Control and Automation*, 2015, pp. 650–654, doi: 10.1109/ICCUBEA.2015.133.

[86] O. Sharifi and M. Eskandari, "Optimal face-iris multimodal fusion scheme," *Symmetry (Basel).*, vol. 8, no. 6, pp. 1–16, 2016, doi: 10.3390/sym8060048.