Ministry of Higher Education and Scientific Research

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba
Faculty of Technology

Electronics Department

جامعة باجي مختار – عنابـــة

كـليـة التكنـولـوجيـا

قســم الإلكترونيك

Thesis

Submitted to obtain the diploma of

Doctorate Third Cycle

Field : Electronics

Speciality : Security and Biometrics

By :

ZEHIR Hatem

Title :

Elaboration d'un système biométrique multimodale hybride

Thèse soutenue le 09/07/2025 devant le jury composé de :

N°	Name and surname	Grade	Establishment	Quality
01	GHERBI Sofiane	Prof.	Badji Mokhtar Annaba University	President
02	**HAFS Toufik**	MCA	Badji Mokhtar Annaba University	Supervisor
03	DAAS Sara	MCB	Badji Mokhtar Annaba University	Co-supervisor
04	ZERMI Narima	MCA	Badji Mokhtar Annaba University	Examiner
05	OUCHTATI Salim	Prof.	Université 20 Août 1955 -Skikda	Examiner
06	BOUROUBA Hocine	Prof.	Université 8 Mai 1945 - Guelma	Examiner

Ministère de l'enseignement Supérieur et de la recherche Scientifique

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba
Faculté de Technologie

Département Electronique

جامعة باجي مختار – عنابـــة

كـليـة التكنـولـوجيـا

قســم الإلكترونيك

Thèse

Présentée pour obtenir le diplôme de

Doctorat Troisième Cycle

Filière : Electronique

Spécialité : Sécurité et biométrie

Par :

ZEHIR Hatem

Thème :

Elaboration d'un système biométrique multimodale hybride

Thèse soutenue le 09/07/2025 devant le jury composé de :

N°	Nom et prénom	Grade	Etablissement	Qualité
01	GHERBI Sofiane	Prof.	Université Badji Mokhtar -Annaba	Président
02	**HAFS Toufik**	MCA	Université Badji Mokhtar -Annaba	Rapporteur
03	DAAS Sara	MCB	Université Badji Mokhtar -Annaba	Co-rapporteur
04	ZERMI Narima	MCA	Université Badji Mokhtar -Annaba	Examinateur
05	OUCHTATI Salim	Prof.	Université 20 Août 1955 -Skikda	Examinateur
06	BOUROUBA Hocine	Prof.	Université 8 Mai 1945 - Guelma	Examinateur

" تطوير نظام بيومتري هجين متعدد الوسائط "

الملخص:

تتناول هذه الأطروحة تطوير نظام بيومتري متعدد الوسائط يجمع بين بيانات تخطيط القلب الكهربائي ECG والصوت بهدف تحسين دقة وموثوقية التعرف على الأفراد. الهدف من البحث هو التغلب على قيود الأنظمة البيومترية أحادية الوسيط من خلال دمج وسائط مكملة، مما يوفر تحسينا للأداء العام. تعتمد الدراسة على الخصائص الفريدة للإشارات الكهربائية للقلب والخصائص الصوتية، مما يجعلها مثالية لأنظمة التحقق البيومترية الآمنة.

شملت المنهجية نظامين أحاديي الوسائط منفصلين للتعرف على تخطيط القلب الكهربائي والصوت. استخدم النظام القائم على تخطيط القلب نموذجًا يعتمد على التعلم العميق، بما في ذلك معماريات GRU وLSTM ، وتم تدريبه على ثلاث قواعد بيانات رئيسية لتخطيط القلبMIT-BIH ، NSRDB، و .PTB أما نظام التعرف الصوتي فقد اعتمد على نموذج يعتمد على الشبكات العصبية التلافيفيةCNN، وتم تدريبه على مجموعة فرعية من47 متحدثا من قاعدة بيانات .LibriSpeech قام كلا النظامين باستخراج ميزات ذات صلة من البيانات مثل MFCCللصوت و IMFs لتخطيط القلب (بعد تنفيذ مراحل المعالجة المسبقة المناسبة. تم بعد ذلك تطوير نظام بيومتري متعدد الوسائط من خلال محاكاة مجموعة بيانات تجمع بين عينات صوتية وبيانات تخطيط القلب، حيث تم استخدام تقنيات دمج مثل Softmax و SVMلدمج مخرجات الأنظمة أحادية الوسائط.

أظهرت النتائج الرئيسية أن نظام تخطيط القلب القائم على GRU حقق دقة بلغت 98.57% على قاعدة بيانات MIT-BIH ، متفوقا على نموذج LSTM. وحقق نظام التعرف الصوتي القائم على CNN دقة بلغت 98.42% ومع ذلك، كانت النتيجة الأكثر أهمية هي النظام متعدد الوسائط، حيث حقق دمج درجات تخطيط القلب والصوت باستخدام Softmax مع قاعدة الجمع دقة عالية بلغت 99.61%، مع معدل خطأ متساوي EER بلغ 0.22%. تؤكد هذه النتائج أن النهج متعدد الوسائط يوفر أداءً أفضل مقارنة بالأنظمة أحادية الوسائط.

كلمات مفتاحية: التعرف البيومتري، القياسات الحيوية متعددة الوسائط، التعرف المستند إلى تخطيط كهربية القلب، التعرف على المتحدث، التعلم العميق، GRU، LSTM، CNN، MFCC، تحليل الوضع التجريبي، دمج النتائج، Softmax، آلة دعم المتجهات

« Elaboration d'un système biométrique multimodale hybride »

Résumé :

Cette thèse examine le développement d'un système biométrique multimodal qui intègre des données d'électrocardiogramme (ECG) et de voix afin d'améliorer la précision et la fiabilité de l'identification des personnes. L'objectif de la recherche est de pallier les limites des systèmes biométriques unimodaux en combinant des modalités complémentaires, offrant une amélioration des performances globales. L'étude exploite les caractéristiques uniques des signaux ECG et des caractéristiques vocales, connus pour leur résilience face aux changements environnementaux et aux conditions de santé, ce qui les rend idéaux pour les systèmes d'authentification biométrique sécurisés.

La méthodologie a impliqué deux systèmes unimodaux distincts pour la reconnaissance ECG et vocale. Le système basé sur l'ECG a utilisé un modèle d'apprentissage profond, notamment les architectures GRU et LSTM, entraîné sur trois bases de données ECG majeures (MIT-BIH, NSRDB, et PTB). Le système de reconnaissance vocale a quant à lui employé un modèle basé sur des CNN, entraîné sur un sous-ensemble de 47 locuteurs de la base de données LibriSpeech. Les deux systèmes ont extrait des caractéristiques pertinentes des données (MFCC pour la voix et IMFs pour l'ECG) après les étapes de prétraitement appropriées. Un système biométrique multimodal a ensuite été développé en simulant un ensemble de données combinant des échantillons vocaux et ECG, avec des techniques de fusion de scores telles que Softmax et SVM pour fusionner les sorties des systèmes unimodaux.

Les principaux résultats montrent que le système ECG basé sur GRU a atteint une précision de 98,57% sur la base de données MIT-BIH, surpassant le modèle LSTM. Le système de reconnaissance vocale basé sur CNN a atteint une précision de 98,42%. Cependant, le résultat le plus significatif provient du système multimodal, où la fusion des scores ECG et voix utilisant Softmax avec la règle de somme a obtenu une haute précision de 99,61%, avec un EER de 0,22%. Ces résultats confirment que l'approche multimodale offre des performances supérieures par rapport aux systèmes unimodaux.

Mots clés : Identification biométrique, Biométrie multimodale, Identification basée sur l'ECG, Reconnaissance du locuteur, Apprentissage profond, GRU, LSTM, CNN, MFCC, Décomposition modale empirique, Fusion de scores, Softmax, Machine à vecteurs de support

« Development of a hybrid multimodal biometric system »

Abstract :

This thesis investigates the development of a multimodal biometric system that integrates electrocardiogram (ECG) and voice data to enhance the accuracy and reliability of person identification. The research aims to address limitations in unimodal biometric systems by combining complementary modalities, offering greater robustness against spoofing and improving overall performance. The study leverages the unique characteristics of both ECG signals and voice features, which are known to be resilient to environmental changes and health conditions, making them ideal for secure biometric authentication systems.

The methodology involved two distinct unimodal systems for ECG and speaker recognition. The ECG-based system utilized a deep learning model, specifically GRU and LSTM architectures, trained on three major ECG databases (MIT-BIH, NSRDB, and PTB). The speaker recognition system employed a CNN-based model trained on a subset of 47 speakers from the LibriSpeech dataset. Both systems extracted meaningful features from the data (MFCC for voice and IMFs for ECG) after appropriate preprocessing steps. A multimodal biometric system was then developed by simulating a dataset combining voice and ECG samples, with score fusion techniques like Softmax and SVM used to merge the outputs of the unimodal systems.

The key findings demonstrate that the GRU-based ECG system achieved an accuracy of 98.57% on the MIT-BIH database, outperforming the LSTM model. The CNN-based speaker recognition system reached an accuracy of 98.42%. However, the most significant result came from the multimodal system, where the fusion of ECG and voice scores using Softmax with the Sum rule yielded a high accuracy of 99.61%, with an EER of 0.22%. These results confirm that the multimodal approach provides superior performance compared to unimodal systems.

Key words : Biometric identification, Multimodal biometrics, ECG-based identification, Speaker recognition, Deep learning, GRU, LSTM, CNN, MFCC, Empirical mode decomposition, Score fusion, Softmax, Support vector machine

I would like to dedicate this thesis to my family, whose unwavering support, patience, and encouragement have been my constant source of strength throughout this journey. To my parents, for their boundless love and belief in my abilities, for their guidance and inspiration. This work is as much a reflection of their contributions as it is of my own.

Acknowledgment

As I reach the end of this journey, I find myself reflecting on the countless individuals who have contributed to its success. Completing this thesis has been a long and challenging process, one that I could never have accomplished alone. It is with immense gratitude that I acknowledge those who have walked beside me, guided me, and supported me throughout these years.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. HAFS Toufik, for his invaluable guidance, unwavering support, constructive suggestions, and patience throughout the entirety of this work. His expertise and encouragement have been instrumental in the completion of this thesis.

I would also like to extend my thanks to my co-supervisor, Dr. DAAS Sara, for her constant support and insightful feedback throughout the entirety of this work during this research. Her advice has greatly enriched the quality of my work.

I extend my sincere thanks to the directors of the LERICA laboratory, to Pr. SAOUCHI Kadour, Pr. BOUGHAZI Mohamed, and Dr. BOUKARI Karima for their commitment to the laboratory. I truly appreciate the support they provided throughout my time at LERICA.

I also wish to extend my appreciation to my fellow members at LERICA, whose camaraderie and shared knowledge have been an invaluable part of this experience.

To my family—my mother, father, brother, and sister—I owe the greatest thanks. Their love, sacrifices, and encouragement have been the bedrock upon which I have built this journey. Without their constant belief in me, this achievement would not have been possible. This accomplishment is as much yours as it is mine.

I am also deeply grateful to my friends who believed in me even when I doubted myself, thank you. Their presence has made this journey more fulfilling.

Finally, I want to thank everyone who has contributed, in one way or another, to the completion of this thesis. Your help, no matter how small, has been appreciated beyond measure.

List of Figures

List of Tables

List of Abbreviations

Acronymes

1D CNN	One-dimensional convolutional neural network
AFIS	Automated fingerprint identification system
ANN	Artificial neural network
BGRU	Bidirectional GRU
CDS	Cosine distance scoring
CGPANN	Cartesian genetic programming evolved artificial neural network
CIR	Correct identification rate
CNIBE	Carte nationale d'identité biométrique electronique
CNN	Convolutional neural networks
CYBHi	Check your biosignals
db8	Daubechies wavelets
dB	Decibel
DCA	Discriminant correlation analysis
DCT	Discrete cosine transform
DEC	Deep embedded clustering
DIHARD	Diverse recordings with varying speakers and conditions
DL	Deep learning
DNA	Deoxyribonucleic acid

DST	Dempster-Shafer theory
DT	Decision tree
DTW	Dynamic time warping
DWT	Discrete wavelet decomposition
ECG	Electrocardiogram
EEG	Electroencephalogram
EER	Equal error rate
E-FLF	Eigen-based feature-Level fusion
EMD	Empirical mode decomposition
FAR	False acceptance rate
FFT	Fast Fourier transform
FIR	Finite impulse response
FRR	False rejection rate
GAR	Genuine acceptance rate
GDF	Global decision fusion
GRU	Gated recurrent unit
HMM	Hidden Markov models
HR	Heartbeat recognition
IC	Integrated circuits
IGR	Information gain ratio
IMF	Intrinsic mode function
KNN	K-nearest neighbor
LDA	Linear discriminant analysis
LDF	Local decision fusion

LGDF	Local-global decision fusion
LP	Linear prediction
LPQ	Local phase quantization
LSTM	Long short-term memory
MFCC	Mel frequency cepstral coefficient
ML	Machine learning
MLP	MultiLayer perceptron
NRC	Normalized relative compression
PCA	Principal component analysis
PIN	Personal identification numbers
PLDA	Probabilistic linear discriminant analysis
PLDA	Probabilistic linear discriminant analysis
RBF	Radial basis function
RNN	Recurrent neural networks
SIMCA	Soft independent modeling of class analogy
SI	Subject identification
SITW	Speakers in the wild
STE	Short-term energy
STZ	Short-term zero-crossing rate
SVM	Support vector machine
UofTDB	University of Toronto ECG database
xaFCMs	Extended-alphabet finite-context models

General Introduction

We live in an era where concerns about security, privacy, and identity theft are paramount, and the development of robust and dependable biometric systems has become increasingly crucial. Traditional unimodal biometric systems, reliant on single physiological or behavioral traits for identification, often encounter significant challenges, including susceptibility to spoofing attacks and limited performance in noisy or adverse environmental conditions. Consequently, there has been a notable surge in interest within the scientific community towards multimodal biometric systems, which leverage multiple modalities to enhance security, accuracy, and reliability in identity verification processes.

Despite the advancements made in the field of multimodal biometrics, significant gaps persist in our understanding of the optimal integration of diverse biometric modalities, particularly regarding the fusion techniques employed, and the application of deep learning methodologies within this context. Thus, the primary research problem addressed in this study revolves around the need to investigate the efficacy of integrating voice and electrocardiogram (ECG) modalities through score-level fusion mechanisms, augmented by deep learning techniques. Bridging this gap is critical for advancing the field of biometrics and fostering the development of more robust and secure identity verification systems.

This study is underpinned by the theoretical framework of biometrics, a multidisciplinary field encompassing principles from electronics, computer science, statistics, and signal processing, among others. Within this framework, the concept of multimodal fusion theory serves as a guiding principle for combining information from diverse sources to enhance identification accuracy. The purpose of this research is to develop and evaluate a multimodal biometric system integrating voice and ECG signals using score-level fusion and deep learning techniques. By exploring the synergistic effects of combining these modalities at the score level and leveraging deep learning models for classification and fusion, this study aims to enhance the accuracy, robustness, and security of biometric identification systems.

This study adopts the following research method, employing experimental tests and

analysis to evaluate the performance of the proposed multimodal biometric system. The primary research question guiding this investigation is: "How does integrating voice and ECG modalities using score-level fusion and deep learning techniques affect the accuracy and reliability of biometric identification?" Hypotheses will be formulated to test the effectiveness of the proposed system in comparison to unimodal and other multimodal approaches. The significance of this study lies in its potential to advance the field of biometrics by providing insights into the effectiveness of multimodal fusion techniques and deep learning methodologies for enhancing biometric identification systems. The findings of this research can inform the development of more secure and reliable identity verification systems, with applications ranging from access control and surveillance to healthcare and finance.

This study focuses specifically on integrating voice and ECG modalities using score-level fusion and deep learning techniques within the context of biometric identification. Other modalities and fusion techniques may offer alternative approaches but are beyond the scope of this research. This study assumes the availability of sufficient and reliable voice and ECG datasets for training and evaluation purposes. Also, it assumes the deep learning models employed in the study will be appropriately configured and optimized for biometric identification.

The thesis contributes to biometrics by exploring the integration of voice and ECG modalities using score-level fusion and deep-learning techniques in multimodal biometric systems. This contribution is substantiated by a series of published and submitted papers and conference presentations, which underscore the innovative strides made in advancing biometric authentication methodologies. The contributions can be summarized as follows:

- [1]: **H. Zehir**, T. Hafs, and S. Daas, “Empirical mode decomposition-based biometric identification using GRU and LSTM deep neural networks on ECG signals”, Evolving Systems, vol. 15, no. 6, pp. 2193–2209, August 2024.

- [2]: **H. Zehir**, T. Hafs, and S. Daas, “Involutional neural networks for ECG spectrogram classification and person identification”, International Journal of Signal and Imaging Systems Engineering, vol. 13, no. 1, pp. 41-53, July 2024.

- [3]: **H. Zehir**, T. Hafs, S. Daas, and A. Nait-ali, “Support Vector Machine for Human Identification Based on Non-Fiducial Features of the ECG”, Journal of Engineering Studies and Research, vol. 29, no. 1, pp. 61-69, May 2023.

- [4]: **H. Zehir**, T. Hafs, and S. Daas, “TinyCNN: An Embedded CNN Model for Speaker Identification Using ESP32”, ICEERES’23: The 1st International Conference on Electrical Engineering & Renewable Energies Systems, Bechar, Algeria,

November 2023.

- [5]: **H. Zehir**, T. Hafs, and S. Daas, "ECG-Based Biometric System using TinyML: Implementation and Performance Evaluation on ESP32", ICAECCT'23: The 1st International Conference on Advances in Electronics, Control and Computer Technologies, Mascara, Algeria, October 2023.

- [6]: **H. Zehir**, T. Hafs, and S. Daas, "Healthcare Decision-Making with an ECG-Based Biometric System", DASA'23: The 2023 International Conference On Decision Aid Sciences And Applications, Annaba, Algeria, September 2023.

- [7]: **H. Zehir**, T. Hafs, S. Daas, and A. Nait-ali, "An ECG Biometric System Based on Empirical Mode Decomposition and Hilbert-Huang Transform for Improved Feature Extraction", BioSMART2023: 5th International Conference on Bio-engineering for Smart Technologies, Paris, France, June 2023.

- [8]: **H. Zehir**, T. Hafs, and S. Daas, "Bidirectional Long Short-term Memory Neural Networks Based Electrocardiogram Biometric System", ICESTI'22: International Conference on Embedded Systems in Telecommunications and Instrumentation, Annaba, Algeria, Dec. 2022.

Collectively, these publications and conference presentations underscore the thesis's main contribution to the field of biometric identification through the integration of voice and ECG modalities using novel signal processing techniques and deep learning methodologies. By disseminating these findings to the academic community, the thesis aims to contribute to further research and development in the aim for more secure, reliable, and efficient biometric identification systems.

Following this introduction, the thesis will proceed with a literature review examining existing research on multimodal biometric systems, score-level fusion techniques, and deep learning applications in biometrics. Subsequent chapters will detail the methodology, present the results of experimental evaluations, discuss the findings, and conclude with recommendations for future research.

Chapter 1

Overview of Biometric Systems

1.1 Introduction

Biometric systems have emerged as a critical technology in enhancing security, identification, and authentication processes in both public and private sectors. As reliance on traditional methods of identity verification, such as passwords or ID cards, has proven to be susceptible to fraud, biometrics offers a more secure and reliable solution. By using physiological or behavioral characteristics unique to each individual, biometric systems promise to improve identity management across diverse applications, including border control, banking, and healthcare.

The field of biometrics encompasses various modalities, each leveraging distinct biological traits, such as fingerprints [9, 10], facial recognition [11, 12], and iris scans [13, 14]. These technologies have evolved over time, driven by advancements in pattern recognition, machine learning (ML) [15], deep learning (DL) [16, 17], and computer vision [18]. The history of biometrics stretches back centuries, but it has seen unprecedented development in recent decades, especially with its integration into digital infrastructures. Globally, countries are increasingly adopting biometric systems for national identification, and Algeria is no exception, demonstrating the growing importance of this technology in governance and security.

This chapter of the thesis explores the comprehensive architecture of biometric systems, from the initial data acquisition to the final decision-making phase. Each phase plays a crucial role in ensuring the accuracy and robustness of biometric identification and authentication. Furthermore, the integration of biometric systems within existing infrastructures requires thoughtful design and evaluation to meet the required standards of performance and security. We will also discuss the different types of biometric recognition, the architecture of biometric systems, and the phases involved in processing biometric data. Moreover, this introduction will delve into how Algeria has incorporated biometrics into its national systems, and we will conclude by exploring the metrics used to evaluate the performance of biometric systems.

1.2 Biometric Systems

Today, with the increase in computer processing power and the incredible progress made in fields such as artificial intelligence [19, 20], DL [19], and sensor capabilities [21], biometric systems have become more accurate than ever before. Additionally, due to the advances in integrated circuits (IC) and semiconductor manufacturing techniques [22], these systems have become a part of our everyday lives. We can find them in most of our devices, such as smartphones and smartwatches.

1.2.1 Definition of Biometrics

We can define biometrics as the science of applying different statistical approaches and computer science techniques on measured human characteristics with identifying or verifying individuals as the main purpose [23]. The term "biometrics" has Greek origins. It is divided into two parts: "bios", which means "life", and "metron", or "metrein", which means "to measure" [24].

1.2.2 A Brief History of Biometrics

The use of biometrics as a means of identification or authentication dates to the Babylonian civilization, where they signed business transactions and contracts by using fingerprints on clay tablets [25] and identified criminals [26]. During the Qin Dynasty (221–206 BCE), which ruled ancient China, handprints were used by the Qin Dynasty administration to authenticate contracts and documents, according to historical archives [27].

In the 1860s, William James Herschel became the first scientist to utilize fingerprints in a practical way [28]. Later, Francis Galton's research in the field of forensic science led him to discover a new method for classifying fingerprints [28].

Biometrics, as we know it today, didn't begin until the electronics revolution started in the 1960s with the development of computer technology, which allowed for automated user recognition from biometric traits. Since that date, the field of biometrics has undergone rapid development:

- In the 1960s: The initial experiments with computer-based facial recognition algorithms began, albeit with limited processing power and accuracy [29].
- In the 1970s: the first Automated Fingerprint Identification System (AFIS) is developed [30].
- In the 1980s: The first speaker identification system was developed [31].
- In the 1990s: the implementation of iris recognition technology was initiated [32].
- In the 2000s: during this decade, a significant focus was placed on establishing national-level security policies due to ongoing security concerns [33, 34]. Biometric systems experienced an upsurge in adoption across diverse sectors, including law enforcement [35], border control [36], and financial services [37]. Additionally, the integration of biometric systems into mobile devices [38, 39], such as smartphones [40–42], became prevalent for user authentication purposes.

- In the 2010s: to overcome the problems of existing unimodal biometric systems, researchers started to focus on combining multiple modalities [43–46].
- In the 2020s: More companies adopted biometric systems due to the COVID-19 pandemic [47].
- Today: the recent innovations in biometrics encompass the incorporation of deep-learning algorithms such as convolutional neural networks (CNN) [4, 48–52], long short-term memory (LSTM) [8, 53, 54], and transformers [55–57], which have contributed to significant improvements in system performance and accuracy.

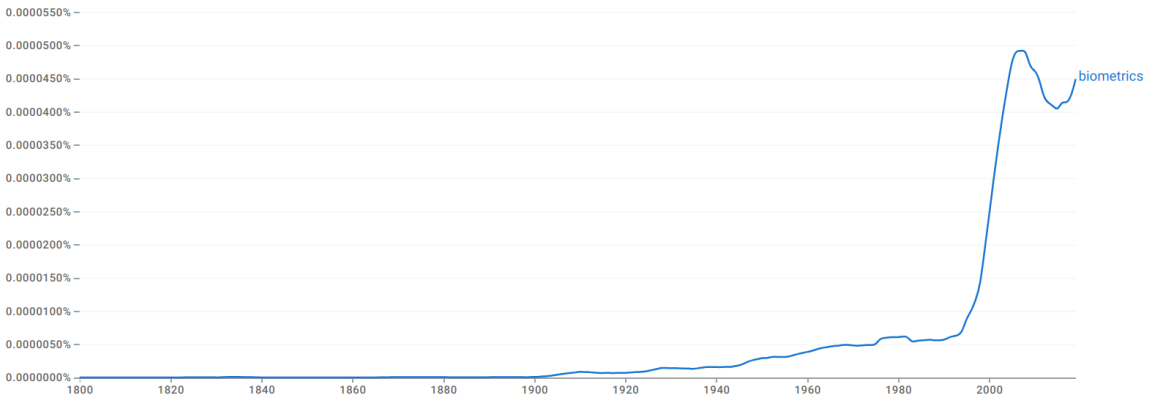


Figure 1.1: The Linguistic Footprint of Biometrics: Examining Historical Trends with Google Ngram (1800-2019) [58].

Figure 1.1 extracted from Google Ngram demonstrates that the popularity of the term "biometrics" started to accelerate in the late 1990s and early 2000s. This is due to several factors:

- Security Concerns: Traditional authentication methods like passwords and personal identification numbers (PIN) are vulnerable to theft, loss, or hacking. Biometrics offers a more secure alternative by using unique biological traits such as fingerprints, iris patterns, or facial features for identity verification.
- Convenience: Users can access devices, systems, or services more conveniently when they utilize biometric authentication, as it does away with the need to carry physical tokens like ID cards or memorize passwords.
- Technological Advancements: Advances in sensor technology, image processing algorithms, and computing power have made biometric systems more accurate, reliable, and cost-effective over time, facilitating their widespread adoption.

- Regulatory Requirements: increasing regulations and requirements relating to security and identification verification, particularly in industries like border control, government, and finance, have driven the adoption of biometric solutions to ensure compliance and protect sensitive data.
- User Acceptance: With the proliferation of biometric-enabled devices like smartphones and laptops, users have become more familiar and comfortable with biometric authentication, leading to greater acceptance and demand for biometric solutions.

