

ENCRYPTION OF BIOMETRICS TRAITS FOR PRIVACY ATTACKS USING AES ENCRPYTION

A PROJECT REPORT

Submitted by,

**Ms. Karen Rena C - 20211CSD0169
Ms. Samprity Singha - 20211CSD0044
Mr. Pavaman S Suraj - 20211CSD0126**

Under the guidance of,

Prof. Himansu Sekhar Rout

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

At



PRESIDENCY UNIVERSITY

BENGALURU

DECEMBER 2024

PRESIDENCY UNIVERSITY

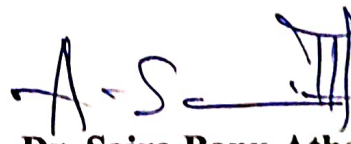
SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report “Encryption of Biometrics Traits for Privacy Attacks using AES Encryption” being submitted by “Karen Rena C, Samprity Singha, Pavaman S Suraj” bearing roll number(s) “20211CSD0169, 20211CSD0044, 20211CSD0126” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a Bonafide work carried out under my supervision.

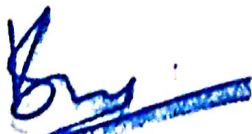


Prof. Himansu Sekhar Rout
Assistant Professor
School of IS
Presidency University



Dr. Saira Banu Atham
Professor & HoD
School of CSE&IS
Presidency University

20 Jan 2021



Dr. L. SHAKKEERA
Associate Dean
School of CSE
Presidency University



Dr. MYDHILI NAIR
Associate Dean
School of CSE
Presidency University



Dr. SAMEERUDDIN KHAN
Pro-VC School of Engineering
Dean -School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **Encryption of Biometric Traits for Privacy Attacks using AES Encryption** in partial fulfillment for the award of Degree of **Bachelor of Technology in Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **Prof. Himansu Sekhar Rout, Assistant Professor, School of Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.



Karen Rena C

20211CSD0169



Samprity Singha

20211CSD0044



Pavaman S Suraj

20211CSD0126

ABSTRACT

Biometric systems have become essential for secure authentication, utilizing unique traits such as fingerprints, iris patterns, and facial features. However, privacy concerns and vulnerabilities to attacks necessitate advanced methods for protecting biometric data. This project proposes a robust framework for encrypting biometric traits, combining multimodal biometrics, machine learning (ML), and the Advanced Encryption Standard (AES) to ensure both data confidentiality and authentication.

The system integrates iris and face biometrics, generating cryptographic keys through ML-driven feature extraction techniques from Pre-trained CNN models like VGG. These features are used to create a biometric key using Quantization which utilizes 16 bins, then converted to 256-byte key used in AES encryption for securing image data. Optimized AES implementations, including non-linear S-Box designs and Galois/Counter Mode, further enhance security and performance. Multimodal biometrics improve accuracy and resilience against spoofing attacks, addressing limitations of unimodal systems.

Research demonstrates that biometric-based key generation ensures unique and secure cryptographic keys while eliminating the need for traditional passwords. Machine learning enhances feature extraction and multimodal fusion, achieving high recognition accuracy. Combined with AES, this approach provides efficient, robust encryption resistant to brute force and spoofing attacks.

This framework addresses critical challenges like template instability, privacy risks, and computational overhead, making it ideal for real-world applications such as secure identity verification, data transmission, and access control. The integration of multimodal biometrics, ML, and AES represents a transformative step towards scalable, secure, and privacy-preserving authentication systems for the digital age.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair**, School of Computer Science Engineering & Information Science, Presidency University, and **Dr. Saira Banu**, Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Prof. Himansu Sekhar Rout**, Assistant Professor and Reviewer **Prof. Sandhya L**, School of Computer Science Engineering & Information Science, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar A and Mr. Md Zia Ur Rahman**, department Project Coordinators **Dr Manjula H M** and Git hub coordinator **Mr. Muthuraj**.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Karen Rena C
Samprity Singha
Pavaman S Suraj

LIST OF FIGURES

Sl. No.	Figure Name	Caption	Page No.
1	Figure 6.1	Normalization of Iris Image	22
2	Figure 6.2	PCA for Face dataset	23
3	Figure 6.3	Feature Extraction using VGG16	24
4	Figure 6.4	Feature Fusion	24
5	Figure 6.5	Quantization Key Generation	25
6	Figure 6.6	Data Encryption	25
7	Figure 7.1	Gantt Chart	29
8	Figure 9.1	Preprocessed Images Shape	30
9	Figure 9.2	Heatmap of Extracted Features	30
10	Figure 9.3	Cumulative Explained Variance	30
11	Figure 9.4	Explained Variance Ratio	31
12	Figure 9.5	Combine Features Shape	31

