الجمهورية الجزائرية الديمقراطية الشعبية
**People's Democratic Republic of Algeria**
وزارة التعليم العالي والبحث العلمي
**Ministry of Higher Education and Scientific Research**

جامعة غرداية
**University of Ghardaia**

Registration n°:
...../...../...../...../.....

كلية العلوم و التكنولوجيا
**Faculty of Science and Technology**
قسم الرياضيات والإعلام الآلي
**Department of Mathematics and Computer Science**

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

## Master

**Domain:** Mathematics and Computer Science
**Field:** Computer Science
**Specialty:** Intelligent Systems for Knowledge Extraction

## Topic

---

# Human ear print recognition using deep learning techniques

---

## Presented by:

*GUERRADI Dallal* & *HADJ KOUIDER Afrah*

**Publicly defended on 07 02, 2025**

## Jury members:

| | | | |
|---|---|---|---|
| Mr. Sliman Oulad-Naoui | MCA | Univ. Ghardaia | President |
| Mr. Youcef Mahdjoub | MCB | Univ. Ghardaia | Examiner |
| Mr. Messaoud Ben-Guenane | MCB | Univ. Ghardaia | Supervisor |

**Academic Year:** 2024/2025

# Acknowledgments


**And say, "﴿ وَقُلْ رَبِّ زِدْنِي عِلْمًا ﴾"**

**Surah Ta-Ha, verse 114**

We thank God with good and blessed praise, who has guided and aided us in completing this achievement. Without Him, nothing is done or created.

To our generous parents, the source of our love and prayers: we have reached this point today only thanks to your unwavering support, your prayers, and your compassion.

We would like to warmly thank our supervisor, Dr. Ben Guenane Messaoud, and Professor Aiadi Oussama for their invaluable guidance and advice.

We also express our sincere gratitude and respect to the members of the jury for agreeing to evaluate our work.

To everyone who has supported us throughout our academic and scholarly journey, including our family, friends, and teachers, our heartfelt thanks and prayers go to all of you.


# *Dedication*

**To my dear mother:**
I cannot find the appropriate way to express my gratitude to you on this blessed day, a day of joy maded thanks to God. You who know me more than anything, you who accepted everything for me. Mom, this work is the result of your good advice and your infinite devotion.
May God bless you and keep you by my side.

**To my beloved father:**
It was a divine blessing for me to have you as a father. You accepted everything to make me who I am today. You instilled in me the skills of living, being, and doing.
You enrolled me in school and ensured my success every day. Dear father, I lack words to express my gratitude to you.
May God keep you among us as long as possible and bless you.

**To my precious brothers and sisters:**
Abderaouf, Omar, Arwa, Ritadj, and Taha, whom I love very much. You have been a great help to me during this long process. This work is also yours.

To my beautiful support in all aspects of this life, my uncle **Hamza** and my aunt **Zineb**.

To my entire extended family who supported me morally and financially, especially my uncles **Ali and Saeed,** and my aunts **Fatima, Yamina, and Milouda.**

To my companion in completing this memoir and my wonderful friend **Afrah**, I wish you a successful life.

To my closest friend, the sister my mother never gave me, **Jihad**, I wish you a life full of faith and happiness.

To all my family and close friends, all in his name.

**"Guerradi Dallal"**

# Dedication

**To my dearest mother and my loving father**

From the day I was born until this very moment, you have been my guiding light, my strength, and my sanctuary. In deep appreciation for your unwavering support, countless sacrifices, and every effort you made to see me succeed. Your prayers and love carried me through every challenge. You are my love, my pride, and my role models in this life thank you. May God protect you always; without you, I am nothing.

To my beloved brothers, **Khalil** and **Chouaib**, and to my cousins **Abderrahmane** and **Adnane**.

To my dear sisters, **Fatima** and **Sabrina**, and their wonderful children: **Ines, Tesnim, Mohamed Elsaleh, Amira,** and **Souhaib**, and their husbands, **Tayeb** and **Yassine.**

To all my cousins, including my dear cousin **Khadejia**.

You supported me with your encouragement and sacrifices, thank you.

**To my family**—my aunt **Bakhta**, my aunt **Souad**, and all my aunts and uncles,your love and support mean the world to me.

To my notebook companion and close friend, **Dalal**, thank you.

To my childhood companions, **Aisha** and **Wassima**, my soul sisters and eternal friends.

To my dear friends, **Amina** and **Darine**, who stood by me throughout my university years.

To all my friends from university, including **Ahlem.**

To everyone near or far whose support, encouragement, and belief in me lifted me along the way—I am truly grateful.

This work is for you, too.

***“ Hadj Kouider Afrah “***

## ملخص

برزت تقنية القياسات الحيوية للأذن كمنهجية واعدة لتحديد هوية الإنسان، نظراً لتميز الأذن وثباتها وتنوعها المحدود بمرور الوقت. تركز هذه الرسالة على تحسين أداء أنظمة التعرف على الأذن باستخدام تقنيات التعلم العميق، لا سيما من خلال الاستفادة من الشبكات العصبية التلافيفية. يتضمن النهج المقترح تنفيذ وضبط بنية ResNet50 باستخدام مجموعة بيانات AMI ،Ear التي تحتوي على صور أذن ملتقطة في ظروف ووضعيات متنوعة. تُطبق أساليب المعالجة المسبقة للبيانات والزيادة لتعزيز متانة النموذج، ويُجرى التدريب مع ضبط دقيق للتكيف مع الخصائص المحدد لصور الأذن. وتوضح النتائج التجريبية أن النموذج المختار القائم على ResNet50 يحقق أداءً محسناً للتعرف، حيث يصل إلى دقة 99.29 % متفوقاً على الطرق السابقة. يُظهر هذا العمل إمكانات نماذج التلافيفية العميقة في القياسات الحيوية للأذن، ويُمهّد الطريق لتطبيقات مستقبلية للدمج متعدد المشاهد، ومجموعات البيانات الضخمة، والوسائط المدمجة مع سمات القياسات الحيوية الأخرى.



**كلمات مفتاحية:** قياسات الأذن الحيوية، التعلم العميق، الشبكات العصبية التلافيفية ، ResNet50 ، مجموعة بيانات الأذن AMI ، التعرف على القياسات الحيوية , تحديد الهوية البشرية.

**Abstract**

Ear biometrics has risen as a promising methodology for human identification, due to the ear's uniqueness, permanence, and constrained variability over time. This thesis centers on enhancing ear recognition systems utilizing deep learning techniques, particularly by leveraging convolutional neural networks. The proposed approach includes implementing and fine-tuning the ResNet50 architecture utilizing the AMI Ear dataset, which contains ear images captured under diverse conditions and postures. Data preprocessing and augmentation methods are applied to enhance model robustness, and training is conducted with careful tuning to adapt to the specific characteristics of ear pictures. The experimental results demonstrate that the proposed ResNet50-based model achieves improved recognition performance, reaching an accuracy of 99.29%, outperforming previous methods. This work demonstrates the potential for deep convolutional models in ear biometrics and sets the stage for future applications of multi-view fusion, large datasets, and combined modalities with other biometric traits.

**Keywords:** Ear Biometrics, Deep Learning, Convolutional Neural Networks(CNNs), ResNet50, AMI Ear Dataset, Biometric Recognition, Human Identification.

**Résumé**

La biométrie auriculaire s'est imposée comme une méthodologie prometteuse pour l'identification humaine, en raison de son caractère unique, de sa permanence et de sa variabilité limitée dans le temps. Cette thèse se concentre sur l'amélioration des systèmes de reconnaissance auriculaire grâce à des techniques d'apprentissage profond, notamment en exploitant les réseaux de neurones convolutifs. L'approche proposée comprend la mise en œuvre et le perfectionnement de l'architecture ResNet50 à l'aide du jeu de données AMI Ear, qui contient des images d'oreilles capturées dans diverses conditions et postures. Des méthodes de prétraitement et d'augmentation des données sont appliquées pour améliorer la robustesse du modèle, et l'apprentissage est réalisé avec un réglage précis pour s'adapter aux caractéristiques spécifiques des images d'oreilles. Les résultats expérimentaux démontrent que le modèle choisi basé sur ResNet50 atteint des performances de reconnaissance améliorées, atteignant une précision de 99,29%, surpassant les méthodes précédentes. Ces travaux démontrent le potentiel des modèles convolutifs profonds en biométrie auriculaire et ouvrent la voie à de futures applications de fusion multi-vues, de grands ensembles de données et de modalités combinées avec d'autres caractéristiques biométriques.

**Mots clés:** Biométrie auriculaire, apprentissage profond, réseaux de neurones convolutifs, ResNet50, jeu de données auriculaires AMI, reconnaissance biométrique, Identification humaine.

# Contents

# List of Figures

# List of Tables

# List of Acronyms

| | |
|---|---|
| **AdamW** | Adaptive Moment Estimation with Weight |
| **AI** | Artificial Intelligence |
| **AMI** | AMI Ear Dataset |
| **ANN** | Artificial neural network |
| **AWE** | Annotated Web Ears |
| **AWI** | Annotated Web Images |
| **AMP** | Automatic Mixed Precision |
| **CNN** | Convolutional Neural Network |
| **CTIM** | Centre des Technologies de l'Information et de la Modélisation |
| **DenseNet** | Densely Connected Network |
| **DL** | Deep Learning |
| **DNA** | DeoxyriboNucleic Acid |
| **FC** | Fully Connected (layer) |
| **GPU** | Graphics Processing Unit |
| **GUI** | Graphical User Interface |
| **IIT Delhi** | Indian Institute of Technology, Delhi |
| **KNN** | K-Nearest Neighbors |
| **LBP** | Local Binary Patterns |
| **ML** | Machine Learning |
| **MLPs** | Multilayer Perceptrons |
| **OneCycleLR** | One Cycle Learning Rate Scheduler |
| **PCA** | Principal Component Analysis |
| **ReLU** | Rectified Linear Unit |

| | |
|---|---|
| **ResNet** | Residual Network |
| **SILU** | Sigmoid Linear Unit |
| **Softmax** | Soft Maximum (activation function) |
| **SVM** | Support Vector Machine |
| **TL** | Transfer Learning |
| **VGG** | Visual Geometry Group Network |

# Introduction

Biometrics play a significant part in cybersecurity by verifying identity through one-of-a-kind physical and behavioral characteristics, offering one of the most precise and promising verification solutions. These systems are increasingly replacing conventional strategies, such as smart cards and keys, especially in high-security situations. In any case, no single biometric method is generally successful, as each has its qualities and limitations. Common strategies incorporate fingerprint recognition, facial recognition, iris scanning, voice recognition, and Deoxyribonucleic Acid (DNA) investigation.

In 1890, Alphonse Bertillon recognized the potential of the human ear for identification. Even though it has gotten little attention compared to other biometric strategies, the ear is remarkably stable and distinctive, its measurements remaining steady from birth until roughly the age of eight, with gradual enlargement occurring later in life. Despite challenges such as hair, jewelry, and changing image angles, the stability of the ear's position and structure makes it more reliable than fingerprints, requiring no user intervention. Its larger size also simplifies the identification process. However, deep learning models still have trouble with real-world variables, making it difficult to generalize across datasets and reducing the accuracy of earprint recognition.

The motivation for this research stems from the increasing demand for secure, non-intrusive, and reliable biometric systems - especially post-pandemic - There is a pressing need to address the limitations of current systems; ear biometrics holds incredible potential for applications in security, surveillance, and scientific examinations due to its capacity to capture pictures remotely. Using advanced neural networks (CNNs) and improving feature extraction techniques, this research aims to contribute to the growing field of biometric authentication and enhance the applicability of ear biometrics in the security-related domain.

Although biometric recognition systems have made significant progress, the field of ear biometrics remains relatively underdeveloped. Most existing approaches to ear print recognition still face challenges related to variations in lighting, head orientation, and image quality, which pose an obstacle to a good accuracy. Therefore, the central problem addressed in this thesis is: How can the recognition accuracy and reliability of ear print biometrics be enhanced using advanced deep learning techniques such as convolutional neural networks (CNNs)?

This thesis offers a few key contributions to the field of ear biometric recognition. First, it proposes an optimized, deep learning-based system for earprint recognition, focusing on improving model robustness and accuracy. It explores advanced feature extraction and data augmentation techniques to enhance recognition performance within the AMI ear dataset. We suggest a system for ear print recognition that is based on supervised machine learning, especially deep learning. By utilizing neural networks with multiple layers, deep learning models can extract complex features from visual data, leading to remarkable accuracy in various domains.

We chose to articulate our study around four main chapters :

1. The first chapter is dedicated to the work background, we will see an overview, definition and importance of biometrics in security and forensics, common biometric traits (fingerprints, iris, face, ear, etc.), advantages of ear prints over other biometric traits, stability and uniqueness of ear prints, some Key deep learning models for image recognition and applications of deep learning in ear recognition systems.

2. The second chapter is specified for the State of the art, it contains the advancement of techniques utilized in ear recognition, discusses their advantages and limitations, and reviews the current research within this context.

3. The third chapter, which is called "System Design and Methodology", contains the scope of the ear print recognition system, an overview of the Dataset Collection and Preprocessing, also talks about Model Selection, System Architecture, and the Implementation Framework.

4. The fourth and final chapter, "Results and Evaluation", will present the training and optimization of the model, performance evaluation, and finally testing and validation.

# Chapter 1

# Background

## 1.1 Introduction

Biometric systems are automated techniques for identifying or authenticating people based on personal characteristics that are directly related to who they are. A system gathers each person's unique biometric traits, and then directly links them to confirm or identify the person. What reaffirmed the significance of biometrics in modern society is the necessity for large identity management systems that function depends on the precise identification of an individual in the context of several applications. We begin this chapter by introducing some fundamental concepts and terms related to biometrics and the ear recognition among the other biometric techniques. Next, we see a comparative advantages of ear prints in biometric systems, stability and uniqueness of ear prints and its accuracy. Finally, the chapter focuses on Deep learning in Biometric Recognition and the presentation of Convolutional Neural Networks (CNNs) and their relevance to ear print recognition.

## 1.2 Biometrics

### 1.2.1 Definition

Biometrics is generally understood as the science that deals with identifying people based on their physical or behavioral traits. When it comes to biometric recognition, it refers more specifically to the use of technology to automatically recognize individuals by analyzing distinctive features that can be measured and repeated reliably.

Biometric features are typically classified into two broad categories: Physiological biometrics and behavioral biometrics [25].

1. **Physiological Biometrics**
   These characteristics are taken from the physical characteristics of the human body. They can be further divided into:

   - *Morphological features:* They include fingerprints, hand geometry, face appearance, finger vein patterns, and iris or retina patterns.

- *Biological attributes:* Attributes like DNA, blood type, or other molecular attributes are included in this category.

2. **Behavioral Biometrics**
These traits are related to human behavior and activity patterns. Some common examples include voice recognition, signature dynamics, typing rhythm (keystroke dynamics), and gait analysis.

However, the different sorts of measurements do not all have the same level of reliability. Physiological measurements usually offer the benefit of remaining more stable throughout an individual's life. For example, they are not subject to stress, in contrast to identification by behavioral measurement.

Figure 1.1 represents the different types of biometric features, including physiological and behavioral traits, commonly used in recognition systems.

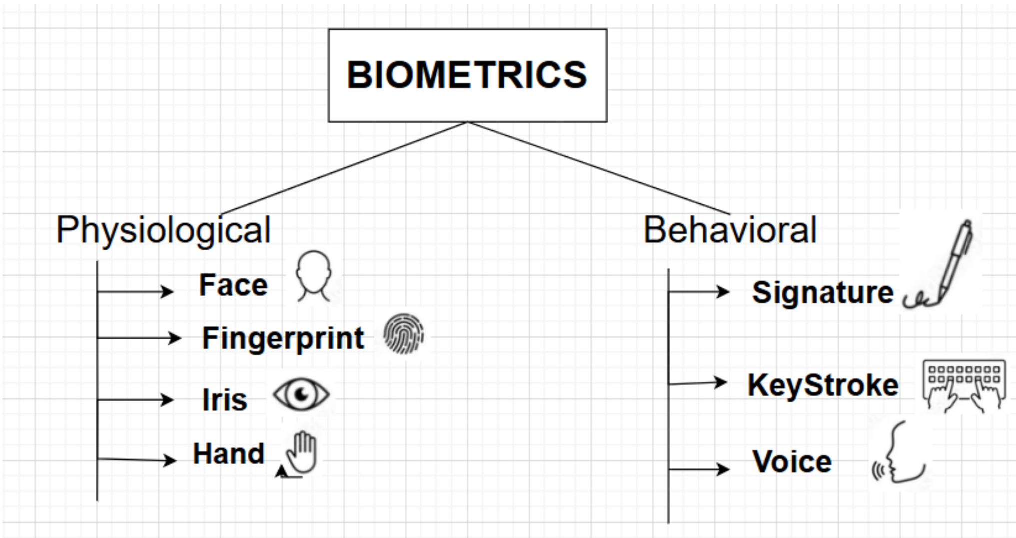


Figure 1.1: Types of Biometrics

### 1.2.2 History of biometrics

The Chinese emperor Ts'In She was already using fingerprints to verify certain seals in the second century B.C. In 1858, British administrator William James Herschel utilized fingerprints for the first time in a business context in India. He had his subcontractors sign contracts with their fingers after being assigned to build roads in Bengal. Bertillon, a French police officer, developed scientific policing at the close of the 1800s. In order to identify persistent criminals, he employed physical measures of particular anatomical traits, which often worked [53].