1.2.2.1 In Algeria

The history of biometrics in Algeria can be traced back to the efforts of the Algerian Ministry of Interior in modernizing and enhancing the security features of identification documents. The following points are a brief overview of the evolution of biometrics in Algeria:

- In 2012, the Algerian Ministry of Interior created the Direction of Secured Titles and Documents. This department was tasked with personalizing secured biometric documents and titles in accordance with international standards and recommendations, such as those outlined in ICAO Doc 9303. It has two data centers: the primary one is located in Algiers, and the backup one is 450 km away in Laghouat [59].
- The first Algerian biometric ePassport was delivered in January 2012 [60]. This marked the beginning of the transition from traditional passports to biometric ePassports in Algeria. By November 2015, all classic passports had been converted into biometric ePassports [61].
- In March 2016, the first Algerian Biometric eID, known as the Carte Nationale d'Identité Biométrique Electronique (CNIBE), was introduced [62].
- Over the years, the Algerian Ministry of Interior has significantly expanded its biometric document delivery capabilities. By 2019, more than 14 million ePassports and 16 million CNIBEs had been delivered to Algerian citizens [63].
- In 2019, the transition from conventional driving licenses to biometric driving licenses was implemented as a component of the Algerian modernization initiative aimed at enhancing data security and information integrity [64].

1.3 Types of Biometric Recognition

Biometric systems are designed to perform two fundamental tasks: identification and authentication. These tasks, while closely related, serve distinct purposes and are essential for the secure and effective use of biometric technology.

1.3.1 Identification

Identification involves determining the identity of an individual from a biometric sample by comparing it against a database of stored biometric templates. In this mode, the system attempts to find a match among all possible users, identifying who the person is based on the provided biometric data. For example, if a user provides a fingerprint, the system will compare it against all fingerprints in the database to find the closest match. Identification is typically used in scenarios where the identity of the user is unknown or needs to be verified from a large set of potential candidates. The accuracy and efficiency of identification systems are critical, as they must quickly and reliably match the biometric data to the correct individual out of potentially millions of entries.

1.3.2 Authentication

Authentication, on the other hand, verifies an individual's claimed identity by comparing their biometric sample against a stored template specific to that individual. This process confirms whether the provided biometric data matches the pre-enrolled sample associated with the claimed identity. For instance, when a user logs into a system using their voice, the system checks the provided biometric data against the template that was previously registered for that user. Authentication is used in scenarios where the identity is already known and needs to be confirmed, such as logging into a personal account or accessing secure areas. It typically involves a one-to-one comparison and focuses on verifying that the individual is who they claim to be, rather than identifying them from a large population.

1.4 Biometric Modalities

Biometric modalities can be classified into three main categories: physiological, behavioral, and hidden traits. Physiological traits encompass characteristics inherent to an individual's anatomy, such as fingerprints [43], palmprints [65], facial features [66], and iris patterns [67]. These traits are relatively stable over time and difficult to alter, making them well-suited for long-term identification. Behavioral traits, on the other hand, comprise patterns of behavior or actions, including voice [68], handwriting [69], gait [70],

and keystroke dynamics [71]. While behavioral traits may exhibit greater variability, they offer the advantage of being non-invasive and easy to capture. Hidden traits, a relatively newer classification, refer to biometric identifiers that are not consciously controlled or observed, such as ECG [6], electroencephalogram (EEG) [72], or deoxyribonucleic acid (DNA) [73]. These traits offer potential advantages in terms of security and accuracy but may present challenges in terms of practical implementation and accessibility. Figure 1.2 illustrates a comprehensive classification of biometric modalities, highlighting the diverse range of physiological, behavioral, and hidden traits utilized in biometric authentication systems.

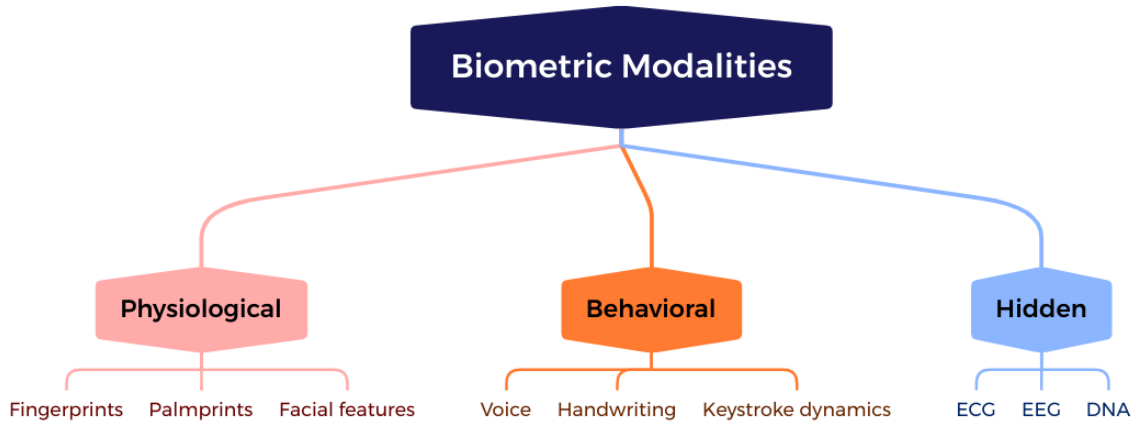


Figure 1.2: Classification of Biometric Modalities: Physiological, Behavioral, and Hidden Traits.

Despite the diversity of biometric modalities discussed earlier, each modality must satisfy certain key properties to be considered effective and reliable for identification and authentication purposes. These properties ensure that the modality can be practically applied in real-world scenarios, providing both security and convenience. The primary properties that each biometric modality must satisfy include the following:

1. **Universality**: This property refers to the requirement that every individual must possess the biometric trait in question. For a biometric modality to be useful, it must be applicable to the entire target population. For instance, fingerprints are nearly universal, as almost everyone has them, with exceptions being extremely rare. Universality ensures that the biometric system can enroll and recognize all users without excluding any individual based on the absence of the biometric trait.

2. **Uniqueness**: This property implies that the biometric trait must be sufficiently distinctive to differentiate one individual from another. Each instance of the biometric trait should be different across individuals, ideally making it impossible for

two people to share the same biometric characteristic. For example, the patterns of ridges and valleys in fingerprints are unique to each person, even among identical twins. High uniqueness is critical for ensuring that biometric systems can accurately distinguish between users.

3. **Permanence**: The biometric trait's long-term stability is defined by this feature. A biometric modality should exhibit minimal changes throughout an individual's life, ensuring that once enrolled, the system can consistently recognize the individual over long periods. For instance, voice tends to be relatively stable but can change due to age, health conditions, or even vocal strain. ECG signals, while generally stable, can also be affected by factors such as heart conditions, medications, or significant lifestyle changes. A high degree of permanence reduces the need for frequent re-enrollment or updating of biometric templates.

4. **Collectability**: This characteristic, which is often referred to as measurability, has to do with how simple it is to measure and capture the biometric trait for processing. The biometric trait should be easily measurable using available technology, providing sufficient quality and consistency for the system to function effectively. For example, voice is highly measurable, requiring only a microphone for data collection, making it accessible and non-invasive. ECG collection is also feasible, especially with modern wearable technology, though it typically requires contact with the skin through electrodes. Measurability is essential for both the enrollment process and the day-to-day operation of the biometric system.

5. **Performance**: This characteristic refers to the biometric system's reliability, speed, and accuracy in real-world applications. A high-performing biometric modality should have low error rates, while also offering quick and reliable identification or verification. The quality of the biometric data that is gathered and the matching algorithms that are employed all have an impact on performance.

6. **Acceptability**: This property refers to the degree to which users are willing to adopt and use the biometric system. For a biometric modality to be viable, it must be non-invasive, culturally appropriate, and not perceived as a threat to privacy. Behavioral modalities like voice recognition tend to have higher acceptability, as they are generally less invasive compared to modalities like retinal scans.

7. **Circumvention Resistance**: It is the biometric modality's ability to withstand attempts at spoofing or bypassing the system. A secure biometric system should be resilient against fraudulent attacks, such as using a fake fingerprint or a recorded

voice. However, the biological nature of ECG makes it challenging to forge, providing strong circumvention resistance.

To guarantee optimal performance and user acceptance, it is imperative to carefully analyse the features of various biometric modalities while creating a biometric system. A comparison of several biometric modalities based on important parameters including Universality, Uniqueness, Permanence, Collectability, Acceptability, and Performance is shown in Table 1.1.

Table 1.1: Comparison of various biometric modalities.

Biometric Modality	**Universality**	**Uniqueness**	**Permanence**	**Collectability**	**Acceptability**	**Performance**
Fingerprint	High	High	Moderate (Scars may affect)	High	High	High
Face	High	High (Except identical twins)	Moderate (Aging, facial hair)	High	High	Moderate (Lighting, pose variations)
Iris	High	High	High	High	High	High
Finger Knuckle Print	High	High	Moderate (Injuries)	High	High	Moderate (Sensor quality)
Voice	High	Moderate (Speaker identification easier than verification)	Low (Voice can change with age, illness)	High	High	Moderate (Background noise, accents)
Handwriting	High	Moderate (Can be similar within families)	Low (Can change with practice)	High	High	Moderate (Depends on quality of sample)
Signature	High	Moderate (Can be forged)	Low (Can change deliberately)	High	High (For legal documents)	Low
DNA	High	Extremely High	High	Moderate (Invasive for some collection methods)	High (For legal purposes)	High
ECG	High	Moderate (Can vary with activity level)	High	High	Low (Not commonly used for identification)	High (For medical diagnosis)
EMG	High	Moderate (Varies depending on muscle used)	High	Moderate (Requires medical equipment)	Low (Not commonly used for identification)	High (For medical diagnosis)
EEG	High	Moderate (Can vary with mental state)	High	Moderate (Requires medical equipment)	Low (Not commonly used for identification)	High (For medical diagnosis)

1.5 General Architecture of Biometric Systems

A biometric system is a framework that incorporate various interconnected components and phases, each one designed to fulfill a specific role within the process of identity verification. This architecture ensures the seamless integration of hardware, software, and

algorithms to facilitate reliable and secure authentication. The architecture of a biometric system is specifically designed to navigate through the complexities of capturing, processing, and analyzing biometric data, resulting in accurate and efficient identification processes. The typical architecture of a biometric system comprises several well-defined phases, providing a structured framework for capturing, analyzing, and validating biometric traits. During these stages, a biometric system converts raw biometric data into actionable insights, enabling authentication and identification processes in a variety of applications and settings. The architecture of a biometric system typically includes the following phases:

1.5.1 Data Acquisition Phase

Biometric recognition systems initiate with the stage of data acquisition. This entails the capture of an individual's unique characteristics by specialized sensors or capture devices. Fingerprints, facial features, iris patterns, voice samples, and keystroke dynamics can all be utilized as biometric identifiers as discussed in section 1.4. The fidelity and efficacy of the entire recognition process are related to two critical aspects: the characteristics of the employed sensors and the environmental conditions during data capture. High-resolution scanners can produce detailed fingerprint images, while low-quality cameras might struggle to capture intricate facial features under poor lighting. Similarly, background noise can significantly lower the accuracy of voice recognition systems. Therefore, meticulous selection of appropriate sensors and ensuring optimal capture conditions are necessary for acquiring reliable and high-quality biometric data, which serves as the foundation for accurate identification or verification within the broader biometric recognition system.

1.5.2 Pre-processing Phase

Following the initial data acquisition, the captured biometric data enters a subsequent phase known as pre-processing. This stage focuses on refining the raw data to enhance its quality and suitability for subsequent analysis within the biometric recognition system. Diverse algorithms and pre-processing techniques are employed to achieve this objective. Noise reduction algorithms tackle unwanted distortions or interferences that may have been introduced during data capture. Normalization techniques address variations in the captured data, such as scaling the amplitude of ECG signals to a standard size or adjusting voice recordings for volume inconsistencies. Segmentation isolates the relevant biometric features from the data stream. For instance, ECG-based recognition systems might segment the captured signal to focus solely on PT interval for heartbeat analysis.

Finally, quality assessment procedures evaluate the pre-processed data to identify and potentially reject samples that fall below established thresholds. By applying these pre-processing techniques, the inherent variability and imperfections within the raw biometric data are avoided, resulting in a standardized and better-quality data representation that facilitates efficient and accurate processing in the subsequent stages of the biometric recognition system.

1.5.3 Feature Extraction Phase

Following the data pre-processing stage, the biometric information undergoes another analysis stage to extract a set of discriminative features. These features act as a unique identifier for the individual. The nature of these extracted features depends entirely on the type of biometric modality employed. Fingerprint recognition systems, for example, focus on minutiae points. Facial recognition, on the other hand, might extract key facial landmarks such as the positions of the mouth, nose, and eyes. Iris recognition leverages texture descriptors to capture the unique patterns within the iris, while ECG-based systems might extract spectral features that reflect the frequency components of the electrical activity of the heart. By effectively isolating these features, the system establishes a unique signature for each individual, enabling accurate identification or verification in later stages.

1.5.4 Matching Phase

The extracted features enter a stage known as the matching phase. During this phase, a comparison is conducted between the extracted features and reference templates stored within a secure database. These reference templates represent the enrolled biometric data of authorized individuals. Matching algorithms assess how closely the features and the saved templates resemble each other. Diverse matching techniques can be employed for this purpose, including distance metrics, similarity measures, ML classifiers, and DL. Distance metrics, such as Euclidean distance, quantify the difference between the query features and each template. Similarity measures, on the other hand, aim to assess the degree of correspondence between them. ML and DL classifiers, trained on a vast dataset of biometric information, can learn complex patterns within the data and make predictions about the likelihood of a match. The ultimate goal of the matching phase is to identify the best match, or a set of potential matches, among the enrolled individuals. By analyzing the results of the comparison, the system can determine the likelihood of a successful authentication or identification. If the features were very close to a specific template, this

would indicate that the match was successful and the person's identification or access could be validated. Conversely, a low degree of similarity might indicate an unauthorized attempt or necessitate further verification steps.

1.5.5 Decision Phase

Following the matching phase, the system arrives at a stage where a definitive decision regarding the individual's identity is made. This decision is supported by the results of the process of comparing the biometric features that were extracted with the templates that were saved in the database. The standard for successful identification or authentication is a predetermined threshold or confidence level. An match is declared when the degree of similarity between the features and a given template is greater than this threshold. In such scenarios, the system might grant access to secure resources, confirm an individual's identity for transactions, or expedite authorized tasks. Conversely, if the similarity score falls below the established threshold, the outcome can vary depending on the security posture of the system. The system might prompt additional verification measures, such as a PIN or security question, to enhance confidence in the identification process. Alternatively, for stricter security protocols, the system might reject the authentication attempt altogether, potentially flagging the attempt for further investigation or requiring the user to enroll again. By evaluating the matching results and employing a predefined threshold, the system makes a clear and decisive judgment regarding the individual's identity. This decision directly influences subsequent actions within the system, ensuring authorized access, upholding security measures, and streamlining the process of identification or verification.

1.5.6 Integration Phase

This stage acts as the bridge between the biometric system's decision-making capabilities and the security infrastructure. The system seamlessly integrates the authentication or identification verdict into the existing framework, enabling it to take appropriate actions based on the individual's identity. This integration can manifest in various forms depending on the specific application. For instance, upon successful authentication, the system might grant access privileges to secure resources, such as unlocking doors, authorizing financial transactions, or granting system permissions. Conversely, failed authentication attempts might trigger alarms or notifications, alerting security personnel of potential unauthorized access attempts. Additionally, the system might log user activities, creating a detailed record of access attempts and interactions for auditing purposes. Furthermore,

integration can pave the way for initiating further actions based on the individual's identity. For example, personalized settings or preferences could be automatically loaded, streamlining the user experience. The significance of this integration phase lies in its ability to ensure seamless interoperability with existing systems and workflows. By effectively integrating the biometric system's decision-making capabilities, biometric technology can significantly enhance security measures across various domains. Additionally, this integration streamlines processes by automating access control, user authentication, and potentially even personalized actions based on identity. Ultimately, this final phase allows biometric systems to seamlessly integrate into existing infrastructure, unlocking their full potential to enhance security and improve user experience.

1.6 Conclusion

As a conclusion to this chapter, biometric systems represent a significant advancement in identity verification, offering a more secure and efficient alternative to traditional methods. By leveraging unique physiological and behavioral traits, these systems provide robust mechanisms for both identification and authentication across various applications. The architecture of biometric systems, comprising phases such as data acquisition, preprocessing, feature extraction, matching, decision-making, and integration, ensures that the process is accurate and reliable. With the growing adoption of biometric technologies worldwide, including in Algeria, these systems are poised to play an increasingly integral role in securing digital and physical environments.

Chapter 2

Review of Related Works

2.1 Introduction

This chapter reviews the extensive body of research surrounding ECG-based biometrics, speaker recognition, and multimodal biometric systems. The focus is on both fiducial and non-fiducial approaches to ECG biometrics, each of which represents a distinct methodology for extracting and analyzing ECG signals for identity recognition. Fiducial techniques rely on specific, manually or algorithmically defined points in the ECG waveform, while non-fiducial methods leverage the entire signal, often employing machine learning models for automated feature extraction. Recent advances in both approaches have significantly improved the accuracy, robustness, and usability of ECG biometrics in real-world applications, bolstered by the availability of various ECG databases—both on-the-person and off-the-person—that provide critical training and testing data.

In parallel, speaker recognition has emerged as another vital biometric modality, used for identifying or verifying individuals based on their voice patterns. This field encompasses several sub-tasks, including speaker identification, verification, and diarization, each with its own techniques and applications. Both text-dependent and text-independent methods have been explored, depending on whether the spoken content is constrained or unrestricted. Speaker recognition systems are often evaluated using closed-set and open-set approaches, with multiple databases developed for benchmarking performance.

Further, multimodal biometric systems that combine two or more biometric modalities are increasingly gaining attention for their ability to enhance system reliability and security. Fusion techniques—whether at the sensor, feature, score, or decision level—have been instrumental in integrating multiple biometric modalities, improving the overall performance by addressing the limitations inherent in unimodal systems. Various approaches to normalization, classification, and combination are essential in optimizing score-level fusion, which plays a pivotal role in achieving robust and accurate multimodal biometric systems.

In this chapter, we examine the evolution and recent innovations across these domains, reviewing key databases, techniques, and the role of neural networks in advancing ECG, voice, and multimodal biometrics. This discussion provides a comprehensive overview of the foundational works and current state of research, offering a basis for understanding the challenges and opportunities in the field.

2.2 ECG Biometrics and Relevant Databases

2.2.1 Fiducial ECG Biometric Systems

The use of ECG as a biometric modality is grounded in its physiological uniqueness; each individual's ECG signal exhibits distinctive patterns due to variations in heart structure and function. This characteristic makes ECG a reliable candidate for biometric systems, particularly in contexts where traditional methods (e.g., passwords, PINs) may be vulnerable to spoofing or forgery. The increasing prevalence of identity theft and unauthorized access has prompted the exploration of physiological biometrics, with ECG standing out for its robustness against such threats [74].

2.2.1.1 Fiducial techniques

Fiducial techniques focus on extracting features from specific points of the ECG signal, primarily the QRS complex, which represents ventricular depolarization. Key studies [75, 76] have highlighted the efficiency of fiducial methods. For instance, Hassan et al. [76] indicated that fiducial methods could achieve higher accuracy than non-fiducial techniques in identifying individuals, particularly in smaller datasets. Fiducial techniques are characterized by:

- Extracting features from specific points (e.g., R-peaks)
- Resulting in higher accuracy and reliability.
- Commonly used in studies focusing on individual identification.

2.2.1.2 Recent Advances in Fiducial ECG Biometrics

ECG signals were first introduced by Biel et al. [77] as a biometric trait. In their paper, The authors investigated the use of ECG signals for human identification. The study utilized a dataset comprising ECG recordings from 20 subjects, with each subject undergoing 4 to 10 measurements. The features extracted from the ECG signals included P-wave onset, QRS wave duration, and T-wave morphology, among others, resulting in a total of 30 features. For classification, the study employed the Soft Independent Modeling of Class Analogy (SIMCA) method. The results demonstrated a high accuracy rate, with 49 out of 50 samples correctly classified using various feature sets.

Tantawi et al. [78] presented another fiducial-based approach to ECG biometrics using a reduced set of fiducial points. The authors utilized datasets from Physionet, including PTB, MIT-BIH, and Fantasia, with a total of approximately 100 subjects. The proposed

method focuses on a set of 23 features derived from five major peaks and valleys (P, Q, R, S, T) instead of the traditional 11 fiducial points. The Radial Basis Function (RBF) neural network was employed as the classifier. The results demonstrated that the proposed PV set achieved comparable subject identification (SI) and heartbeat recognition (HR) accuracies to the full set of 36 features, with 100% SI accuracy and up to 96% HR accuracy on test sets.

In another paper by Tantawi et al. [79], the authors utilized datasets from the PhysioNet ECG repository, including 51 subjects from the PTB database, 18 subjects from the MIT-BIH Normal Sinus Rhythm database, and 40 subjects from the Fantasia database. The preprocessing involved noise reduction and baseline wandering elimination using a second-order Butterworth filter. Features extracted included 28 fiducial points, encompassing temporal, amplitude, and angle features. Feature reduction was performed using principal component analysis (PCA), linear discriminant analysis (LDA), information gain ratio (IGR), and rough sets with the PASH algorithm. The classification was conducted using an RBF neural network. The study achieved high classification accuracy, with the PCA method yielding 100% subject identification accuracy.

Gargiulo et al. [80] investigated the influence of QT interval correction on ECG-fiducial-based biometric identification systems. The study utilized ECG signals from the Physionet database, specifically the 39 subjects from the Fantasia dataset. Preprocessing involved notch filtering to remove 60 Hz powerline interference and fiducial point detection using the ECGPUWAVE detector [81]. The features extracted were both temporal and amplitude-based, with QT interval correction applied using various models. The classifiers evaluated included MultiLayer Perceptron (MLP), SVM, and DT, along with ensemble methods like Random Forest, Bagging, and AdaBoost. The results demonstrated that QT correction significantly improves identification performance, with identification rates ranging from 0.97 to 0.99. The study found that MLP and SVM classifiers provided better generalization capabilities compared to DT-based classifiers.

2.2.2 Non-Fiducial ECG Biometric Systems

Non-fiducial ECG biometrics represents a more recent advancement in the field of biometric identification than fiducial techniques, utilizing the unique characteristics of ECG signals without relying on specific fiducial points such as R-peaks. The next lines of this thesis represent a literature review that explores various methodologies, findings, and challenges associated with non-fiducial ECG biometrics.

2.2.2.1 Non-Fiducial Techniques Overview

Unlike fiducial approaches, Non-fiducial ECG biometrics primarily focuses on analyzing the entire ECG waveform rather than specific points. This approach can enhance robustness against noise and variability in signal acquisition. Key studies have highlighted various techniques and their effectiveness:

- **Compression-Based Methods**: Carvalho et al. [82] presented a compression-based non-fiducial method for ECG biometric identification, utilizing the Normalized Relative Compression (NRC) measure. The dataset comprises ECG signals from 25 participants, collected over three different days at the University of Aveiro. Preprocessing involved using a Butterworth low-pass filter and first-order differentiation of the ECG signals. The features were extracted using Lloyd-Max quantization on the first-order derivatives, and the classification was performed using extended-alphabet finite-context models (xaFCMs). The method achieved an accuracy of 89.3%. The results also included a confusion matrix with an F1-score of 0.88.

- **ML Approaches**: Kim et al. [83] introduced a user identification system based on one-dimensional shallow neural networks using non-fiducial segmented ECG signals. The datasets used include self-acquired data from 100 subjects and publicly available datasets such as MIT-BIH, ECG-ID, and PTB-XL. Preprocessing techniques involve downsampling to 256 Hz and noise removal using a bandpass filter. The features used are non-fiducial segmented ECG signals divided into 1-second, 2-second, and 3-second windows. The classifier employed is a one-dimensional convolutional neural network (1D CNN). The proposed system achieved a user identification accuracy of 95.51% on self-acquired data and over 94% on public datasets. The study also reported a 25.48% improvement in accuracy compared to fiducial-based segmentation methods.

- **Wavelet Transform Techniques**: Elshahed [84] presented a biometric authentication system using non-fiducial ECG features. The study utilized two datasets: the ECG-ID Database and the MIT-BIH Arrhythmia Database, involving a total of 90 subjects, with 72 subjects used for testing. Preprocessing included filtering signals with a Butterworth filter and applying the Pan and Tompkins algorithm for R peak detection [85]. Features were extracted using discrete wavelet decomposition (DWT) with daubechies wavelets (db8). The Euclidean Distance algorithm was employed for verification. The system achieved a verification rate of 94.44%, with a sensitivity of 95.08%, specificity of 90.9%, precision of 98.3%, and an F-score of 96.66%.

- **Empirical mode decomposition (EMD)**: Aziz et al. [86] presented a methodology for ECG-based biometric authentication. The study utilizes ECG data from 14 subjects (8 males and 6 females) collected using BIOPAC systems. The raw ECG signals are preprocessed using EMD to denoise and extract the region of interest. Five features are extracted from the time, frequency, and statistical domains: Shannon energy, skewness, variance, occupied bandwidth, and median frequency. Various classifiers, including SVM with different kernels, K-nearest neighbor (KNN), and DT, are evaluated. The SVM with cubic kernel achieves the best performance with an accuracy of 98.72%, sensitivity of 100%, and specificity of 99.82%.