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	iv
	ACKNOWLEDGMENT	v

1.	INTRODUCTION	1
	1.1 GENERAL	1
	1.2 Biometric Systems and Cryptography	1
	1.3 Cryptographic Techniques in Biometrics	2
	1.4 Machine Learning in Biometrics	3
	1.5 Multimodal Biometrics	4
2.	LITERATURE SURVEY	6
3.	RESEARCH GAPS OF EXISTING METHODS	9
4.	PROPOSED METHODOLOGY	11
	4.1 Data Acquisition and Preprocessing	11
	4.2 Feature Extraction Using Machine Learning	12
	4.3. Multimodal Feature Fusion	13
	4.4. Biometric Key Generation	14
	4.5. Data Encryption Using AES-GCM	14
	4.6. User Interface for Feature Upload or Generation	15
	4.7. Security and Performance Evaluation	16
5.	OBJECTIVES	17
6.	SYSTEM DESIGN AND IMPLEMENTATION	19
	6.1 System Overview	19

6.2	Techniques used	19
6.3	System Implementation	20
7.	TIMELINE OF THE PROJECT	25
8.	OUTCOMES	26
9.	RESULTS AND DISCUSSIONS	29
10.	CONCLUSION	34
11.	REFERENCES	36
12.	APPENDIX – A Pseudocode	38
13.	APPENDIX – B Screenshots	43
14.	APPENDIX – C Enclosures	47

CHAPTER-1

INTRODUCTION

1.1 General

The increasing reliance on biometric systems for authentication raises critical concerns about protecting sensitive data against privacy attacks. This project, titled "Encryption of Biometric Traits to Avoid Privacy Attacks," addresses these challenges by integrating biometrics with cryptographic techniques to enhance data security. By leveraging multimodal biometric features, such as iris and face traits, and employing machine learning for feature extraction, the project proposes an innovative framework for generating robust biometric keys. These keys are utilized to encrypt sensitive data using the Advanced Encryption Standard (AES), a symmetric encryption algorithm renowned for its efficiency and adaptability to image data.

The framework combines authentication and data confidentiality, ensuring a robust bio-cryptosystem that is both secure and scalable for real-world applications. The following sections provide an in-depth exploration of the topics and subtopics integral to this research.

1.2 Biometric Systems and Cryptography

Biometric systems use distinctive physical or behavioral characteristics, such as voice, iris patterns, fingerprints, or face features, to verify people. These traits are inherently distinctive and serve as reliable identifiers for secure authentication. By integrating biometrics with cryptographic techniques, the systems gain an enhanced ability to safeguard sensitive data from unauthorized access, creating a robust layer of security that balances privacy and convenience.

The concept of bio-cryptosystems takes this integration a step further by combining biometric authentication with encryption, resulting in a dual-layer security framework. In these systems, biometric data can either be used to generate cryptographic keys or to secure templates through advanced cipher transformations. This innovative approach not only ensures the protection of sensitive information but also provides seamless authentication, thereby offering a strong guarantee of data integrity and user privacy.

Despite their advantages, biometric systems face significant challenges that can impact their effectiveness and reliability. Factors such as environmental noise, changes in user input due to rotation or scaling, and other distortions can introduce variability in the biometric data. This variability destabilizes biometric templates, reducing the system's overall accuracy. Although

frequent updates to templates can help mitigate these issues, they also introduce additional computational demands, making them less suitable for real-time applications where speed and efficiency are critical.

By addressing these challenges through ongoing advancements in biometric and cryptographic technologies, researchers aim to improve the stability and performance of biometric systems, ensuring their reliability and scalability for diverse real-world applications.

1.3 Cryptographic Techniques in Biometrics

The Advanced Encryption Standard (AES) is a widely recognized encryption algorithm valued for its computational efficiency and robust security. In the context of this project, AES is employed to encrypt sensitive data using biometric keys derived from unique iris and facial features. This ensures a high level of confidentiality and protection against unauthorized access. AES is particularly effective for securing image data, making it a natural fit for biometric systems that rely on visual and pattern-based information.