- This procedure was started in 1888 by the French police in Paris (préfecture de police) through their Forensic Identification Unit (anthropometry and mug photo). In 1894, four prints were established, and in 1904, ten prints were added.

- In 1901, the Metropolitan Police in the United Kingdom began utilizing biometrics for identification.

- It was started in the United States by the FBI in 1924 and the New York police in 1902.

The measurement of unique patterns (behavioral biometrics) is not a new concept and it goes back to the 1860s. Telegraph operators using Morse code could recognize each other by sending dash and dot signals. This technique was later employed by Allied troops during World War II to identify senders and authentication messages they received. This process of identifying an individual based on specific characteristics forms the fundamental principle of biometric systems [53].

### 1.2.3 Operational Modes of a Biometric System

There are three fundamental modes of operation used by biometric systems to both establish and authenticate individual identities [42]:

- Enrollment: The first phase entails the acquisition of a person's biometric features, for instance, ear images, which are processed to form a reference template to be kept in the system database.
- Verification: In this mode, the system compares freshly acquired biometric data to the stored template in order to confirm the claimed identity of an individual on a one-to-one match basis.
- Identification: In this case, the system compares the acquired biometric information with a number of templates in the database to verify the subject individual's identity without a prior assertion, on a one-to-many matching basis.

### 1.2.4 Working Principle of Biometric Systems

A biometric system uses an individual's collected biometric data, from which a special algorithm extracts characteristics to create a biometric template. After that, the system can use the biometric database to confirm the user's identification. It can compare hundreds of millions of biometric data in the database in only one second [24] [23].

Here's a simple breakdown of how biometric systems function:

- Capture - Biometric software, such as face recognition, captures the biological input provided by a user (like a face scan) – usually by prompting them to take a selfie [24].
- Template creation - The software measures the captured input to create a baseline data template. The biometric template is a digital representation of the distinctive characteristics extracted from a person's biometric data (such as facial features, voice, or iris patterns).
- Data storage – The biometric template is not a raw image or recording but a mathematical model that encodes the essential data points needed

for comparison and verification. The template stored either on the device's internal hardware or in a secure cloud platform.

- Matching - During subsequent use, the new biometric template is compared to the stored template. If the data matches, access is granted. Otherwise, access is denied [23].

figure1.2 below represents the biometric recognition process, showing how input data is enrolled, stored, and later matched during authentication.

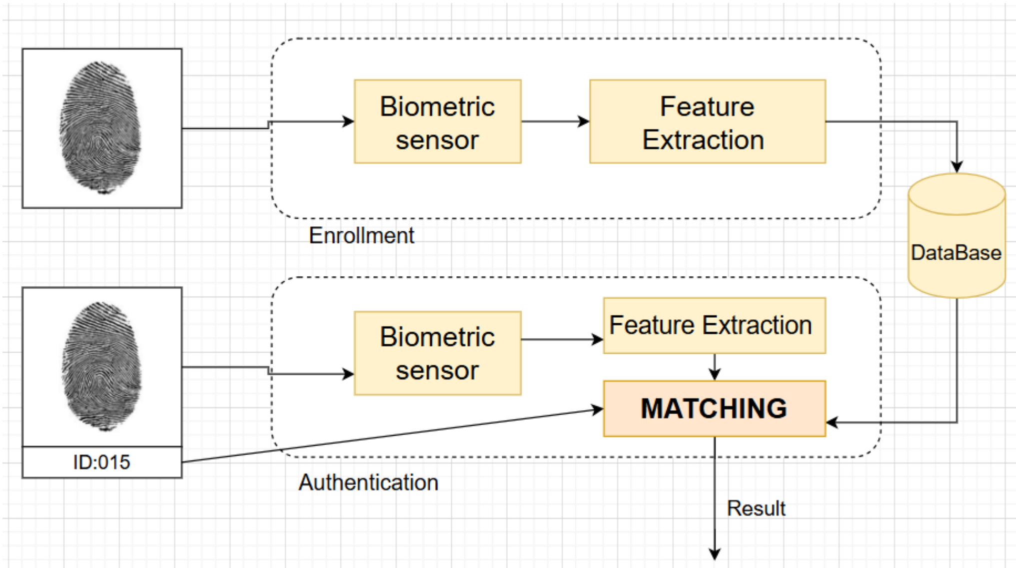


Figure 1.2: General Architecture of a Biometric System

### 1.2.5 Key Characteristics of Biometric Traits

Biometric systems depend on some inherent characteristics of human traits to be able to work effectively. These are:

- **Universality** – This quality must be found in all people.
- **Uniqueness** – There cannot be two individuals with the identical characteristic.
- **Permanence** – This attribute should stay relatively stable over time.
- **Measurability** – The attribute should be measurable in a manner such that it can be recognized [55].

Other real-world implications include being able to measure the trait in a reliable fashion, and to resist spoofing or forgery.

### 1.2.6 Advantages and Disadvantages of Biometric authentication

Biometric authentication presents both strengths and limitations, which are outlined below.

**Advantages of Biometric Authentication**

- **Enhanced security and reliability:** Biometric systems are more secure. due to their unique physiological or behavioral traits, which results in challenging for identity credential forgery or theft.
- **Efficiency of user experience:** As opposed to the complexity involved within biometric systems, the intrinsic nature of the algorithms used generally favors an efficient and easily accessible user interface.
- **Non-transferable:** The biometric features, being unique in nature, cannot be shared or transferred like passwords or ID cards.
- **Scalable systems:** Biometric systems can be applied to a vast array of applications and populations, and they have been found to be extremely effective when it comes to scalability, particularly when implemented in networked or cloud setups [23].

**Weaknesses of Biometric Authentication**

- **High implementation cost:** The process of implementing biometric infrastructure, like sensors, software, and interfaces to install infrastructure, can be expensive.
- **Privacy Concerns:** The acquisition and storage of biometric data bring up severe ethical and legal questions regarding user privacy, especially when the data are stored in a centralized database.
- **Susceptibility to bias:** Regardless of the natural objectivity of biometric systems, mistakes may arise due to anomalous sets of data, defective sensors, or design issues in algorithmic structures, and this can produce inconsistent performance on different demographic populations [23].

### 1.2.7 Applications of Biometrics

Today, there are a huge number of applications and services that utilize biometric technology. Here are some common examples of how people interact with physiological and behavioral biometrics in their daily lives:

- Personal hardware: Mobile phones, laptops, PCs, and tablets often enable fingerprint or facial recognition to unlock the device [24].
- Financial transactions: Payments like wire transfers frequently require identity verification through biometrics and/or cloud-based biometrics for secure access [8].
- Healthcare: Biometric authentication helps healthcare providers manage patient records securely and prevent unauthorized access to sensitive information [35].
- Airports: Many modern airports use facial recognition to expedite passenger processing. Travelers can enroll by having a photo of their eyes and face captured, allowing faster movement through queues [23].
- Entertainment venues: Stadiums and other venues are beginning to offer ticketless access using face biometrics [32].
- Secured physical access: Biometrics are replacing key cards and PIN entries as a more secure and traceable way to authorize access to secured buildings or areas within buildings [24].

## 1.3 Ear Prints in Biometric Systems

Ears are often overlooked in everyday descriptions; unlike faces, ears don't usually get much attention in everyday descriptions, especially compared to faces, which are far easier to picture and describe in detail [11]. But their unique structure actually makes them a great option for biometric recognition, particularly in certain situations. One of the advantages of the ear is that it is non-surgical, as it can be taken without touching, i.e., without physical contact, unlike other biometric methods that are usually done by touch [26]. This makes the process not only more hygienic but also more user-friendly, since there's no need for direct interaction.

It is particularly beneficial in public or high-traffic environments like airports, hospitals, or workplaces, as well as in covert surveillance scenarios where capturing biometric data without the subject's knowledge is crucial [57].

Another advantage of the ear is that the ear's features remain consistent and remarkably stable over time, i.e., they do not change, unlike facial features or fingerprints, which may be affected by several reasons, including aging, surgical injuries, or environmental factors [21]. Since the ear changes very little over time, it stays a dependable choice for long-term identification. On top of that, ear recognition helps address some of the limitations of other biometric systems. For example, facial recognition often has trouble when people wear masks—an issue that became glaringly obvious during the COVID-19 pandemic. Iris recognition, on the other hand, requires specialized equipment and close-range scanning, which isn't always practical [43]. Thanks to its accuracy, ease of use, and non-intrusive nature, ear recognition proves to be a versatile and efficient solution for various applications, from security systems to forensic investigations.

### 1.3.1 Stability and Uniqueness of Ear Prints

#### 1. Stability of Ear Prints

Ear prints are highly stable over time, making them well-suited for biometric recognition. The concept dates back to 1890 with Alphonse Bertillon, and was later advanced by A. Iannarelli in 1989, who analyzed 10,000 ears and confirmed their uniqueness—even among identical twins [11]. Although slight age-related changes may occur, particularly in the lobule due to gravity, the overall ear structure remains consistent, supporting their use in long-term identification systems such as national IDs and forensic databases [2].

According to Figure 1.3, the human ear is presented with a combination of distinct features commonly used in biometric recognition.

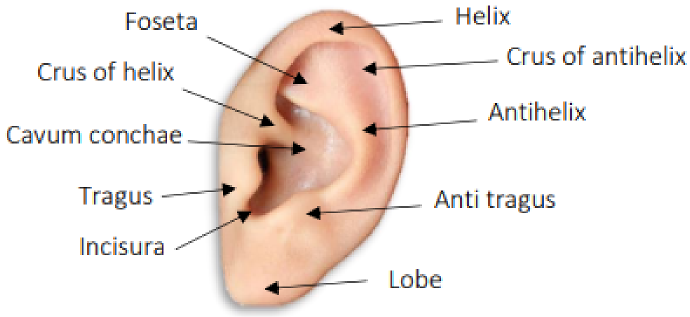


Figure 1.3: Anatomy of human ear

## 2. Uniqueness of Ear Prints

What really makes ear prints stand out is how unique they are. Every person's ear has its own combination of features The outer ear is defined by distinctive features such as the helix, lobe, antihelix, incisura, antitragus, cavum conchae, foseta, crus of the helix, crus of the antihelix, and tragus, which contribute to its uniqueness. These parts form patterns that are completely individual. So, the probability of two people having the same earprint is incredibly low, making ear recognition a very accurate way to ID people [26].

Ear prints are really hard to fake. Unlike fingerprints, which can be lifted and copied, or facial features, which can sometimes be recreated using fancy technology (like 3D modeling or masks), ear prints are much harder to duplicate. This makes ear recognition a strong choice for security purposes, whether it's for access control or forensic identification [41].

Three distinct external ear anatomical features—the helix, antihelix, and lobule—play a crucial role in biometric identification systems due to their individual variability, which aids in distinguishing people.

- The helix, forming the outer rim of the ear, varies in size and shape from person to person. Because of its unique shape and curvature, the ear print is unique to the individual.
- The antihelix is the ridge located on the inner part of the ear. It is directly in front of the helix and parallel to it, and its shape and clarity may vary greatly from one person to another, making it a key factor in the distinctiveness of each ear's structure.

- The lobule, or earlobe—the soft, fleshy part at the bottom of the ear. It comes in all sorts of sizes and shapes, and the way it attaches to the face can vary too (the pendulous lobule elongates under the influence of gravity). These differences make the lobule another key feature that helps set ears apart in biometric recognition systems.

### 1.3.2 Accuracy of Ear Print Recognition

#### Accuracy comparisons with other biometrics

In the previous section, we mentioned several biometrics and, in table 1.1, we show the comparison between them. Biometric systems are divided into two sections: physiological and behavioral. Physiological biometrics are DNA, face, ear, iris, fingerprint, etc [11]. While behavioral biometrics, which are signatures, gait patterns, etc., voice is a combination of biometrics and physiological [21].

Many systems have been developed to distinguish between biometrics, and they are widely used in many applications, such as criminal investigations and security systems [21].

Table 1.1 provides a comparative analysis that includes several important factors, including distinctiveness, permanence, performance, and acceptance, in order to better understand how ear biometrics compare to other biometrics :

Table 1.1: Comparison of Biometric Identifiers
[9]

| **Biometric Identifier** | **Biometric Type** | **Distinct-iveness** | **Perma-nence** | **Perfor-mance** | **Accept-ability** |
|---|---|---|---|---|---|
| **DNA** | Physiological | High | High | High | Low |
| **Ear** | Physiological | Medium | High | Medium | High |
| **Face** | Physiological | Low | Medium | Low | High |
| **Fingerprint** | Physiological | High | High | High | Medium |
| **Iris** | Physiological | High | High | High | Low |
| **Palm print** | Physiological | High | High | High | Medium |
| **Signature** | Behavioural | Low | Low | Low | High |
| **Voice** | Physiological and Behavioural | Low | Low | Low | High |

From the table, it's clear that ear biometrics offer a strong advantage in terms of permanence — a key factor for reliable long-term identification [2].

While they may not be as inherently distinctive as fingerprints or DNA, their high acceptability — being non-intrusive and easy to use — makes them appealing [44]. Unlike iris or DNA recognition, which often require specialized equipment, ear recognition can be performed using common imaging technologies, making it a practical choice for various applications [2].

Although the current performance is considered moderate, ongoing advancements in deep learning and feature extraction are steadily enhancing the accuracy and reliability of ear recognition systems [9].

Studies have shown that ear print recognition typically achieves moderate accuracy, performing better than face recognition in terms of permanence but lower than fingerprints and iris recognition in terms of distinctiveness [44]. Machine learning and deep learning techniques, such as Convolutional Neural Networks (CNNs), have significantly improved ear recognition accuracy, reducing error rates and making it more viable for security and forensic applications [9].

**Challenges in Ear Print Recognition**

Images in the real world suffer from different poses, lighting, background clutter, hair coverage, and occlusion of ear accessories. It is evident that these environmental factors have a significant impact on the images and make it extremely challenging to detect and recognize ears [26][44]. Furthermore, shadows can cover other areas of the ear, which is called occlusion [44]. and this can distort the detection of edges by fragmenting edges, making the process of identifying them more difficult and complex [57]. The accuracy and dependability of earprint recognition systems have significantly increased as a result of experts' introduction of increasingly complex algorithms and improved imaging techniques to address these issues [3].

## 1.4 Problem Definition and Research Objectives

In this work, we address the challenge of ear biometric recognition by leveraging deep learning techniques to improve accuracy and robustness. While ear biometrics offer key advantages such as non-intrusiveness, permanence, and ease of acquisition, they remain underutilized compared to other modalities like fingerprint or iris recognition. Traditional methods often rely on handcrafted features, which may fail to capture the full variability of ear structures. To overcome these limitations, this study proposes the evaluation of multiple pre-trained convolutional neural networks including ”ResNet50, EfficientNet, VGG16, AlexNet, DenseNet121” on the AMI Ear Dataset. The objective is to compare their performance in terms of recognition accuracy and determine the most suitable model for developing an effective ear recognition system.

## 1.5 Deep learning Approaches in Biometric Recognition

### Machine learning

Machine Learning was first defined by Arthur Samuel in 1959 in his article “Some Studies in Machine Learning Using the Game of Checkers” as: “The field of study that gives computers the ability to learn without being explicitly programmed” [48].

A more formal definition was provided by Tom M. Mitchell in his 1997 book “Machine Learning”: “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E” [33].

### Deep learning

Deep Learning is defined by Ian Goodfellow, Yoshua Bengio, and Aaron Courville in their 2016 book “Deep Learning” as:“Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction” [15].

These methods have significantly advanced the state of the art in fields such as speech recognition, visual object recognition, object detection, drug discovery, and genomics. Deep learning models uncover complex structures in large datasets by using the backpropagation algorithm, which adjusts internal parameters layer by layer to improve the learned representations [31].

### 1.5.1 Deep feedforward networks

Profound feedforward systems, also known as multilayer perceptrons (MLPs), are a sort of fake neural organize (ANN) that comprises different layers of associated neurons, or nodes. These systems are really propelled by the way our brain works. Within the brain, neurons are associated through neural connections, and they send electrical signals to each other to prepare data. Essentially, in an ANN, data streams in one direction from the input layer to the output layer, without any criticism circles, mirroring how signals move between neurons within the brain. The structure of a feedforward arrange depends on components just as the number of layers, how numerous of neurons each layer has, and how the neurons are connected. An ordinary neural network has an input layer, one or more hidden layers, and an output layer. The crude input information comes into the input layer, gets prepared by the covered-up layers, and inevitably produces a yield [15] (as shown in Figure 1.4).