2.2.3 Hybrid Techniques Overview

Hybrid ECG biometric systems combine fiducial and non-fiducial techniques to leverage the strengths of both methods. One such method presented by Carvalho and Brás [87] explored a hybrid classification system to mitigate intra-subject variability in ECG-based biometric systems. The study utilizes two private databases, EMOTE_1 (30 subjects) and EMOTE_2 (53 subjects), as well as the public MIT-BIH database (47 subjects). Preprocessing techniques include Butterworth bandpass filtering, R-peak detection, normalization, and outlier removal. The hybrid architecture leverages both fiducial and non-fiducial features, such as autocorrelation, statistical, and wavelet features. The system achieved an accuracy of 99.98% on the MIT-BIH database and improved the F1-score by up to 12% on private databases.

2.2.4 The Role of Neural Networks in ECG Biometrics

The ability of neural networks to learn complex patterns and features from data makes them well-suited for processing the intricate and variable nature of ECG signals. The next sections of this thesis examine the various applications of neural networks in ECG biometrics, highlighting their contributions to identification accuracy and overall system performance.

2.2.4.1 CNNs

CNNs are a powerful tool for processing and classifying ECG signals, particularly when transformed into two-dimensional representations. A study by Ciocoiu and Cleju [88] comparing different spatial representations of ECG signals found that CNNs could effectively classify and identify individuals based on their ECG data. The research employed various transformations, including a modified version of the Continuous Wavelet Trans-

form and Gramian Angular Field, to convert 1D ECG signals into 2D images, achieving high identification accuracy and low (EER) in off-person scenarios.

In another study, Byeon and Kwak [89] evaluated pre-configured CNN models combined with various time-frequency representations of ECG signals. This study demonstrated that CNNs could effectively classify noisy ECG data, which can be applied to real-world applications where signal quality may vary. The integration of time-frequency transformations, such as mel spectrograms and scalograms, with CNN architectures like VGGNet and ResNet was shown to enhance classification performance.

2.2.4.2 Recurrent Neural Networks (RNN)

RNNs, particularly those utilizing LSTMs and gated recurrent unit (GRU) architectures, have also been applied to ECG biometrics. Unlike CNNs, RNNs can process sequential data without the need for feature extraction.

In a paper by Salloum and Kuo [90], the authors propose using RNNs for ECG-based biometric identification and authentication. They utilized two publicly available datasets: ECG-ID with 90 subjects and MIT-BIH with 47 subjects. The preprocessing involved segmenting ECG recordings into individual heartbeat waveforms using the Pan-Tompkins algorithm and standardizing them. The study employed raw ECG data as features, directly fed into various RNN architectures, including traditional RNNs, LSTM, and GRU networks. The LSTM-based RNN achieved nearly 100% classification accuracy on both datasets for identification tasks. For authentication, the EER dropped to 0% when 80% of subjects were used for training. The study concluded that LSTM-based RNNs are highly effective for ECG-based biometric identification and authentication.

Another paper by Lynn et al. [91] leveraged RNN-based techniques. The authors utilized two publicly available datasets: the ECG-ID Database, containing 310 recordings from 90 subjects, and the MIT-BIH Arrhythmia Database, with recordings from 47 subjects. Preprocessing involved detrending, filtering using a 6th-order Butterworth filter, and R-peak detection. The study explored various features, including raw ECG signals and heartbeat waveforms. The deep learning algorithms employed were 1D CNN and RNN with LSTM and GRU cells, both in unidirectional and bidirectional configurations. The proposed bidirectional GRU (BGRU) model achieved the highest classification accuracy of 98.55%. The results demonstrated that the BGRU model outperformed other models, highlighting its effectiveness in capturing temporal dependencies in ECG signals for biometric identification. The study concludes that bidirectional RNNs, particularly with GRU cells, offer significant advantages in ECG-based biometric systems.

2.2.5 Transformers and Vision Transformers

Transformers [55], originally proposed for natural language processing tasks, have shown promising results when applied to ECG signal classification. D'angelis et al. [92] presented an approach to ECG-based biometric recognition using vision transformers. The study utilizes two datasets: CYBHi, with 63 subjects, and Heartprint, with 199 subjects. Preprocessing involves filtering ECG signals using a zero-phase Butterworth bandpass filter and segmenting the signals based on R peaks. The features used are raw ECG segments converted into 2D images. The core deep learning algorithm is a fine-tuned vision transformer model, which processes these 2D images for classification. The vision transformer model, pre-trained on ImageNet, is fine-tuned with a cross-entropy loss function for 1000 epochs. The results show a single sample-based identification accuracy of over 70% and an EER of 0.48% on the CYBHi dataset. The study concludes that vision transformers significantly enhance the robustness and accuracy of ECG-based biometric systems, particularly in long-term identification scenarios.

2.2.5.1 Autoencoders

Deep autoencoders have been employed for feature learning in ECG biometrics, enabling the extraction of lower-dimensional representations of heartbeat templates while maintaining high identification performance. Eduardo et al. [93] explored the use of autoencoders for ECG biometric identification. The dataset comprises 960 10-second records from 709 subjects. Preprocessing involved filtering the raw signals with a 150th-order bandpass finite impulse response (FIR) filter and transforming them into heartbeat templates. The features used were lower-dimensional representations of these templates, learned via a deep autoencoder. The autoencoder's architecture included multiple hidden layers, with topologies such as [300, 100, 50, 100, 300] and [300, 150, 50, 150, 300]. The results demonstrated superior identification performance, with the deep autoencoder achieving a lower identification error compared to baseline models. The study concluded that the deep autoencoder effectively learns expressive representations, making it suitable for ECG-based biometric systems, even in transfer learning settings.

2.2.5.2 Hybrid Models

The integration of various deep learning techniques has also been explored to capture both spatial and temporal features of ECG signals. Min Keun and Kim [94] explored an approach to dog identification using ECG signals. The study utilized two datasets: the PhysioZoo database with 17 subjects and a Holter monitoring database with 16 subjects,

integrating them for a combined dataset of 33 subjects. Preprocessing involved noise removal using a fourth-order Butterworth bandpass filter and normalization. Features were extracted using R-peak based and blind segmentation methods1. The deep learning model employed a 1D CNN-LSTM architecture, combining convolutional layers for feature extraction and LSTM layers for sequential information2. The proposed model achieved up to 98.7% accuracy on a separate database and 96.3% accuracy on the integrated dataset.

2.2.6 ECG Databases

2.2.6.1 On-the-Person ECG Databases

On-the-person ECG databases typically involve the use of electrodes placed directly on the skin to capture ECG signals. This method has been predominant in clinical settings and offers high-quality data, but it presents several challenges:

- **Intrusiveness**: The requirement for physical contact can deter users, especially in non-clinical environments.
- **Variability**: Factors such as electrode placement, skin condition, and movement can introduce variability in the data, complicating the authentication process.
- **Limited Scalability**: The need for specialized equipment and controlled environments limits the scalability of on-the-person systems for widespread use.

Here are some examples of notable On-the-Person ECG databases commonly used in research:

- **PTB Diagnostic ECG Database**: The PTB Diagnostic ECG Database [95] is a comprehensive collection of high-resolution 15-lead ECGs, including 12 standard leads and 3 Frank XYZ leads. This database, hosted on PhysioNet [96], comprises 549 ECG records from 290 subjects, ranging in age from 17 to 87 years, with a mean age of 57.2 years. The subjects include 209 men with a mean age of 55.5 years and 81 women with a mean age of 61.6 years. The ECGs were acquired using a non-commercial PTB prototype recorder, which digitized the signals at a sampling frequency of 1000 samples per second with 16-bit resolution over a range of $\pm$16.384 mV. On special request, recordings may be available at sampling rates up to 10 KHz. Each record in the database includes a detailed clinical summary within the header (.hea) file, providing information on age, gender, diagnosis, medical history, medication, interventions, coronary artery pathology, ventriculography, echocardiography, and hemodynamics. However, clinical summaries are not available for 22

subjects. The database was prepared by experts from the Physikalisch-Technische Bundesanstalt (PTB) and the Charité Medical Center, and it serves as a valuable resource for research, algorithmic benchmarking, and teaching purposes. The ECGs were collected from both healthy volunteers and patients with various heart diseases, ensuring a diverse dataset for comprehensive analysis.

- **MIT-BIH Arrhythmia Database**: The MIT-BIH Arrhythmia Database [97] is a collection of ECG recordings designed for the evaluation of arrhythmia detectors and basic research into cardiac dynamics and it is widely used in biometric applications. The database was created by the BIH Arrhythmia Laboratory between 1975 and 1979 and contains 48 half-hour excerpts of two-channel ambulatory ECG recordings. These recordings were obtained from 47 subjects, with 23 recordings chosen randomly from a set of 4000 24-hour ambulatory ECG recordings collected from a mixed population of inpatients (about 60%) and outpatients (about 40%) at Boston's BIH. The remaining 25 recordings were selected to include less common but clinically significant arrhythmias. The acquisition technique involved digitizing the recordings at a sampling frequency of 360 samples per second per channel with 11-bit resolution over a 10 mV range. Two or more cardiologists independently annotated each record, and disagreements were resolved to obtain the computer-readable reference annotations for each beat, resulting in approximately 110,000 annotations in total. The database includes a clinical summary of the subjects, detailing their age, gender, and diagnosis.

- **MIT-BIH Normal Sinus Rhythm Database**: This database comprises 18 long-term ECG recordings from subjects referred to the Arrhythmia Laboratory at Boston's BIH, now known as the Beth Israel Deaconess Medical Center. The acquisition technique involved continuous ECG monitoring, capturing the electrical activity of the heart over extended periods. Each recording includes data from multiple leads. The database includes recordings from 18 subjects, consisting of 5 men and 13 women. The age range of the male subjects is between 26 and 45 years, while the female subjects are aged between 20 and 50 years. All subjects included in this database were found to have no significant arrhythmias, providing a baseline of normal sinus rhythm ECGs for research purposes.

- **ECG-ID**: The ECG-ID [98] Database was created to support research in biometric human identification based on ECG signals. The acquisition technique involved recording single-lead ECGs from 90 volunteers using limb clamp electrodes, specifically focusing on Lead I, which measures the potential difference between the left

and right hands. This choice was made for its ease of measurement and insensitivity to minor variations in electrode placement. The database comprises 310 ECG recordings, each 20 seconds long, sampled at a frequency of 500 Hz with 12-bit precision.

- **Fantasia Database**: The Fantasia Database [99] is a collection of physiological signals, specifically designed for research purposes. It includes ECG and respiration recordings from a total of 40 subjects, divided equally between young (21-34 years old) and elderly (68-85 years old) participants. Each group consists of 20 individuals, with an equal distribution of men and women. The data acquisition involved continuous monitoring of ECG and respiration signals for a duration of two hours while the subjects were in a resting state. Additionally, half of the recordings from each age group include continuous, albeit uncalibrated, non-invasive blood pressure signals. The ECG signals were digitized at a sampling frequency of 250 Hz. Each heartbeat was annotated using an automated arrhythmia detection algorithm, followed by manual verification to ensure accuracy. The recordings were made while the subjects watched the movie "Fantasia" (Disney, 1940) to help maintain wakefulness during the resting state. This setup provided a controlled environment for capturing the physiological signals, minimizing external influences that could affect the data quality.

2.2.6.2 Off-the-Person ECG Databases

Off-the-person ECG databases utilize non-invasive methods to capture ECG signals, often through wearable devices or sensors placed at a distance from the body. This approach has gained traction due to its potential for seamless integration into daily life. Key advantages include:

- **User Acceptance**: The non-intrusive nature of off-the-person methods enhances user comfort and acceptance, making it suitable for biometric applications.
- **Reduced Variability**: Off-the-person systems can minimize some sources of variability associated with direct contact methods, although they still face challenges related to noise and signal quality.

Here are some notable examples of Off-the-Person ECG databases that have been used in research:

- **Check Your Biosignals Here Database (CYBHi)**: The CYBHi [100] was developed to create a dataset and a consistent acquisition framework for ECG biometrics,

particularly focusing on off-the-person data acquisition. The dataset includes ECG data collected at the hand palms and fingers using dry Ag/AgCl electrodes and Electrolycras. The acquisition technique involved a custom, two-lead differential sensor design with virtual ground, ensuring high-quality signal capture. The data was collected using a bioPLUX research unit with a 12-bit resolution and a sampling frequency of 1 kHz. The dataset comprises two parts: a short-term dataset and a long-term dataset. The short-term dataset includes data from 65 participants, predominantly engineering students and researchers, with 49 males and 16 females, averaging 31.1 years of age. The long-term dataset was collected over several days and included data from 63 subjects, primarily nursing and health technologies students, with 14 males and 49 females, averaging 20.68 years of age. The clinical summary of the dataset indicates that the participants were healthy individuals, with no reported health issues. The demographic information collected includes age and gender, ensuring a diverse sample for robust biometric analysis. The dataset is publicly available and aims to facilitate benchmarking and comparison of ECG-based biometric algorithms across different research teams.

- **University of Toronto ECG Database (UofTDB)**: The UofTDB [101] comprises recordings from 1020 subjects, captured from fingertips in a configuration similar to Lead I. A subset of fewer than 100 subjects was recorded for up to six sessions over a period of six months, under different postures and exercise conditions. The Vernier EKG sensors were used for data acquisition, with a sampling rate of 200 Hz. Each recording session lasted between 2 to 5 minutes.

2.3 Speaker Recognition

Speaker recognition is a field of biometrics that focuses on identifying or verifying an individual based on their unique voice characteristics. This field has witnessed significant advancements over the years, driven by the need for secure and efficient identity verification methods. The next sections will cover the various aspects of speaker recognition, including identification, verification, and diarization, along with text-dependent and text-independent methods, and closed-set versus open-set conditions which are interconnected within the broader framework of speaker recognition systems as shown in Figure 2.1.

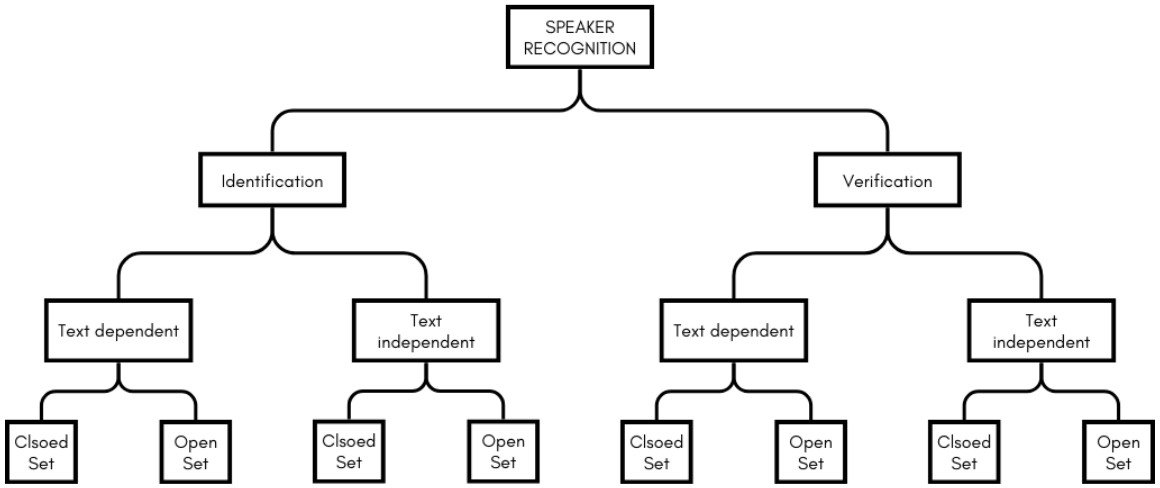


Figure 2.1: The concepts of closed sets and open sets, text-dependent and text-independent systems, as well as identification and verification, are all interconnected and can be understood as layers of a speaker recognition system.

2.3.1 Speaker Recognition Types

2.3.1.1 Speaker Identification

Speaker Identification is the process of determining the identity of an unknown speaker from a set of known speakers. In identification tasks, the system compares the voice of an unknown speaker with a pre-stored database of voice prints and selects the most likely match. The study by Fredouille and Charlet [102] investigates the application of the I-vector framework for speaker identification in TV shows, particularly within the REPERE challenge. The study utilizes a training corpus of 47 hours, a development set of 3 hours, and a test set of 10 hours, covering 533 speakers. Preprocessing techniques include feature extraction using 19 LFCC augmented with delta coefficients, delta energy, and double delta coefficients, followed by cepstral mean subtraction and variance normalization. The I-vector extraction is performed using a 200-dimension total variability space. The Cosine Distance Scoring (CDS) and Probabilistic Linear Discriminant Analysis (PLDA) are employed as classifiers. The results demonstrate the effectiveness of the I-vector framework, achieving a Correct Identification Rate (CIR) of up to 97.5% in closed-set tasks and F-measure of 80.7% in open-set tasks.

2.3.1.2 Speaker Verification

Speaker Verification involves confirming or denying a claimed identity based on the speaker's voice. This process is generally more complex than identification, as it re-

quires the system to authenticate whether the voice matches the claimed identity rather than finding the closest match. The proposed system by by Ullah et al. [103] utilized a dataset comprising voice samples from four speakers, each contributing 200 voice signals (100 valid and 100 from other speakers). The preprocessing techniques included pre-emphasis, framing, windowing, and Fast Fourier Transform (FFT). The features were extracted using the Mel Frequency Cepstral Coefficient (MFCC). The classifier employed was a Cartesian Genetic Programming Evolved Artificial Neural Network (CGPANN). The system demonstrated high accuracy, achieving up to 100% training efficiency and 99% testing efficiency in two-fold testing.

2.3.1.3 Speaker Diarization

Speaker Diarization is the task of segmenting an audio stream into homogeneous segments, each associated with a different speaker. Diarization is essential for scenarios where multiple speakers are present, such as meetings or telephone conversations. The paper by Dimitriadis [104]

presents enhancements for audio-only speaker diarization systems, focusing on the speaker clustering component. The author evaluates his methods on three datasets: AMI (166 meetings with 4 speakers each), diverse recordings with varying speakers and conditions (DIHARD), and internal meeting data (two meetings with 6 and 4 participants). Preprocessing includes temporal smoothing and median filtering of d-vectors. Features are extracted using d-vectors from a deep neural network's bottleneck layer. The clustering algorithms compared include k-means, Spectral Clustering, x-means, and an improved Deep Embedded Clustering (DEC) algorithm. The proposed methods show significant improvements, with the DEC algorithm yielding up to 31% better clustering performance.

2.3.2 Overview on Text-Dependent and Text-Independent Speaker Recognition

2.3.2.1 Text-Dependent Speaker Recognition

Text-dependent speaker recognition relies on the speaker uttering a specific phrase during both the enrollment and verification phases. This approach typically achieves higher accuracy because the recognition system can focus on specific speech features that are consistent across sessions:

- **Accuracy and Applications**: Text-dependent systems are known for their accuracy, making them suitable for high-security applications such as banking or access

control. The fixed phrase requirement enables the system to use techniques such as Dynamic Time Warping (DTW) [105] and Hidden Markov Models (HMM) [106] to effectively model the speech dynamics.

- **Challenges**: Despite the high accuracy, these systems face challenges such as vulnerability to replay attacks and the need for the user to remember the specific phrase. Additionally, the performance of text-dependent systems can degrade significantly if the phrase is spoken differently during enrollment and verification phases [107].

2.3.2.2 Text-Independent Speaker Recognition

Text-independent speaker recognition does not require the speaker to use a specific phrase, allowing more flexibility in user interaction. However, this flexibility comes at the cost of generally lower accuracy, especially in varying acoustic environments:

- **Flexibility and Challenges**: The major advantage of text-independent systems is their ability to recognize a speaker regardless of what is being said, which is useful in applications like forensic investigations or continuous authentication. However, the lack of fixed phrases makes it difficult to achieve high accuracy due to the variability in speech content and environmental noise [108].

- **Techniques and Performance**: Techniques such as MFCCs combined with GMMs or LSTM have been developed to improve the robustness and accuracy of text-independent systems. These approaches have demonstrated high recognition rates under controlled conditions but still face challenges in more diverse environments.

2.3.3 Closed Sets and Open Sets

2.3.3.1 Closed Set Speaker Recognition

Closed-set speaker recognition is traditionally more straightforward, where the task is to match an input speech sample with one of the known speakers in the database. The performance in closed-set systems is generally high due to the fixed and known set of possible speakers. For instance, Dutta et al. [109] presented a closed-set text-independent speaker identification system using multiple Artificial Neural Network (ANN) classifiers. The dataset comprises continuous speech samples from 16 native Assamese speakers, including both male and female, covering four dialects: Eastern, Central, Kamrupi, and Goalpariya. Each speaker provided 90 speech samples recorded in three different moods: normal, loud, and angry. Preprocessing involved extracting voiced parts from the speech using short-term energy (STE) and short-term zero-crossing rate (STZ). The features

used for classification included pitch, Linear Prediction (LP) residuals, and EMD residuals. Three classifiers were designed: one using MLP and two using RNNs. The hybrid classifier combining all three achieved a speaker identification accuracy of 98% and a dialect classification accuracy of 71%.

2.3.3.2 Open Set Speaker Recognition

Open set recognition introduces the challenge of identifying speakers who may not belong to the known set, which is more reflective of real-world scenarios. To address this, researchers have developed algorithms to enhance the robustness of open-set systems. In the study presented by Wilkinghoff [110] an open-set speaker identification system based on i-vectors is introduced. The system was evaluated using the MCE 2018 dataset, which includes 600-dimensional i-vectors from 3631 blacklist speakers and an unknown number of non-blacklist speakers. The preprocessing techniques employed include linear alignment and length normalization. The features used are i-vectors, and the classifiers include LDA and Probabilistic LDA (PLDA). The system achieved a 37.5% improvement in top-S EER and a 50% reduction in top-1 EER compared to the baseline system. Additionally, the system's performance surpassed all other published results on the same dataset

2.3.4 Speaker Recognition Databases

Two primary categories of datasets are critical for evaluating speaker recognition systems: clean speech datasets, typically collected under controlled conditions, and "in the wild" datasets, which encompass real-world challenges such as background noise and speaker variability.

2.3.4.1 Clean Speech Datasets

These datasets are typically recorded in controlled environments with minimal background noise and consistent recording conditions:

- **TIMIT**: The TIMIT [111] is designed to support acoustic-phonetic studies and the development and evaluation of automatic speech recognition systems. It includes approximately five hours of English speech, featuring broadband recordings from 630 speakers representing eight major dialects of American English. Each speaker reads ten phonetically rich sentences, providing a diverse and balanced dataset. The speech waveform files are single-channel, 16-bit, and sampled at a frequency of 16 kHz. The dataset is divided into test and training subsets, balanced for both phonetic and dialectal coverage. Additionally, the corpus includes extensive speaker

metadata, such as gender, dialect, birth date, height, race, and education level. Of the 630 speakers, approximately 70% are men and 30% are women, ensuring a representative sample of the population.

- **LibriSpeech**: The LibriSpeech [112] corpus is a comprehensive dataset designed for training and evaluating speech recognition systems. It is derived from audiobooks that are part of the LibriVox project, which is a volunteer effort responsible for creating approximately 8000 public domain audiobooks, primarily in English. The corpus contains 1000 hours of read English speech, sampled at a frequency of 16 kHz. The acquisition technique involves aligning the audio recordings with their corresponding texts and splitting them into shorter segments to ensure that each segment has an accurate transcript. This process includes text preprocessing, lexicon and language model creation, and a two-stage alignment procedure to filter out segments with potential inaccuracies. The dataset is structured to ensure a balance between male and female speakers. Specifically, the training portion of the corpus is divided into three subsets: train-clean-100, train-clean-360, and train-other-500, with approximate sizes of 100, 360, and 500 hours, respectively. The development and test sets are also balanced, with each set containing approximately 5.4 hours of speech from 20 male and 20 female speakers. In total, the corpus includes recordings from 251 speakers in the train-clean-100 subset, 921 speakers in the train-clean-360 subset, and 1166 speakers in the train-other-500 subset.

2.3.4.2 In the Wild Datasets

In contrast to clean speech datasets, "in the wild" datasets are designed to challenge speaker recognition systems with the variability encountered in real-world environments. These datasets typically include a wide range of acoustic conditions, including background noise, reverberation, and speaker variability.

- **Speakers in the Wild (SITW)**: The SITW database [113] is a comprehensive collection of hand-annotated speech samples sourced from open-source media. This database is specifically designed to benchmark text-independent speaker recognition technology under real-world conditions. The acquisition technique involves collecting audio recordings from various media sources, ensuring that the data reflects natural, unconstrained environments. This approach introduces real-world challenges such as noise, reverberation, intra-speaker variability, and compression artifacts, which are often encountered in practical applications. The SITW database comprises recordings from 299 speakers, with each speaker contributing an average

of eight different sessions. This diverse dataset includes both male and female speakers. The recordings are captured at a sampling frequency 16 kHz.

- **VoxCeleb**: The VoxCeleb dataset [114] is a comprehensive collection of speech data from a diverse range of speakers, encompassing various ethnicities, accents, professions, and ages. All the recordings include natural background noises such as chatter, laughter, overlapping speech, pose variations, and different lighting conditions, providing a realistic and challenging environment for speaker recognition tasks. VoxCeleb is divided into two versions: VoxCeleb1 and VoxCeleb2. VoxCeleb1 contains over 150,000 utterances from 1,251 celebrities, while VoxCeleb2 expands significantly with more than 1,000,000 utterances from 6,112 celebrities. The dataset includes both audio and video segments, with each segment being at least 3 seconds long. 61% of the speakers are male while the remaining 39% are female.

2.4 Multimodal Biometric Systems

Multimodal biometrics refers to the integration of multiple biometric modalities within a single authentication or identification system. Unlike unimodal biometric systems, which rely on a single physiological, behavioral, or hidden trait for identity verification, multimodal biometric systems explores the power of multiple modalities to enhance security, accuracy, and reliability.