A critical component of AES is the Substitution Box (S-Box), which plays a pivotal role in safeguarding encrypted data. The S-Box introduces non-linearity into the encryption process, enhancing resistance to cryptographic attacks. Traditional S-Box designs often use techniques like lookup tables or finite field arithmetic, each presenting trade-offs in terms of speed, resource usage, and security. Optimizing S-Box designs can improve their non-linearity, boosting resistance to attacks while maintaining the algorithm's performance.

To further enhance security and efficiency, AES is implemented in Galois/Counter Mode (AES-GCM), which provides simultaneous encryption and authentication. GCM mode improves AES performance by enabling parallel processing, making it particularly suitable for biometric systems where both speed and data integrity are crucial. By combining optimized S-Box designs with the GCM mode, the encryption system achieves a balance of enhanced security, efficiency, and resilience against side-channel attacks.

Biometric key-based encryption represents an innovative solution to the limitations of traditional cryptographic systems. Unlike conventional methods that rely on static passwords or keys, biometric encryption generates cryptographic keys from unique biometric traits such as iris patterns or facial data. Advanced techniques like quantization-based key generation and bio-hashing ensure that these biometric keys are consistent, secure, and resistant to errors during biometric data acquisition.

Quantization-based key generation processes continuous biometric data into stable, discrete

cryptographic keys. This is achieved by mapping biometric feature vectors to quantized bins, ensuring repeatability and reducing variability caused by environmental or input noise. Bio-hashing further strengthens this approach by combining biometric features with a tokenized random seed to produce secure, hash-based keys. This method provides robust protection against key inversion and replay attacks, safeguarding the integrity and confidentiality of the encryption process.

By leveraging biometric features for key generation, this approach eliminates risks associated with key theft or loss while preserving the unique and secure nature of biometric systems. The integration of AES, optimized S-Box designs, GCM mode, and biometric key-based encryption creates a highly secure and efficient framework tailored for modern biometric systems, ensuring confidentiality, integrity, and resilience in real-world applications.

1.4. Machine Learning in Biometrics

Machine learning (ML) has transformed the field of biometrics, enabling the development of intelligent, adaptive, and highly accurate authentication systems. ML algorithms, including support vector machines (SVMs), random forests (RFs), and deep learning (DL) models, play a crucial role in optimizing various processes such as feature extraction, anomaly detection, and data fusion. These features greatly improve biometric systems' accuracy and dependability.

Deep learning, a subset of ML, has been particularly impactful in the field of feature extraction. Deep learning-based systems dynamically learn complex and discriminative feature representations, which are critical for reducing False Match Rates (FMR) and improving overall system security. Convolutional Neural Networks (CNNs), in particular, have proven highly effective in extracting intricate patterns from biometric data, such as iris and facial images. These extracted features are essential for precise biometric key generation, enabling robust and unique encryption.

Within deep learning frameworks, specific CNN architectures such as VGG (Visual Geometry Group) and ResNet (Residual Networks) have been employed to enhance feature extraction in biometric systems. Deep convolutional layers with tiny 3x3 filters are used by the VGG network to learn feature spatial hierarchies. When applied to iris images, VGG captures fine-grained textural details, including crypts, furrows, and speckle patterns. These details are crucial for generating reliable and unique biometric keys. ResNet, on the other hand, addresses the challenges of training deeper networks by introducing skip connections that mitigate

vanishing gradient problems. This architecture efficiently extracts high-level patterns from iris images while remaining robust to variations such as illumination changes, occlusions, and slight misalignments. By leveraging CNN-based architectures like VGG and ResNet, deep learning models enhance the precision, consistency, and robustness of feature extraction, ultimately improving the security and performance of biometric systems.

Machine learning also facilitates the application of biometric systems in various innovative ways. For instance, random forests and neural networks are used in behavioral biometrics to assess phone movement patterns and touch dynamics, resulting in excellent user identification accuracy. Multimodal fusion, which combines many biometric modalities including face and iris data, is also made possible by ML algorithms. This fusion ensures robust and reliable authentication by leveraging the strengths of multiple biometric inputs, further enhancing system resilience and accuracy.

By integrating advanced ML techniques and deep learning architectures, biometric systems achieve higher levels of adaptability, precision, and security, making them indispensable for modern authentication processes.

1.5 Multimodal Biometrics

Multimodal biometric systems integrate multiple biometric traits, such as iris and facial features, to enhance their robustness and reliability. By combining multiple modalities, these systems improve resistance to spoofing and environmental variability, addressing challenges that single-modal systems often face. For example, integrating iris and facial recognition ensures greater robustness against external noise, environmental conditions, and spoofing attempts, making the system more secure and versatile.