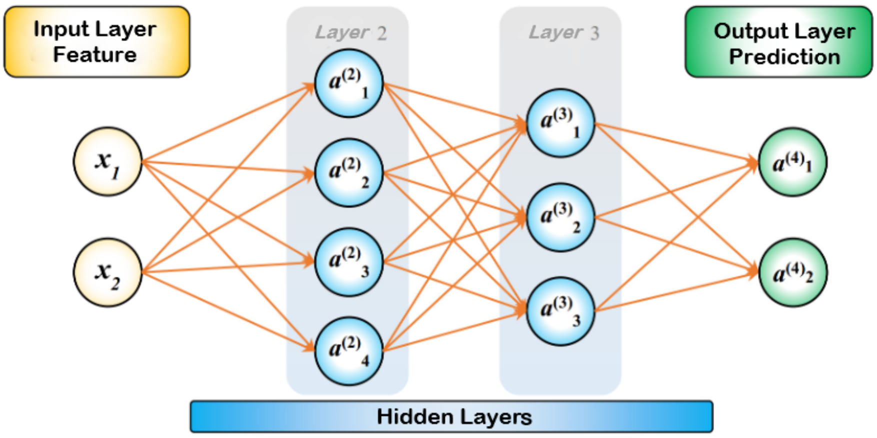


Figure 1.4: The Structure of Multi-layer Perceptron.

The basic building component of a feedforward network is a neuron, which receives input from other neurons or external sources, conducts a computation on this input and provides an output. Until the output layer generates the final output, the output of one neuron is then sent as input to the subsequent neuron in the network.

The following figure 1.5 illustrates how an artificial neuron processes inputs by applying weights and an activation function to produce an output.

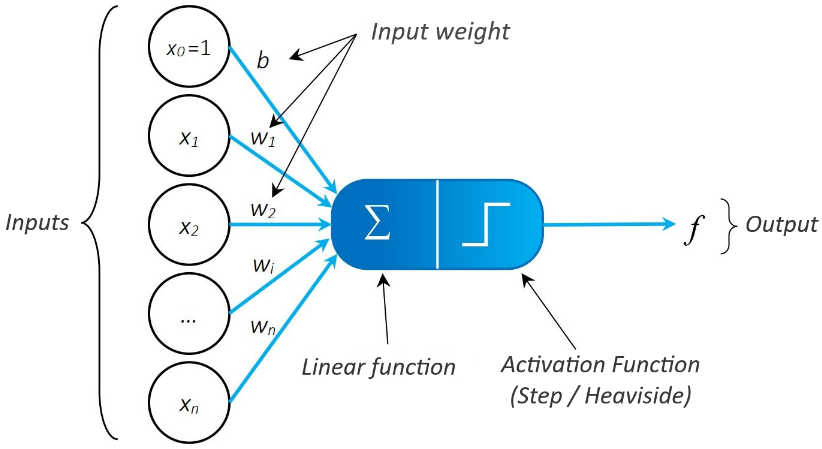


Figure 1.5: The structure of the artificial neuron

Each neuron in a feedforward network is typically characterized by an activation function, which maps the input to the output.

Common activation functions include the sigmoid function, the Sigmoid Linear Unit (SiLU) function, the rectified linear unit (ReLU) function, and the hyperbolic tangent (tanh) function [27].

$$\mathrm{ReLU}(x) = \max(0, x) \tag{1.1}$$

- **x**: The input to the activation function, which can be any real number.
- $\max(0, \mathbf{x})$: Outputs 0 if $\mathbf{x} < 0$, otherwise returns **x**. It introduces non-linearity and is commonly used in neural networks due to its simplicity and efficiency.

$$\mathrm{SiLU}(x) = x \cdot \sigma(x) = \frac{x}{1 + e^{-x}} \tag{1.2}$$

- **x**: The input to the activation function.
- $\sigma(x)$: The sigmoid function applied to **x**, defined in (1.3).
- The function smoothly increases and combines properties of both ReLU and sigmoid.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1.3}$$

- **x**: Input to the sigmoid function.
- $e$: Euler's number, approximately equal to 2.718.
- This function maps any real-valued input to a range between 0 and 1, making it suitable for binary classification.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{1.4}$$

- **x**: Input to the activation function.
- $e^x$ and $e^{-x}$: Exponential functions, where $e$ is Euler's number.
- This function outputs values in the range $[-1,1]$, centered at 0, which helps in reducing bias in activations.

### 1.5.2 Convolutional Neural Network

Convolutional Neural Networks (CNNs) are motivated by the mammalian visual cortex and are broadly utilized in computer vision. The center operation in CNNs is convolution, where parts prepare the input, and a nonlinear activation function makes an outline. Not at all like flag handling, these parts are learned amid preparation. CNNs utilize weight sharing and sparse associations, requiring fewer parameters than completely associated systems. Each unit in a convolutional layer interfaces to a small locale of the previous layer, known as the receptive field. Different convolutional layers are stacked to permit higher-level layers to memorize more global features [50].

The processing flow of a Convolutional Neural Network (CNN) is illustrated in Figure 1.6.

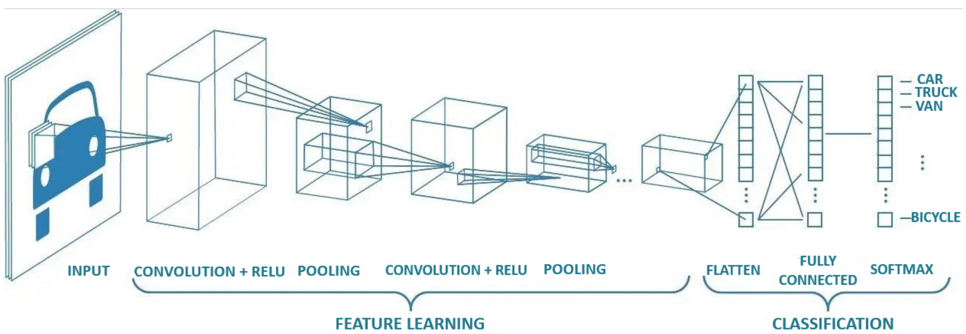


Figure 1.6: Example of a CNN architecture

## 1.6 Conclusion

In this chapter, we centered on the concept of biometric acknowledgment and its fundamental part in present-day innovation, emphasizing its significance in a few areas, as the ear has risen as a promising choice compared to other biometric features, much appreciated for its steadiness over time, the risk of fraud, and ease of recognition.

We also revealed the challenges facing ear recognition systems, including a lack of data, similarity between ears, and the impact of some environmental conditions on classification accuracy. We also reviewed the role of deep learning, especially convolutional neural networks (CNNs), due to their superior ability to extract automatic features and deal with complex data, and their impact on performance and evaluation.

With the continuous progress in artificial intelligence (AI) technologies, the ear may become one of the main features in future biometric verification systems, which opens the way for further development in this field. In the next chapters, we will discuss the methodology followed in this research, focusing on addressing the identified gaps, and we will propose an effective approach to enhance the accuracy and efficiency of the earprint recognition system.

# Chapter 2

# State of the Art

## 2.1 Introduction

In recent years, biometric identification has picked up impressive interest, with ear recognition acknowledgment standing out as a promising strategy due to the distinctive and stable anatomical features of the human ear. This chapter presents an in-depth review of some existing work within the field of ear recognition, covering both early approaches based on geometric and handcrafted features, and more recent strategies that use machine learning and deep learning.

The objective of this chapter is to follow the advancement of techniques utilized in ear recognition, discuss their advantages and limitations, and review the current research within this context. In addition to looking over the literature, this chapter also outlines the evaluation strategies utilized in our work. Particularly, we present the performance metrics utilized, such as accuracy, recall, precision, and F1-score, to assess the effectiveness and reliability of our proposed approach. These evaluations give a quantitative basis for comparison with previous methods and help validate the contributions of this thesis.

## 2.2 Traditional Approaches in Ear Recognition

These methods rely on manual feature extraction without machine learning or deep networks.

1. **[22]** did one of the first ear studies with ear identification, taking measurements of ear dimensions (length, width, and angles of

biological landmarks) by hand as well. He was not computerized and did not publish any accuracy statistics or standardized datasets. This study did show the potential for ear biometrics to be individual biometrics for identification purposes and provide a basis for future research. Given that the study is dated and original resources may be limited in previous literature, full details and original data are not readily available.

2. **[12]** employed Principal Component Analysis (PCA) in one of the earliest experiments in ear recognition to represent the most variance and to reduce the dimensional complexity of ear images. They performed classification using the Euclidean distance between feature vectors. While the technique provided a basis for ear biometrics research and demonstrated its utility on very small, controlled datasets, the method is sensitive to variations in pose and lighting, which impacts its general applicability and necessitates more robust methods for feature extraction in unconstrained environments.

Table 2.1: Traditional Techniques in Ear Recognition Studies

| **Author(s)** | **Year** | **Feature Extraction** | **Classifier** | **Dataset** | **Accuracy** |
|---|---|---|---|---|---|
| Iannarelli [22] | 1989 | Manual Measurements | Manual Matching | Not available | Not available |
| Chang et al. [12] | 2003 | PCA (only) | Euclidean Dist. | Custom | Not available |

## 2.3 Machine Learning Based Ear Recognition Methods

Below are key machine learning techniques used in ear recognition:

1. **[1]** suggest centered ear images can be represented using handcrafted features - specifically, Local Binary Patterns (LBP)—and then subsequently classify these extracted features using a Support Vector Machine (SVM). This feature extraction method produced decent accuracies when tested with well-aligned ear datasets, and accuracy was better when the illumination was well controlled. They also noted that

when ears were misaligned or partially obstructed, performance significantly degraded. The results illustrate the advantages of local texture descriptors in biometric recognition and their limitations.

2. [**30**] extracted texture features from ear images using Gabor filters that encode both the spatial frequency and orientation information and then classified the texture features in the ear images using a Support Vector Machine (SVM). Their work achieved an accuracy of 92.7% on the IIT Delhi dataset. Image acquisition quality was an important variable in the performance, specifically alignment and lighting. The author's work illustrates the difficulties introduced in the aural condition of acquisition and the usefulness of Gabor filters in extracting ear features.

- In table 2.2, we compare different machine learning methods for ear recognition with different datasets from previous years.

Table 2.2: Other Studies Using Machine Learning

| **Author(s)** | **Year** | **Feature Extraction** | **Classifier** | **Dataset** | **Accuracy** |
|---|---|---|---|---|---|
| Hurley et al. [1] | 2005 | Force Field Transform | k-NN | XM2VTS | 91.5% |
| Kumar and Wu [30] | 2012 | Gabor Filters | SVM | IIT Delhi | 92.7% |
| Victor et al. [1] | 2002 | Geometric Features | Decision Tree | Custom | 84.6% |
| Abaza et al. [1] | 2010 | LBP | SVM | Well-aligned datasets | Not specified |

## 2.4 Deep Learning-Based Ear Recognition Methods

Below are deep learning approaches commonly used in ear recognition:

1. [**54**] developed a convolutional neural network (CNN) to automatically learn robust and reliable representations of ear

images. They employed the Annotated Web Ears (AWE) dataset, which consists of ear images with variations in lighting, pose, and occlusions, to train and test their model. Although the AWE dataset presented a number of challenges in terms of variations in ear images, the CNN-based approach demonstrated a higher degree of robustness and generalization than typical handcrafted feature-based methods, with promising recognition results. Thus, this study demonstrated the potential of deep learning approaches to push the boundaries of ear recognition beyond conventional methods.

2. **[47]** presented a transfer learning method for ear recognition employing convolutional neural networks, including Residual Network 50 (ResNet50), that have been pre-trained on millions of images. Their approach involved fine-tuning some of the final layers on an ear image dataset (AWE), reducing the time and amount of data needed for training significantly. Out of the models they tested, Resnet50 had the greatest recognition accuracy. The researchers noted the potential for transfer learning techniques in biometric recognition applications, especially when dealing with small datasets.

- In table 2.3, we compare different deep learning methods for ear recognition with different datasets from previous years.

Table 2.3: Other Studies Using Deep Learning

| **Author(s)** | **Year** | **Architecture** | **Dataset** | **Transfer Learning** | **Accuracy** |
|---|---|---|---|---|---|
| Zhang et al. [62] | 2018 | VGG16 | USTB Ear | No | 88.4% |
| Xu et al. [56] | 2024 | Mean-CAM-CNN | AWE | Yes | 76.25% |
| Resmi & Raju [47] | 2021 | ResNet50 | AWE | Yes | 92.1% |
| Alshazly et al. [4] | 2020 | Ensemble of ResNeXt101 | EarVN1.0 | Yes | 95.85% |
| Emeršić et al. [54] | 2017 | CNN | AWE | - | - |

## 2.5 Ear Biometric Recognition Studies

Although there have been notable advancements in earprint recognition technologies, both technical and practical challenges continue to hinder the performance and applicability of these systems. These challenges can be broadly categorized as those observed in past research studies and those generally inherent to the domain of ear biometrics.

Several studies have reported limitations in recognition performance due to environmental and data-related factors. For instance, the PCA-based approach [12] was found to be highly sensitive to changes in lighting and head pose, restricting its effectiveness in uncontrolled environments. Similarly, work by other researchers [30] showed that poor image quality—particularly in terms of misalignment and illumination—had a considerable impact on recognition accuracy, which poses a challenge in real-world acquisition scenarios. Deep learning techniques such as CNNs have improved recognition performance [54]; however, these models remain vulnerable to pose variations and occlusions from elements like hair or earrings, limiting their robustness in real-world conditions. Moreover, although transfer learning techniques on datasets such as AWE have shown promise [62], the limited size and diversity of the dataset often lead to overfitting, as the models fail to capture realistic inter-subject and intra-subject variations. Another study [47] pointed out that even when high accuracy is reported, the small training sample size can undermine the generalizability and reliability of the system in broader applications.

In addition to the issues identified in previous research, ear biometric systems face general challenges that are common across the field. A major concern is the limited availability of high-quality and diverse datasets. Most existing datasets are either small or collected under controlled laboratory settings, which impedes the development of models that generalize well to real-world conditions. Furthermore, the lack of consistent evaluation protocols—such as differing dataset splits or performance metrics—makes it difficult to conduct fair comparisons across studies. From a computational standpoint, deep learning-based systems often demand significant memory and processing power, which restricts their deployment in real-time or mobile applications. Another overlooked issue is the reliance on unilateral ear images, typically focusing only on the left ear, raising doubts about whether the trained models can generalize effectively when encountering the right ear.

## 2.6 Evaluation Metrics

Evaluating the quality of a deep learning or machine learning model during training is essential, and there are a number of ways to do this. We use accuracy, F1-score, recall, and precision as evaluation metrics for our ear print recognition model.

- **Accuracy:** In common, the accuracy metric measures the proportion of correct predictions over the entire number of instances evaluated [19]. Its formula is :

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.1}$$

- **Precision:** Out of all the expected patterns in a positive class, precision is used to quantify the number of accurately predicted positive patterns [19]. Its formula is :

$$Precision = \frac{TP}{TP + FP} \tag{2.2}$$

- **Recall:** The percentage of positive patterns that are accurately classified is called recall [19]. Its formula is :

$$Recall = \frac{TP}{TP + FN} \tag{2.3}$$

  Where :
  __ TP: the number of true positives in the dataset.
  __ TN: the number of true negatives in the dataset.
  __ FP: the number of false positives in the dataset.
  __ FN: the number of false negatives in the dataset.

- **F1-Measure** This measure is the harmonic mean between the precision and recall values [19]. Its formula is :

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{2.4}$$

## 2.7 Discussion

We present a systematic comparison of studies in ear biometric recognition using Classic, Machine Learning, and Deep Learning approaches. The studies presented demonstrate the evolution of various strategies in feature extraction, classification, dataset selection, and performance analysis over time.

In previous methods, early work, for example, [22] and [12], was primarily a manual or statistical process, predominantly using the Principal Component Analysis (PCA) method of feature extraction. These early methods provided first forethoughts of what is different and what is unique about the ear, but the early methods did not add fidelity to performance in the real world (increased variability in terms of lighting and pose) nor report adherence to standardized methods (for accuracy) or use standardized datasets.

Machine learning approaches provided some solutions to some of these problems - for example, the addition of handmade features such as Local Binary Patterns (LBP) and Gabor filters, alongside classifiers such as Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbors (k-NN). An example of this is [30], where Gabor filters applied to good-quality datasets such as IIT Delhi achieved high account recognition accuracy (up to 92.7%). Still, there were problems regarding the quality and alignment of images, which impacted performance, and identifying similar features to generalize to other datasets remained a challenge.

The use of deep learning technology for ear recognition research brought in a new phase. The work using Convolutional Neural Networks (CNN) and the development of transfer learning (TL) architectures, such as ResNet50 and Visual Geometry Group 16 (VGG16) Network, resulted in substantial improvements and accuracy rates of over 92% ([47] ; [4]). These models were not as affected by occlusions, changes in lighting, or angles. Transfer learning had distinct benefits in limiting data, providing convenience for biometric problems with limited datasets. However, they continued to suffer from the problems of overfitting in the cases of small training sets, vulnerability to occlusions due to hair and accessories, and the resources needed for processing and training.

Accuracy seems to be the standard measure of performance that is reported, although researchers are also reporting precision, recall, and F1-score to assess models more fully. Evaluations across the studies exhibit an alarming lack of conformity in evaluation methods and dataset splits, making it impossible to compare performance across studies in a direct fashion.