In a multimodal biometric system, two or more biometric traits are combined to create a comprehensive biometric profile of an individual. The integration of multiple modalities enables multimodal biometric systems to overcome the limitations of individual modalities, such as susceptibility to spoofing attacks, variability in performance under different conditions, and limitations in universality or uniqueness. By leveraging complementary information from multiple modalities, multimodal systems enhance the accuracy, robustness, and security of identity verification processes. Multimodal biometrics can be implemented at various levels of fusion, including sensor-level fusion, feature-level fusion, score-level fusion, and decision-level fusion as shown by Figure 2.2. Each fusion level offers distinct advantages and challenges, and the choice of fusion strategy depends on factors such as the characteristics of the biometric traits, system requirements, and deployment scenarios.

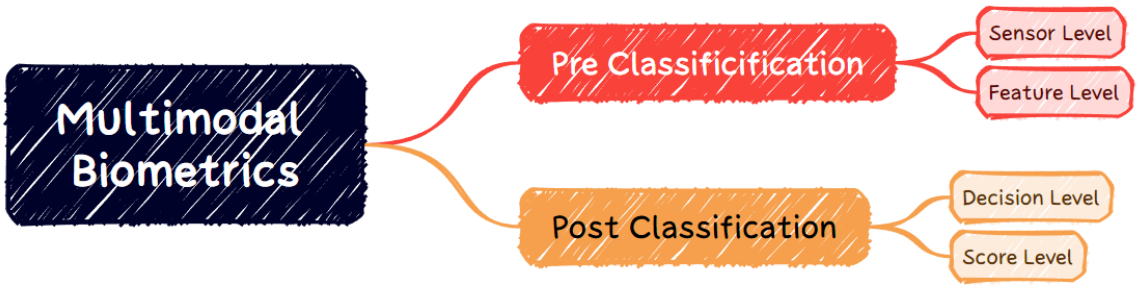


Figure 2.2: The multiple levels of biometric fusion.

2.4.1 Pre-classification fusion

Pre-classification fusion in multimodal biometric systems can occur at different levels, each offering unique advantages and challenges. These fusion levels determine how biometric data from multiple modalities is combined before undergoing classification or matching processes. The main pre-classification fusion levels include:

2.4.1.1 Sensor-Level Fusion

Sensor-level fusion, one of the earliest stages at which data from different modalities can be combined, involves merging the raw data captured by different sensors before feature extraction. This approach aims to leverage the complementary nature of various sensors to improve the overall recognition accuracy and system reliability.

Sensor-level fusion offers distinct advantages. Firstly, it facilitates the simultaneous capture of multiple biometric traits, significantly reducing acquisition time compared to sequential acquisition methods. This efficiency improvement proves beneficial in user experience and application scenarios where rapid identification is crucial. Secondly, by combining sensors into a single unit, sensor-level fusion simplifies system design by minimizing the number of independent components and associated complexities.

However, sensor-level fusion has been explored less frequently compared to other fusion levels, such as score-level or decision-level fusion, due to its inherent challenges. One significant issue is sensor interoperability, where different sensors may produce data in incompatible formats, making direct fusion difficult. Another challenge is the high dimensionality and redundancy of the combined sensor data, which can lead to increased computational complexity and the potential for reduced system performance. Fusing data from sensors with vastly different operating principles or data formats might require complex pre-conditioning steps. Additionally, the cost of integrating multiple high-performance

sensors can be substantial, impacting the overall system affordability. Furthermore, physical space constraints may limit the feasibility of sensor-level fusion in certain applications, particularly when dealing with wearable or portable devices [115].

Despite these challenges, several studies have proposed methods to enhance the effectiveness of sensor-level fusion. For example, Gowda et al. [116] presented a multimodal biometric verification system using various fusion strategies. The datasets used include AR-Face, PolyU-Palmprint, PolyU-Finger Knuckle Print, and Cluj-Handvein, with 100 subjects each. Preprocessing involved extracting texture features using Local Phase Quantization (LPQ) and employing HAAR wavelets for sensor-level fusion. The features extracted were texture-based, leveraging the LPQ operator. The system's performance was evaluated using metrics such as Genuine Acceptance Rate (GAR), (FAR), and (FRR). At the sensor level, the system achieved a GAR of 83.5% at a 1% FAR when fusing face, palmprint, finger knuckle print, and handvein modalities.

2.4.1.2 Feature-Level Fusion

Feature-level fusion in multimodal biometric systems involves combining the features extracted from multiple biometric traits before the matching process. This fusion technique is considered to offer richer information than other fusion levels, such as score-level or decision-level fusion because it preserves the detailed data obtained from the different modalities.

Feature-level fusion has been shown to significantly enhance the performance and security of multimodal biometric systems. By integrating feature sets from different modalities, such as face, fingerprint, and iris, the combined feature vector can leverage complementary information, leading to more robust recognition systems. For instance, a study by Nadheen and Poornima [117] demonstrated that feature-level fusion between iris and ear biometrics resulted in a 93% success rate, highlighting the improved performance compared to unimodal systems.

Several techniques have been proposed to optimize feature-level fusion. For example, Haghighat et al. [118] introduced Discriminant Correlation Analysis (DCA), a technique that maximizes the correlation within classes while minimizing it between classes, thereby improving the discriminative power of the fused feature vector.

Another approach is the Eigen-based Feature-Level Fusion (E-FLF) proposed by Chen et al. [119]. This method employs eigen analysis to find an optimal projection in the cross-energy space, improving the recognition accuracy by effectively combining local and global features from multiple modalities like iris, palmprint, and face.

Despite its advantages, feature-level fusion also faces significant challenges. One of

the main issues is the high dimensionality of the fused feature vector, which can lead to increased computational complexity and the risk of overfitting. Moreover, ensuring compatibility between the feature sets extracted from different modalities is another critical challenge. Soviany et al. [120] addressed this by proposing a model that combines intra-modal and inter-modal feature fusion, optimizing both the same biometric trait and multiple traits simultaneously to enhance human identification accuracy.

Another challenge is the integration of nonlinear relations between features. To address this, Safavipour et al. [121] proposed a hybrid approach using kernel methods to map features into a higher-dimensional space where nonlinear relationships become linear, thereby simplifying the fusion process.

2.4.2 Post-classification fusion

2.4.2.1 Score-Level Fusion

Score-level fusion has become one of the most commonly used techniques in multimodal biometric systems due to its balance between complexity and effectiveness. Unlike sensor- or feature-level fusion, score-level fusion combines the matching scores from multiple biometric modalities after each modality independently processes its data.

One of the key benefits of score-level fusion is its simplicity and ease of implementation. The matching scores generated by different biometric modalities can be accessed and combined without the need to directly process raw data or extracted features. Research by Jain et al. [122] highlighted that score-level fusion provides better recognition performance compared to single-modality systems, especially when applied to modalities such as face, fingerprint, and hand geometry. They also identified the importance of score normalization to ensure that the scores from different modalities are comparable.

Another advantage is its ability to integrate multiple modalities while allowing flexibility in the combination of scores through various fusion rules, such as sum, max, and min. Hanmandlu et al. [123] demonstrated the use of triangular norms (t-norms) for score-level fusion, showing that their method outperformed traditional rules, such as the sum and max fusion, when tested on multimodal datasets.

Despite its advantages, score-level fusion faces several challenges. One of the primary issues is dealing with conflicting or uncertain scores from different modalities. Systems must carefully manage discrepancies in the matching process to avoid degrading performance. Research by Mukherjee et al. [124] addressed this issue by applying differential evolution-based optimization to minimize the overlapping of genuine and imposter scores, thereby improving decision accuracy.

Another area of improvement lies in the adaptive adjustment of fusion rules based on

system conditions. Kumar et al. [125] proposed an adaptive fusion framework using particle swarm optimization to dynamically adjust score-level fusion rules based on security requirements and environmental conditions.

2.4.2.2 Decision-Level Fusion

Decision-level fusion is one of the most common techniques used to integrate information in biometric systems. At this level, each modality independently makes a decision, which is then aggregated to arrive at a final conclusion.

Decision-level fusion offers simplicity in terms of implementation since each biometric modality can function independently, producing a decision or classification, which is then combined with decisions from other modalities. Research by Garg et al. [126] highlights the ability of decision-level fusion to enhance system security by combining multiple traits, such as fingerprints and iris data, and achieving higher recognition accuracy while reducing FAR and FRR.

Various decision fusion strategies have been proposed. Devi and Rao [127] proposed three decision-level fusion schemes: Local Decision Fusion (LDF), Global Decision Fusion (GDF), and Local-Global Decision Fusion (LGDF), that utilize both local and global wavelet features. Other innovative approaches include the use of Support Vector Machines (SVM) for decision-level fusion, as explored by Agrawal et al. [128]. In their system, SVMs were trained on matching scores to classify users as genuine or imposters, resulting in an accuracy rate of 98.42%.

Despite its benefits, decision-level fusion is not without challenges. Conflicting decisions from different modalities can pose challenges in fusion. Szczuko et al. [129] tackled this issue by using Dempster-Shafer Theory (DST) in a multimodal biometric system for banking applications. DST helped manage uncertainty by combining decision evidence from multiple modalities, enhancing the robustness of the system, particularly in cases of conflicting information.

2.4.3 Score-Level Fusion in Biometric Systems

In biometric systems, fusion can be performed at various stages, including sensor, feature, score, and decision levels as discussed earlier. Among these, score-level fusion is the one that offers better performances. Score-level fusion, the focus of this thesis, involves combining the matching scores generated by different biometric modalities or matchers to make a final decision regarding identification or authentication.

2.4.3.1 Normalization in Score-Level Fusion

Normalization is an important step in score-level fusion, ensuring that the scores from different biometric matchers are on a comparable scale. This process addresses discrepancies in the range, distribution, and scale of scores produced by different systems, preventing any single source from disproportionately influencing the final decision. Several normalization techniques are commonly employed:

- **Min-Max Normalization**: This method scales the scores to a fixed range, typically [0, 1], making them comparable across different classifiers [122, 130–132]. The normalized score S_i' for a matcher is computed as:

$$S_i' = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}} \tag{2.1}$$

  Where S_i is the original score, $S_{\min}$ is the minimum score, and $S_{\max}$ is the maximum score observed from that classifier.

- **Z-Score Normalization**: Z-score normalization standardizes the scores by centering them around zero and scaling them based on their standard deviation [122, 131, 132]. The normalized score S_i' is given by:

$$S_i' = \frac{S_i - \mu}{\sigma} \tag{2.2}$$

  Where μ is the mean score and σ is the standard deviation of the scores from that classifier.

- **Tanh Normalization**: This technique normalizes scores using a hyperbolic tangent function, which is particularly useful for handling outliers and skewed distributions. Tanh normalization transforms scores into a range between [0, 1], helping to manage extreme values [122, 132]. The normalized score S_i' is computed as:

$$S_i' = \frac{1}{2}\left(\tanh\left(0.01 \cdot \frac{S_i - \mu}{\sigma}\right) + 1\right) \tag{2.3}$$

2.4.3.2 Classification Approaches

Classification approaches for score-level fusion treat the task as a classification problem, where ML algorithms [133–135] are used to combine scores from different matchers. Key classification-based methods include:

- **Decision Trees (DT)**: Decision trees classify combined scores by creating a model that splits data based on score thresholds. Each branch represents a decision rule based on normalized scores, leading to the final classification.
- **SVM**: SVMs create a hyperplane in a multi-dimensional space that separates the genuine and impostor classes based on the input features. SVMs are effective in handling nonlinear relationships between different biometric modalities.
- **Neural Networks**: This approach can be used to model complex relationships between scores from different modalities. A neural network with hidden layers learns the non-linear mappings between input scores and the final classification.

2.4.3.3 Combination Approaches

Combination approaches in score-level fusion directly integrate the scores from different biometric sources using mathematical or statistical techniques. These methods do not rely on learning algorithms but instead apply predefined rules to combine scores effectively. Key combination approaches include:

- **Sum Rule**: This simple rule is one of the most straightforward methods of score-level fusion. It involves summing the scores from different biometric classifiers and comparing the resulting score against a predefined threshold to make the final decision [122, 132]:

$$S_{\text{sum}} = \sum_{i=1}^{n} S'_i \tag{2.4}$$

  Where S_i represents the normalized score from the i-th biometric classifier, and n is the number of classifiers. The final decision is based on whether S_{sum} exceeds a threshold T.

- **Weighted Sum Rule**: The weighted sum rule introduces weights to the scores from different matchers based on their reliability or performance [136]. The combined score is calculated as:

$$S_{\text{weighted}} = \sum_{i=1}^{n} w_i \cdot S'_i \tag{2.5}$$

  Where w_i is the weight assigned to the i-th classifier.

- **Product Rule**: The product rule multiplies the normalized scores from different matchers [137, 138]:

$$S_{\text{product}} = \prod_{i=1}^{n} S'_i \tag{2.6}$$

- **Min and Max Rules**: Either uses the minimum or maximum score from the set of normalized scores [122, 139]:

$$S_{\text{min}} = \min(S'_1, S'_2, \ldots, S'_n) \tag{2.7}$$

$$S_{\text{max}} = \max(S'_1, S'_2, \ldots, S'_n) \tag{2.8}$$

  The min rule is useful in conservative systems where a low score from any matcher should prevent a positive identification, while the max rule is useful in systems where a high confidence score from any matcher is sufficient for identification.

2.5 Conclusion

In conclusion, the related works reviewed in this section highlight significant advancements in ECG biometrics, speaker recognition, and multimodal biometric systems. Both fiducial and non-fiducial techniques have evolved, with recent innovations leveraging neural networks and hybrid approaches to improve accuracy and applicability. Speaker recognition has expanded with sophisticated methods for identification, verification, and diarization, while multimodal systems, through various fusion techniques, provide enhanced security and reliability. The availability of extensive databases and the development of new classification and fusion strategies have propelled the field forward, setting the stage for continued progress in biometric technologies.

Chapter 3

Proposed System and Implementation

3.1 Introduction

This section outlines the design and implementation of the proposed biometric recognition system, it is built around two distinct unimodal systems: an ECG-based system and a speaker recognition system, each of which undergoes specific preprocessing and feature extraction steps before DL models are applied for classification. The ECG system utilizes EMD for feature extraction, followed by separate models based on LSTM and GRU for classification. The speaker recognition system, on the other hand, processes speech using MFCCs and their derivatives, and applies a CNN for speaker identification.

In addition to the unimodal systems, a multimodal biometric system is proposed by combining both ECG and voice data. As there are no existing multimodal datasets containing both ECG and voice recordings from the same individuals, a simulated multimodal database was created by merging data from separate ECG and voice datasets. To enhance the system's performance, score-level fusion techniques are applied, using both Softmax and SVM methods, with three distinct fusion rules: sum, product, and max. This fusion strategy aims to improve accuracy and robustness by integrating the complementary strengths of the ECG and voice modalities.

The following subsections will provide detailed descriptions of each phase of the unimodal and multimodal systems, including the preprocessing and feature extraction methods, the deep learning architectures employed, and the fusion techniques used in the final decision-making process.

3.2 Unimodal ECG System

The implemented ECG unimodal biometric system utilized three publicly available databases: MIT-BIH, NSRDB, and the PTB Database. These databases provided ECG recordings from a diverse set of individuals, forming the basis for model training and evaluation.

Figure 3.1 provides an overview of the implemented ECG unimodal biometric system, illustrating the key stages from data preprocessing through to model training and identification.

3.2.1 Preprocessing Phase

In the preprocessing phase, a 4th order bandpass filter with a Butterworth response was applied to the raw ECG signals. The filter was designed with cutoff frequencies set at 1 Hz and 40 Hz to remove low-frequency drift and high-frequency noise, respectively, while preserving the essential features of the ECG signal. Figure 3.2 presents three distinct ECG

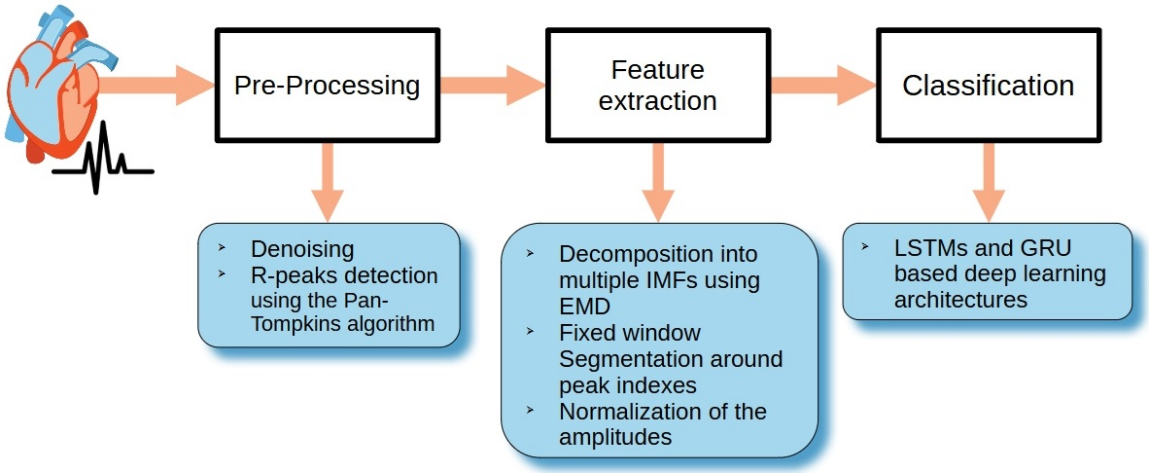


Figure 3.1: The steps of the proposed unimodal ECG system from Raw Signal to subject identification.

signals, each originating from a different database (MITBIH, NSRDB, and PTB), both before and after the denoising process. The original raw signals, shown in the left panels, exhibit various noise artifacts, including baseline wander and high-frequency interference. These distortions are effectively removed in the denoised signals, presented in the right panels, following the application of the 4th-order Butterworth bandpass filter.

Following the filtering process, the Pan-Tompkins algorithm was employed to detect R-peaks in the ECG waveforms (see figure 3.3). This algorithm provided precise identification of the QRS complexes, which is critical for the following segmentation and feature extraction.

3.2.2 Feature Extraction

The feature extraction process began by decomposing the filtered ECG signals into multiple intrinsic mode functions (IMF) using EMD. Figure 3.4 illustrates the first five IMFs and the residual signal obtained after applying EMD to the ECG signal of subject 16795 from the NSRDB. The IMFs capture oscillatory components of the signal at different frequency bands, with the first two IMFs containing the highest frequency details and the subsequent IMFs representing progressively lower-frequency components. The residual signal, displayed at the bottom, represents the slowly varying baseline of the ECG after extracting all IMFs. These decomposed components are critical in the feature extraction process, as only the first two IMFs are retained for biometric identification, while the others and the residual are discarded.

The EMD algorithm can be described as follows: First, the signal $x(t)$ is analyzed

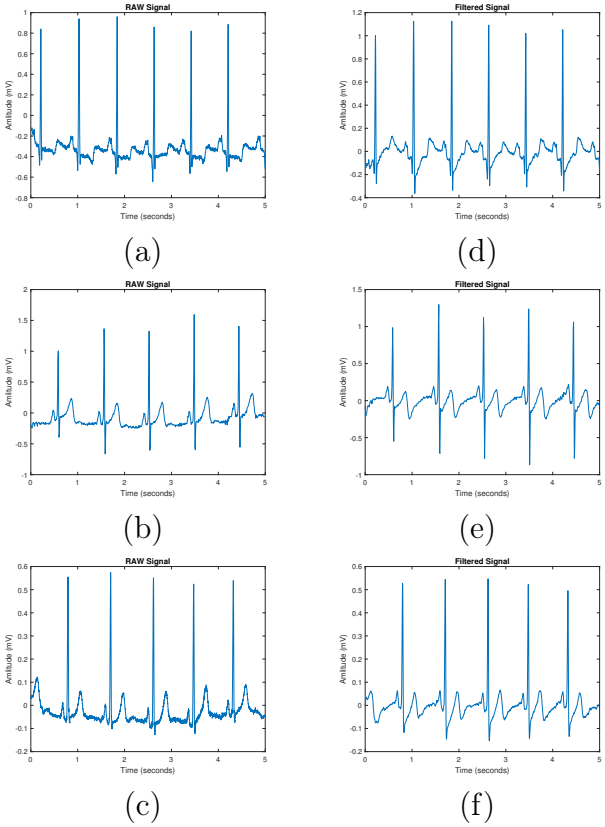


Figure 3.2: A comparative analysis of ECG signals from different databases is presented, highlighting their raw and filtered characteristics. Subfigure (a) displays an unprocessed 5-second ECG signal from the MIT-BIH database, while subfigure (d) shows its filtered counterpart. In contrast, subfigures (b) and (e) illustrate a raw and filtered ECG signal from the NSR database, respectively. Similarly, subfigures (c) and (f) compare an unprocessed and processed 5-second ECG signal from the PTB database.

by identifying all its local maxima and minima. Using these extrema, an upper envelope $e_{upper}(t)$ is constructed by interpolating through the local maxima, and a lower envelope $e_{lower}(t)$ is created by interpolating through the local minima. The mean envelope $m(t)$ then computed by averaging the upper and lower envelopes:

$$m(t) = \frac{e_{\text{upper}}(t) + e_{\text{lower}}(t)}{2} \tag{3.1}$$

Next, the mean envelope $m(t)$ is subtracted from the original signal $x(t)$, resulting in a new signal $h(t)$, called the proto-IMF:

$$h(t) = x(t) - m(t) \tag{3.2}$$

At this stage, the algorithm checks whether $h(t)$ satisfies the conditions to be considered an IMF. The first condition requires that the number of extrema and zero-crossings

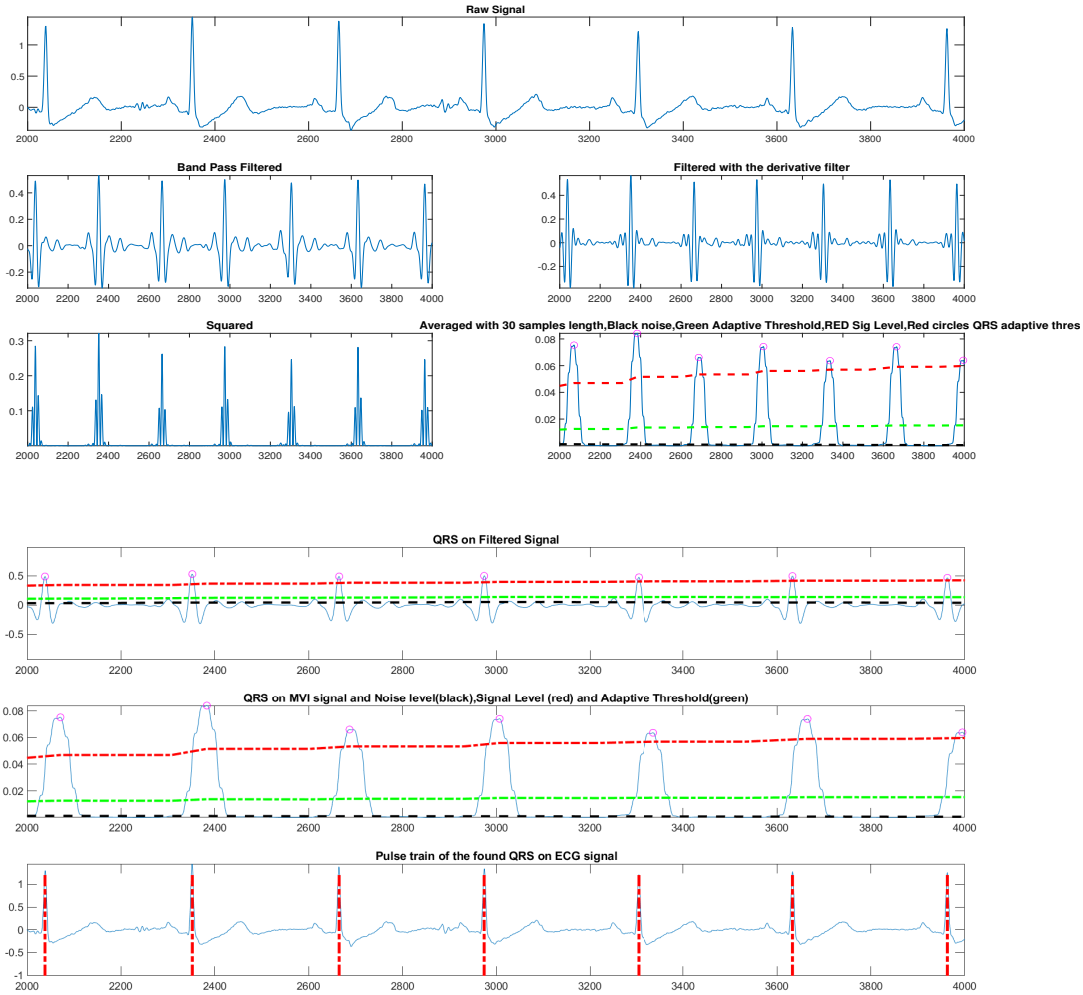


Figure 3.3: A step-by-step illustration of the Pan-Tompkins algorithm's implementation for identifying R peaks in ECG signals.

are equal or differ by at most one. The second condition requires that the mean of the upper and lower envelopes is approximately zero. If these conditions are not met, $h(t)$ undergoes further iterations called the "sifting process," where the mean envelope is recalculated and subtracted again:

$$h_k(t) = h_{k-1}(t) - m_{k-1}(t) \tag{3.3}$$

Once $h(t)$ satisfies the IMF conditions, it is extracted as the first IMF, denoted as $imf_1(dt)$:

$$imf_1(dt) = h_k(t) \tag{3.4}$$

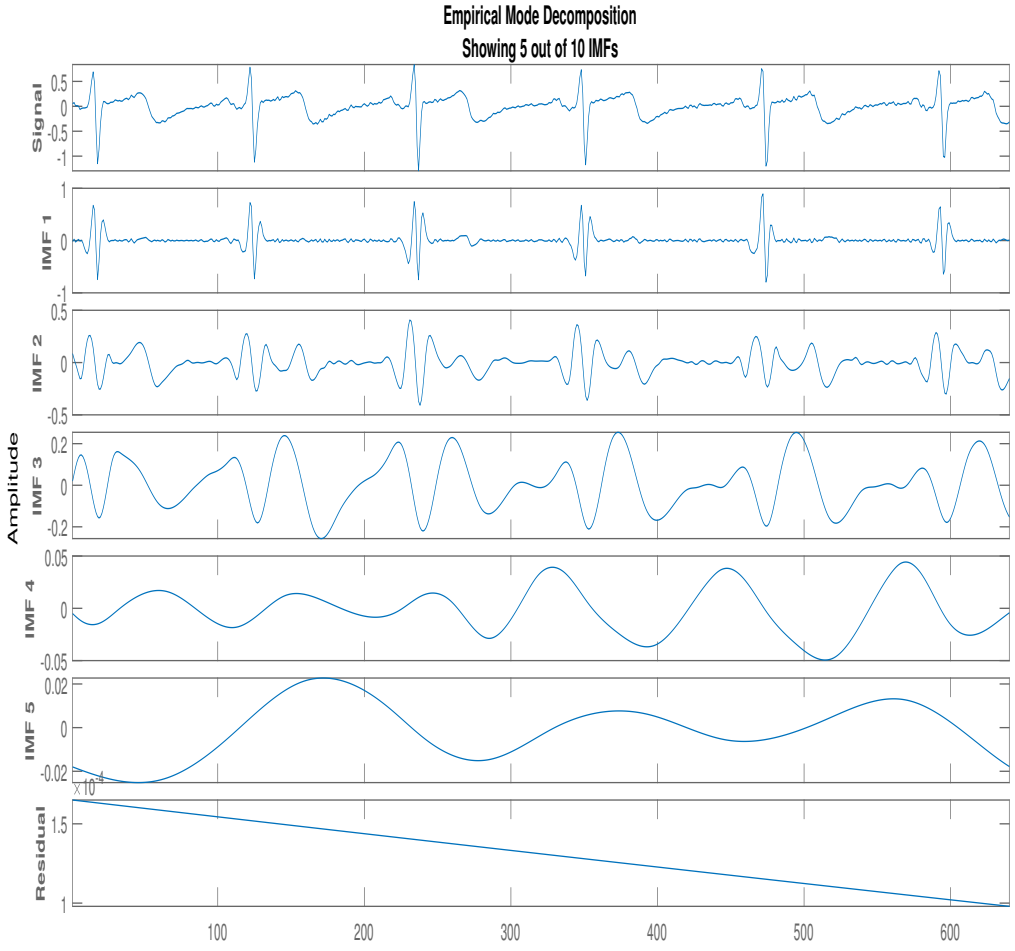


Figure 3.4: The first 5 IMFs and residual signal resulting from the application of EMD to a single-lead ECG signal from subject 16795 in the NSRDB database.