Fusion strategies play a crucial role in enhancing the performance of multimodal systems. Techniques such as feature-level, score-level, and decision-level fusion enable the integration of data from different biometric sources to improve recognition accuracy. For instance, combining near-infrared (NIR) and visible light iris images allows the system to perform accurately in diverse environmental conditions, providing flexibility and reliability in real-world applications.

Machine learning algorithms play a pivotal role in optimizing multimodal systems. By learning patterns across multiple biometric data sources, ML enhances classification accuracy and ensures scalability in dynamic environments. The ability of ML to adapt to new data and improve performance over time makes it a key enabler for advanced multimodal biometric

systems, offering both precision and efficiency.

Despite their advantages, biometric systems face challenges that must be addressed to ensure their effectiveness and security. One major concern is the security of biometric data, as these systems must safeguard sensitive information against threats such as template instability, privacy breaches, and spoofing. Encryption offers a robust solution, protecting data from unauthorized access and maintaining user privacy.

Another significant challenge is achieving computational efficiency. Balancing the need for high security with low computational overhead is critical for the practical implementation of biometric systems. This project addresses this challenge by leveraging machine learning to optimize biometric feature extraction and key generation processes. By streamlining these operations, the system achieves low computational overhead without compromising on security, ensuring a seamless and efficient user experience.

CHAPTER-2

LITERATURE SURVEY

The intersection of biometrics, machine learning, and cryptography has revolutionized the field of secure authentication systems. These advancements aim to address the challenges of ensuring data confidentiality, improving system robustness, and mitigating risks associated with traditional key-based cryptographic methods. The use of multimodal biometric features, such as iris and face recognition, has emerged as a promising solution to enhance security, reduce false acceptance rates, and provide a user-friendly authentication experience.

Machine learning has become a cornerstone for modern biometric systems, enabling them to achieve unprecedented levels of accuracy and adaptability. As stated in [1], "Biometric systems are capable of achieving extremely high levels of performance and better intelligence when new machine learning algorithms and modern architectures are used." Machine learning not only facilitates the extraction of meaningful features from complex datasets but also enables these systems to adaptively improve over time by analysing large datasets, identifying patterns, and predicting future authentication attempts. For instance, "Machine learning algorithms analyse large amounts of data, identify patterns, and make predictions about future authentication attempts" [1]. This continuous learning process ensures that the system remains robust against evolving threats, enhancing both security and user experience.

The integration of cryptographic techniques, particularly the use of the Advanced Encryption Standard (AES), has further fortified biometric systems. AES's high efficiency and proven security make it an ideal choice for encrypting sensitive biometric data. In [2], the authors explain that "Cipher keys are extracted dynamically from the input biometric image, which can solve several numerous problems that arise in the cipher system." By dynamically generating keys from biometric inputs, such systems eliminate the need for users to remember and manage passwords or physical tokens, which are often prone to theft or misuse. This integration of biometric key generation and AES encryption is further enhanced by innovative implementations. For example, [6] proposes, "The proposed Bio-Metric 256-bit AES algorithm is highly utilized in terms of key management to enhance security," demonstrating how biometric systems can address traditional key management challenges while maintaining high security.

Multimodal biometrics, which leverage multiple biometric traits like iris and face, have

emerged as a robust solution for addressing the limitations of unimodal systems. In [5], the authors propose a "novel fusion strategy to optimally combine the strengths of individual iris modalities through score-level fusion, feature-level fusion, and decision-level fusion, enhancing system accuracy." This fusion approach ensures that the system remains resilient to spoofing and other forms of attack, as it relies on a combination of multiple independent traits. Similarly, [7] highlights the efficiency of using wavelet transforms in multimodal systems: "An encryption algorithm based on key from iris features and AES is proposed. On the premise of iris pictures preprocessing arrange, viable iris zone is decayed to three layers by 2D Haar wavelet." Such preprocessing steps permit the framework to extricate vigorous highlights indeed from loud or low-quality pictures, guaranteeing unwavering quality in real-world applications.