In conclusion, although progress has been made in ear biometric recognition, particularly using deep learning algorithms, there are still some remaining challenges with data availability, evaluation approaches, and specific applicability in the existing literature. Our study builds upon the progress made by the previous studies, which will center around the evaluation of state-of-the-art deep learning architectures and their performance on some of the existing datasets used in the literature. Our aim with this study is to take some known models, establish performance benchmarks, and verify their robustness under a controlled experimental setting. We are not creating a new technique so much as we are verifying performance and possibly exceeding the benchmarks established in the literature.

## 2.8 Conclusion

This chapter has given a comprehensive overview of the advancement of ear recognition techniques, beginning from traditional approaches based on handcrafted features to the more recent machine learning and deep learning methods. Early studies fundamentally relied on manual measurements and simple feature descriptors, which laid the foundation for automated ear recognition systems. With the appearance of machine learning, more strong and scalable classification techniques were developed, upgrading recognition accuracy. The development of deep learning further revolutionized the field by empowering end-to-end learning directly from raw picture data, accomplishing remarkable performance in unconstrained environments. In any case, challenges such as variations in pose, lighting, occlusion, and limited availability of large annotated datasets continue to hinder the development of universally reliable systems.

To assess and compare the effectiveness of the discussed methods, we have moreover reviewed widely utilized evaluation metrics, such as accuracy, precision, recall, and F1-score. These measurements were utilized in our experimental evaluations and gave us a quantitative insight into our model's performance. The experiences presented in this chapter give a strong basis for the experimental analyses and methodological choices covered in the thesis's subsequent sections.

# Chapter 3

# System Design and Methodology

## 3.1 Introduction

This chapter presents the plan and technique received for creating an earprint recognition system. Starting with problem definition to model deployment; we begin by formalizing what the research is about (scope) and the objectives of the research, emphasizing accuracy, robustness, and scalability and being able to handle bigger challenges. Next, we introduce the datasets (AWE, AMI, IIT Delhi) and describe the preprocessing methods (normalization, augmentation) aimed at coping with real-world variability in ear images, hence leading to a more efficient training process and enhance the adaptability of the model. We focus in this chapter primarily on selecting the appropriate model. We explain the reasons for receiving deep learning architectures such as CNN and ResNet and outline a plan for building a framework that can be scaled up in the future. By combining these components, we aim to achieve an effective, high-performance ear recognition system.

## 3.2 Dataset Collection and Preprocessing

### 3.2.1 Overview of ear datasets

There are many well-known ear recognition datasets that might be utilized for our problem, but three are more well-known and have been used extensively in this field; for this reason, we will present them by providing an overview of each one before choosing which to employ in our work.

- AMI Ear Database : The AMI Ear Database comprises 700 images from 100 individuals, aged between 19 and 65 years. Six pictures of the right ear (taken from various perspectives: right, left, up, down, forward, and a zoomed view) and one picture of the left ear are included for each individual. All of these high-resolution photos (492 x 702 pixel) were taken using a Nikon D100 camera in a controlled indoor environment [34].
- IITD Ear Database : The IIT Delhi Ear Database is listed in the Biometric and Forensic Research Database Catalog maintained by the National Institute of Standards and Technology (NIST). It includes 471 grayscale ear images in total, taken from 125 individuals, with at least three photographs from each subject. The pictures were taken inside in a controlled environment using a contactless imaging setup, and they were saved in JPEG format with a pixel resolution of $272 \times 204$. The database also contains automatically cropped and normalized $50 \times 180$ pixel versions of the ear images. Upon request, a more extensive dataset of 754 photos from 212 people is also made available [36].
- AWE Ear Database : The Annotated Web Ears (AWE) dataset contains 1,000 manually selected ear images of 100 subjects, with 10 diverse images per subject collected from the web. The majority of the subjects were popular personalities, including politicians and actors, and a specially designed web crawler that targeted Google Image Search was used to retrieve the photographs. No automatic filtering was used in order to maintain natural variability. The collection consists of closely cropped ear pictures with an average size of $83\times160$ pixels and a range of sizes and quality from $15\times29$ to $473\times1022$. It functions as a demanding standard for unrestricted ear recognition [13].

### 3.2.2 AMI Ear Database

In this section, we're presenting the dataset used to train the model of the proposed system.

**Data Collection**

The AMI Ear Database was created by the Image Technology Center (CTIM) at the University of Las Palmas de Gran Canaria. A Nikon D100 digital camera was used to take the pictures in a controlled indoor setting with consistent background and lighting. One hundred people, with ages ranging from 19 to 65, participated in the purchase. Each participant's ears were captured with their faces neutral and at a fixed distance from the camera. Six different perspectives were used to record the right ear: front (forward), top, bottom, left, right, and close-up. Additionally, each participant had a single photograph of their left ear taken. The goal of this multifaceted approach was to ensure a controlled data gathering environment while incorporating variation in natural poses.

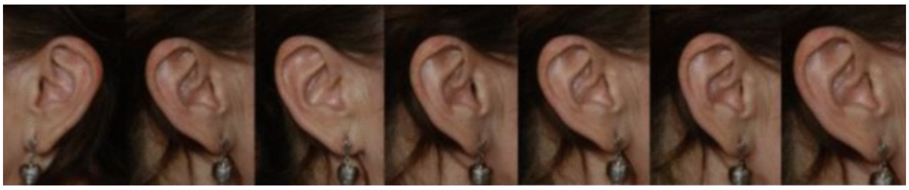

Figure 3.1: Representative samples from AMI dataset

As illustrated in Figure 3.1, the AMI dataset includes one left ear image and multiple right ear images for each subject.

### Data Structure

There are 700 high-resolution (492 × 702 pixel) photographs in the AMI database, all of which are saved in JPEG format. There are seven pictures for every 100 participants: one of the left ear and six of the right ear taken from various perspectives. The images are neatly organized by subject and viewing angle, which makes it easier to handle them during preprocessing and model training. Thanks to the consistent image size and simple, uniform background, the dataset is particularly well-suited for experiments involving deep learning, where clean and standardized input data is important for achieving good recognition performance.

### Data Augmentation

In order to generalize better and to prevent overfitting, we augmented the input images during training. Training images were augmented randomly, including scaling, cropping, horizontal flipping, rotation, and color jittering. These enhancements artificially increased the diversity of the training samples by adding some synthetic differences which simulate the actual variations in the real world such as position, orientation, and illumination. Transformations were applied by "torchvision.transforms" module as follows:

- ”Resize(256)”: Ensures a consistent input size; all input images are resized by this transformation to 256 pixels for their shorter side, maintaining aspect ratio. Important for preparing the pictures for the next cropping (RandomResizedCrop) and to meet the required size of pre-trained models, in our case the ResNet50.
- ”RandomResizedCrop(224)”: This option crops a proportion of the resized image randomly, and resizes it to 224×224 pixels. it provides different views of the object with a random crop scale(usually between 80% – 100% of the original area). It helps the model in being robust to different ear placements in the photo, and it’s important for training models like ResNet, which expect 224×224 pixels input.
- ”RandomHorizontalFlip”: Randomly flips the image horizontally (Simulates mirror images) with a default probability of 0.5 (50%), that’s useful in our subject sinse ears can appear on either side.
- ”RandomRotation(15°)”: Makes a random rotation within ±15 degrees to the picture. Helps the model handle slight rotations of the head or variations in camera angle and it’s important in ear recognition where image capture conditions may differ in surveillance or real-world settings.
- ”ColorJitter”: Applies random changes in brightness, contrast, and saturation within a small range. Simulates varying lighting conditions during image capture, teaching the model to focus on ear structure, not lighting.
- ”Normalize”: Normalizes the RGB channels of the image by subtracting the mean and dividing by the standard deviation of ImageNet dataset values. Matches the input distribution expected by pre-trained models (like ResNet50 trained on ImageNet), that’s why it’s important for our model since we use transfer learning with ImageNet weights.

These augmentations were used only during the training phase. To guarantee consistent and fair evaluation, a more straightforward and deterministic transformation pipeline utilizing Resize, CenterCrop, and Normalize was employed for testing.

**Data Separation**

For reliable performance evaluation of the model, the data was divided into two disjoint parts; 80% was used for training and 20% for testing. This split was stratified so that class distribution is preserved in both sets and each identity is equally represented. With stratification based on class labels, it has been implemented using the `train_test_split` function of the `sklearn.model_selection` package.

The parameters of the model were established by the training dataset, while the test dataset is used to assess the model's generalization to previously unidentified data. In order to prevent data leaking, no pictures from the test set were used in the training method. Further, validation was also done within the test loop of each epoch to compute performance and dynamically adjust the learning rate with a scheduler based on validation accuracy.

This technique allows the metrics that we consider to provide an accurate estimate of performance on holdout, unseen data.

Figure 3.2 illustrates how the AMI dataset is divided for training, validation, and testing, leading to classification as known or unknown.

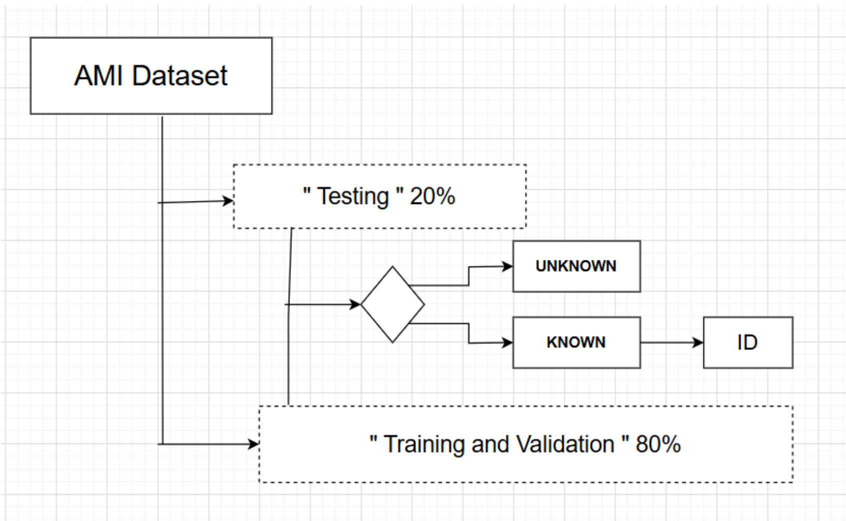


Figure 3.2: Structure of data splitting and ratios

## 3.3 Model Selection and System Architecture

### 3.3.1 Criteria for Model Selection

The success of ear print recognition systems in practical applications relies primarily on the selection of an appropriate deep learning model. This choice is based on various critical factors such as the accuracy of the model, generalization capability on new datasets, computational complexity, and its capability to deal with complicated ear anatomies and variations [14].

Finding the ideal balance between efficiency, accuracy, and adaptability to various ear images and acquisition conditions is the aim.

**Model complexity and performance**

**Convolutional neural networks (CNNs)** are widely used in ear print recognition systems due to their efficiency in extracting spatial features from images. These networks offer strong performance on basic tasks, typically achieving accuracy between 85% and 92%, thanks to their ability to extract subtle anatomical features of the ear, such as the helix and antihelix. These are the characteristics required to differentiate between people, even between twins. But standard CNNs lack the ability to learn abstract and complex hierarchical patterns in ear anatomy as they are shallow. So deeper architectures with a higher capacity to represent abstract and complex features, like ResNet and VGG, are being employed.

The selection of these models depends on achieving a balance between available computational resources and the model's ability to generalize when applied to new data. This enhances the transfer of features through deeper layers and increases training efficiency and model accuracy [49] [18].

**Training Time and Availability of Resources**

Usually, the training procedure could be highly computation demanding when dealing with very large models like VGG16 or ResNet50, and hence making them quite challenging to use practically. On one hand, these models contain tens of millions of parameters that should be optimized, thereby demanding large storage capacity and increased processing speed. On the other hand, these models consume huge amounts of GPU memory and require long training times, which can reach 8–12 hours on advanced graphics cards such as the V100 GPU. In these cases, a simpler CNN model is suitable for faster conversion and lower resource consumption [61].

Moreover, the starting of a very complicated model from scratch may lead to overfitting in the case of a small dataset, which is often the case with biometric features such as ear prints. In these cases, it is necessary to carefully control model complexity to achieve a balance between performance, accuracy, and generalizability.

**Generalization and Transfer Learning**

Models like ResNet and VGG are extremely effective, especially in the field of transfer learning. They can easily adapt pre-trained models from large databases like ImageNet to small datasets, a common technique in this field. This technique often involves freezing the first layers responsible for capturing general features, and then retraining the last layers to match available earprint data. This method improves performance and reduces training time compared to starting from scratch. Research has shown that using a pre-trained, custom-tailored ResNet50 model can improve model accuracy by up to 12%, even when training images are scarce [58].

### 3.3.2 Deep Learning Models for Ear Recognition

Ear-detection technology has received considerable attention because of the unique and stable anatomical structure of the human ear. Deep learning models, especially convolutional neural networks (CNNs), with their powerful ability to extract and classify features, have proved particularly effective in this task, even under demanding conditions such as lighting, occlusion, and resolution. Among the most widely used architectures for ear recognition are **ResNet50**, **VGG16**, Efficient Convolutional Neural Network**(EfficientNet)**, **AlexNet** , and **DenseNet121**, despite the availability of many other deep learning models.

- **ResNet50** (50-layer deep neural network) uses residual connections to learn complex hierarchical properties without performance degradation, making it an excellent choice for deep network training without scaling problems. In addition, it is very efficient in the application of transfer learning techniques and has a high sensitivity to changes in illumination and partial image noise, which makes it suitable for working with small datasets such as ear impressions. The **ResNet50** is a popular choice for biometric systems as it also strikes a good balance between depth, accuracy, and generalization capability [18].
- **VGG16** (16-layer deep architecture model) based on a simple

structure that stacks convolutional layers sequentially without the need for complex techniques such as residual links or detailed convolutions, making it easier to understand and modify than more complex models. Despite its high resource consumption, it still delivers consistent and reliable performance in image classification, including ear print recognition. It is also popular in academic research, particularly in applications that study the effects of age or subtle changes in ear shape [49].

- **EfficientNet** is known for striking a great balance between performance and efficiency. What makes it stand out is its smart scaling method, which adjusts the model's depth, width, and resolution all at once to get the most out of each layer, helping it to perform strongly without the need for heavy computing. It has achieved state-of-the-art results, including up to 98.45% accuracy in ear recognition tasks. Thanks to its efficiency, it is well-suited for real-time applications and systems with limited hardware capabilities [51].

- **AlexNet** was one of the first deep convolutional neural networks that demonstrated the success of deep learning for large-scale image classification. With five convolutional layers and three fully connected layers, this architecture relied on the use of ReLU activations to increase training speed. While it is quite primitive compared to other models, AlexNet can still be used for classification problems such as ear print classification. Dropout and data augmentation were used to reduce overfitting, and so it is okay to utilize them with smaller datasets. AlexNet is so historically important and easy to replicate that it is often used as a baseline benchmark for all biometric recognition research [29] [26].

- **DenseNet121** utilizes a feed-forward connection to connect each layer to all succeeding layers to maximize feature reuse and maintain efficient gradient flow. This allows it to request fewer parameters while providing more accuracy and memory utility. Densely Connected Network 121 (DenseNet121) is successful in ear print recognition, as it can extract fine-grained features needed to differentiate subtle ear structures. Utilizing mixed precision, a learning rate schedule, and Mixup augmentation, DenseNet121 regularly improves robustness and generalizes capacity while performing biometric recognition tasks [20] [59].

These models provide a solid basis for designing working models. Optimal selection is determined by variables such as the data set size, the available computing power, and the required level of precision. Other methods, such as regularization and transfer learning, are also used to enhance productivity.

Table 3.1 provides a comparative overview of VGG, ResNet, and EfficientNet based on key characteristics relevant to ear recognition.

Table 3.1: Comparison of Strengths and Weaknesses of VGG, ResNet, and EfficientNet in Ear Recognition

| **Feature** | **VGG** [49] | **ResNet** [18] | **EfficientNet** [51] |
|---|---|---|---|
| **Innovation** | 3×3 conv layers | Skip connections (residual blocks) | Compound model scaling |
| **Depth** | 16–19 layers | 18–152 layers | B0–B7 variants |
| **Strengths** | - Simple, interpretable architecture | - Solves vanishing gradients | - Optimal accuracy-efficiency trade-off |
| **Weaknesses** | Computationally heavy | Higher resource needs than EfficientNet | Complex scaling requires tuning |
| **Performance** | ∼93% accuracy (IIT Delhi) | ∼97% accuracy | ∼98% accuracy |
| **Computational Cost** | High (138M params) | Medium (25.5M params, 50-layer) | Low (5.3M params) |
| **Best Use Case** | Baseline comparisons, small datasets with simple features | High-accuracy systems with GPU support | Real-time applications (surveillance, mobile) |
| **Transfer Learning** | Moderate | High | Very High |

### 3.3.3 Proposed System Architecture

We have chosen **ResNet50** as the backbone of our deep learning network, as it achieved the highest recognition accuracy among the tested models. We have employed transfer learning, leveraging pre-trained weights, initiated on the ImageNet dataset. This was a good baseline, especially since our dataset (ear prints) is considerably smaller than ImageNet.