The residual signal $r_1(t)$ is then computed by subtracting $imf_1(dt)$ from the original signal xt:

$$r_1(t) = x(t) - imf_1(dt) \tag{3.5}$$

The same process is then applied to the residual signal $r_1(t)$ to extract further IMFs. This process continues iteratively:

$$r_{i+1}(t) = r_i(t) - imf_{i+1}(t) \tag{3.6}$$

The decomposition terminates when the residual signal rn becomes a monotonic function or contains no significant oscillatory components. The original signal $x(t)$ is finally

expressed as the sum of all extracted IMFs and the residual:

$$x(t) = \sum_{i=1}^{n} imf_i(t) + r_n(t) \tag{3.7}$$

Based on the R-peaks detected during the preprocessing phase, the IMFs were segmented into fixed-length windows to ensure that each segment was centered around an R-peak. This approach ensures that the model captures the QRS complex, which is highly informative for ECG-based biometric systems. Specifically, considering the maximum duration of the QRS complex, a window of 100 ms was chosen to segment the IMFs. For the PTB database, 50 samples before and 50 samples after each R-peak were extracted, capturing the full duration of the QRS complex. For the MIT-BIH database, given its different sampling rate, 18 samples before and 18 samples after each R-peak were selected. For the MIT-BIH Atrial Fibrillation database, where a higher sampling frequency is used, 5 samples before and 6 samples after each R-peak were selected.

To further standardize the features and facilitate the model's training, the amplitude of each segmented IMF was normalized within the range [0 - 1] according to 3.8. This normalization process ensures that the input data fed into the deep learning models is uniform and eliminates potential biases due to amplitude variations across different segments or subjects, allowing the models to focus on temporal patterns and distinguishing features in the signal.

$$x_{norm} = \frac{x - min_x}{max_x - min_x} \tag{3.8}$$

3.2.3 Proposed DL Model

Two separate deep learning models were developed and trained for individual identification: one based on LSTM networks (Table 3.1) and the other on GRU (Figure 3.5). Both models were designed to capture the temporal dependencies and sequential nature of the ECG signal, leveraging their respective architectures to model the dynamic behavior of ECG features over time.

The proposed LSTM architecture, detailed in Table 3.1, is designed for ECG-based biometric identification. The model begins with a sequence input layer, where 2-dimensional ECG sequences are fed into the network. This is followed by the first LSTM layer, which contains 100 neurons to capture temporal dependencies in the ECG signal. To prevent overfitting, a dropout layer with a rate of 20% is applied. Another LSTM layer with 100 neurons is added to further capture temporal patterns. A fully connected layer follows, with N hidden units corresponding to the number of output classes (i.e., individuals to

Table 3.1: The proposed LSTM architecture.

Layer #	Type	Description
1	Sequence Input	2-dimensional sequence input
2	LSTM	an LSTM layer with 100 neurons
3	Dropout	Dropout at 20%
4	LSTM	an LSTM layer with 100 neurons
5	fully connected	Fully connected layer with N hidden units
6	Softmax	Implementation of Softmax
7	Classification Output	Cross Entropy Function

be identified). The network ends with a Softmax layer to convert the outputs into probabilities, and the classification output is computed using a cross-entropy loss function. This architecture effectively leverages the recurrent properties of LSTM layers to model the sequential nature of ECG signals for biometric identification.

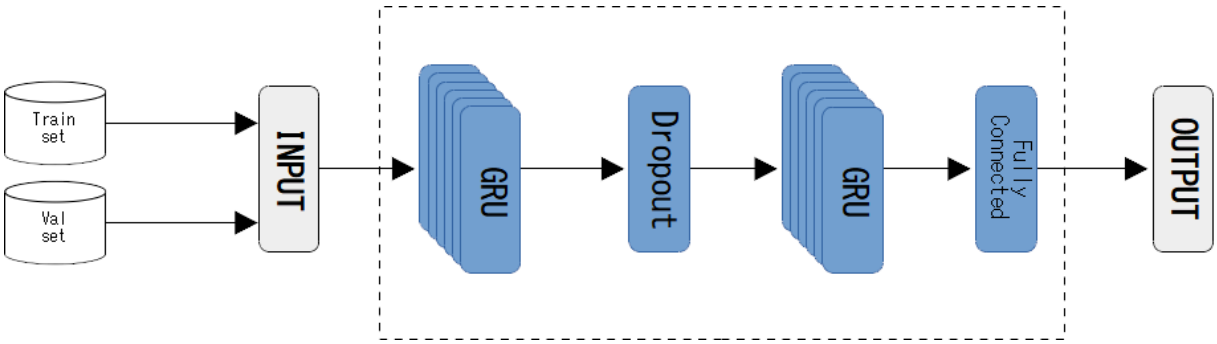


Figure 3.5: The proposed GRU deep neural network model architecture.

Table 3.2: The proposed GRU architecture.

Layer #	Type	Description
1	Sequence Input	2-dimensional sequence input
2	GRU	a GRU layer with 100 neurons
3	Dropout	Dropout at 20%
4	GRU	a GRU layer with 100 neurons
5	fully connected	Fully connected layer with N hidden units
6	Softmax	Implementation of Softmax
7	Classification Output	Cross Entropy Function

The proposed GRU architecture, shown in Figure 3.5 and Table 3.2, is structurally similar to the LSTM-based model, with the key difference being the use of GRUs in place of LSTM layers. The model starts by taking 1-dimensional ECG sequences as input from

both the training and validation sets. The first GRU layer processes the input sequence, followed by a dropout layer with a 20% rate to prevent overfitting. This is followed by a second GRU layer, which continues to capture the temporal dependencies in the ECG data. A fully connected layer with N units (corresponding to the number of output classes) is then applied, and the final output is generated.

The LSTM and GRU models were trained separately on the extracted and normalized features, allowing for a comparison of their effectiveness in identifying individuals from ECG data.

3.3 Unimodal Speaker Recognition System

The implemented speaker recognition system was developed using a subset of 47 speakers from the LibriSpeech database. The system follows a structured pipeline that begins with preprocessing the audio signals, followed by feature extraction, and finally, classification using a CNN as shown in figure 3.6.

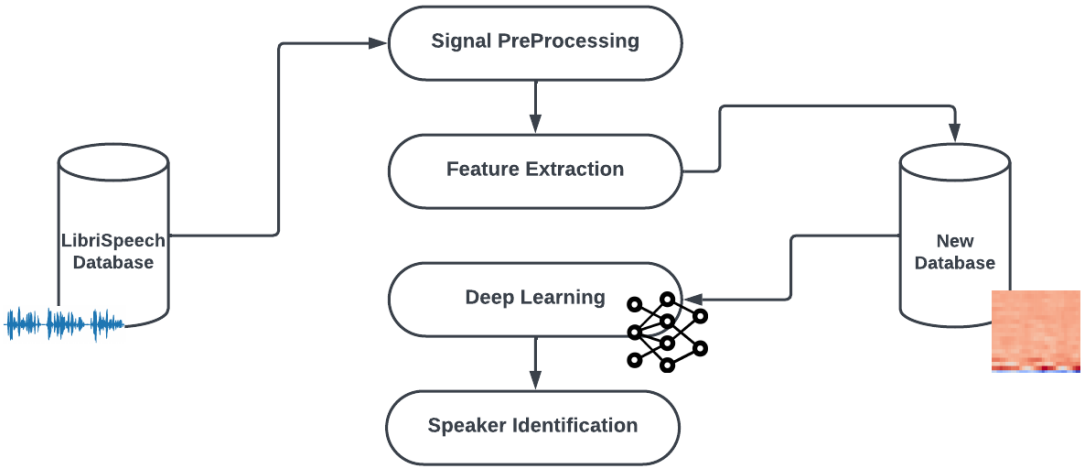


Figure 3.6: A visual representation of the proposed Deep Learning-Based Speaker Identification System: A Comprehensive Pipeline from Signal Preprocessing to Model Evaluation.

3.3.1 Preprocessing Phase

During the preprocessing phase, the raw audio recordings undergo a critical step of silence removal to filter out non-informative sections of the speech signal. These silent portions, which often carry little to no relevant speaker-specific information, can introduce unnecessary noise and computational overhead during the subsequent stages of the system. The

silence removal process is performed by calculating a dynamic threshold based on the mean and standard deviation of the decibel (dB) levels across the entire audio signal.

First, the mean decibel level $mean_{db}$ of the signal is computed, as shown in equation (3.9), where x_i represents the dB level of each audio frame, and n is the total number of frames:

$$mean_{db} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.9}$$

Next, the standard deviation std_{db} of the decibel levels is calculated, reflecting the variability in audio intensity throughout the recording. This is described by equation (3.10), where N is the number of frames, and $\overline{x}$ is the mean level:

$$std_{db} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{3.10}$$

To ensure that all silent sections of the audio are properly identified, a threshold $threshold_{db}$ is defined. Any audio segment with a decibel level below this threshold is considered silent and is removed from the signal. The threshold is computed as the difference between the mean and the standard deviation of the decibel values:

$$threshold_{db} = mean_{db} - std_{db} \tag{3.11}$$

By utilizing this dynamic threshold, the system is capable of adapting to varying noise levels within different recordings, ensuring that only truly silent segments are eliminated. This method effectively preserves the informative portions of the speech while removing segments that would otherwise not contribute to speaker identification.

After the silence removal, the processed audio is segmented into fixed-length sound windows of 300 milliseconds, with an overlap of 150 milliseconds between consecutive segments. The use of overlapping windows ensures temporal continuity, which is essential for capturing subtle transitions in speech and preserving speaker-specific characteristics. The overlapping also reduces the loss of valuable information at the segment boundaries, thus improving the overall system performance. Figure 3.7 illustrates the resulting signal after the silence removal step, highlighting the non-silent regions retained for further processing.

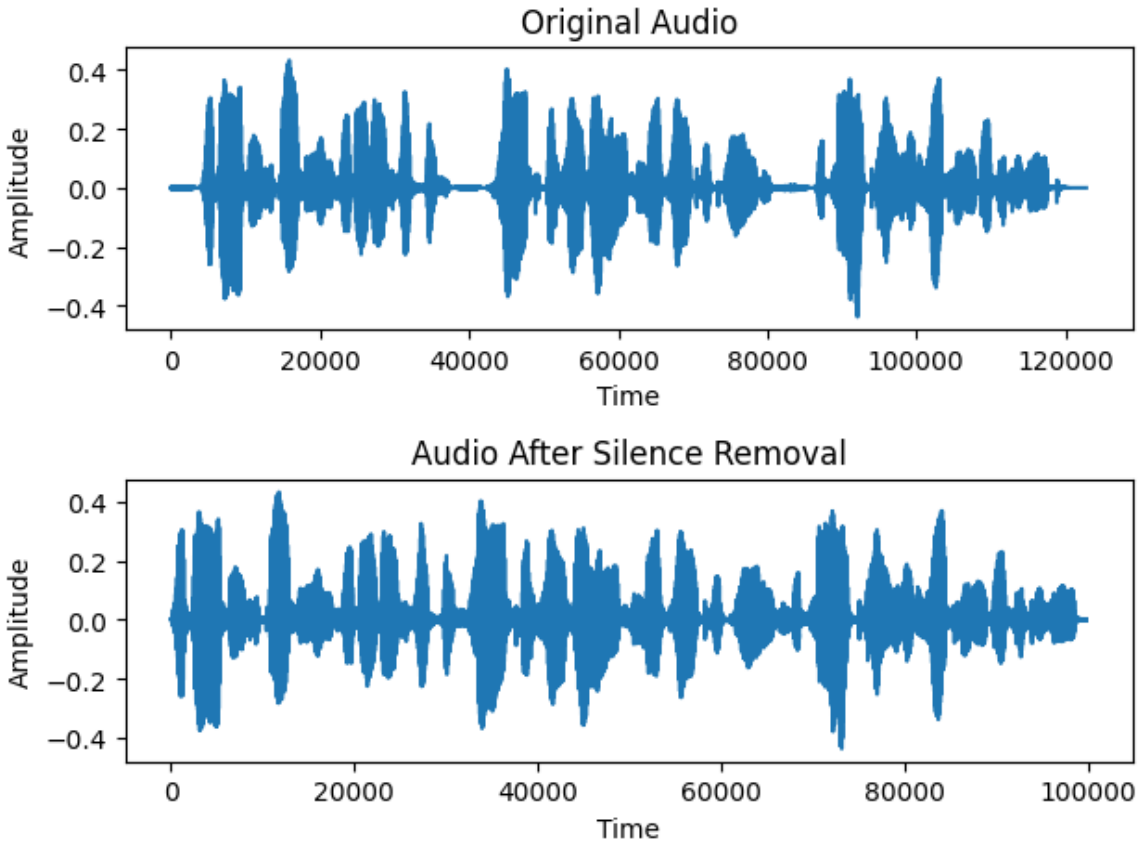


Figure 3.7: Comparative analysis of original and silence-removed audio waveforms for the proposed speaker identification system.

3.3.2 Feature Extraction

In the feature extraction phase, the primary task is to distill meaningful information from the audio that can effectively represent speaker characteristics. The technique employed is the computation of MFCCs, which are widely regarded as one of the most effective features for speech and speaker recognition tasks. MFCCs are computed by transforming the audio signal into the Mel scale, a perceptual scale of pitches that closely aligns with the way humans perceive sound, particularly speech. This transformation allows the system to focus on the frequency bands most important for distinguishing between different speakers. By converting the signal from the time domain to the frequency domain and emphasizing critical speech frequencies, MFCCs provide a compact, speaker-specific representation that captures the unique spectral qualities of each individual's voice.

The process begins by taking each segmented frame of audio and applying the Fourier transform to convert the time-domain signal into the frequency domain. Following this, the frequency components are mapped onto the Mel scale using a series of triangular filters spaced logarithmically to reflect the non-linear perception of frequency in human

hearing. After applying a logarithmic function to the Mel-filtered frequencies, the discrete cosine transform (DCT) is used to compress the signal by decorrelating the coefficients and retaining only the most important information in a set of low-dimensional MFCC features. The first 40 coefficients are retained for each audio frame, representing the most critical speaker-specific spectral characteristics.

In addition to the MFCCs, the system also computes the first two derivatives, known as delta and delta-delta (or acceleration) coefficients. These derivatives capture the temporal dynamics of the speech signal, which are crucial for improving speaker discrimination. While the MFCCs themselves capture static spectral information for each frame, the delta coefficients represent the rate of change (velocity) of the MFCCs over time, and the delta-delta coefficients capture the acceleration or rate of change of the delta coefficients. This dynamic information allows the system to model subtle fluctuations in the speech patterns, which are often speaker-specific. Together, the MFCCs, delta, and delta-delta coefficients form a robust feature set that combines both static and dynamic information, providing a rich characterization of the speaker's vocal identity over time.

By incorporating these features, the process ensures that the system can effectively capture and distinguish between the nuanced vocal patterns of different speakers. This enriched feature set serves as the input for the subsequent classification phase, where a DL model can learn to identify speakers based on the distinctive characteristics encoded within these features.

3.3.3 Proposed DL Model

The speaker identification model designed for this task is a CNN architecture (Table 3.3)tailored to capture the spatial and temporal patterns in the extracted features that are described in section 3.3.2. The model begins with an input layer that processes spectrogram data with a shape corresponding to the dimensions of the preprocessed audio feature matrix, specifically for 2D convolution.

The first convolutional layer uses 96 filters with a kernel size of 3×3 and applies a linear activation function, followed by batch normalization to stabilize the learning process and speed up convergence. A Leaky ReLU activation function (with $\alpha = 0.2$) is applied to introduce non-linearity, followed by a max-pooling layer that reduces the spatial dimensions by pooling over a 3×1 window, allowing the network to focus on the most salient features. This is followed by a dropout layer with a probability of 25% to prevent overfitting by randomly dropping units during training.

The next two convolutional blocks follow a similar structure, each containing 64 filters, with 3×3 kernels, batch normalization, Leaky ReLU activation, and max-pooling

Table 3.3: CNN architecture for speaker identification.

Layer #	Layer Type	Description
Block 1: Initial Convolution and Pooling		
1	Input	Input shape: $(120, 10, 1)$
2	Conv2D	96 filters, kernel size: 3×3, linear activation
3	Batch Normalization	Batch normalization
4	Leaky ReLU	Leaky ReLU with $\alpha = 0.2$
5	MaxPooling2D	Pool size: 3×1, padding: same
6	Dropout	Dropout rate: 25%
Block 2: Second Convolution and Pooling		
7	Conv2D	64 filters, kernel size: 3×3, linear activation
8	Batch Normalization	Batch normalization
9	Leaky ReLU	Leaky ReLU with $\alpha = 0.2$
10	MaxPooling2D	Pool size: 1×3, padding: same
11	Dropout	Dropout rate: 25%
Block 3: Final Convolution and Pooling		
12	Conv2D	64 filters, kernel size: 3×3, linear activation
13	Batch Normalization	Batch normalization
14	Leaky ReLU	Leaky ReLU with $\alpha = 0.2$
15	MaxPooling2D	Pool size: 1×3, padding: same
16	Dropout	Dropout rate: 50%
Block 4: Classification Layers		
17	Flatten	Flatten output
18	Dense	128 units, linear activation
19	Batch Normalization	Batch normalization
20	Leaky ReLU	Leaky ReLU with $\alpha = 0.2$
21	Dense	Number of output classes (softmax)

operations. The second pooling operation applies a pooling size of 1×3, reducing the temporal dimension, while the dropout layers are kept at 25% for these blocks to further regularize the model. After the third convolutional block, the dropout rate is doubled to 50% to enhance regularization.

The flattening layer transforms the 2D feature maps into a 1D vector, which is passed through a fully connected layer with 128 units and linear activation. Batch normalization and Leaky ReLU are applied once more to introduce non-linearity and maintain the stability of activations. Finally, the output layer employs a softmax activation to produce a probability distribution over the target classes, corresponding to the 47 speakers.

3.4 Multimodal System

To overcome the limitations of the unimodal systems proposed in section 3.2 and section 3.3, we have implemented a multimodal biometric system that integrates both proposed systems to improve the accuracy and robustness of biometric recognition. Figure 3.8 illustrates the overall architecture of the multimodal biometric system, including preprocessing, feature extraction, and classification phases, showcasing the fusion of both ECG and voice signals at the score level.

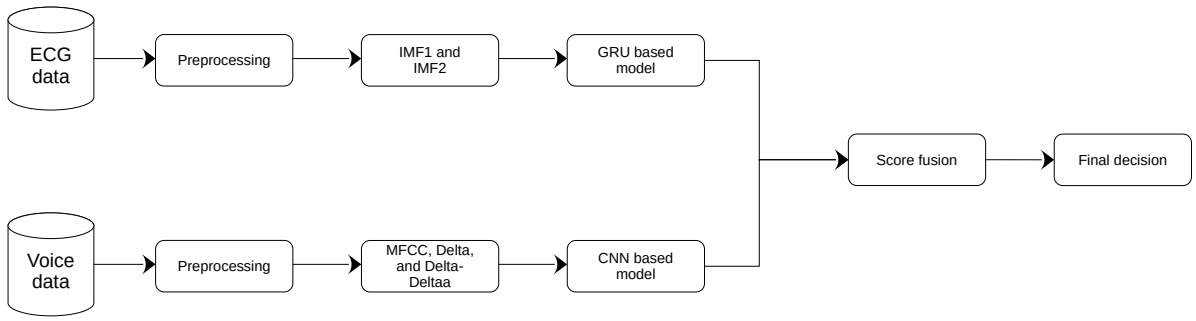


Figure 3.8: An Illustrated Diagram of the Overall Structure of the Proposed Multimodal System.

3.4.1 Multimodal Database Simulation

As there is no known publicly available database that contains both ECG and voice recordings from the same subjects, we simulated a multimodal dataset by combining segments from MITBIH, which contains ECG recordings, with segments from the LibriSpeech dataset, which provides speech data. Specifically, the 47 subjects from the MIT-BIH database and a subset of 47 speakers from the LibriSpeech database was selected for this purpose.

For each subject, 900 samples were created, with an equal number of samples coming from each modality. This allowed us to simulate a balanced multimodal dataset, where each subject is represented by both ECG and voice data. By merging these datasets, the multimodal system leverages the complementary nature of physiological (ECG) and behavioral (voice) biometrics to enhance identification accuracy.

3.4.2 Scores Fusion

Our proposed multimodal biometric system applies score-level fusion to combine the outputs from both ECG and voice modalities. Score-level fusion operates by integrating the

matching scores from individual modalities to produce a single score, which is then used for the final decision of identification or verification as discussed in section 2.4.3. This approach offers flexibility and ease of implementation compared to other fusion strategies like feature-level or decision-level fusions. The used fusion technique are illustrated in Figure 3.9.

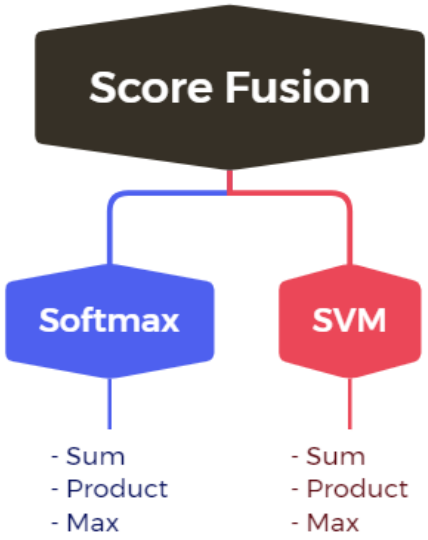


Figure 3.9: The proposed multimodal system employs softmax and SVM for score fusion.

3.4.2.1 Softmax-Based Fusion

In our system, one of the methods employed for score-level fusion is based on the Softmax function. Softmax normalizes the scores from each modality into a probability distribution, ensuring that the outputs are comparable across different scales. We employ three different rules for fusing the scores:

- **Sum Rule**: The Softmax scores from both ECG and voice modalities are summed together. The final decision is made based on the highest cumulative score across all classes.
- **Product Rule**: Here, the Softmax scores are multiplied together. This method is particularly effective when scores from the two modalities are independent, as the product reduces the impact of any single, unreliable score.
- **Max Rule**: For this rule, the maximum score from either modality is taken as the final score. This approach emphasizes the strongest evidence from either modality,

which may be beneficial in cases where one modality performs significantly better than the other.