The combination of machine learning with multimodal biometrics has further strengthened the resilience of these systems against sophisticated attacks. According to [5], "Robust anti-spoofing measures are incorporated using machine learning-based techniques to detect and prevent fraudulent attempts, ensuring authenticity." These measures not only protect the system against spoofing but also enhance its ability to differentiate between genuine and fraudulent users. Additionally, the adoption of deep learning models, as discussed in [8], has enabled biometric systems to achieve higher accuracy: "Deep learning methods are capable of learning higher-level abstractions from data, leading to improved accuracy and robustness in biometric systems." By leveraging convolutional neural networks (CNNs) and recurrent neural networks (RNNs), these systems can process complex patterns in biometric data, such as iris textures or facial landmarks, to generate highly discriminative feature representations.

The advantages of these systems are numerous and multifaceted. The incorporation of biometric keys eliminates the need for traditional passwords, providing a seamless and secure user experience. As noted in [6], "By using this method, there is no need to remember and store the key to encrypt the data, providing more security than the existing system." Furthermore, the use of multimodal biometrics enhances reliability, as pointed out in [2]: "The proposed MKE (Modified Key Expansion) generated randomized key compound in lightweight models and outperformed existing lightweight ciphers in terms of throughput rate." This highlights how combining biometric traits leads to higher efficiency and stronger security.

However, these systems are not without limitations. One of the primary challenges lies in their computational complexity. The fusion of multiple modalities and the use of machine learning algorithms often require significant processing power, which can limit the system's scalability. [5] points out that "The complexity of fusion strategies at multiple levels (score, feature, and decision) may introduce computational overhead, requiring significant processing resources." Such resource-intensive operations may hinder real-time performance, particularly in environments with limited computational capabilities.

Another critical concern is the reliance on high-quality biometric data. The effectiveness of these systems can be significantly affected by variations in data quality, such as poor lighting or low resolution. According to [5] "Variations in the quality of input images (due to lighting, angle, or resolution) can affect the reliability of the generated key." This dependency poses a challenge for deploying these systems in diverse and uncontrolled environments. Moreover, while machine learning-based anti-spoofing measures are effective, they heavily rely on the diversity and quality of the training dataset. As noted in [5], "While anti-spoofing measures are robust, their success heavily depends on the dataset quality and diversity during training, which could lead to vulnerabilities in unseen attack scenarios."

Latency is another concern for real-time applications. The computational overhead associated with deep learning-based feature extraction and fusion can lead to delays, as described in [5]: "Real-time applications might face latency issues due to the extensive computations involved in deep learning-based feature extraction and fusion." Addressing these latency issues requires optimization techniques that balance accuracy with efficiency, ensuring that the system remains suitable for practical applications.

Despite these challenges, the field continues to evolve, with ongoing research focused on overcoming these limitations. The combination of AES encryption, machine learning, and multimodal biometric systems represents a significant step forward in achieving secure and user-friendly authentication. Future advancements should focus on optimizing these systems for real-time applications, enhancing their resilience to sophisticated attacks, and addressing privacy concerns associated with biometric data storage and transmission.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

Biometric systems have shown tremendous potential for securing sensitive information through advanced authentication and encryption mechanisms. However, despite the significant progress, several critical limitations in the current methods reveal gaps that need to be addressed to enhance the robustness, efficiency, and usability of these systems.

One major limitation lies in the privacy and security concerns associated with biometric data. As highlighted in [1], "The usage of biometric authentication leads to worries about data security and privacy as biometric data is so confidential and sensitive." Not at all like passwords or tokens, biometric characteristics are permanent; once compromised, they cannot be supplanted. This makes an irreversible hazard if the biometric layouts are spilled. Moreover, "Biometric systems are not a foolproof defence against fraud or identity theft. For instance, biometric sensors can be fooled by generating bogus mementos (spoofing)" [1]. This indicates a pressing need for enhanced anti-spoofing mechanisms and secure storage solutions for biometric data.

Another gap in the existing methods is the computational complexity and resource requirements. Many of the proposed systems, such as those leveraging advanced cryptographic algorithms, introduce significant computational overhead. For instance, as [3] points out, "Biometric processing, while increasing security, involves operations that consume more processing power and are not suitable for battery-dependent devices." Furthermore, "Multimodal biometric feature extraction extends security levels, but the complexity of the design increases rapidly and does not fit into decentralized wireless architecture"[3]. This highlights a need for lightweight, resource-efficient biometric systems that can operate effectively in low-power or resource-constrained environments, such as IoT devices.