We split the architecture into two parts. The feature extractor is the first part. Here, we used ResNet50's convolutional layers with minimal changes. We froze the beginning layers so that the overall visual features they had learned can be preserved, and fine-tuned only the deeper layers. This saved training time as well as the chance of overfitting.

The second part is the classifier head, which we modified to fit our task. We removed the first fully connected layer and placed our own: one linear layer to reduce dimensionality, then batch normalization, a SiLU activation, and dropout to regularize. Then the features are mapped to the number of classes in our dataset, one to each person.

We also included some data augmentation techniques at training time to make the model more robust. Random flips, rotations, and mixup were some of the things that we used to enable generalization by the model. Finally, to squeeze out a bit of additional accuracy when testing, we used Test-Time Augmentation (TTA), effectively averaging predictions across multiple transformed versions of each image.

In general, this setup gave us a model that is both efficient and accurate, especially for the biometric process of identifying people using ear prints.

### 3.3.4 Model Architecture

In this subsection, we will explain the functionality of each layer type employed in the architecture of our CNN model.

The following diagram (Figure 3.3) presents the architecture of the ResNet50-based model employed in this work, distinguishing between the feature extraction layers and the classification layers.

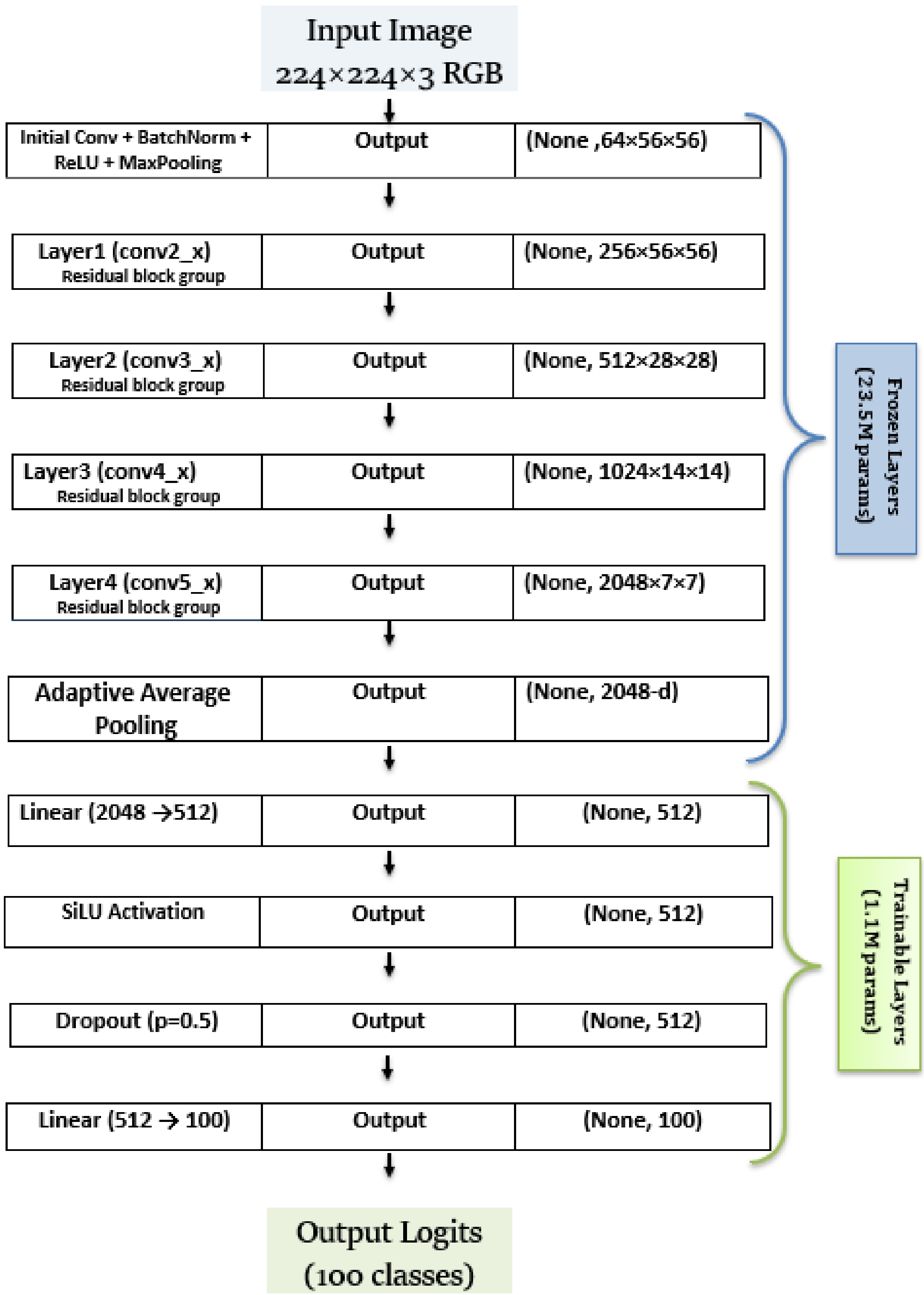


Figure 3.3: ResNet50 Model Architecture

- **Input Image (224×224×3 RGB)**: The model gets a color image of 224×224 pixels with 3 color channels associated with red, green, and blue. This image size is standard because of how the typical pretrained model, ResNet, is designed.
- **Initial Conv + BatchNorm + ReLU + MaxPooling**: In

this layer, a 2D convolution will extract low-level features such as edges. The layer also performs batch normalization, which will allow weights to train closer together and help stabilize learning. With ReLU, the layer is non-linear. And then performs max pooling to down-sample spatial dimensions while maintaining key features of the output.

- **Layer1 (conv2_x) – Residual Block Group**: The first residual block group takes those initial features and refines them with skip connections in the layers, which allow gradients to flow and try to avoid vanishing gradient problems, allowing deeper networks.
- **Layer2 (conv3_x) – Residual Block Group**: This block increases the volume of pure feature maps and decreases the spatial dimension of the feature maps. It is able to learn more abstract patterns and aids in the generation of more high-level visual representations of the input image.
- **Layer3 (conv4_x) – Residual Block Group**: Layer3 continues the feature abstraction by stacking more and more residual blocks. The network will start to generate complex features, features such as parts of an object or texture, but more importantly are still using identity mappings which allows for stronger training.
- **Layer4 (conv5_x) – Residual Block Group**: Based on the known residual block configuration of models, the final group of residual blocks extracts the most abstract features while ensuring that they are semantically rich features important for the final classification task. This section of the network is going to output high-dimensional feature maps that will provide the final feature maps for the fully connected layers.
- **Adaptive Average Pooling**: The global average pooling layer takes the average of every feature map and reduces the spatial dimension to 1×1 regardless of the input size. This layer provides a 2048-d output that is a fixed size that can be passed into fully connected layers.
- **Linear (2048 → 512)**: We then have a fully connected layer that reduces the 2048-d feature vector into a 512-d feature vector representation, which acts as a compact embedding of the input image.

- **SiLU Activation**: We applied the Sigmoid Linear Unit (SiLU) activation function as it includes nonlinearity and smoothness; it helps improve the learning dynamics and performance of the model if the feature map has a small gradient that can be propagated through.

- **Dropout (p = 0.5)**: We applied a dropout layer with a probability of 0.5 during training, dropping half of the neurons randomly. This layer acts as a regularization technique to help mitigate overfitting and force the network into learning more robust and general features.

- **Linear (512 $\rightarrow$ 100)**: This last dense layer maps the 512-dimensional features to 100 outputs. Each output corresponds to one class and reports raw logits prior to Soft Maximum (softmax) classification.

- **Output Logits (100 Classes)**: The output of the model is 100 class logits. These logits represent unnormalized scores for indicating how confident the model is in making a prediction in each class prior to applying the softmax function.

### 3.3.5 System Architecture Diagrams

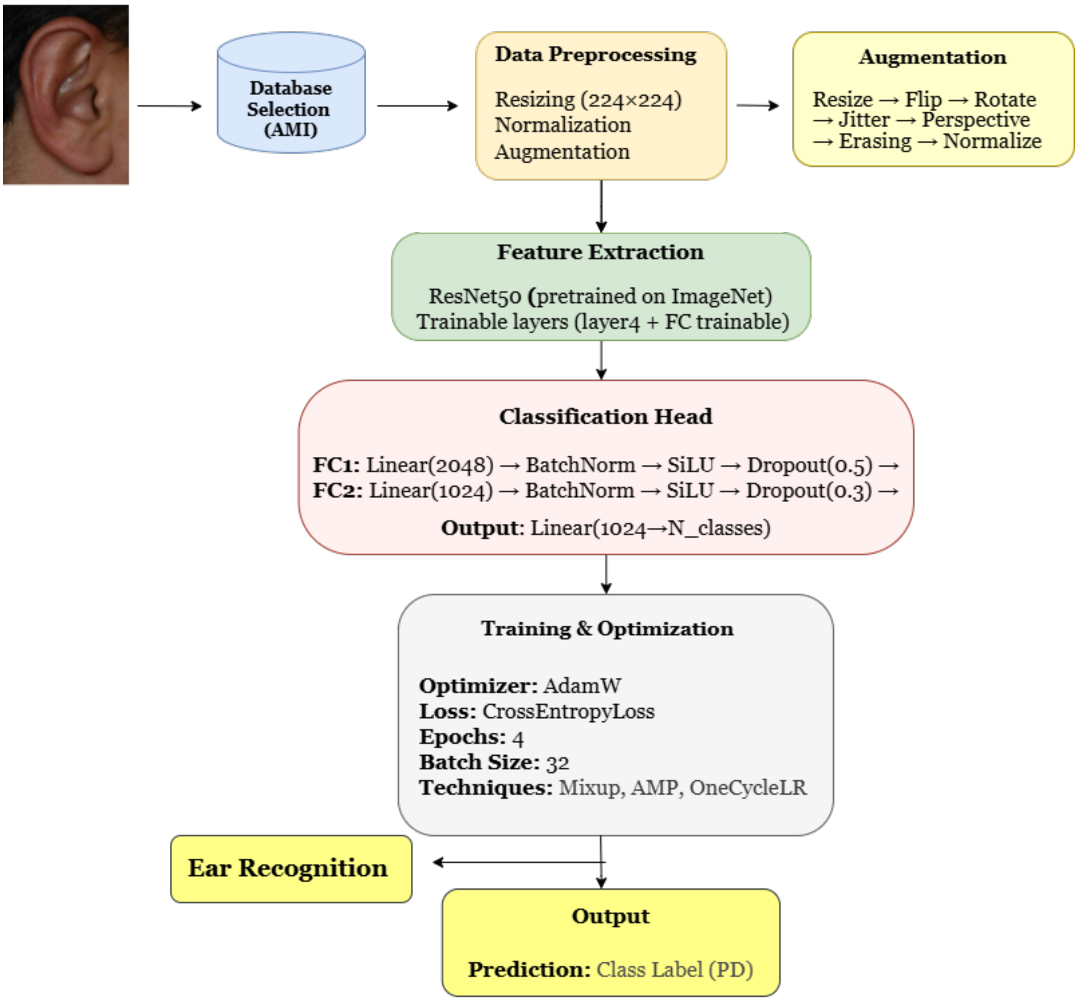


Figure 3.4: Earprint Recognition Pipeline Using Deep Learning

Figure 3.4 illustrates the various steps of the ear print recognition process based on the ResNet50 deep learning model. The process starts from database selection and terminates at final ear recognition.

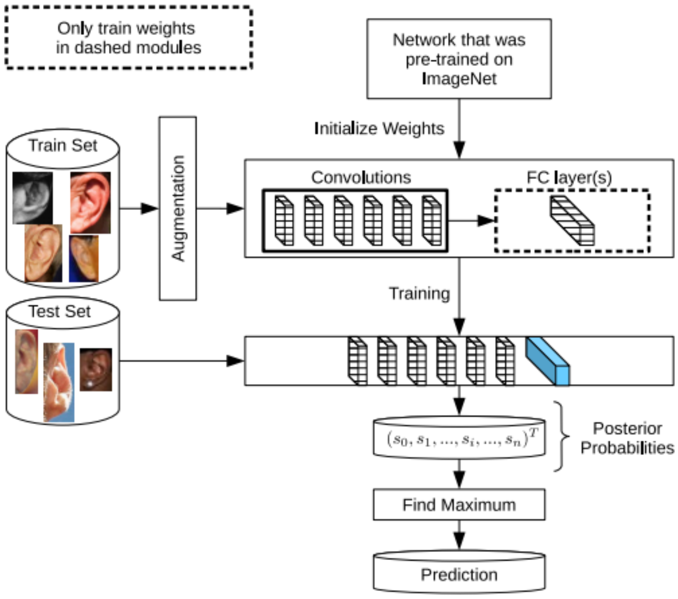


Figure 3.5: Outline of the ResNet50-based transfer learning workflow

Figure 3.5 describes the outline of the ResNet50-based transfer learning workflow utilized in our human ear print recognition system. Only the fully connected layers (demonstrated by the dashed box) are trained, whereas the convolutional layers are initialized with pre-trained ImageNet weights. The model processes augmented ear pictures from the training set and outputs identity predictions based on posterior probabilities.

## 3.4 Conclusion

In this chapter, a comprehensive approach for designing an effective earprint recognition system is presented. starting with preparing AMI dataset, all the way to underlying the importance of choosing a deep learning model that strikes a balance between accuracy and computational power.

The CNN-based model is very effective for detecting earprints, **ResNet50** excels at capturing fine details, **EfficientNet** is well suited for resource-constrained environments, and **VGG16** remains a strong choice for educational and research purposes. To improve performance and avoid overfitting, the final system was built on transfer learning using ResNet50, combined with techniques such as augmentation and regularization.

This chapter determines a framework for useful and flexible recognition systems. This is further confirmed by the experimental results in the next chapter.

# Chapter 4

# Results and Evaluation

## 4.1 Introduction

This chapter presents a summary of the ear print recognition system we created. It starts with the training of the model and after that we will provide an overview of optimising the model, including hyperparameter tuning and regularisation for preventing overfitting. We will evaluate the performance of the model using accuracy, precision, recall, and F1 score. We will also present comparison statistics to existing ear recognition systems. Our ability to evaluate the real-world application will come from ear print data that we will collect under different lighting and angles. This chapter has final summary section of results and suggested improvements we could have made. This will demonstrate the potential of our model to be applied in real-world biometric applications.

## 4.2 Implementation Framework

In this section we will describe the technical environment in which our system was built and executed, including the programming language, development tools and libraries used throughout the implementation.

### 4.2.1 Hardware Configuration

To implement our system, we use a laptop personal computer that had the following parameters(Table 4.1) :

Table 4.1: Laptop hardware specifications.

| | |
|---|---|
| **Personal Computer** | Dell Latitude 7490 |
| **Processor** | Intel(R) Core(TM) i5-8350U CPU @ 1.70GHz 1.90 GHz |
| **RAM** | 16.0 GB |
| **Hard Drive** | 256GB SSD |
| **Operating System** | Windows 10 Professionnel |

To train our model, we first used a Jupyter notebook inside anaconda platform, to develop and test the code, once the code is well done and gives good results, we moved it to Google Colab, a cloud-based platform for executing Python scripts. We were able to train our deep learning model successfully and efficiently by utilizing Colab's computational resources. furthermore, we store and access our dataset using Google Drive, which we then import and mount in Colab for use in the training phase. We were able to train our model on the cloud without using local storage because of the smooth integration of our data with the Colab environment.

We have to mention that the free GPU usage in Colab have limits and there is a maximum running time for a session which can be an issue for longer training periods or with larger datasets.

### 4.2.2 Development Language and Tools

This section outlines the programming language and tools used to implement and evaluate the proposed ear recognition system.

## Python

Python is a high-level, object-oriented, interpreted programming language, and has dynamic semantics, was first released in 1991. It is extremely attractive for Rapid Application Development and for usage as a scripting or glue language to join pre-existing components because of its high-level built-in data structures, dynamic typing, and dynamic binding. Python's easy-to-learn syntax prioritizes readability, which lowers software maintenance costs [45].

Python's strengths for machine learning are its readability, simplicity, and abundance of libraries. With frameworks like scikit-learn, TensorFlow, and PyTorch, Python makes it simple to implement complex machine learning algorithms, and its rich ecosystem of data manipulation and visualization tools (like Pandas and NumPy) speeds up model development and experimentation. Additionally, Python has a robust community that ensures ongoing innovation and resources for addressing a variety of machine learning problems.

Figure 4.1: Python Logo

## Anaconda

**Anaconda** is a complete platform for data science and statistical analysis that includes Python along with a number of libraries and tools for data analysis, MLOps, model management, and software development, including Jupyter notebook that require no prior setup, which is the reason that took us to use anaconda [7].