3.4.2.2 SVM-Based Fusion

In addition to Softmax-based fusion, we employ SVM classifiers for score fusion. The SVM learns an optimal decision boundary in a high-dimensional feature space, combining the scores from both modalities for classification. Similar to Softmax, we implement the following fusion rules using the SVM model:

- **Sum Rule**: The SVM takes the sum of the scores from both modalities as input and learns an optimal separation between classes based on this fused score.
- **Product Rule**: The product of the scores from the ECG and voice modalities is passed to the SVM, allowing the classifier to learn from the joint distribution of scores.
- **Max Rule**: The SVM is trained on the maximum score between the two modalities, focusing on the stronger signal for making decisions.

These fusion techniques aim to enhance system performance by combining the complementary nature of ECG and voice modalities. By leveraging both Softmax and SVM fusion methods with various rules, the system achieves greater robustness and adaptability to varying conditions across different subjects.

3.5 Implementation

The implementation of the proposed system is a critical phase of the research, translating theoretical frameworks into practical applications. This section provides a detailed description of the technical setup and environment in which the experiments were conducted, as well as the metrics used to evaluate the performance of the models. By outlining the software and hardware components, this section offers insight into the computational resources that supported the development and testing of the system.

The evaluation metrics chosen for this study are essential in assessing the model's accuracy and robustness across multiple datasets and modalities. The metrics are carefully selected to reflect both the overall system performance and specific characteristics, such as precision, recall, and error rates, in each of the evaluated models. Understanding these metrics is crucial for interpreting the results and validating the effectiveness of the

proposed multimodal system. Together, these aspects form the foundation of a comprehensive and rigorous evaluation process, ensuring that the results are both reliable and replicable.

3.5.1 Implementation Environment

All code implementations for the proposed biometric system were developed and executed on a machine equipped with an Intel Core i3-10100F processor, a quad-core CPU that can handle moderate computational tasks efficiently. Initially, the machine was configured with 8GB of RAM, which was sufficient for basic operations and initial phases of development. However, as the project progressed and the complexity of the models increased, particularly with voice processing algorithms requiring significant memory for training, the RAM was upgraded to 16GB. This upgrade was essential for handling larger datasets and running more computationally demanding tasks, such as managing the multimodal fusion processes. The system was executed on Windows 11, the latest version of the Windows operating system at the time of writing, offering a stable and reliable environment for software development.

3.5.1.1 Implementation of the ECG-based Unimodal System

For the ECG-based unimodal system, the primary software environment utilized was MATLAB, chosen due to its robust suite of tools for advanced signal processing and feature extraction. MATLAB was particularly well-suited for handling complex ECG processing tasks such as the detection of R-peaks, segmentation, and the extraction of IMFs from using EMD better than python. These capabilities made MATLAB an essential tool in the preprocessing and feature extraction phases of the ECG unimodal system.

Initially, the development work was conducted using MATLAB 2022a, which provided all the necessary functionality for signal analysis and manipulation. However, as the project progressed and new features were introduced, especially in later stages involving more intricate model evaluation and optimization, an upgrade to MATLAB 2024a was implemented. The newer version offered enhanced machine learning toolboxes and additional resources that supported the development of more scalable models. The transition to MATLAB 2024a played a critical role in fine-tuning the ECG-based system and advancing the project towards its final stages.

3.5.1.2 Implementation of the Voice-based Unimodal System

In contrast to the ECG-based system, the voice-based speaker recognition system was developed using Python, a widely-adopted programming language known for its versatility and extensive ecosystem of libraries dedicated to machine learning and deep learning. The implementation specifically leveraged the TensorFlow framework, which is renowned for its scalability, flexibility, and support for deep neural network architectures. TensorFlow facilitated the efficient design, training, and evaluation of the CNNs that formed the core of the speaker recognition model.

Python's robust libraries for speech processing, such as Librosa for feature extraction, were employed to preprocess audio data and compute key acoustic features, including MFCCs and their derivatives. These features were crucial for capturing speaker-specific characteristics within the audio signals. TensorFlow's powerful deep learning framework enabled the rapid prototyping and optimization of CNN models, which were used to classify the speech data by learning complex patterns within the MFCC features.

3.5.1.3 Implementation of the Multimodal System

The implementation of the multimodal biometric system and the subsequent score fusion was primarily carried out using Python, with the support of libraries such as TensorFlow and NumPy for model integration, training, and evaluation. The system leverages the strengths of both ECG and voice modalities, and each model was handled separately before being fused at the score level.

The ECG-based model was initially developed and trained using MATLAB 2022a [1], which provided the sufficient tools for signal processing and deep learning. After training, the model was saved in the MATLAB .mat format. As part of the multimodal system's integration, MATLAB 2024a was used to export the trained ECG model to TensorFlow's .h5 format, ensuring compatibility with the Python-based framework used for multimodal fusion.

On the other hand, the voice-based model was trained directly using TensorFlow in Python. Once both models were finalized, their outputs were combined in the multimodal system through various score-level fusion techniques, fully implemented in Python using TensorFlow and NumPy for efficient computation and integration. This approach allowed for seamless fusion of ECG and voice data, allowing for enhanced system's overall recognition performance.

The choice of Python and TensorFlow for this part of the project was driven by their ease of integrating diverse libraries for both audio processing and neural network training. This environment proved instrumental in developing a highly efficient and accurate

speaker recognition system that complemented the ECG-based unimodal system.

3.5.2 Evaluation Metrics

The evaluation of biometric systems is essential to assess their performance, accuracy, and reliability in real-world applications. Metrics provide quantitative measures to gauge the effectiveness and efficiency of biometric recognition algorithms, helping researchers, developers, and end-users make informed decisions regarding system design, deployment, and optimization. In this section, we delve into the various metrics used to evaluate biometric systems comprehensively.

Biometric systems are evaluated based on several key performance metrics, which encompass different aspects of system performance, including accuracy, security, robustness, and usability. These metrics provide insights into the system's ability to correctly identify or verify individuals, mitigate security threats, withstand adversarial attacks, and accommodate diverse user populations.

To assess the performance of the proposed multimodal system combining, we employed a comprehensive set of evaluation metrics. These metrics provide a holistic view of the system's effectiveness, covering various aspects of its performance. The following metrics were used to evaluate our model:

3.5.2.1 Accuracy

In biometric systems, accuracy is a fundamental metric used to evaluate the overall performance of the system in correctly identifying or verifying individuals. It measures the proportion of correct identifications or verifications made by the system out of the total number of attempts. It directly reflects the system's reliability in distinguishing between genuine users and impostors. It is calculated using the equation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3.12}$$

where TP (True Positives) represents the number of correctly accepted genuine users, TN (True Negatives) denotes the number of correctly rejected impostors, FP (False Positives) is the number of impostors incorrectly accepted as genuine users, and FN (False Negatives) indicates the number of genuine users incorrectly rejected.

3.5.2.2 Precision

Precision quantifies the model's ability to avoid labeling negative instances as positive. It is the ratio of correctly identified positive instances to the total instances labeled as

positive by the model:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.13}$$

3.5.2.3 Recall

Also known as sensitivity, recall measures the model's ability to find all positive instances. It is the ratio of correctly identified positive instances to the total actual positive instances:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.14}$$

3.5.2.4 F1-score

The F1-score provides a single score that balances both precision and recall. It is the harmonic mean of precision and recall, offering a more robust measure of the model's performance, especially in cases of imbalanced datasets:

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.15}$$

3.5.2.5 Specificity

This metric measures the model's ability to correctly identify negative instances. It is the ratio of correctly identified negative instances to the total actual negative instances:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3.16}$$

3.5.2.6 FAR

FAR measures the probability of erroneously accepting an impostor as a genuine user during authentication. It quantifies the system's susceptibility to false positive errors, where unauthorized individuals gain access to secured resources or facilities. A lower FAR indicates a higher level of security and reliability in rejecting impostors. It is calculated using the equation:

$$\text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{3.17}$$

3.5.2.7 FRR

FRR measures the probability of erroneously rejecting a genuine user during authentication. It quantifies the system's susceptibility to false negative errors, where legitimate

users are denied access due to authentication failures. A lower FRR indicates a higher level of user convenience and acceptance, as genuine users experience fewer authentication failures. It is calculated using the equation:

$$\text{FRR} = \frac{\text{FN}}{\text{TP} + \text{FN}} \tag{3.18}$$

3.5.2.8 EER

EER represents the point where FAR and FRR are equal, indicating the system's balanced performance in distinguishing between genuine and impostor samples. It serves as a critical threshold for evaluating the overall accuracy and effectiveness of biometric systems. A lower EER signifies higher discriminative power and reliability in biometric recognition tasks. It is calculated using the equation:

$$\text{EER} = \frac{\text{FAR} + \text{FRR}}{2} \tag{3.19}$$

These metrics were chosen to provide a comprehensive evaluation of our multimodal system's performance. By using this diverse set of metrics, we can assess the system's overall accuracy, its ability to correctly identify both positive and negative instances, and its performance in terms of false acceptances and rejections. The combination of these metrics allows for a nuanced understanding of the system's strengths and potential areas for improvement. The results obtained from these metrics enable us to:

- Gauge the effectiveness of our approach in combining ECG and voice data
- Compare the performance of our multimodal system to unimodal approaches
- Assess our system's performance against existing state-of-the-art methods in the field
- Identify any biases or imbalances in the system's performance across different aspects of classification and identification tasks

3.6 Conclusion

In conclusion, the proposed system integrates multimodal approaches to biometric recognition, leveraging distinct ECG and speaker recognition systems for individual classification and user identification. By employing preprocessing and feature extraction techniques, such as EMD for ECG signals and MFCC feature extraction for voice data, along with

deep learning models like GRU, LSTM, and CN. Furthermore, the implementation of score-level fusion using Softmax and SVM methods provides a comprehensive solution for biometric authentication by capitalizing on the complementary strengths of ECG and voice data.

Additionally, the implementation of the proposed system was successfully carried appropriate hardware and software resources to ensure efficient model training and evaluation. The use of diverse and robust evaluation metrics allowed for a comprehensive assessment of the system's performance, highlighting its strengths in handling both unimodal and multimodal inputs. The chosen environment and metrics contributed significantly to validating the proposed models and demonstrating their effectiveness in the context of speaker recognition and ECG analysis.

Chapter 4

Experimental Results and Discussion

4.1 Introduction

This section provides a comprehensive analysis of the outcomes of the proposed biometric system, evaluating its performance across both unimodal and multimodal configurations. This section discusses the findings from the ECG-based unimodal system, the voice-based speaker recognition system, and the multimodal fusion of ECG and voice data. Each of these systems is examined in terms of accuracy, robustness, and the efficiency of their respective deep learning models, with key performance metrics such as accuracy, precision, recall, F1-score, and EER as discussed in section 3.5.2.

The unimodal ECG system explores the classification results derived from the ECG signals of multiple databases, demonstrating the capability of GRU and LSTM-based architectures to effectively process and classify cardiac data for biometric application. The performance of the models is analyzed across several metrics, with the results compared to state-of-the-art methods. This discussion helps validate the efficacy of ECG signals as a reliable biometric modality for identity verification.

The unimodal speaker recognition system examines the voice-based biometric system, where CNN models are applied to MFCC and its derivatives to extract speaker-specific features. The system's results are evaluated to determine how well the proposed model identifies individuals based on voice patterns, providing insights into the strengths and limitations of voice as a singular modality for speaker identification.

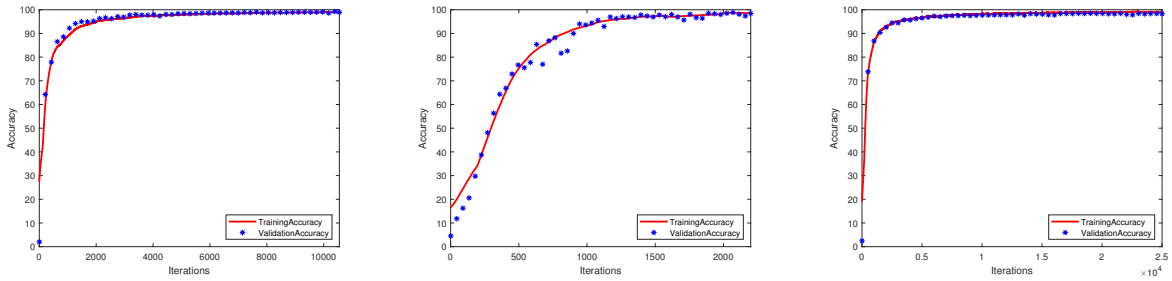
Finally, the multimodal system section delves into the fusion of ECG and voice data, which was implemented to enhance recognition accuracy and reduce error rates. By combining the two modalities at the score level using various fusion techniques (Softmax and SVM with Sum, Product, and Max rules), this section presents how multimodal biometrics can outperform unimodal approaches. The discussion also highlights how the proposed fusion methods effectively balance the trade-offs between false acceptance and rejection rates, providing a more robust solution for biometric identification.

4.2 Unimodal ECG System

Our proposed ECG-based unimodal system [1] was designed to classify individuals based on their unique heart signals, using deep learning models for feature extraction and classification. The dataset used for this system consists of ECG recordings from three databases: MITBIH, PTB, and NSRDB (described in Section 2.2.6). The data from each database was randomly split into three sets: 70% for training, 20% for testing, and 10% for validation. This ensures that the models are properly trained and tested on unseen data, while the validation set is used to monitor performance during training and prevent overfitting.

For training the deep learning models, both the GRU and LSTM architectures were employed. The training process was conducted with a learning rate of 0.001, a batch size of 150, and for a total of 50 epochs. The Adam optimizer [140] was used, as it provides adaptive learning rates and is well-suited for complex, high-dimensional data like ECG signals. The performance of the network during training was evaluated based on the validation loss, and the model with the best validation loss was selected as the final model for testing.

Figure 4.1 presents the training plots for the GRU model, showing the loss and accuracy curves across the MITBIH, PTB, and NSRDB databases. Similarly, Figure 4.2 provides the corresponding training plots for the LSTM model on the same datasets. These figures illustrate the convergence of both models and highlight the differences in performance across the three databases.

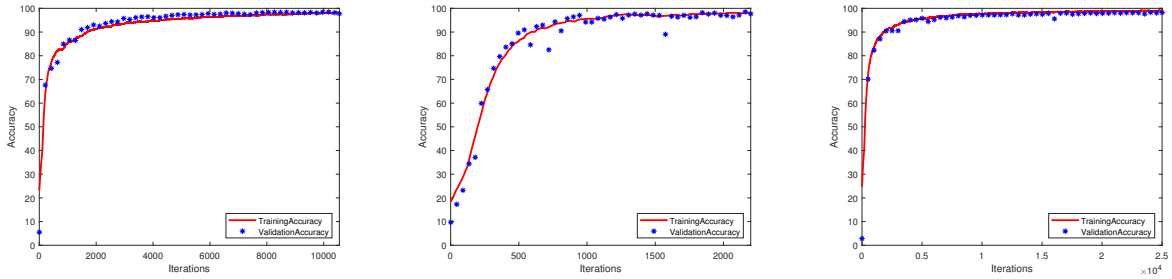


(a) NSRDB **(b)** PTB **(c)** MITBIH

Figure 4.1: Training and validation accuracies of the GRU model on (a) NSRDB (b) PTB (c) MIT-BIH.

In Figure 4.1, which presents the performance of the GRU model, we observe rapid convergence in all three datasets. For the NSRDB dataset, the GRU model quickly reaches near-perfect accuracy with both training and validation curves following a nearly identical path, indicating a strong ability to generalize to unseen data without overfitting. The performance on the PTB dataset is similarly impressive, although the convergence is slightly slower than in NSRDB. Nevertheless, both accuracies reach close to 100%, reflecting the model's robustness. On the MIT-BIH dataset, the GRU model again shows rapid learning, with the training and validation accuracies achieving almost identical values, indicating excellent performance and minimal overfitting.

In Figure 4.2, which depicts the performance of the LSTM model, we see similar trends in terms of accuracy. For the NSRDB dataset, the LSTM model also reaches near-perfect accuracy very quickly, and both the training and validation accuracies are closely aligned, suggesting a high level of generalization. On the PTB dataset, the LSTM model shows strong performance, comparable to the GRU, with both accuracies converging to near



(a) NSRDB **(b)** PTB **(c)** MITBIH

Figure 4.2: Training and validation accuracies of the LSTM model on (a) NSRDB (b) PTB (c) MIT-BIH.

100%. This further highlights the robustness of LSTM for this dataset. For the MIT-BIH dataset, the LSTM model demonstrates fast convergence, with almost no noticeable difference between training and validation accuracies, confirming the model's ability to generalize well across different subjects within the dataset.

Tables 4.1 and 4.2 summarize the performance of the system across the three databases for both the GRU and LSTM models, providing a detailed comparison of their effectiveness in classifying ECG signals from different subjects.

Table 4.1: The classification results of the proposed GRU model.

Dataset	Accuracy	Precision	Recall	Specificity	F1-score	FAR	FRR
MIT-BIH	98.57%	98.58%	98.62%	99.97%	98.60%	0.031%	1.42%
NSRDB	99.17%	99.16%	99.14%	99.95%	99.15%	0.048%	0.84%
PTB	98.26%	98.18%	98.18%	99.96%	98.14%	0.037%	1.82%

The results in Table 4.1 provide a comprehensive evaluation of the proposed GRU model across three ECG datasets: MIT-BIH, NSRDB, and PTB. For the MIT-BIH dataset, the model achieves an accuracy of 98.57%, indicating its strong overall classification ability. The precision (98.58%) and recall (98.62%) are nearly identical, suggesting that the model is equally proficient in correctly identifying true positives and minimizing false negatives. Additionally, the specificity is extremely high at 99.97%, reflecting the model's ability to reject impostor samples effectively. The F1-score of 98.60% reinforces the balance between precision and recall. Notably, the FAR is exceptionally low at 0.031%, and the FRR is 1.42%, indicating that the model is highly effective in both rejecting impostors and correctly identifying genuine subjects.

For the NSRDB dataset, the GRU model shows even higher performance, achieving an accuracy of 99.17%, which surpasses the MIT-BIH results. Similarly, the precision (99.16%), recall (99.14%), and F1-score (99.15%) are all remarkably close, reflecting consistent model performance. The specificity is slightly lower than MIT-BIH at 99.95%, but

still outstanding. The FAR is marginally higher at 0.048%, but this remains a low rate. Meanwhile, the FRR of 0.84% is the lowest across all three datasets, demonstrating that the model is least likely to incorrectly reject genuine subjects in this dataset.

For the PTB dataset, the model achieves a slightly lower but still impressive accuracy of 98.26%. The precision and recall are both 98.18%, and the F1-score is 98.14%, indicating a balanced classification performance. The specificity remains high at 99.96%, similar to the other datasets. The FAR of 0.037% and FRR of 1.82% show that while the model performs well on this dataset, it is slightly less effective than on MIT-BIH and NSRDB. This minor drop in performance could be attributed to the unique characteristics of the PTB dataset or the signal quality variations across its samples.

Table 4.2: The classification metrics of the proposed LSTM model.

Dataset	Accuracy	Precision	Recall	Specificity	F1-score	FAR	FRR
MIT-BIH	98.33%	98.39%	98.43%	99.96%	98.40%	0.036%	1.61%
NSRDB	98.27%	98.27%	98.23%	99.90%	98.24%	0.101%	1.73%
PTB	97.89%	97.71%	97.83%	99.96%	97.70%	0.045%	2.29%

Table 4.2 presents the classification metrics of the proposed LSTM model across the three ECG datasets: MIT-BIH, NSRDB, and PTB. For the MIT-BIH dataset, the LSTM model achieves an accuracy of 98.33%, reflecting its strong classification capability. The precision (98.39%) and recall (98.43%) are closely aligned, suggesting a good balance between correctly identifying true positives and minimizing false negatives. The model's specificity is high at 99.96%, demonstrating its ability to accurately reject non-genuine subjects. The F1-score, at 98.40%, indicates that the model strikes a good balance between precision and recall. The FAR is very low at 0.036%, while the FRR of 1.61% suggests a slightly higher rejection rate of genuine subjects compared to the GRU model.

For the NSRDB dataset, the LSTM model shows slightly lower performance with an accuracy of 98.27%. The precision and recall are identical at 98.27% and 98.23%, respectively, maintaining a similar performance to MIT-BIH. The specificity is somewhat lower at 99.90%, though still strong. However, the FAR is significantly higher at 0.101%, indicating a greater likelihood of accepting impostor subjects compared to MIT-BIH. The FRR of 1.73% also indicates a slight increase in the rejection of genuine subjects relative to the MIT-BIH dataset.

In the PTB dataset, the LSTM model achieves the lowest accuracy, at 97.89%, among the three datasets. The precision is 97.71%, and the recall is slightly higher at 97.83%, suggesting that the model is effective but slightly less robust in identifying positive samples compared to the other datasets. The specificity is high at 99.96%, comparable to MIT-BIH, but the F1-score of 97.70% reflects a small drop in overall performance. The FAR

of 0.045% is relatively low, but the FRR of 2.29% is the highest among the datasets, indicating that the model is more likely to incorrectly reject genuine subjects in the PTB dataset.

The LSTM model shows strong classification performance across all datasets, with consistently high accuracy, precision, recall, and specificity. However, compared to the GRU model, it generally exhibits slightly higher FAR and FRR, particularly on the NSRDB and PTB datasets. Despite these variations, the LSTM model remains a highly effective system for ECG-based biometric identification.

The EMD played an important role in enhancing the performance of both the GRU and LSTM models, it is designed to decompose non-linear and non-stationary signals, like ECG data, into IMFs. These IMFs isolate different frequency components of the signal, which in turn allow deep learning models to learn richer and more discriminative features from the underlying data.

For both GRU and LSTM models, the application of EMD enhances feature extraction by simplifying the ECG signals into more manageable components. Instead of relying on raw ECG signals, the models can process frequency-specific signal components, which likely leads to more effective learning of temporal patterns specific to individual subjects. The impact of EMD on both models is evident in the high accuracy, precision, recall, and specificity across all datasets. By transforming the raw ECG data into meaningful signal components, EMD enables both GRU and LSTM models to better capture person-specific traits in the ECG patterns, contributing to the consistently strong results. However, the GRU appears to utilize the EMD-enhanced signals slightly more effectively than the LSTM, as evidenced by marginally better performance metrics.

GRU achieved better performance than LSTM likely due to several potential reasons:

1. **Simpler Architecture**: GRU has a simpler structure compared to LSTM, with fewer gates (GRU has two gates, while LSTM has three). This reduced complexity leads to faster convergence during training (figures 4.1 and 4.2), making GRU more efficient in capturing temporal dependencies without overfitting, especially in cases where the amount of data is moderate, as is the case with the ECG datasets used.

2. **Efficient Learning with Smaller Data**: GRUs are known to generalize better in scenarios with limited data [141] since they require fewer parameters to be learned. ECG-based datasets are often not as large as other datasets used in deep learning tasks, and the GRU's capacity to handle relatively small datasets might give it an edge over LSTM, which might overfit or struggle to optimize the additional parameters.

3. **Better Handling of Short-Term Dependencies**: The nature of the ECG signal, which is periodic and repetitive (i.e., the regular occurrence of R-peaks and QRS complexes), may favor the GRU architecture's ability to handle short-term dependencies efficiently. While LSTM excels in longer sequence dependencies, ECG signals and the short segments used to train the proposed models of this system do not typically require the long-term memory benefits that LSTM provides, allowing GRU to perform better in this task.

4. **Data Characteristics**: The QRS complexes may contain more short-term changes than long-term dependencies, making the GRU more adept at capturing these short-term features while still effectively managing the temporal nature of the data. The R-peak-centered segmentation, combined with normalization of the amplitude, might also favor the GRU's simpler update mechanism.

Table 4.3: Comparaison of the ECG unimodal system with state-of-the-art results

Paper	Database	Max. accuracy
Jyotishi et al., 2020 [142]	MIT-BIH	96.81%
	PTB	97.30%
Belo et al., 2020 [143]	MIT-BIH	96.40%
El Boujnouni et al., 2022 [144]	MIT-BIH	98.10%
Chee et al., 2022 [145]	PTB	98.10%
Hamza et al., 2022 [146]	PTB	95.40%
Fatimah et al., 2022 [147]	MIT-BIH	97.92%
Patro et al., 2022 [148]	PTB	95.30%
Li et al., 2022 [149]	PTB	95.77%
Yi et al., 2023 [150]	PTB	73.79%
Wang et al., 2023 [151]	MIT-BIH	97.66%
Fuster-Barcel et al., 2023 [152]	MIT-BIH	97.89%
	PTB	97.09%
Zehir et al., 2023 [3]	MIT-BIH	97.00%
Proposed method	MIT-BIH	98.57%
	PTB	98.26%
	NSRDB	99.17%

Table 4.3 presents a comparison between the proposed ECG-based unimodal system and state-of-the-art methods across various databases, including MIT-BIH, PTB, and NSRDB. The results demonstrate that the proposed system achieves superior accuracy on all three databases when compared to existing methods.

For the MIT-BIH database, the proposed method attains an accuracy of 98.57%, which surpasses previous works such as El Boujnouni et al. [144] with 98.10%, and Wang

et al. [151] with 97.66%. Notably, this result also exceeds other state-of-the-art approaches like Jyotishi et al. [142], who reported an accuracy of 96.81%.

On the PTB database, the proposed method achieves an accuracy of 98.26%, again outperforming methods like Chee et al. [145] at 98.10% and Jyotishi et al. [142], who achieved 97.30%. The system also shows marked improvement over other works, such as Hamza et al. [146] with 95.40%, highlighting the robustness of the proposed GRU-based approach.

For the NSRDB database, the proposed system achieves the highest reported accuracy, reaching 99.17%, further reinforcing its efficacy.

Table 4.3 demonstrates the superior performance of the proposed method compared to a wide range of state-of-the-art techniques, particularly in handling both healthy and diseased ECG signals, as discussed in earlier sections. The use of advanced models and robust training processes clearly contributes to these notable improvements in accuracy across all datasets.