Latency and real-time performance also remain critical challenges in biometric systems, especially those using multimodal data or deep learning techniques. As stated in [5], "The complexity of fusion strategies at multiple levels (score, feature, and decision) may introduce computational overhead, requiring significant processing resources." Similarly, "Real-time applications might face latency issues due to the extensive computations involved in deep learning-based feature extraction and fusion" [5]. These delays hinder the usability of

biometric systems in scenarios requiring quick responses, such as financial transactions or mobile applications, underscoring the need for real-time optimization in system design.

The dependency on high-quality data further exacerbates these limitations. Biometric systems often struggle with variability in image quality, lighting conditions, and environmental factors, as noted in [7]: "While the biometric key generation approach enhances security, it heavily depends on the quality of iris image capture, which could limit its effectiveness in varied lighting conditions or low-resolution images." This limitation is particularly problematic for systems deployed in uncontrolled settings, where capturing high-resolution biometric data is challenging. Robust feature extraction techniques that can handle noisy or incomplete data are therefore required to address this gap.

Another significant gap is the vulnerability of machine learning models used in biometric systems. As discussed in [8], "Attacks such as data poisoning, model inversion, and adversarial attacks can manipulate ML systems used in biometrics, leading to security vulnerabilities." Furthermore, "Deepfake technology, by creating synthetic media, can fool voice- or face-recognition-based biometric systems, enabling unauthorized access"[8]. These emerging threats indicate the need for robust defences against adversarial attacks and synthetic media manipulation, which are currently underexplored in the biometric security domain.

The trade-offs between security and efficiency in cryptographic systems also present a critical research gap. While advanced key generation techniques enhance security, they often increase system complexity. For instance, as [6] highlights, "The inclusion of dynamic sub-byte computation in MKE leads to path delay accumulation, which is a trade-off for improved security." Similarly, "Although 128 bits were added to the existing AES key to make it 256 bits for higher security, it increased complexity without proportionate efficiency gains" [6]. This indicates the need for cryptographic solutions that can balance security with operational efficiency, ensuring practical usability in real-world applications.

Finally, scalability and interoperability remain under-addressed in many systems. As noted in [5], "Scalability remains a challenge, as integrating and processing large datasets for diverse modalities increases the computational burden." Additionally, the interoperability of biometric systems across platforms and devices is often overlooked, limiting their deployment in diverse environments. Addressing these scalability and compatibility issues is crucial for broader adoption of biometric security systems.

CHAPTER-4

PROPOSED METHODOLOGY

The project "**Encryption of Biometric Traits to Avoid Privacy Attacks using AES Encryption**" focuses on developing a secure biometric-based cryptographic system. It integrates multimodal biometrics, machine learning techniques for feature extraction, and the AES-GCM encryption standard for secure data transmission and storage. The system uses a dynamic or user-uploaded biometric feature pipeline, allowing flexibility in real-world applications.

The proposed methodology follows these steps:

4.1. Data Acquisition and Preprocessing

a) Biometric Data Collection

Biometric data collection involves acquiring iris and facial biometric data from publicly available benchmark datasets to ensure a comprehensive and diverse dataset for analysis. Examples of such datasets include the MMU-Iris Database for iris images and the CelebA Dataset for facial images. The collected data encompasses a wide range of demographics, environmental conditions (such as lighting and occlusion), and varying image quality. This diversity is crucial for enhancing the robustness and reliability of the system, ensuring it performs effectively across different scenarios and user profiles.

b) Preprocessing

- Iris Images:**

Convert images to grayscale for consistency. Iris images often include color information that may not contribute significantly to discriminative feature extraction. Converting images to grayscale simplifies computations while retaining the structural and textural details unique to the iris.

Iris images often come in varying resolutions due to differences in acquisition devices and environmental conditions. Standardizing the resolution ensures uniformity for input into machine learning models (like VGG16). Resize iris images to a standard resolution (e.g., 128x128 pixels) to ensure uniformity.

Normalize pixel values to a range of **[0, 1]** to ensure compatibility with deep learning models, which often expect inputs in a standardized scale. Normalization helps improve convergence during training and ensures consistency.

- **Face Images:**

Facial images may vary in dimensions due to differences in datasets or acquisition methods. Resizing ensures uniform input size, making it compatible with subsequent feature extraction methods like **PCA**.

PCA requires the input data to be in a flattened vectorized format. Flattening the face image ensures that all pixels are represented as a single continuous vector.