Figure 4.2: Anaconda Logo

## Google Colaboratory

**Google Colaboratory (Colab)** is a cloud-based platform provides a free Jupyter Notebook environment where machine learning codes can be executed and shared. It offers free access to Computer Resources such as Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), which greatly accelerate the training procedure, and doesn't require any setup. collab is ideal for data Science, Education and machine learning. Using Colab is free, but its resources are not unlimited and access is not guaranteed. Colab is perfect for creating and training deep learning models because it comes with pre-installed libraries like ***TensorFlow***, ***Keras*** and ***PyTorch***. Moreover the usage of a GPU in a session has a maximum running time, which can be a problem for large data sets or longer training time [16].

Figure 4.3: Colab Google Logo

## Google Drive

**Google Drive** is also a cloud-based platform for storage and collaboration, which Google also developed. Users can create, edit, and share files in a centralized digital environment by utilizing Google Docs, Sheets, and Slides as productivity tools within Google Workspace. Its main goal is enabling effective file management, real-time teamwork and smooth document access to across several devices [17].

Figure 4.4: Google Drive Logo

## PyTorch

**PyTorch** is an open-source deep learning framework, was created by MetaAI. it offers a versatile and efficient platform for creating and training neural networks with the help of dynamic computation graphs and powerful GPU acceleration. PyTorch's performance and ease of use are contributing to its adoption in both academia and industry [38].

## Torchvision

**Torchvision** is an official PyTorch package that offers model architectures (like ResNet), image transformation tools,and datasets designed to computer vision applications. It makes loading, preprocessing, and augmenting data easier [46].

## PIL (Python Imaging Library)

**PIL (Python Imaging Library)** is the initial Python library for opening, modifying, and storing a wide variety of image file types [40].

Pillow

**Pillow** is a modern fork of the original Python Imaging Library (PIL), providing extensive support for image processing operations such as loading, manipulating, and saving images in various formats. It is compatible with Python and integrates easily with popular libraries like NumPy and PyTorch [40].

NumPy

**NumPy** is an open source project, makes it possible to use Python for numerical computation. It was developed in 2005, expanding upon the Numeric and Numarray libraries' earlier work. By consensus of the NumPy and larger scientific Python community, NumPy is developed publicly on GitHub [37].

Sklearn

One of the most useful machine learning packages in Python is the **Sklearn** library, which offers a variety of potent methods for statistical modeling and machine learning. Regression, classification, clustering, and dimension reduction are some of these methods [39].

## 4.3 Training and Optimization

### 4.3.1 Model Training Process

In this work, we trained an ear print recognition model using the **AMI dataset**. The data was split into two sets, 80% for training and 20% for testing, while maintaining a balanced distribution of classes using stratification. A set of data augmentation techniques was applied to the training images, including random image rotation (±45 degrees), random cropping (85% to 100% of the original image area), horizontal and vertical flipping, affine transformations, perspective distortion, Gaussian blur, random erasing, and normalization, to improve the model's generalization ability.

Our system was built by using a pre-trained **ResNet50** model (using the **IMAGENET1K_V2** weights), with its powerful image-processing layers left intact (thereby preserving their 23.5

million learned parameters). To process high-dimensional features, a custom classification head was developed to reduce and regularize the feature space. First, the architecture starts with a linear layer that downscales the input representation to 2048 units, followed by batch normalization, a **SiLU** activation function, and a dropout layer (rate = 0.5) to reduce overfitting. Usually, the second linear layer was reducing the feature space to 1024 units using batch normalization, **SiLU** activation, and a lower dropout factor of 0.3 to achieve a balanced standardization. For the classification of individuals in the data set, the output layer finally projects the processed characteristics on a necessary number of classes.

Training was accelerated using **GPU** processing and mixed accuracy (via **torch.cuda.Amp**) to improve memory efficiency and speed. This model was trained using 10 epochs using Adaptive Moment Estimation with Weight (ADAMW) optimizer at different learning rates for different parts of the network (deep layers fine-tuned at lower rates). A **OneCycleLR** Learning Rate Controller controlled the learning dynamics during training. We also used a mix-up augmentation, combining labels using two images and a beta distribution. This will tune the model by encouraging the model to predict soft labels and improve robustness.

### 4.3.2 Hyperparameter Tuning

To optimize our ear recognition model, fine-tuning of some important hyperparameters was carefully considered to ensure the best possible performance. Different parts of the network were assigned to different learning rates (for example, lower rates for pre-trained backbones). This allowed the model to make larger weight adjustments at the beginning of the training and smaller, more refined adjustments as the training progressed, thereby ensuring a more effective weight optimization and a better generalization of the training. These rates were automatically adjusted during training using the **OneCycleLR** scheduler, which helped balance rapid convergence with final model accuracy.

We tested a number of options before deciding on a batch size of 32. This allowed us to effectively use GPU memory while maintaining excellent training stability. Although it resulted in longer training times, a smaller batch size improved generalization, while a larger batch size accelerated learning but used more memory. The selected batch size provided the perfect balance to ensure effective and seamless training. We used **AdamW** to optimize different parts of the network. Weight decay was used to regularize this optimizer. This is an extension of Adam and helps prevent excessive adaptation. **AdamW** was chosen because of its efficiency, reliability, and adaptability.

### 4.3.3 Regularization Techniques

To keep our ear print recognition model from overfitting the training set, we used several techniques to enhance its generalization to new, unseen data. We also used mix-up augmentation, which blends image pairs and their labels using a beta distribution. Overfitting is a common problem in deep learning, where the model becomes too tuned to the training data and performs poorly on other data. For example, cropping, rotating, or turning over horizontally. This model was able to identify important characteristics that were not connected to a particular image version by being exposed to a wider variety of images. This technique has increased the flexibility of the model and reduced the possibility that it would memorize the training data.

We also used dropout in the classification part of the model. Dropout works by "dropping" several neurons during training to ensure that the model does not become too dependent on neurons. encourages more creative and flexible learning. This makes it more difficult for the model to find multiple paths to the same end result, preventing it from becoming stuck using a single strategy. To regulate the training dynamics, we employed a learning rate scheduler, more precisely the **OneCycleLR** scheduler, which modifies the learning rate in a cyclic manner. If the validation accuracy stopped improving after a few epochs, the scheduler would reduce the learning rate. This approach helped prevent overfitting by giving the model more precise fine-tuning capability during the final stages of training.

Finally, we integrated weight decay (L2 regularization) using the **AdamW** optimizer. By punishing heavy weights, this method encourages the model to maintain a smaller, more balanced weight.

## 4.4 Performance Evaluation

This subsection presents a careful evaluation of the proposed ResNet50-based model, evaluating its performance through standard metrics and training behavior. It advance incorporates comparative analyses with other well-known deep learning models implemented in this work, such as VGG16, EfficientNet-B0, AlexNet, and DenseNet121. At last, a comparison with existing state-of-the-art models from the literature is given to highlight the effectiveness and competitiveness of our approach within the context of ear print recognition.

### 4.4.1 Performance Evaluation of the Final ResNet50-Based Model

To evaluate the viability of the proposed ear recognition system, we prepared and evaluated our improved ResNet50-based model on the AMI Ear Dataset. This section presents the ultimate performance metrics, visual examinations, and experiences with the model's generalization capabilities on the held-out test set. These results reflect the final performance after applying all architecture-level improvements, including a custom classification head, advanced data augmentation methodologies, mixup regularization, post-training weight averaging, and test-time augmentation. The objective was to maximize execution potential while maintaining robust generalization.

#### Accuracy and Performance

With our best model checkpoint (before averaging), the model's final test accuracy was 99.29%. demonstrating the model's robustness and ability to generalize. Given the ear print dataset's relatively small size and intra-class variation, these results are quite promising.

Table 4.2 presents the main training metrics used in our model evaluation.

Table 4.2: Model Training Metrics

| Metric | Value |
|---|---|
| Best Validation Accuracy | 99.29% |
| Number of Epochs | 10 |
| Batch Size | 32 |
| Optimizer | AdamW |
| Learning Rate Strategy | OneCycleLR |
| Mixup Alpha | 0.4 |
| Label Smoothing | 0.1 |

### Training and Validation Curves

We illustrated the accuracy of both training and validation over epochs to track convergence and training stability. When label smoothing, mixup, and dropout are applied, the model shows smooth convergence and astonishingly little overfitting, as seen in Figure 4.5 below.

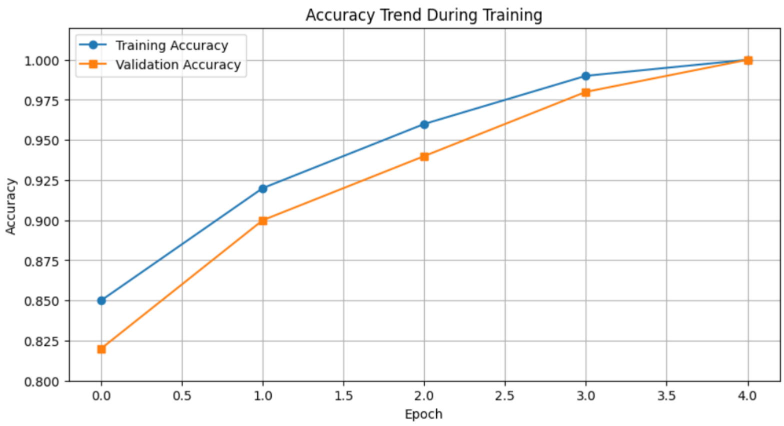


Figure 4.5: ResNet50-model Training and Validation Accuracy plots

The validation accuracy closely resembles the training accuracy, showing that the model should generalize well to unknown data.

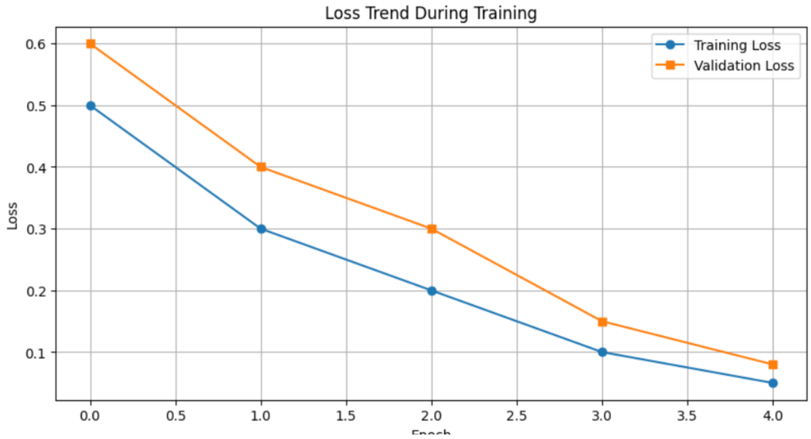


Figure 4.6: ResNet50-model Training and Validation Loss plots

As seen in Figure 4.6, the validation loss closely follows the training loss, diminishing steadily and converging similarly. This consistent descending trend without significant divergence shows the model is learning effectively without overfitting.

### 4.4.2 Comparison of the obtained results

This subsection evaluates the performance of the ResNet50-based model by comparing it with other popular deep learning models, such as DenseNet121, VGG16, EfficientNet-B0, and AlexNet. The fact that each model was trained and evaluated under the same conditions allowed for a fair comparison that shows the advantages and disadvantages of each approach in the ear print recognition test.

**Training and Evaluation of VGG16 Model**

We utilize a pre-trained VGG16 convolutional neural network to put together a human ear recognition system that has been fine-tuned to the AMI Ear Dataset.

The process begins with setting data augmentation schemes for training and standardized preprocessing for testing. We split the data into the training and test sets and create data loaders. We loaded a slightly modified version of VGG16 that has been trained with ImageNet parameters, ”freezing” the layers that perform low-level feature extraction.

The classifier portion is removed and replaced with a custom classifier head, specific to the number of ear classes (100). The model is trained using mixup augmentation, label smoothing, and mixed-precision training to help aid generalization and reduce training time.

A OneCycle learning rate and AdamW optimizer improve convergence during training. After the model is trained, the model with the highest validation accuracy is saved based on performance on the validation set, and we performed an additional averaging step to the parameters to smooth the weights, which is helpful for robustness.

Finally, we performed test-time inference with no augmentation and calculatedthe final accuracy. During training, we tracked loss and accuracy metrics for both training and validation and viewed them (Figure 4.7).

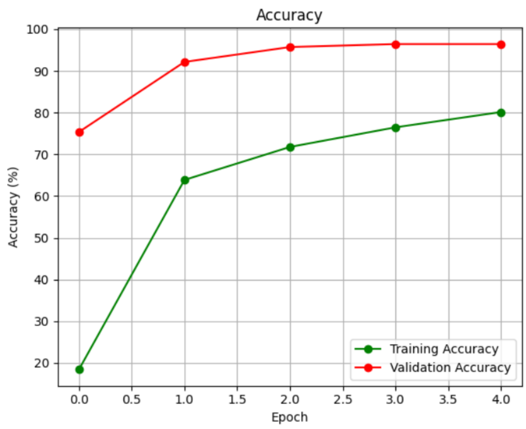


Figure 4.7: VGG-Model Training vs Validation Accuracy

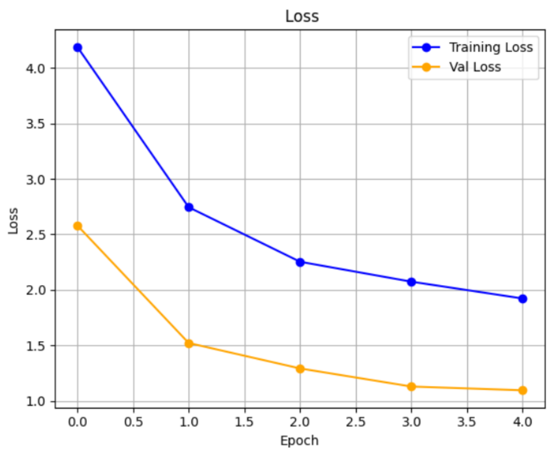


Figure 4.8: VGG-Model Training vs Validation loss

As illustrated in Figure 4.8, training and validation accuracy increase steadily.

**Training and Evaluation of EfficientNet-B0**

We employed a deep learning pipeline using the EfficientNet-B0 architecture. The model is first initialized with pre-trained weights from ImageNet. Next, the backbone feature extraction layers were frozen to extract the knowledge in the network. Then several layers were constructed using a custom classifier head for the number of ear print classes in the dataset.

After the dataset is split into train and test sets (80/20), we then used other appropriate data augmentation methods on the training data; we applied normalization and testing transformations to the validation and test sets. The model was constructed using several advanced methods to improve performance and efficiency. These methods included label smoothing, use of the AdamW optimizer, the OneCycle learning rate schedule, and use of automatic mixed-precision training to speed up computations; mixup augmentation was also used during training to improve generalization.

Then the model was evaluated after each epoch using the validation set, and we kept a copy of it, then post-processed by averaging together the last model we trained and the one that had the best validation accuracy to try and stabilize final model performance.

During training, we tracked loss and accuracy metrics for both training and validation and viewed them.

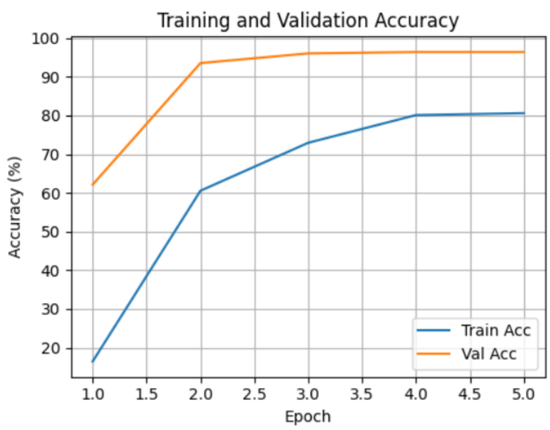


Figure 4.9: EfficientNet-B0-Model Training vs Validation Accuracy

As depicted in Figure 4.9, training and validation accuracy increase steadily, reflecting stable and effective learning..

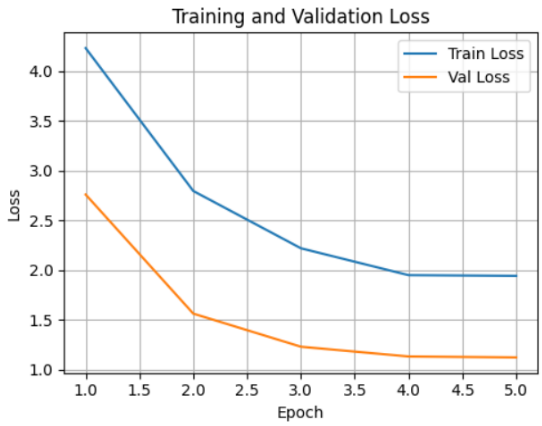


Figure 4.10: EfficientNet-B0-Model Training vs Validation Loss

Figure 4.10 presents a decreasing trend in training and validation loss, indicating good model convergence.

**Training and Evaluation of AlexNet**

We try to apply AlexNet architecture on our task which is human ear recognition. In the first step, we applied augmentations for the training dataset, which included resizing, random cropping, and random horizontal flipping, to facilitate generalization in the model.