4.3 Unimodal Speaker Recognition System

In this section, a speaker identification system is proposed, leveraging a subset of 47 speakers extracted from the LibriSpeech database. The dataset was randomly partitioned into three sets: 70% of the data was utilized for training, 20% for testing, and the remaining 10% for validation purposes.

The feature extraction process involved computing 40 MFCC coefficients from each audio sample, alongside their first and second derivatives. These features, as described in detail in Section 3.3.2, capture important speaker-specific characteristics related to the spectral properties and dynamic changes in speech, which are essential for accurate identification of speakers.

The CNN-based classification model employed for this task was trained using a learning rate of 0.001, a batch size of 32, and the Adam optimizer for effective weight updates. The training spanned 50 epochs to ensure the model sufficiently learned the underlying patterns in the data.

Figure 4.3 illustrates the training and validation accuracy of the CNN model over the course of the training epochs, while Figure 4.4 depicts the corresponding loss values. These plots highlight the model's convergence behavior and indicate its performance on both the training and validation sets.

Figure 4.3 demonstrates the training and validation accuracies of the speaker recognition model over 50 epochs. The blue solid line represents the training accuracy, while

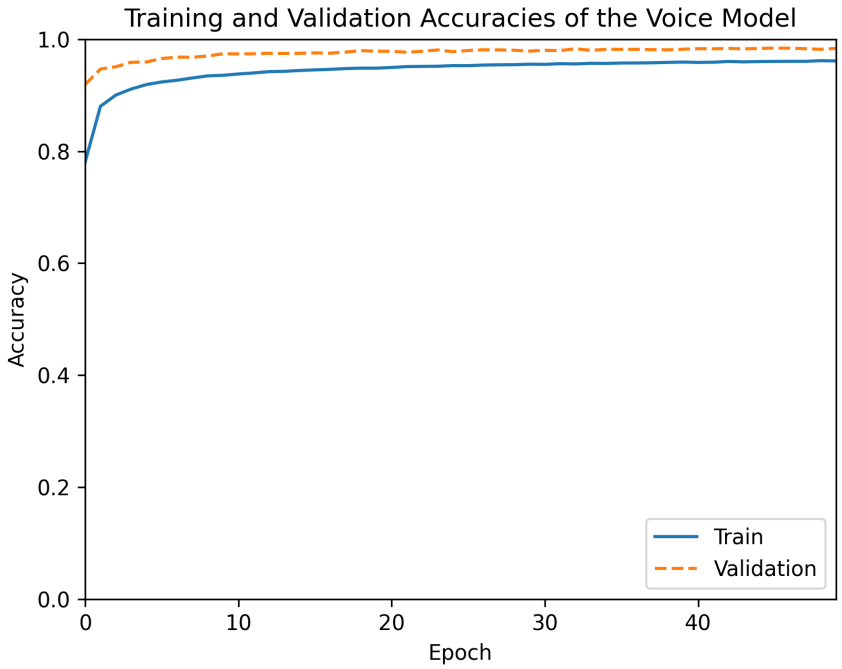


Figure 4.3: Training and validation accuracies of the CNN-based speaker recognition model over 50 epochs.

the orange dashed line corresponds to the validation accuracy. Both the training and validation accuracies exhibit an upward trend in the initial epochs, reaching approximately 98-99% accuracy by the end of the training.

The rapid convergence of the model is evident from the steep rise in accuracy during the first 10 epochs. After this point, the accuracies continue to increase at a slower rate, eventually stabilizing around the 40th epoch, indicating that the model has learned the underlying patterns in the data and is no longer improving significantly.

The close alignment between the training and validation accuracies suggests that the model is not overfitting, as the validation accuracy follows the training accuracy closely. This implies good generalization capability, meaning the model performs similarly on unseen data.

Figure 4.4 illustrates the training and validation losses of the CNN-based speaker recognition model over 50 epochs. The solid blue line represents the training loss, while the dashed orange line corresponds to the validation loss.

Both losses decrease significantly within the first few epochs, indicating that the model is learning effectively. The training loss starts around 0.7 and steadily decreases, reaching a value below 0.1 by the 50th epoch. The validation loss follows a similar trend, though it consistently remains lower than the training loss. This suggests that the model is

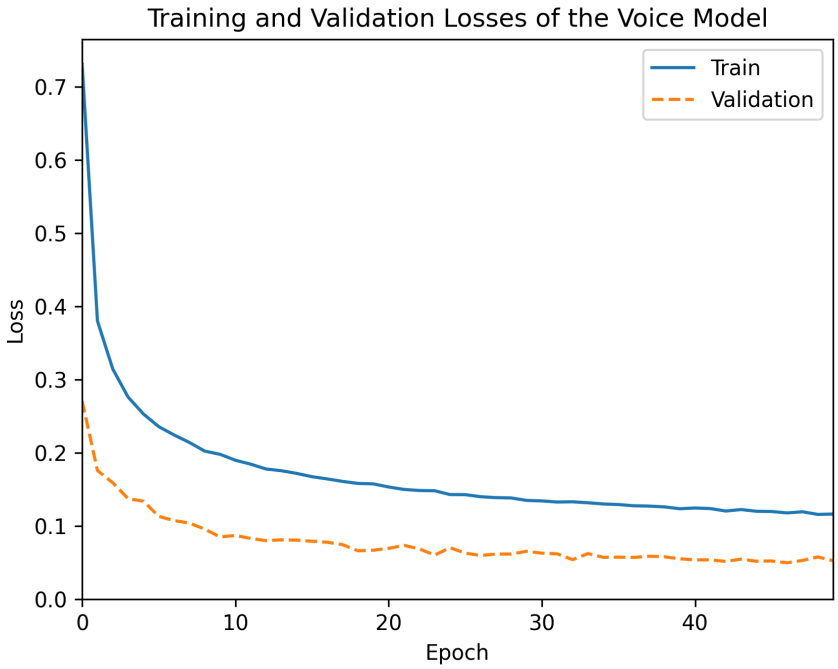


Figure 4.4: Training and validation losses of the CNN-based speaker recognition model over 50 epochs.

well-regularized and not overfitting to the training data, as the validation loss does not increase or deviate significantly from the training loss.

The stabilization of the validation loss after approximately 20 epochs indicates that the model has reached a plateau, with minimal further improvements in performance. The low final validation loss, along with the previous accuracy plot, demonstrates that the model has effectively learned to differentiate between speakers.

The classification results are summarized in Table 4.4, which presents the model's performance in terms of accuracy, precision, recall, and other relevant metrics. These results demonstrate the system's ability to effectively distinguish between different speakers using the extracted MFCC features and CNN-based classification.

Table 4.4: The classification results of the proposed speaker recognition system.

Accuracy	Precision	Recall	F1-score	FAR	FRR	EER
98.42%	98.46%	98.45%	98.45%	0.03%	1.55%	0.79%

The classification results of the proposed speaker recognition system, as shown in table 4.4, demonstrate strong overall performance. The accuracy of the model is 98.42%, indicating that the system correctly identified speakers in a significant majority of cases. This high accuracy is further supported by similarly high precision, recall, and F1-score

values, all around 98.45% to 98.46%, suggesting a balanced model that performs well across different evaluation metrics.

The FAR of 0.03% is particularly notable, indicating that the system rarely misclassifies a non-authorized speaker as an authorized one. This low FAR is critical in biometric tasks where security is paramount. On the other hand, the FRR stands at 1.55%, representing the percentage of times the system incorrectly rejected an authorized speaker. Although slightly higher than the FAR, this value is still within an acceptable range, showing the system's overall reliability.

Finally, the EER of 0.79% reinforces the robustness of the model. The EER is an important metric in biometric systems, representing the point where the FAR and FRR are equal. A low EER value is a strong indicator of a well-balanced system, minimizing both false acceptances and false rejections.

Table 4.5: Comparison with other speaker recognition state-of-the-art methods.

Method	Accuracy	Precision	Recall	F1-score
Caiet al. [153]	89.90%	-	-	-
Pentapati et al. [154]	91.19%	-	-	-
An et al. [155]	90.80%	-	-	-
Proposed Method	98.42%	98.46%	98.45%	98.45%

Table 4.5 presents a comparison of the proposed speaker recognition method with several state-of-the-art techniques. The proposed method achieves an accuracy of 98.42%, significantly outperforming prior approaches.

For example, An et al. [155] reported an accuracy of 90.80%, and Pentapati et al. [154] achieved 91.19%. Similarly, Cai et al. [153] obtained 89.90% accuracy, all of which are notably lower than the results obtained by the proposed system. Additionally, while earlier works did not provide precision, recall, or F1-scores, the proposed method excels in these metrics with 98.46% precision, 98.45% recall, and a 98.45% F1-score.

The superior performance of the proposed method highlights its robustness and effectiveness in speaker recognition tasks, likely due to the use of CNNs and advanced feature extraction techniques (such as MFCCs and derivatives). The results demonstrate the system's ability to handle variations in voice data with higher accuracy and efficiency compared to previous methods.

4.4 Multimodal System

In the proposed multimodal system, data from both ECG signals and voice recordings are integrated to enhance the biometric identification task. A total of 47 subjects were included, with 900 samples selected for each subject as discussed in section 3.4.1. These samples were chosen from the MIT-BIH database for ECG data and the LibriSpeech database for voice data, ensuring that every subject has corresponding ECG and voice inputs. This balanced selection of 900 samples per subject is important to prevent any biases and maintain a balanced dataset for training, testing, and validation.

To ensure that the system has enough data for both model training and performance evaluation, the dataset was partitioned into three sets: 630 samples (70%) for training, 180 samples (20%) for testing, and 90 samples (10%) for validation.

ECG data was specifically extracted from the MIT-BIH database for this multimodal system for several key reasons. First, as opposed to the NSRDB, the MIT-BIH dataset provides ECG recordings from both healthy and diseased subjects, as detailed in section 2.2.6, offering a broader and more diverse set of features, which is critical for ensuring the system's robustness across various health conditions. This diversity is especially valuable when building a multimodal system that aims to generalize well in real-world applications, where subjects may exhibit varying cardiac health conditions.

Second, the MIT-BIH database contains a significantly larger number of QRS complexes compared to the PTB database, which makes it more suitable for this fusion system. The higher number of annotated QRS complexes provides more comprehensive data for training the deep learning models, thus improving the system's ability to accurately identify and classify ECG signals. The abundance of QRS complexes also ensures that the model is exposed to a wide range of ECG variations during training, thereby enhancing its learning capabilities and generalization performance.

The scores used for the ECG component in this multimodal fusion system are derived from the GRU model that was trained on the MIT-BIH dataset. The GRU model demonstrated superior performance in the unimodal ECG system as can be seen in Table 4.1 and Table 4.2, making it the optimal choice for contributing to the multimodal fusion system.

To enhance system performance of the unimodal systems, score-level fusion was employed. Two fusion techniques were implemented: Softmax and SVM. Both techniques were tested with three fusion rules: the sum rule, the product rule, and the max rule. For Softmax, the probabilities generated by the ECG and voice models were fused according to these rules, while in the SVM-based fusion, the classifier was trained on the individual scores from each modality.

Table 4.6 summarizes the classification results of the ECG-Voice multimodal system using these score fusion techniques, demonstrating the improved identification accuracy obtained through the combination of both biometric modalities.

Table 4.6: Results of the proposed Voice-ECG Multimodal System on the MIT-BIH and LibriSpeech Databases.

Method	Accuracy	FAR	FRR	EER
Softmax + Sum	99.61%	0.01%	0.43%	0.22%
Softmax + Product	98.32%	0.06%	2.59%	1.32%
Softmax + Max	99.55%	0.01%	0.45%	0.23%
SVM + Sum	99.33%	0.01%	0.67%	0.67%
SVM + Product	97.53%	0.05%	2.47%	1.26%
SVM + Max	99.36%	0.01%	0.64%	0.33%

Table 4.6 presents the performance results of the proposed Voice-ECG multimodal biometric system on the MIT-BIH and LibriSpeech databases using various score-level fusion methods. The metrics evaluated include accuracy, FAR, FRR, and EER, providing a comprehensive view of the system's performance.

The Softmax-based fusion methods show consistently high performance, with the "Softmax + Sum" rule achieving the best overall results. It records an accuracy of 99.61%, the lowest FAR of 0.01%, an FRR of 0.43%, and the lowest EER of 0.22%. The high accuracy and minimal error rates suggest that summing the Softmax scores from both modalities leads to a balanced and effective fusion. The "Softmax + Max" rule also performs well, closely following the "Sum" rule with an accuracy of 99.55%, FAR of 0.01%, FRR of 0.45%, and an EER of 0.23%. The slightly higher FRR and EER compared to the Sum rule imply that while the max operation is effective, it is not as reliable as summing the probabilities. Conversely, the "Softmax + Product" rule shows a noticeable drop in performance with an accuracy of 98.32%, a higher FAR of 0.06%, an FRR of 2.59%, and an EER of 1.32%. The multiplication of Softmax scores appears to introduce more error, particularly in rejecting genuine users, as seen in the FRR.

For the SVM-based fusion methods, the "SVM + Sum" and "SVM + Max" rules offer strong results. The "SVM + Max" method yields an accuracy of 99.36%, a FAR of 0.01%, FRR of 0.64%, and an EER of 0.33%. Similarly, the "SVM + Sum" approach provides an accuracy of 99.33% with a FAR of 0.01% but has a slightly higher FRR of 0.67% and an EER of 0.67%. The "SVM + Product" rule, like its Softmax counterpart, performs poorly with the lowest accuracy of 97.53%, a FAR of 0.05%, an FRR of 2.47%, and an EER of 1.26%. This again highlights that product-based fusion introduces higher error rates, making it less favorable for the multimodal system.

Overall, the "Softmax + Sum" rule delivers the best performance, suggesting that

summing the normalized scores from both modalities is the most effective strategy for fusion. Both Softmax and SVM perform well with sum and max rules, but the product rule consistently introduces higher error rates, suggesting it may be less suitable for this type of multimodal biometric system. The superior results of score fusion demonstrate the advantages of combining ECG and voice modalities, leveraging the strengths of each to enhance the system's robustness and accuracy.

Table 4.7: Comparaison With Multimodal State-of-the-art Methods.

Method	Modalities	Accuracy	FRR	FAR	EER
Rabab A Rasool [156]	Iris and Face	97.53%	0.24%	0.24%	-
Joshi et al. [157]	Face, Fingerprint, Signature, and Iris	-	1.66%	0.00%	0.4%
Tharewal et al. [158]	Ear and Face	99.25%	-	-	-
Ammour et al. [45]	Iris and Face	99.33%	-	-	-
Bugdol and Mitas [159]	ECG and Voice	77%	-	-	-
Proposed Method	ECG and Voice	99.61%	0.43%	0.01%	0.22%

Table 4.7 provides a comparison of the proposed ECG and voice-based multimodal biometric system with several state-of-the-art multimodal methods. The comparison includes various biometric modalities and evaluates key performance metrics such as accuracy, FRR, FAR, and EER.

The proposed method, which integrates ECG and voice data, achieves the highest accuracy at 99.61%, outperforming systems like Ammour et al.'s [45] face-iris fusion and Tharewal et al.'s [158] face-ear combination, which reported accuracies of 99.33% and 99.25% respectively. Furthermore, the proposed system shows a notably low FAR of 0.01%, indicating a strong resistance to false acceptances, while maintaining a reasonable FRR of 0.43%. The EER of 0.22% achieved by the ECG-Voice system is also among the lowest in the comparison, demonstrating its balanced performance in managing both false positives and false negatives.

Compared to other methods, such as the one by Joshi et al. [157], which involved multiple modalities like face, iris, signature, and fingerprint with an EER of 0.4%, the proposed system's EER of 0.22% reflects its effectiveness despite using fewer modalities. Similarly, Rasool's [156] face-iris system, while achieving a relatively high accuracy of 97.53%, reported higher FAR and FRR values than the proposed method, further illustrating the superior reliability and robustness of the ECG-Voice fusion approach. Overall, the proposed method demonstrates competitive, if not superior, performance across all evaluation metrics.

4.5 Conclusion

In conclusion, the results presented in this section demonstrate the effectiveness of both the unimodal and multimodal systems for biometric recognition, with the multimodal approach yielding superior performance. The unimodal ECG and speaker recognition systems achieved high accuracy and low error rates, but the fusion of these modalities through score-level fusion significantly enhanced the robustness and reliability of the system. Softmax-based fusion methods, particularly using the Sum rule, consistently outperformed others, providing the best overall accuracy, FAR, FRR, and EER. These findings indicate that combining ECG and voice data offers a more resilient biometric identification solution, minimizing the limitations of individual modalities and enhancing security in biometric systems.

General Conclusion

In this thesis, we have explored the design and development of a multimodal biometric recognition system, specifically leveraging ECG signals and voice data. The results illustrate that while unimodal systems such as the ECG-based system using GRU/LSTM and the speaker recognition system utilizing CNN exhibited high accuracy and reliability in their respective domains, the integration of multiple modalities through score-level fusion provided superior performance. By combining the strengths of both ECG and voice biometrics, the multimodal system demonstrated improved robustness, and accuracy, as well as a marked reduction in false acceptance and rejection rates.

The key findings of this thesis highlight the effectiveness of combining ECG and voice modalities for biometric recognition. Through experiments on both unimodal and multimodal systems, the ECG-based system demonstrated high performance, particularly when using the GRU model, which outperformed LSTM across various metrics such as accuracy, precision, and recall. The GRU achieved a high classification accuracy on the MIT-BIH database, owing to its ability to capture temporal dependencies in ECG signals more efficiently than LSTM.

The speaker recognition system based on MFCC feature extraction and a CNN architecture also yielded promising results. The system achieved high accuracy in identifying speakers from the Librispeech database, further validating the potential of voice as an effective biometric. The use of 40 MFCC coefficients, along with their first and second derivatives, was important in capturing the unique characteristics of each speaker's voice, contributing to the system's robust performance.

In the multimodal system, score-level fusion techniques, including Softmax and SVM with different combination rules (Sum, Product, and Max), were applied to the fused ECG and voice data. The Softmax fusion using the Sum rule resulted in the highest overall accuracy of 99.61%, with the lowest EER of 0.22%. This demonstrates that combining ECG and voice significantly enhances biometric recognition performance, providing better accuracy and reduced error rates compared to unimodal systems. The fusion strategies, particularly those involving Softmax, proved to be highly effective, confirming the benefits

of multimodal approaches in biometric applications.

This thesis contributes to the growing field of multimodal biometrics by demonstrating that fusing physiological and behavioral traits can enhance system performance. These findings provide a strong foundation for future research in biometric security, particularly in applications that require higher levels of accuracy and robustness.

For future work, a natural extension of this research would be the hardware implementation of the proposed multimodal biometric recognition system. Deploying the system on embedded platforms such as STM32 microcontrollers or other resource-constrained devices could enhance its practical applications, particularly in scenarios requiring portability, low power consumption, and real-time processing. This would involve optimizing the existing DL models for embedded systems by reducing their computational complexity and memory footprint through techniques such as model pruning, quantization, and compression. A potential future direction could also involve designing custom hardware accelerators (e.g., FPGA) for faster inference of DL models in real-time applications like access control systems or personal authentication on wearable devices.

Another promising area for future development is testing and validating the system in various real-world environments with diverse populations to enhance generalizability and robustness. This could be particularly important in mobile or IoT-based applications, where environmental noise, hardware limitations, and diverse user characteristics may affect the system's performance.

Bibliography

[1] Hatem Zehir, Toufik Hafs, and Sara Daas. Empirical mode decomposition-based biometric identification using gru and lstm deep neural networks on ecg signals. *Evolving Systems*, pages 1–17, 2024.

[2] Hatem Zehir, Hafs Toufik, and Sara Daas. Involutional neural networks for ecg spectrogram classification and person identification. *International Journal of Signal and Imaging Systems Engineering*, 13, 2024.

[3] Hatem Zehir, Toufik Hafs, Sara Daas, and Amine Nait-Ali. Support vector machine for human identification based on non-fiducial features of the ecg. *Journal of Engineering Studies and Research*, 29(1):61–69, 2023.

[4] Hatem Zehir, Hafs Toufik, and Sara Daas. Tinycnn: An embedded cnn model for speaker identification using esp32. 11 2023.

[5] Hatem Zehir, Hafs Toufik, and Sara Daas. Ecg-based biometric system using tinyml: Implementation and performance evaluation on esp32. 11 2023.

[6] Hatem Zehir, Toufik Hafs, and Sara Daas. Healthcare decision-making with an ecg-based biometric system. In *2023 International Conference on Decision Aid Sciences and Applications (DASA)*, pages 88–92. IEEE, 2023.

[7] Hatem Zehir, Toufik Hafs, Sara Daas, and Amine Nait-Ali. An ecg biometric system based on empirical mode decomposition and hilbert-huang transform for improved feature extraction. In *2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART)*, pages 1–4. IEEE, 2023.

[8] Hatem Zehir, Hafs Toufik, and Sara Daas. Bidirectional long short-term memory neural networks based electrocardiogram biometric system. 12 2022.

[9] U Sumalatha, K Krishna Prakasha, Srikanth Prabhu, and Vinod C Nayak. A comprehensive review of unimodal and multimodal fingerprint biometric authentication systems: Fusion, attacks, and template protection. *IEEE Access*, 2024.

[10] Diptadip Maiti, Madhuchhanda Basak, and Debashis Das. A review on fingerprint based authentication-its challenges and applications. *Computer Science Review*, 57:100735, 2025.

[11] Zheyu Chen, Zhiqiang Yao, Biao Jin, Mingwei Lin, and Jianting Ning. Fibnet: privacy-enhancing approach for face biometrics based on the information bottleneck principle. *IEEE Transactions on Information Forensics and Security*, 2024.

[12] Alamgir Sardar, Saiyed Umer, Ranjeet Kumar Rout, Kshira Sagar Sahoo, and Amir H Gandomi. Enhanced biometric template protection schemes for securing face recognition in iot environment. *IEEE Internet of Things Journal*, 2024.

[13] Sushil Bhatt, Jagmahender Singh Sehrawat, and Vishali Gupta. A systematic review of iris biometrics in forensic science: applications and challenges. *Egyptian Journal of Forensic Sciences*, 15(1):12, 2025.

[14] Kien Nguyen, Hugo Proença, and Fernando Alonso-Fernandez. Deep learning for iris recognition: A survey. *ACM Computing Surveys*, 56(9):1–35, 2024.

[15] Milkias Ghilom and Shahram Latifi. The role of machine learning in advanced biometric systems. *Electronics*, 13(13):2667, 2024.

[16] Saif Mohanad Kadhim, Johnny Koh Siaw Paw, Yaw Chong Tak, and Shahad Ameen. Deep learning models for biometric recognition based on face, finger vein, fingerprint, and iris: A survey. *Journal of Smart Internet of Things (JSIoT)*, 2024(01):117–157, 2024.

[17] Kashif Shaheed, Piotr Szczuko, Munish Kumar, Imran Qureshi, Qaisar Abbas, and Ihsan Ullah. Deep learning techniques for biometric security: A systematic review of presentation attack detection systems. *Engineering Applications of Artificial Intelligence*, 129:107569, 2024.

[18] Vivek Upadhyaya. Advancements in computer vision for biometrics enhancing security and identification. In *Leveraging Computer Vision to Biometric Applications*, pages 260–292. Chapman and Hall/CRC, 2025.

[19] Nada Alay and Heyam H Al-Baity. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. *Sensors*, 20(19):5523, 2020.

[20] Ali Ismail Awad, Aiswarya Babu, Ezedin Barka, and Khaled Shuaib. Ai-powered biometrics for internet of things security: A review and future vision. *Journal of Information Security and Applications*, 82:103748, 2024.

[21] Mohd Javaid, Abid Haleem, Ravi Pratap Singh, Shanay Rab, and Rajiv Suman. Significance of sensors for industry 4.0: Roles, capabilities, and applications. *Sensors International*, 2:100110, 2021.

[22] Ekta Sharma, Reena Rathi, Jaya Misharwal, Bhavya Sinhmar, Suman Kumari, Jasvir Dalal, and Anand Kumar. Evolution in lithography techniques: microlithography to nanolithography. *Nanomaterials*, 12(16):2754, 2022.

[23] Sara Daas. *Multimodal data fusion : Biometric application.* Phd thesis, Badji Mokhtar – Annaba University, 2021. Available at https://biblio.univ-annaba.dz/wp-content/uploads/2022/12/These-Daas-Sara.pdf.

[24] Amine Naït-Ali and Regis Fournier. *Signal and image processing for biometrics.* John Wiley & Sons, 2012.

[25] Marcel Segrist. The cuneiform tablets at syracuse university. 1980.

[26] Manahil Shakeel and Shahzada Khurram Syed. A review on fingerprint as an identification tool in the discipline of forensics. *Forensic Insights and Health Sciences Bulletin*, 1(1):6–10, 2023.

[27] Facedapter. "The Evolution of Biometrics: From Ancient Times to Modern Technology" — facedapter.medium.com. https://facedapter.medium.com/the-evolution-of-biometrics-from-ancient-times-to-modern-technology-9410113c4f31. [Accessed 28-09-2024].

[28] Ajantika Ghosh and Indranil Pahari. Fingerprinting: The unique tool for identification in forensic science. *Advanced Research in Veterinary Sciences*, 22:22, 2021.

[29] Om Pradyumana Gupta, Arun Prakash Agrawal, and Om Pal. A study on evolution of facial recognition technology. In *2023 International Conference on Disruptive Technologies (ICDT)*, pages 769–775. IEEE, 2023.

[30] Foudil Belhadj. *Biometric system for identification and authentication.* PhD thesis, Ecole nationale Supérieure en Informatique Alger, 2017.