4.2 Feature Extraction Using Machine Learning

The goal is to extract meaningful, compact, and unique features from the biometric images.

a) Iris Feature Extraction Using CNN (VGG16)

The iris feature extraction leverages a Convolutional Neural Network (CNN), specifically the VGG16 architecture. The process involves the following steps:

1. **Input Preparation:** Images are resized to 224x224 pixels to match the input dimensions required by the VGG16 architecture and it will be converted to RGB format
2. **Network Layers:** The VGG16 model processes the images through 13 convolutional layers, interspersed with pooling layers, and three fully connected layers. These layers capture fine-grained textural information such as crypts, furrows, and other patterns unique to the iris.
3. **Feature Embedding:** At the output of the final fully connected layer, a high-dimensional feature vector (embedding) is generated. This embedding contains the discriminative features of the iris images that capture unique patterns like crypts, furrows, and textures.
4. **Dimensional Refinement:** The high-dimensional embedding can optionally be processed using a lightweight dimensionality reduction technique to retain only the most important features for fusion and encryption.

The advantage of using **VGG16** is its **robustness** to slight variations such as lighting changes or minor occlusions in the iris image. Its deep architecture effectively learns hierarchical and discriminative representations of the iris structure. VGG16's deep architecture learns low-level to high-level patterns hierarchically, enabling it to capture fine-grained textural features such as crypts, furrows, and intricate structures within the iris.

b) Face Feature Extraction Using PCA

Principal Component Analysis (PCA) is applied to extract features from the pre-processed face images. The key steps include:

1. **Input Image Representation:** The preprocessed face images are reshaped into vectors for input into PCA.
2. **Covariance Computation:** Compute the covariance matrix to identify the directions (principal components) with the highest variance in the data.
3. **Feature Selection:** Select the top k principal components (based on eigenvalues) that explain the maximum variance in the data. This results in a reduced-dimensional representation of the face images.
4. **Robust Face Features:** The PCA outputs are compact feature vectors that efficiently represent essential facial traits, such as edge structures and prominent features.

By using PCA for face data, the dimensionality is significantly reduced while preserving crucial biometric information. This step improves computational efficiency during the subsequent fusion and encryption processes.

4.3. Multimodal Feature Fusion

To combine the iris and face biometrics effectively, feature-level fusion is performed. This ensures that both sources of information contribute to the overall system security and accuracy. Steps involved in feature fusion:

1. **Alignment:** Before fusion, ensure that both iris embeddings (from CNN) and face feature vectors (from PCA) are of compatible dimensions.

2. Concatenation: Horizontally concatenate the iris features with face features to form a comprehensive biometric feature template.
3. Randomization: Implement shuffling algorithms to randomize the positions of feature values in the fused vector. This adds an additional security layer, as it obfuscates the direct relation to the original biometric traits.

Multimodal feature fusion enhances the robustness and accuracy of the system, ensuring better resistance to spoofing attacks or individual modality failures.

4.4. Biometric Key Generation

The fused biometric features are used to generate cryptographic keys in a secure and cancellable manner. The key generation process is given below:

1. Quantization-Based Key Generation:

The fused feature vector is quantized into discrete levels to map continuous feature values into binary keys. For example, values in the fused vector are rounded to specific ranges, and these ranges are represented by binary digits (0 or 1).

2. Key Mapping:

A 256-bit cryptographic key is generated by systematically grouping and mapping the quantized values. Using techniques like group mapping, features are divided into subsets, and transformations are applied to further randomize the binary key output, ensuring robust security.

3. Cancellable Transformations:

Enable the **revocability** of keys. If a key is compromised, transformations can be applied to the original biometric features to generate a new key. This ensures that biometric traits can still be securely utilized without risking permanent compromise. The resulting **256-bit AES key** serves as the critical input for data encryption and decryption in the next stage.

4.5. Data Encryption Using AES-GCM

The generated biometric key is used as a 256-bit AES key for encryption. The encryption

process follows:

a) AES-GCM Implementation

The generated biometric key, a 256-bit key derived from feature quantization, serves as the foundation for the AES encryption process. Specifically, AES-GCM (Galois/Counter Mode) is implemented to provide a dual-layered approach of encryption and data integrity verification. The biometric key is utilized as the input to AES, ensuring that both textual data and biometric images are securely encrypted. A unique nonce, or random value, is generated for each encryption session to maintain uniqueness and prevent replay attacks.