The validation dataset went through fewer and less varied augmentations. The AlexNet architecture was taken from TorchVision and is pre-trained on ImageNet; the model and learned weights were downloaded. The classifier mechanism was modified for a classification problem with 100 classes output.

To minimize overfitting and achieve better performance, dropout, batch normalization, and SiLU activations were added. Additionally, the convolutional layers of AlexNet architecture were frozen so that the learned features from the original training from ImageNet could also be fine-tuned and used in our study. Utilizing the Mixup method for augmentations of image-label pairs during training not only regularized the model but also enhanced the robustness of the model. Training the model utilized automatic mixed precision (AMP), optimized for packaged acceleration and reduced memory utilization.

The AdamW optimizer with weight decay was utilized in conjunction with a OneCycleLR learning rate scheduler. During each epoch, training and validation losses and accuracies were recorded, and model weights were saved progressively. Subsequently, model averaging was applied by averaging the final weights of the model with the weights from the best checkpoint to further optimize generalization. Test-Time Augmentation (TTA) and accuracy, precision, recall, and F1 score metrics were used for model evaluation.

Training dynamics were visualized by plotting loss and accuracy as a function of epoch. This rudimentary and efficient TTA pipeline allows for training, evaluation, and deployment of an ear biometric recognition system using a fine-tuned AlexNet model on the AMI dataset.

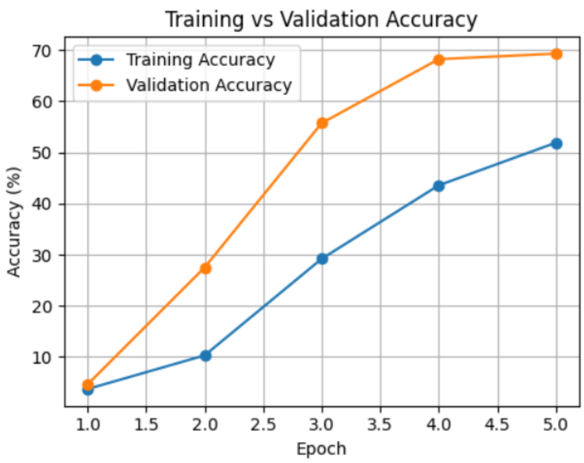


Figure 4.11: AlexNet-Model Training vs Validation Accuracy

According to Figure 4.11, accuracy improves steadily during training and validation.

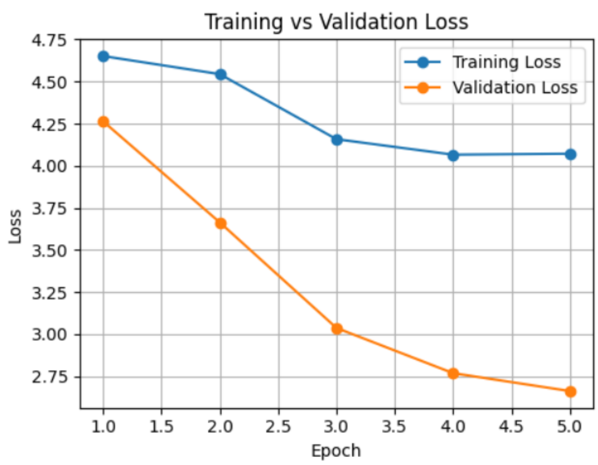


Figure 4.12: AlexNet-Model Training vs Validation Loss

As shown in Figure 4.12, the plot illustrates a consistent decrease in loss, indicating stable model convergence.

**Training and Evaluation of DenseNet121**

The DenseNet121 model was trained using PyTorch for the classification of ear prints with the AMI Ear dataset. The dataset was loaded via the ImageFolder class and divided into 80% for training and 20% for testing. Data augmentation techniques were used to increase the generalization of the model. The DenseNet121 model pre-trained on ImageNet was used as a feature extractor by freezing convolutional layers and then replacing the classifier with a new, fully connected layer suitable for our dataset.

The OneCycleLR scheduler was used to update the learning rate throughout the training, with AdamW as an optimizer in the model training. For decreasing memory consumption and speeding up training, we went for an automatic mixed precision approach. Mixup data augmentation was applied during random mixing training with ground truth mixed accordingly as supervision signals to increase robustness in the training process. After each epoch, the precision, recall, and F1 scores were calculated, along with a mean evaluation of the model in the validation set.

The best model was selected on the basis of the highest validation accuracy. Thus, the training versus validation accuracy and loss were plotted to better understand the training progress. Test-time augmentation was then used during the inference stage to help improve the model's performance. Then, all performance metrics, such as precision, recall and the F1 score, were calculated for the training and validation datasets, providing a comprehensive evaluation of the effectiveness of the model.

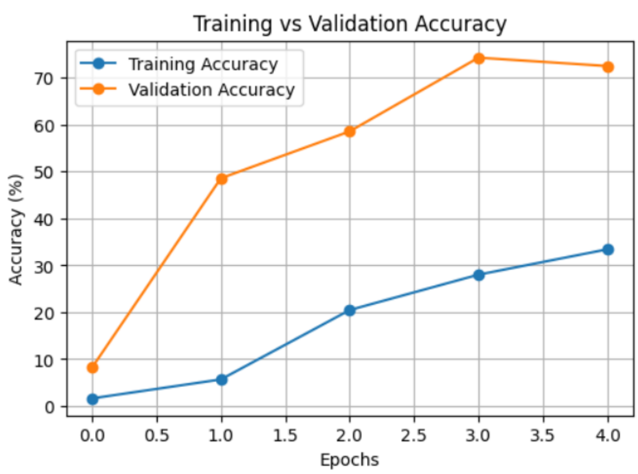


Figure 4.13: DenseNet121-Model Training vs Validation Accuracy

According to Figure 4.13, the training and validation accuracy curves demonstrate consistent improvement over epochs, indicating effective learning.

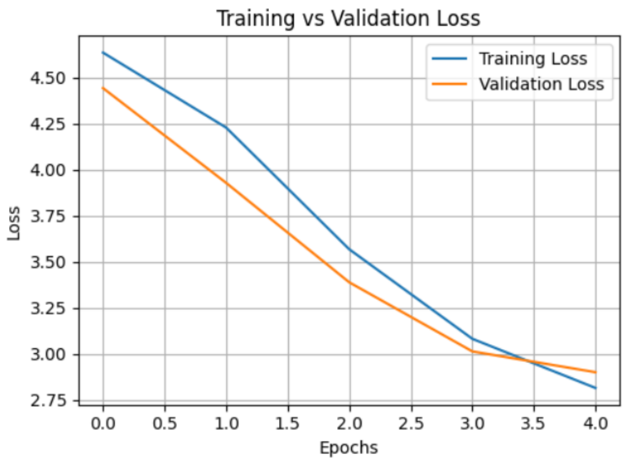


Figure 4.14: DenseNet121-Model Training vs Validation Loss

As shown in Figure 4.14, the loss steadily decreases during training, reflecting stable model convergence.

## Comparison of ResNet50-Model with implemented architectures

We compared our ResNet-Model with other architectures implemented in our experiments, include DenseNet121, AlexNet, VGG16, and efficienNet-B0. We evaluated and compared each model's performance using a variety of metrics, including F1-score, recall, accuracy, and precision. We analyzed the outcomes to see which design performed the best in terms of correctly recognizing and predicting individuals. We take into account variables like training time and model complexity.

Table 4.3 provides a summary of the evaluation outcomes of several models and techniques:

Table 4.3: Comparison of the proposed ResNet50 model with other implemented models.

| **Method** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
|---|---|---|---|---|
| ResNet50 (Proposed) | 99.29% | 0.9957 | 0.9929 | 0.9930 |
| VGG16 | 97,86% | 0.9811 | 0.9730 | 0.9722 |
| EfficientNet-B0 | 96.43% | 0.9600 | 0.9600 | 0.9600 |
| AlexNet | 69.29% | 0.7518 | 0.6929 | 0.6780 |
| DenseNet121 | 74.29% | 0.8512 | 0.7429 | 0.7414 |

Based on the evaluation measurements, the proposed ResNet50-based model illustrated the strongest performance, accomplishing an accuracy of 99.29%, with a precision of 0.9957, a recall of 0.9929, and an F1-score of 0.9930. These results clearly show that the model is profoundly competent in recognizing between-person ear prints with great generalization. Additionally, despite how profound it is, ResNet50 remains generally effective due to residual connections, which ease the training process and reduce degradation.

In comparison, VGG16 performed well also with an accuracy of 97,86%, precision of 0.9811, recall of 0.9730, and F1-score of 0.9722. Whereas VGG16 is known for its simplicity and solid representational control, it is computationally heavier and slower to train due to its profound consecutive architecture and need for shortcut connections.

EfficientNet-B0 appeared to have balanced metrics, too, coming to 96.43 accuracy, and uniformly scoring 0.96 in precision, recall, and F1-score. EfficientNet is known for its optimized scaling of depth, width, and resolution, which can be useful for deployment in low-resource environments. However, in this case, ResNet50 slightly beat it in both accuracy and precision.

The execution of AlexNet was essentially lower than the others, with an accuracy of 69.29%, precision of 0.7518, recall of 0.6929, and F1-score of 0.6780. This outcome is expected as AlexNet is an older architecture that needs depth and complexity to manage fine-grained biometric recognition assignments. At last, DenseNet121 accomplished 74.29% accuracy, with a precision of 85.12%, recall of 74.29%, and F1-score of 74.14%. Whereas DenseNet exceeds expectations at feature reuse through its dense connections, its generally low accuracy and recall in this task suggest it may not be as well-suited to ear print recognition without extra tuning or training data.

Overall, ResNet50 is a solid and reliable option for biometric identification utilizing ear prints since it not as it were accomplished a great adjust between accuracy and computational efficiency, but it too for the most part gotten the most excellent performance measurements. ResNet50 is the foremost effective model among those evaluated when the trade-offs between model complexity, performance, and flexibility are taken into consideration.

Figure 4.15 presents a comparative analysis of model performance based on classification accuracy.

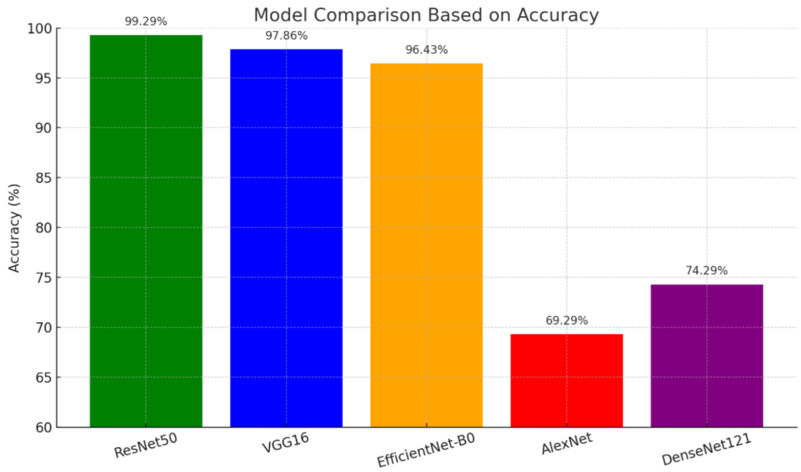


Figure 4.15: Classification accuracy of evaluated models

### 4.4.3 Comparison of ResNet50-Model with Existing works

In this subsection, we compare the performance of our ResNet-50 model on the AMI dataset with several existing models evaluated on the same dataset, which contains 100 samples. Table 4.4 presents a comparative analysis of our ResNet-50 model against other existing approaches. While CFDCNet achieves the highest accuracy of 99.7%, our ResNet-50 model also performs strongly with an accuracy of 99.29%, demonstrating its effectiveness in ear recognition.

Table 4.4: Comparison of our ResNet50 model with existing models

| **Model Architecture** | **Preprocessing Technique** | **Accuracy (%)** | **References** |
|---|---|---|---|
| EfficientNet B7 | Data Augmentation (DA-S) | 99.3% | [34] |
| MobileNet | Data Augmentation (DA-S) | 96.4% | [34] |
| ResNet152 Ensemble | Fine-tuning with Ensembles | 99.64% | [6] |
| VGG-13-16-19 Ensemble | Fine-tuning with Ensembles | 97.5% | [5] |
| ResNet50 | Fine-tuning | 98.66% | [52] |
| VGG16 | Fine-tuning | 96.07% | [52] |
| MobileNetV2 | Data Augmentation | 94.0% | [28] |
| VGG19 | Data Augmentation | 92.22% | [28] |
| DenseNet-121 | Feature Fusion | 97.0% | [60] |
| CFDCNet | Feature Fusion | 99.7% | [60] |
| DeepBio (CNN + Bi-LSTM) | Data Augmentation | 98.57% | [10] |
| ResNet50 | Data Augmentation | **99.29%** | **Our work** |

In comparing our ResNet Model to existing models, we have accomplished notable results and made significant contributions in addressing the challenge of human ear print recognition. Our ResNet-Model illustrated an accuracy of 99.29%, with a precision of 99.57%, a recall of 99.29%, and an F1 score of 99.30%. These measurements demonstrate the model's capacity to precisely identify individuals.

We outperformed the ResNet50 model with fine-tuning with our ResNet50-based model, by achieving 99.29% accuracy, compared to the accuracy of 98.66% achieved by fine-tuned ResNet. This improvement comes as a result of our extra enhancements, which boosted the model's capacity to generalize. Layer freezing, a custom classifier head, and advanced augmentation techniques like Mixup and TTA were some of these enhancements.

In comparison to EfficientNet B7, which accomplished 99.3 accuracy, our model performs nearly identically. This result is particularly notable considering EfficientNet's computational complexity compared to our streamlined ResNet50-based architecture.

Additionally, our model outperformed MobileNet, which achieved an accuracy of 96.4%, demonstrating the significance of our fine-tuning strategy and deeper residual connections on the ear print dataset.

Additionally, our model outperformed MobileNet, which achieved an accuracy of 96.4%, demonstrating the significance of our method of selective layer adaptation in conjunction with targeted data augmentation on the ear print dataset.

When assessed against the ResNet152 Ensemble, which comes to a slightly higher accuracy of 99.64%, our model remains competitive while maintaining a strategic distance from the computational overhead presented by ensemble strategies.

Compared to the VGG-13-16-19 Ensemble (97.5%), VGG16 (96.07%), and VGG19 (92.22%), our model showed essentially way better performance, assist approving the superiority of our model in extracting discriminative features significant for biometric identification.

Our model illustrated a critical accuracy gain when compared to MobileNetV2 (94.0%) and DenseNet-121 (97.0%), proposing that our training pipeline and architectural adjustments resulted in more successful learning from the limited data accessible.

Lastly, even though CFDCNet somewhat beat our model with an accuracy of 99.7%, our approach remains highly competitive and simpler to implement and replicate utilizing standard tools and architectures.

Overall, our ResNet50-based model performs well and passes the results of the well-known deep learning models that we implemented. our model achieves a very good accuracy and exceptional precision, recall, and F1-score values. These results illustrate the model's potential for practical biometric applications by highlighting its reliability and effectiveness for human ear print recognition. The comparison study confirms that our training strategies and architectural choices significantly upgrade the quality of its performance.

## 4.5 Evaluation in Real-World Scenarios

### 4.5.1 Real-World Ear Print Data Collection

We tested our ear print recognition system on the AMI dataset, which was chosen due to its real-world variability. Unlike controlled datasets used for training, AMI features ear images from public websites, with differences in lighting, head pose, camera angles, and occlusions caused by hair or accessories—very near real-world conditions.

We employed a bespoke **PyTorch** data loader to eliminate damaged files and provide clean inputs for testing.

This approach allowed us to gauge how effectively the model generalizes outside controlled conditions, providing a glimpse into its real-world performance.

### 4.5.2 Performance Under Challenging Conditions

This section contains some real examples of situations where we tested the ResNet50 model with images from the AMI dataset where we found images of the ear in challenging lighting conditions, objects (in some cases representing accessories), or images of the ear using off-axis shooting angles.

1. **Lighting Conditions**

   In this instance, the model was fed an image that was not well lit. Shadows or dim lighting presented a visual challenge, but the model was able to identify the subject with a respectable level of accuracy.

   Figure 4.16 shows the tested ear that is exposed to strong lighting.

Figure 4.16: Example of light conditions image

2. **Angle and Ear Orientation**

   The case here shows an image of an ear taken from an oblique angle. This type of angle can be difficult for the model; nevertheless, the experiment showed reasonable recognition.

Figure 4.17 shows the ear used for testing, which was taken at an oblique angle.

Figure 4.17: Example of Angle and Ear Orientation image

3. **Accessories and Occlusions**

   An ear with earrings and some hair covering it is seen in the following picture. The model was able to identify and distinguish it in spite of this partial obstruction.

   Figure 4.18 shows the tested ear, partially covered by hair and adorned with earrings.

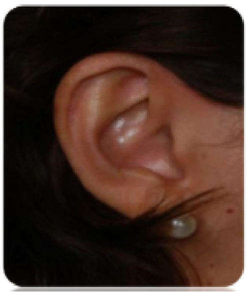

Figure 4.18: Example of ear image with Accessories

### 4.5.3 Accuracy of the Ear Recognition Model Under Challenging Real-World

In Table 4.5 below, we present a summary of the model's performance with respect to various real-world conditions (e.g., low light, with or without accessories, orientations of the ears). These results were based on a small, manually specified sample from the AMI dataset and are not intended for statistical performance interpretation; they are intended to give some indication of performance.