[31] Aaron E Rosenberg. Fifty years of progress in speaker verification. *The Journal of the Acoustical Society of America*, 116(4_Supplement):2497–2497, 2004.

[32] James R Matey, Oleg Naroditsky, Keith Hanna, Ray Kolczynski, Dominick J LoIacono, Shakuntala Mangru, Michael Tinker, Thomas M Zappia, and Wenyi Y Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, 94(11):1936–1947, 2006.

[33] Daniel Morgan and William Krouse. Biometric identifiers and border security: 9/11 commission recommendations and related issues. Congressional Information Service, Library of Congress, 2005.

[34] Nissan Moradoff. Biometrics: Proliferation and constraints to emerging and new technologies. *Security Journal*, 23:276–298, 2010.

[35] Fieke Jansen, Javier Sánchez-Monedero, and Lina Dencik. Biometric identity systems in law enforcement and the politics of (voice) recognition: The case of siip. *Big Data & Society*, 8(2):20539517211063604, 2021.

[36] Mohamed Abomhara, Sule Yildirim Yayilgan, Anne Hilde Nymoen, Marina Shalaginova, Zoltán Székely, and Ogerta Elezaj. How to do it right: a framework for biometrics supported border control. In *E-Democracy–Safeguarding Democracy and Human Rights in the Digital Age: 8th International Conference, e-Democracy 2019, Athens, Greece, December 12-13, 2019, Proceedings 8*, pages 94–109. Springer, 2020.

[37] Anusha Bodepudi and Manjunath Reddy. Cloud-based biometric authentication techniques for secure financial transactions: A review. *International Journal of Information and Cybersecurity*, 4(1):1–18, 2020.

[38] Ioannis Stylios, Spyros Kokolakis, Olga Thanou, and Sotirios Chatzis. Behavioral biometrics & continuous user authentication on mobile devices: A survey. *Information Fusion*, 66:76–99, 2021.

[39] Shuqi Liu, Wei Shao, Tan Li, Weitao Xu, and Linqi Song. Recent advances in biometrics-based user authentication for wearable devices: A contemporary survey. *Digital Signal Processing*, 125:103120, 2022.

[40] Attaullah Buriro, Bruno Crispo, Filippo Delfrari, and Konrad Wrona. Hold and sign: A novel behavioral biometrics for smartphone user authentication. In *2016 IEEE security and privacy workshops (SPW)*, pages 276–285. IEEE, 2016.

[41] Attaullah Buriro, Bruno Crispo, Filippo Del Frari, and Konrad Wrona. Touchstroke: Smartphone user authentication based on touch-typing biometrics. In *New*

Trends in Image Analysis and Processing–ICIAP 2015 Workshops: ICIAP 2015 International Workshops, BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings 18, pages 27–34. Springer, 2015.

[42] Gary M Weiss, Kenichi Yoneda, and Thaier Hayajneh. Smartphone and smartwatch-based biometrics using activities of daily living. *Ieee Access*, 7:133190–133202, 2019.

[43] Toufik Hafs, Hatem Zehir, Ali Hafs, and Amine Nait-Ali. Multimodal biometric system based on the fusion in score of fingerprint and online handwritten signature. *Applied Computer Systems*, 28(1):58–65, 2023.

[44] Basma Ammour, Toufik Bouden, and Larbi Boubchir. Face-iris multimodal biometric system based on hybrid level fusion. In *2018 41st international conference on telecommunications and signal processing (TSP)*, pages 1–5. IEEE, 2018.

[45] Basma Ammour, Larbi Boubchir, Toufik Bouden, and Messaoud Ramdani. Face–iris multimodal biometric identification system. *Electronics*, 9(1):85, 2020.

[46] Sara Daas, Amira Yahi, Toufik Bakir, Mouna Sedhane, Mohamed Boughazi, and El-Bay Bourennane. Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion. *IET Image Processing*, 14(15):3859–3868, 2020.

[47] Help Net Security. COVID-19 creates a boom in biometric adoption - Help Net Security — helpnetsecurity.com. https://www.helpnetsecurity.com/2021/04/23/biometric-adoption-boom/. [Accessed 28-09-2024].

[48] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[49] Aweem Ashar, Muhammad Shahid Bhatti, and Usama Mushtaq. Speaker identification using a hybrid cnn-mfcc approach. In *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, pages 1–4. IEEE, 2020.

[50] Mandana Fasounaki, Emirhan Burak Yüce, Serkan Öncül, and Gökhan İnce. Cnn-based text-independent automatic speaker identification using short utterances. In *2021 6th international conference on computer science and engineering (UBMK)*, pages 413–418. IEEE, 2021.

[51] Hana Ben Fredj, Safa Bouguezzi, and Chokri Souani. Face recognition in unconstrained environment with cnn. *The Visual Computer*, 37(2):217–226, 2021.

[52] Lei Zong, Chen Xu, and HongLin Yuan. A rf fingerprint recognition method based on deeply convolutional neural network. In *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pages 1778–1781. IEEE, 2020.

[53] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[54] Debasish Jyotishi and Samarendra Dandapat. An lstm-based model for person identification using ecg signal. *IEEE Sensors Letters*, 4(8):1–4, 2020.

[55] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[56] Xiang Zhang, Lina Yao, Salil S Kanhere, Yunhao Liu, Tao Gu, and Kaixuan Chen. Mindid: Person identification from brain waves through attention-based recurrent neural network. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–23, 2018.

[57] Debasish Jyotishi and Samarendra Dandapat. An ecg biometric system using hierarchical lstm with attention mechanism. *IEEE Sensors Journal*, 22(6):6052–6061, 2021.

[58] Google Books Ngram Viewer — books.google.com. https://books.google.com/ngrams/graph?content=biometrics&year_start=1800&year_end=2019&corpus=en-2019&smoothing=3. [Accessed 28-09-2024].

[59] Hichem Chaya. The algerian biometric and electronic national identity card "cnibe". Montréal, Canada, 2016. Presented at TWELFTH SYMPOSIUM AND EXHIBITION ON ICAO TRAVELLER IDENTIFICATION PROGRAMME (TRIP).

[60] Algerian passport - Wikipedia — en.wikipedia.org. https://en.wikipedia.org/wiki/Algerian_passport. [Accessed 28-09-2024].

[61] Passeport algérien — Wikipédia — fr.wikipedia.org. https://fr.wikipedia.org/wiki/Passeport_alg%C3%A9rien. [Accessed 28-09-2024].

[62] Carte nationale d'identité en Algérie — Wikipédia — fr.wikipedia.org. https://fr.wikipedia.org/wiki/Carte_nationale_d%27identit%C3%A9_en_Alg%C3%A9rie. [Accessed 28-09-2024].

[63] https://www.aps.dz/ar/algerie/78609-14-16. [Accessed 29-09-2024].

[64] https://radioalgerie.dz/news/fr/article/20190707/174077.html. [Accessed 29-09-2024].

[65] Zineb Maaref, Abdelouahab Attia, and Foudil Belhadj. Generating cancelable multispectral palmprint templates based on cartesian transformation. In *2023 5th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*, pages 1–7. IEEE, 2023.

[66] Murat Taskiran, Nihan Kahraman, and Cigdem Eroglu Erdem. Face recognition: Past, present and future (a review). *Digital Signal Processing*, 106:102809, 2020.

[67] Farmanullah Jan, Nasro Min-Allah, Shahrukh Agha, Imran Usman, and Irfanullah Khan. A robust iris localization scheme for the iris recognition. *Multimedia Tools and Applications*, 80:4579–4605, 2021.

[68] Amjad Hassan Khan MK and PS Aithal. Voice biometric systems for user identification and authentication–a literature review. *International Journal of Applied Engineering and Management Letters (IJAEML)*, 6(1):198–209, 2022.

[69] Marcos Faundez-Zanuy, Julian Fierrez, Miguel A Ferrer, Moises Diaz, Ruben Tolosana, and Réjean Plamondon. Handwriting biometrics: Applications and future trends in e-security and e-health. *Cognitive Computation*, 12:940–953, 2020.

[70] Anubha Parashar, Apoorva Parashar, and Imad Rida. Journey into gait biometrics: Integrating deep learning for enhanced pattern recognition. *Digital Signal Processing*, 147:104393, 2024.

[71] Pawel Kasprowski, Zaneta Borowska, and Katarzyna Harezlak. Biometric identification based on keystroke dynamics. *Sensors*, 22(9):3158, 2022.

[72] Patrizio Campisi and Emanuele Maiorana. Eeg biometrics. In *Encyclopedia of Cryptography, Security and Privacy*, pages 1–6. Springer, 2021.

[73] Ali Z Ghazi Zahid, Ibrahim Hasan Mohammed Salih Al-Kharsan, Hesham A Bakarman, Muntadher Faisal Ghazi, Hanan Abbas Salman, and Feras N Hasoon. Biometric authentication security system using human dna. In *2019 First International*

Conference of Intelligent Computing and Engineering (ICOICE), pages 1–7. IEEE, 2019.

[74] João Ribeiro Pinto, Jaime S Cardoso, and André Lourenço. Evolution, current challenges, and future possibilities in ecg biometrics. *Ieee Access*, 6:34746–34776, 2018.

[75] Amel Benabdallah and Abdelghani Djebbari. Biometric individual authentication system using high performance ecg fiducial features. In *2022 5th International Symposium on Informatics and its Applications (ISIA)*, pages 1–6. IEEE, 2022.

[76] Zeeshan Hassan, Syed Omer Gilani, and Mohsin Jamil. Review of fiducial and non-fiducial techniques of feature extraction in ecg based biometric systems. *Indian J. Sci. Technol*, 9(21):850–855, 2016.

[77] Lena Biel, Ola Pettersson, Lennart Philipson, and Peter Wide. Ecg analysis: a new approach in human identification. *IEEE transactions on instrumentation and measurement*, 50(3):808–812, 2001.

[78] M Tantawi, A Salem, and Mohamed Fahmy Tolba. Fiducial based approach to ecg biometrics using limited fiducial points. In *Advanced Machine Learning Technologies and Applications: Second International Conference, AMLTA 2014, Cairo, Egypt, November 28-30, 2014. Proceedings 2*, pages 199–210. Springer, 2014.

[79] Manal M Tantawi, Kenneth Revett, A Salem, and Mohamed Fahmy Tolba. Fiducial feature reduction analysis for electrocardiogram (ecg) based biometric recognition. *Journal of Intelligent Information Systems*, 40:17–39, 2013.

[80] Francesco Gargiulo, Antonio Fratini, Mario Sansone, and Carlo Sansone. Subject identification via ecg fiducial-based systems: Influence of the type of qt interval correction. *Computer methods and programs in biomedicine*, 121(3):127–136, 2015.

[81] Pablo Laguna, Raimon Jané, and Pere Caminal. Automatic detection of wave boundaries in multilead ecg signals: Validation with the cse database. *Computers and biomedical research*, 27(1):45–60, 1994.

[82] Joao M Carvalho, Susana Brás, and Armando J Pinho. Compression-based ecg biometric identification using a non-fiducial approach. *arXiv preprint arXiv:1804.00959*, 2018.

[83] Yejin Kim, Gyuho Choi, and Chang Choi. One-dimensional shallow neural network using non-fiducial based segmented electrocardiogram for user identification system. *IEEE Access*, 2023.

[84] Marwa A Elshahed. Personal identity verification based ecg biometric using non-fiducial features. *International Journal of Electrical and Computer Engineering*, 10(3):3007, 2020.

[85] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236, 1985.

[86] Sumair Aziz, Muhammad Umar Khan, Zainoor Ahmad Choudhry, Afeefa Aymin, and Adil Usman. Ecg-based biometric authentication using empirical mode decomposition and support vector machines. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 0906–0912. IEEE, 2019.

[87] Miguel Carvalho and Susana Brás. Addressing intra-subject variability in electrocardiogram-based biometric systems through a hybrid architecture. *Biomedical Signal Processing and Control*, 87:105465, 2024.

[88] Iulian B Ciocoiu and Nicolae Cleju. Off-person ecg biometrics using spatial representations and convolutional neural networks. *IEEE Access*, 8:218966–218981, 2020.

[89] Yeong-Hyeon Byeon and Keun-Chang Kwak. Pre-configured deep convolutional neural networks with various time-frequency representations for biometrics from ecg signals. *Applied Sciences*, 9(22):4810, 2019.

[90] Ronald Salloum and C-C Jay Kuo. Ecg-based biometrics using recurrent neural networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2062–2066. IEEE, 2017.

[91] Htet Myet Lynn, Sung Bum Pan, and Pankoo Kim. A deep bidirectional gru network model for biometric electrocardiogram classification based on recurrent neural networks. *IEEE Access*, 7:145395–145405, 2019.

[92] Onorato D'angelis, Luca Bacco, Luca Vollero, and Mario Merone. Advancing ecg biometrics through vision transformers: A confidence-driven approach. *IEEE Access*, 2023.

[93] Afonso Eduardo, Helena Aidos, and Ana Fred. Ecg-based biometrics using a deep autoencoder for feature learning-an empirical study on transferability. In *International conference on pattern recognition applications and methods*, volume 2, pages 463–470. SciTePress, 2017.

[94] Min Keun Cho and Tae Seon Kim. Canine biometric identification using ecg signals and cnn-lstm neural networks. *IEEE Access*, 2023.

[95] Ralf Bousseljot, Dieter Kreiseler, and Allard Schnabel. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. 1995.

[96] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

[97] George B Moody and Roger G Mark. The impact of the mit-bih arrhythmia database. *IEEE engineering in medicine and biology magazine*, 20(3):45–50, 2001.

[98] Tatiana S Lugovaya. Biometric human identification based on ecg. *PhysioNet*, 2005.

[99] Nikhil Iyengar, CK Peng, Raymond Morin, Ary L Goldberger, and Lewis A Lipsitz. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 271(4):R1078–R1084, 1996.

[100] Hugo Plácido Da Silva, André Lourenço, Ana Fred, Nuno Raposo, and Marta Aires-de Sousa. Check your biosignals here: A new dataset for off-the-person ecg biometrics. *Computer methods and programs in biomedicine*, 113(2):503–514, 2014.

[101] Saeid Wahabi, Shahrzad Pouryayevali, Siddarth Hari, and Dimitrios Hatzinakos. On evaluating ecg biometric systems: Session-dependence and body posture. *IEEE Transactions on Information Forensics and Security*, 9(11):2002–2013, 2014.

[102] Corinne Fredouille and Delphine Charlet. Analysis of i-vector framework for speaker identification in tv-shows. In *Interspeech'2014*, 2014.

[103] Farman Ullah, Muhammad Israr, Atif Jan, Arbab Masood Ahmad, Irsha Dullah, and Faheem Ullah. Development of a novel system for speaker verification. In *2020 International conference on intelligent engineering and management (ICIEM)*, pages 12–16. IEEE, 2020.

[104] Dimitrios Dimitriadis. Enhancements for audio-only diarization systems. *arXiv preprint arXiv:1909.00082*, 2019.

[105] Talal Bin Amin and Iftekhar Mahmood. Speech recognition using dynamic time warping. In *2008 2nd international conference on advances in space technologies*, pages 74–79. IEEE, 2008.

[106] Akshay Madhav Deshmukh. Comparison of hidden markov model and recurrent neural network in automatic speech recognition. *European Journal of Engineering and Technology Research*, 5(8):958–965, 2020.

[107] Wanli Chen, Qingyang Hong, and Ximin Li. Gmm-ubm for text-dependent speaker recognition. In *2012 International Conference on Audio, Language and Image Processing*, pages 432–435. IEEE, 2012.

[108] Samia Abd El-Moneim, MA Nassar, Moawad I Dessouky, Nabil A Ismail, Adel S El-Fishawy, and Fathi E Abd El-Samie. Text-independent speaker recognition using lstm-rnn and speech enhancement. *Multimedia Tools and Applications*, 79:24013–24028, 2020.

[109] Munmi Dutta, Chayashree Patgiri, Mousmita Sarma, and Kandarpa Kumar Sarma. Closed-set text-independent speaker identification system using multiple ann classifiers. In *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014: Volume 1*, pages 377–385. Springer, 2015.

[110] Kevin Wilkinghoff. On open-set speaker identification with i-vectors. In *Odyssey*, pages 408–414, 2020.

[111] John S Garofolo. Timit acoustic phonetic continuous speech corpus. *Linguistic Data Consortium, 1993*, 1993.

[112] Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5206–5210. IEEE, 2015.

[113] Mitchell McLaren, Luciana Ferrer, Diego Castan, and Aaron Lawson. The speakers in the wild (sitw) speaker recognition database. In *Interspeech*, pages 818–822, 2016.

[114] Arsha Nagrani, Joon Son Chung, and Andrew Zisserman. Voxceleb: a large-scale speaker identification dataset. *arXiv preprint arXiv:1706.08612*, 2017.

[115] Shiraz Anwar. Comparative analysis of multiple fusion approaches for multimodal biometric systems. 2017.

[116] H. S. Gowda, G. Kumar, and Mohammad Imran. Multi-modal biometric system on various levels of fusion using lpq features. *Journal of Information and Optimization Sciences*, 39:169 – 181, 2018.

[117] M. F. Nadheen and S. Poornima. Feature level fusion in multimodal biometric authentication system. *International Journal of Computer Applications*, 69:36–40, 2013.

[118] Mohammad Haghighat, M. Abdel-Mottaleb, and W. Alhalabi. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Transactions on Information Forensics and Security*, 11:1984–1996, 2016.

[119] Wen-Shiung Chen, Ren-He Jeng, and Yen-Feng Chen. A feature-level fusion scheme based on eigen theory for multimodal biometrics. *IETE Technical Review*, 39:1081 – 1091, 2021.

[120] S. Soviany, V. Sandulescu, S. Puscoci, C. Soviany, and M. Jurian. An optimized biometric system with intra-and inter-modal feature-level fusion. *2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pages 1–8, 2017.

[121] Mohammad H. Safavipour, M. Doostari, and H. Sadjedi. A hybrid approach to multimodal biometric recognition based on feature-level fusion of face, two irises, and both thumbprints. *Journal of Medical Signals and Sensors*, 12:177 – 191, 2022.

[122] Anil K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognit.*, 38:2270–2285, 2005.

[123] M. Hanmandlu, J. Grover, Ankit Gureja, and H. Gupta. Score level fusion of multimodal biometrics using triangular norms. *Pattern Recognit. Lett.*, 32:1843–1850, 2011.

[124] Satrajit Mukherjee, Kunal Pal, Bodhisattwa Prasad Majumder, Chiranjib Saha, B. K. Panigrahi, and Sanjoy Das. Differential evolution based score level fusion for multi-modal biometric systems. *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pages 38–44, 2014.

[125] Ajay Kumar, Vivek Kanhangad, and David Zhang. A new framework for adaptive multimodal biometrics management. *IEEE Transactions on Information Forensics and Security*, 5:92–102, 2010.

[126] Suneet Narula Garg, R. Vig, and Savita Gupta. Multimodal biometric system based on decision level fusion. *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pages 753–758, 2016.

[127] D. V. R. Devi and K. N. Rao. Decision level fusion schemes for a multimodal biometric system using local and global wavelet features. *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–6, 2020.

[128] Nitin Agrawal, H. Mehrotra, Phalguni Gupta, and C. Hwang. An efficient fusion strategy for multimodal biometric system. pages 178–183, 2007.

[129] P. Szczuko, Arkadiusz Harasimiuk, and A. Czyżewski. Evaluation of decision fusion methods for multimodal biometrics in the banking application. *Sensors (Basel, Switzerland)*, 22, 2022.

[130] Mingxing He, S. Horng, P. Fan, R. Run, Rong-Jian Chen, Jui-Lin Lai, M. Khan, and Kevin Octavius Sentosa. Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recognit.*, 43:1789–1800, 2010.

[131] S. Ribaric and I. Fratric. Experimental evaluation of matching-score normalization techniques on different multimodal biometric systems. *MELECON 2006 - 2006 IEEE Mediterranean Electrotechnical Conference*, pages 498–501, 2006.

[132] Messaoud Bengherabi, L. Mezai, F. Harizi, A. Guessoum, and M. Cheriet. Robust authentication using likelihood ratio based score fusion of voice and face. pages 57–61, 2009.

[133] Amioy Kumar and Ajay Kumar. Adaptive management of multimodal biometrics fusion using ant colony optimization. *Inf. Fusion*, 32:49–63, 2016.

[134] S. Horng, Yuan-Hsin Chen, R. Run, Rong-Jian Chen, Jui-Lin Lai, and Kevin Octavius Sentosa. An improved score level fusion in multimodal biometric systems. *2009 International Conference on Parallel and Distributed Computing, Applications and Technologies*, pages 239–246, 2009.

[135] J. Aravinth and S. Valarmathy. Score-level fusion technique for multi-modal biometric recognition using abc-based neural network. *International Review on Computers and Software*, 8:1889–1900, 2013.

[136] Stuti Srivastava and P. S. Sudhish. Continuous multi-biometric user authentication fusion of face recognition and keystoke dynamics. *2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pages 1–7, 2016.

[137] Salah ud-din Ghulam Mohi-ud Din, A. Mansoor, Hassan Masood, and Mustafa Mumtaz. Personal identification using feature and score level fusion of palm- and fingerprints. *Signal, Image and Video Processing*, 5:477–483, 2011.

[138] Yanqiang Zhang, Dongmei Sun, and Z. Qiu. A novel method for fusion operators evaluating at score-level fusion in biometrie authentication. *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*, pages 1339–1342, 2010.

[139] Supreetha Gowda H D, H. G., and Mohammad Imran. Multimodal biometric verification system: Evaluation of various score level fusion rules. *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pages 1–4, 2019.

[140] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[141] Shudong Yang, Xueying Yu, and Ying Zhou. Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example. In *2020 International workshop on electronic communication and artificial intelligence (IWECAI)*, pages 98–101. IEEE, 2020.

[142] Debasish Jyotishi and Samarendra Dandapat. An lstm-based model for person identification using ecg signal. *IEEE Sensors Letters*, 4(8):1–4, 2020.

[143] David Belo, Nuno Bento, Hugo Silva, Ana Fred, and Hugo Gamboa. Ecg biometrics using deep learning and relative score threshold classification. *Sensors*, 20(15):4078, 2020.

[144] Imane El Boujnouni, Hassan Zili, Abdelhak Tali, Tarik Tali, and Yassin Laaziz. A wavelet-based capsule neural network for ecg biometric identification. *Biomedical Signal Processing and Control*, 76:103692, 2022.

[145] Kai Jye Chee and Dzati Athiar Ramli. Electrocardiogram biometrics using transformer's self-attention mechanism for sequence pair feature extractor and flexible enrollment scope identification. *Sensors*, 22(9):3446, 2022.

[146] Sihem Hamza and Yassine Ben Ayed. Recognition of person using ecg signals based on single heartbeat. In *Intelligent Systems Design and Applications: 21st International Conference on Intelligent Systems Design and Applications (ISDA 2021) Held During December 13–15, 2021*, pages 452–460. Springer, 2022.

[147] Binish Fatimah, Pushpendra Singh, Amit Singhal, and Ram Bilas Pachori. Biometric identification from ecg signals using fourier decomposition and machine learning. *IEEE Transactions on Instrumentation and Measurement*, 71:1–9, 2022.

[148] Kiran Kumar Patro, Allam Jaya Prakash, M Jayamanmadha Rao, and P Rajesh Kumar. An efficient optimized feature selection with machine learning approach for ecg biometric recognition. *IETE Journal of Research*, 68(4):2743–2754, 2022.

[149] Meiling Li, Yujuan Si, Weiyi Yang, and Yongheng Yu. Et-umap integration feature for ecg biometrics using stacking. *Biomedical Signal Processing and Control*, 71:103159, 2022.

[150] Pan Yi, Yujuan Si, Wei Fan, and Yang Zhang. Ecg biometrics based on attention enhanced domain adaptive feature fusion network. *IEEE Access*, 2023.

[151] Xuan Wang, Wenjie Cai, and Mingjie Wang. A novel approach for biometric recognition based on ecg feature vectors. *Biomedical Signal Processing and Control*, 86:104922, 2023.

[152] Caterina Fuster-Barceló, Carmen Cámara, and Pedro Peris-López. Unleashing the power of electrocardiograms: A novel approach for patient identification in healthcare systems with ecg signals. *arXiv preprint arXiv:2302.06529*, 2023.

[153] Weicheng Cai, Jinkun Chen, and Ming Li. Exploring the encoding layer and loss function in end-to-end speaker and language recognition system. *arXiv preprint arXiv:1804.05160*, 2018.

[154] Hema Kumar Pentapati and K Sridevi. Log-melspectrum and excitation features based speaker identification using deep learning. In *2022 International Conference on Industry 4.0 Technology (I4Tech)*, pages 1–6. IEEE, 2022.

[155] Nguyen Nang An, Nguyen Quang Thanh, and Yanbing Liu. Deep cnns with self-attention for speaker identification. *IEEE access*, 7:85327–85337, 2019.

[156] Rabab A Rasool. Feature-level vs. score-level fusion in the human identification system. *Applied Computational Intelligence and Soft Computing*, 2021(1):6621772, 2021.

[157] Suvarna Joshi and Abhay Kumar. Multimodal biometrics system design using score level fusion approach. *Int. J. Emerg. Technol*, 11(3):1005–1014, 2020.

[158] Sumegh Tharewal, Timothy Malche, Pradeep Kumar Tiwari, Mohamed Yaseen Jabarulla, Abeer Ali Alnuaim, Almetwally M Mostafa, and Mohammad Aman Ullah. Score-level fusion of 3d face and 3d ear for multimodal biometric human recognition. *Computational Intelligence and Neuroscience*, 2022(1):3019194, 2022.

[159] Marcin D Bugdol and Andrzej W Mitas. Multimodal biometric system combining ecg and sound signals. *Pattern Recognition Letters*, 38:107–112, 2014.