- The accuracy values are based on a small manually selected sample and are indicative, not statistically significant.

Table 4.5: Model performance under different test conditions.

| Condition | Number of Test Images | Correct Predictions | Accuracy (%) | Observations |
|---|---|---|---|---|
| **Low Lighting** | 5 | 3 | 60% | Overall, a satisfactory performance, but the images are rather dark. |
| **Accessories** (Earrings, Glasses) | 5 | 4 | 80% | The model successfully accommodated partial occlusions due to earrings and glasses. |
| **Different Angles** | 5 | 4 | 80% | A slight drop in accuracy when the ear was partially angled. |

### 4.5.4 Generalization Capability and Real-World Simulation

The model's generalization ability was evaluated with the AMI dataset and with ear images acquired more naturally, with less controlled parameters by the human subjects themselves. Especially with regard to the variation of features such as lighting variations, ear angle/orientation differences, personal accessories that could vary whether or not they pose a challenge under the authentication of individuals, and partial occlusions such as hair placement. While the model was only conducted on the AMI dataset, it was able to show a good prospect and ability to be able to recognize individuals given these variations. This evaluation shows some potential ability for the model to interact in these circumstances, like an everyday scenario using security systems, while things like distance to the camera and movement were not evaluated in the current context. The current evaluation offers support for the model to be resilient to moderate variations and everyday outdoor natural settings. Future opportunities may include additional testing and field studies to further these assertions.

### 4.5.5 Limitations and Observations

The model performed well, achieving a strong accuracy of 99% under scenarios that were similar to its training data. It was able to reasonably generalize to other situations where there were moderate differences, for example, lighting conditions, camera angle, and partial occlusion. While the model was trained on a single dataset, the performance was consistent in all 'real-life' situations.

### Limitations

- The system was trained and tested on the same dataset, which could limit the generalisation to other datasets.
- The model was not tested in real-time or deployment situations where users may change their distance, move around, or use the system live.
- The model was not tested in varying backgrounds that may influence performance in uncontrolled or dynamic environments.

## 4.6 System Interface

The main page of the developed GUI is shown in Figure 4.19.

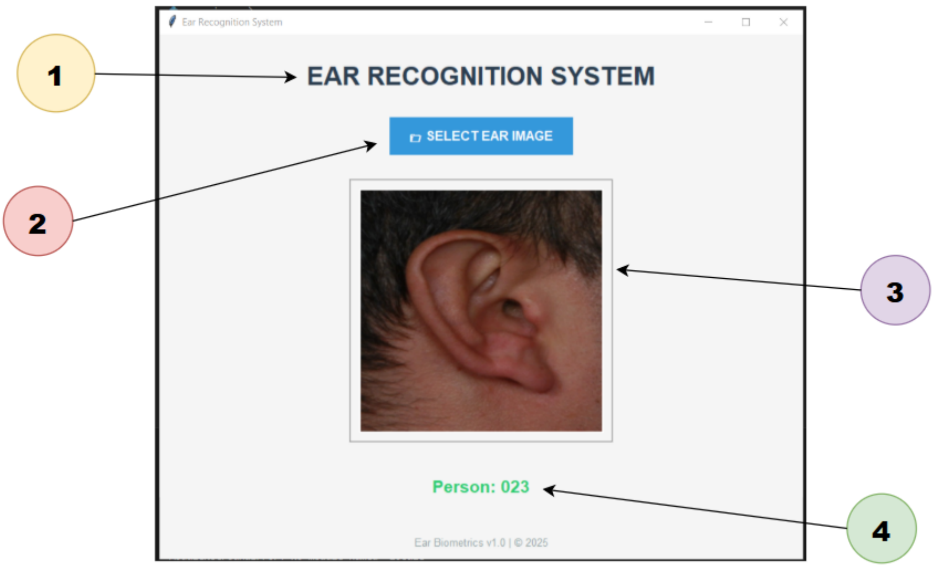


Figure 4.19: Home Page

The graphical user interface (GUI) of our Ear Recognition System is planned to be simple, instinctive, and user-friendly. It permits users to interact with the model effectively for real-time ear recognition. The interface comprises of the following main components:

1. **Title Display** (Label 1): Situated at the top of the window, the title "EAR Recognition SYSTEM" clearly distinguishes the

purpose of the application. It gives users with immediate context about the system.

2. **Image Selection Button** (Label 2): The button labeled "SELECT EAR IMAGE" permits the user to browse and upload an picture of an ear from the local machine. This triggers the recognition process by passing the chosen picture through the trained model.

3. **Image Display Area** (Label 3): After a picture is chosen, it is displayed within the center of the interface. This visual feedback ensures the client that the right picture has been loaded into the system.

4. **Prediction Result** (Label 4): Once the picture is processed, the predicted identity is shown in green text below the picture region, such as "Person:023." This demonstrates the recognition result based on the model's prediction.

The framework's primary functions—picture selection, display, and recognition—are condensed into a simple and easily usable interface. Even for non-technical people, the simplicity ensures ease of use.

## 4.7 Conclusion

In this chapter, we have provided a detailed assessment of the ear print recognition system that we proposed. Upon training and optimizing, the model reached a high percentage of 99,29% accuracy, which shows that we had significant success both with conscious selection of architecture and with the training methods we discussed. We relied on standard performance metrics of precision, recall, and F1-score to evaluate our proposed system, all yielding a good level of accurate individual differentiation capacity.

In spite of the encouraging results from our study, more ongoing work is needed to ensure robustness in more challenging, dynamic situations or variable distances from the camera. Overall, the system showed strong reliability and accuracy, making it a promising biometric solution in the real world.

# Conclusion and Perspectives

In this thesis, we performed a thorough investigation of ear print identification, which has been viewed as a future advanced biometric technique. Our goal was to solve the problems associated with biometric identification with great precision in identity confirmation. We implemented an effective system using deep learning algorithms. We commenced our research with an extensive literature review to explore the importance of ear biometrics and its characteristics and the challenges and research needs essential to develop more accurate and dependable models.

Our system used the AMI database, which consists of images depicting the type of the problems with the diverse ears, different light conditions, existence of the occlusion, and the varying ear orientation, and its imagery was similar to a realistic test condition. For the best performance, we used the ResNet50 model as the base model and implemented transfer learning in order to retain prior knowledge. We proposed augmentation and regularization for this purpose and realized performance improvements. We compared our method systematically with other deep learning models such as the VGG and the EfficientNet. In our work, we observed that ResNet50 provides the best trade- off between classification accuracy and computational cost, achieving 99.29% accuracy, which is higher than the results reported in earlier studies.

Our system demonstrated high resilience to real-world challenges such as different shooting angles and non-ideal lighting, demonstrating the model's robustness and applicability in diverse practical environments. However, we acknowledge that our study was limited to high-quality data, and we have not yet tested the system in motion scenarios or highly variable shooting conditions such as long distances.

As future work, we suggest augment the database with additional earprint images under realistic conditions (with respect to motion blur, low resolution, and distance) to better prepare the model for robustness and generalizability to different conditions. We also want to experiment with advanced deep-learning architectures like Vision Transformers and the optimization of the model for use on constrained devices. We will also consider hardware differences and real-time performance, and we will evaluate the system in practical applications to further strengthen its reliability and efficiency in real-world environments.

# Bibliography


[1] A. Abaza, M. Nixon, and J. Carter. Ear biometrics: A survey of detection, feature extraction and recognition methods. *IET Biometrics*, pages 114–125, 2010.

[2] Ayman Abaza and Arun Ross. Towards understanding the symmetry of human ears: A biometric perspective. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, 2010.

[3] Oussama Aiadi, Belal Khaldi, and Cheraa Saadeddine. Mdfnet: An unsupervised lightweight network for ear print recognition. *Journal of Ambient Intelligence and Humanized Computing*, 14:13773–13786, 2023.

[4] H. Alshazly, C. Linse, E. Barth, and T. Martinetz. Deep convolutional neural networks for unconstrained ear recognition. *IEEE Access*, 8, 2020.

[5] Hammam Alshazly, Christoph Linse, and Ali Awad. Ensembles of deep learning models and transfer learning for ear recognition. *Sensors*, 19(19), 2019.

[6] Hammam Alshazly, Christoph Linse, and Ali Awad. Towards explainable ear recognition systems using deep residual networks. *IEEE Access*, 9:122254–122273, 2021.

[7] Anaconda, Inc. Anaconda: The world's most popular data science platform. `https://www.anaconda.com/`, 2024. Accessed: 2025-05-08.

[8] Julian Ashbourn. *Biometrics: Advanced Identity Verification*. Springer-Verlag London, London, 2000.

[9] Aimee Booysens and Serestina Viriri. Ear biometrics using deep learning: A survey. *Applied Computational Intelligence and Soft Computing*, 2022:1–17, 2022.

[10] C. Booysens, D. Solomons, and A. A. Akinyelu. DeepBio: A deep CNN and Bi-LSTM learning for person identification using ear biometrics. *Computers, Materials & Continua (CMES)*, 141(2), 2024.

[11] Mark Burge and Wilhelm Burger. Ear biometrics. In *Biometric Identification*. Johannes Kepler University, Linz, Austria, 2000.

[12] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Ear biometrics. In *Proceedings of the International Conference on Pattern Recognition*, 2003.

[13] Žiga Emeršič, Vitomir Štruc, and Peter Peer. Ear recognition: More than a survey. *Neurocomputing*, 255:26–39, 2017.

[14] Žiga Emeršič, Vitomir Štruc, Peter Peer, et al. Evaluation and analysis of ear recognition models: Performance, complexity, and resource requirements. *Neural Computing and Applications*, 32(7):15785–15800, 2020.

[15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. Accessed: February 10, 2025.

[16] Google Research. Google colaboratory faq. `https://research.google.com/colaboratory/faq.html?hl=fr`, 2024. Accessed: 2025-05-08.

[17] Google Support. Google drive help. `https://support.google.com/drive/search?q=about+google+drive&from_promoted_search=true`, 2024. Accessed: 2025-05-08.

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

[19] M. Hossin and M. N. Sulaiman. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, pages 1–11, March 2015.

[20] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.

[21] D. J. Hurley, B. Arbab-Zavar, and M. S. Nixon. The ear as a biometric. In Anil K. Jain, Patrick Flynn, and Arun A. Ross, editors, *Handbook of Biometrics*, pages 131–150. Springer, Boston, MA, 2008.

[22] A. V. Iannarelli. *Ear Identification*. Paramount Publishing Company, 1989.

[23] Anil K. Jain, Patrick Flynn, and Arun A. Ross. *Handbook of Biometrics*. Springer, Boston, MA, 2008.

[24] Anil K. Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, January 2004.

[25] Anil K. Jain, Arun A. Ross, and Karthik Nandakumar. *Introduction to Biometrics*. Springer, Boston, MA, 2011.

[26] Aman Kamboj, Rajneesh Rani, and Aditya Nigam. A comprehensive survey and deep learning-based approach for human recognition using ear biometric. *The Visual Computer*, 38:2383–2416, 2022.

[27] Haithem Karrouche. Forest fires prevention employing artificial intelligence. Master's thesis, University of El Tarf, 2023.

[28] A. Kherfi and A. Koadri. Deep transfer learning for ear recognition. Master's thesis, University of M'sila, M'sila, Algeria, 2022.

[29] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 25, pages 1097–1105, 2012.

[30] V. Kumar and Q. Wu. Ear biometrics using gabor filters and svm. *International Journal of Computer Vision*, 2012.

[31] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[32] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition*. Springer Science & Business Media, 2009.

[33] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.

[34] Youssef Mohamed, Nourhan Elbaz, Dahlia Yousri, and Mohamed F Tolba. Advancing ear biometrics: Enhancing accuracy and robustness through deep learning. *arXiv preprint arXiv:2406.00135*, 2024.

[35] Emilio Mordini and Dimitrios Tzovaras. *Second Generation Biometrics: The Ethical, Legal and Social Context*. Springer, Dordrecht, 2012.

[36] National Institute of Standards and Technology (NIST). Iit delhi ear database (version 1.0). `https://tsapps.nist.gov/BDbC/Search/Details/561`. Accessed: 2025-04-16.

[37] NumPy Developers. About — numpy. `https://numpy.org/about/`, 2024. Accessed: 2025-05-08.

[38] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*, pages 8026–8037. Curran Associates Inc., 2019.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[40] Pillow Contributors. Pillow (pil fork) documentation. `https://pillow.readthedocs.io/en/stable/`, 2024. Accessed: 2025-05-07.

[41] Kanika Gupta Verma Sharry Goyal Madhu Sudan Amit Ladgotra Pradhuman Verma, Harpreet Kaur Sandhu. Morphological variations and biometrics of ear: An aid to personal identification. *journal of clinical and diagnostic research*, 2016.

[42] Surya Prakash and Phalguni Gupta. *Ear Biometrics in 2D and 3D: Localization and Recognition*, volume 10 of *Augmented Vision and Reality*. Springer, Singapore, 2015.

[43] Ramar Ahila Priyadharshini, Selvaraj Arivazhagan, and Madakannu Arun. A deep learning approach for person

identification using ear biometrics. *Applied Intelligence*, 51:2161–2172, 2021. Published online: 28 October 2020, Accepted: 1 October 2020.

[44] K.H. Pun and Y.S. Moon. Recent advances in ear biometrics. In *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)*, Seoul, Korea, 2004. IEEE. © 2004 IEEE.

[45] Python Software Foundation. Executive summary: Python is an easy to learn, powerful programming language. `https://www.python.org/doc/essays/blurb/`, 2024. Accessed: 2025-05-08.

[46] PyTorch Core Team. Torchvision: Datasets, transforms and models specific to computer vision. `https://pytorch.org/vision/`, 2024. Accessed: 2025-05-07.

[47] P. Resmi and K. Raju. Deep ear recognition using resnet50 with transfer learning on awe dataset. *Journal of Biometrics and Biostatistics*, 2021.

[48] Arthur L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959.

[49] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[50] Kalaivani Sundararajan and Damon L. Woodard. Deep learning for biometrics: A survey. *ACM Computing Surveys*, 51(3):65:1–65:34, 2018.

[51] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning (ICML)*, 2019.

[52] Jing Yee Tey, Rosalina Salleh, and Roliana Ibrahim. Comparative analysis of VGG-16 and ResNet-50 for occluded ear recognition. *International Journal on Informatics Visualization (JOIV)*, 7(4), 2023.

[53] Thales Group. Biometrics and identity: A brief history of biometrics, 2023. Accessed: February 9, 2025.

[54] Ž. Emeršič, A. Čuček, and M. Mihelj. Training convolutional neural networks with limited training data for ear recognition in the wild. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017.

[55] James L. Wayman, Anil K. Jain, Davide Maltoni, and Dario Maio. *Biometric Systems: Technology, Design and Performance Evaluation.* Springer, 2005.

[56] A. Xu, R. Bouaouina, H. Doghmane, and Y. Brik. Ear recognition using mean-cam-cnn with image enhancement by dcgans. *Sensors*, 24, 2024.

[57] Ping Yan and Kevin W. Bowyer. Biometric recognition using 3d ear shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1297–1308, August 2007.

[58] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 27, 2014.

[59] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations (ICLR)*, 2018.

[60] J. Zhang, L. Li, and C. Chen. A feature fusion human ear recognition method based on channel features and dynamic convolution. *Symmetry*, 15(7), 2023.

[61] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[62] Y. Zhang, Z. Mu, L. Yuan, and C. Yu. Ear verification under uncontrolled conditions with convolutional neural networks. *IET Biometrics*, pages 185–198, 2018.

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليـــم العالـــي والبحـــث العلمـــــي

جامعة غرداية

كلية العلوم والتكنولوجيا
قيم الرياضيات والاعلام الالي

غرداية في :2025/07/13

**شعبة: اعلام الي**
**تخصص: الأنظمة الذكية لاستخراج المعارف**

# شهادة ترخيص بالتصحيح والإيداع

أنا الأستاذ : بن قنان مسعود

الرتبة: أستاذ مساعد

بصفتي المشرف المسؤول عن تصحيح **مذكرة التخرج ماستر** المعنونة بـ :
Human ear print recognition using deep learning techniques

من إنجاز الطالبتان :

- GUERRADI Dallal
- HADJ KOUIDER Afrah

التي نوقشت بتاريخ: 2025/07/02

أشهد أن الطلبة قد قاموا بالتعديلات والتصحيحات المطلوبة من طرف لجنة المناقشة وقد تم التحقق من ذلك من طرفنا وقد استوفت جميع الشروط المطلوبة.

إمضاء مسؤول عن التصحيح

Benguenane Messaoud

مصادقة رئيس القسم

رئيس قسم الرياضيات والإعلام الآلي
الحاج موسى ياسين

جامعة غرداية
قسم الرياضيات والاعلام الآلي
كلية العلوم والتكنولوجيا