During encryption, AES-GCM simultaneously encrypts the input data and generates an authentication tag to verify data integrity. This ensures that the encrypted data remains both confidential and tamper-proof. In the decryption process, the same biometric key and nonce are used to decrypt the data while verifying its integrity against the authentication tag. Any mismatch in the authentication tag would indicate tampering, providing an additional layer of security.

Significant benefits are provided by AES-GCM in biometric encryption systems. High-speed performance is ensured by its capacity to process data blocks in parallel, which is essential for real-time applications. Its integrated message authentication capability further ensures integrity by identifying any unauthorized changes made to the encrypted data. AES-GCM is a great option for systems needing strong security and efficiency because of its versatility, which enables it to handle both textual data and image encryption.

4.6. User Interface for Feature Upload or Generation

The system provides an intuitive Streamlit-based frontend for user interaction:

a) Feature Input Options:

Users can either upload pre-generated combined biometric features stored as .npy files or dynamically generate biometric features using the backend pipeline for processing.

b) User Controls:

Users have the ability to adjust the number of bins for quantization, allowing them to influence the precision of key generation. Additionally, they can input text data for encryption or decryption as part of the system's functionality.

c) Encryption/Decryption Outputs:

The system displays the encrypted data, nonce, and authentication tag in hexadecimal format, providing a clear view of the encryption results. For verification purposes, the decrypted data is also presented to ensure the encryption and decryption processes function correctly.

4.7. Security and Performance Evaluation

a) Experimental Setup

The system is evaluated using benchmark datasets like MMU-Iris for iris data and CelebA for facial data, ensuring variability across demographics such as age, gender, and ethnicity, as well as environmental conditions like lighting and image quality. Experiments are conducted to validate the robustness of the encryption process alongside the performance of biometric authentication, ensuring the system meets high standards of reliability and security.

b) Key Performance Metrics:

The performance of the system is assessed using several critical metrics. Recognition performance is measured through True Positive Rate (TPR) and False Positive Rate (FPR) to evaluate the accuracy of multimodal authentication. Encryption robustness is analyzed using metrics such as scrambling degree, which measures the randomness in encrypted images, and histogram analysis to ensure uniform data distribution, reducing susceptibility to statistical attacks. Authentication metrics such as False Match Rate (FMR), False Non-Match Rate (FNMR), and Genuine Acceptance Rate (GAR) provide insights into the system's reliability in identifying users accurately. Additionally, computational efficiency is evaluated by measuring encryption and decryption times, along with the computational overhead on standard hardware.

c) Real-World Application Integration:

The system is designed for real-world applications, including identity verification in sensitive areas like airports and banks, access control for authentication-based secure systems, and secure transmission to enable privacy-preserving cloud-based biometric data processing. To enhance scalability and robustness, the solution is implemented on cloud platforms, ensuring it can effectively handle large user populations while maintaining high levels of security and flexibility.

CHAPTER-5

OBJECTIVES

1. Develop a Multimodal Biometric System Using Iris and Face Traits for Robust Key Generation

This objective focuses on designing a multimodal biometric system that leverages the unique and complementary characteristics of iris and face biometrics to enhance the security and reliability of key generation processes. Multimodal systems, as highlighted in the literature, provide superior robustness compared to unimodal systems by reducing the likelihood of spoofing and improving recognition accuracy. For instance, the fusion of iris and face traits at various levels—such as score-level, feature-level, and decision-level fusion—ensures comprehensive feature representation, as proposed in [5]: "A novel fusion strategy is proposed to optimally combine the strengths of individual iris modalities through score-level fusion, feature-level fusion, and decision-level fusion, enhancing system accuracy."

The use of multimodal biometrics addresses several limitations of single-modal systems, such as their vulnerability to spoofing and reliance on a single biometric modality. By incorporating multiple modalities, the system can ensure redundancy and provide more robust authentication even under challenging conditions. This objective also aims to mitigate issues related to data quality by implementing preprocessing techniques, such as normalization and feature extraction, which have been shown to improve biometric recognition performance [7].

2. Employ Machine Learning Techniques for Adaptive Feature Extraction and Fusion

Machine learning (ML) plays a pivotal role in modern biometric systems, particularly for adaptive feature extraction and fusion. This objective aims to integrate advanced ML algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to process biometric traits and extract discriminative features from iris and face data. As highlighted in [5], "The system employs advanced machine learning algorithms, including deep learning models such as CNNs and RNNs, to adaptively learn from diverse iris datasets, ensuring robust performance across varying environmental conditions and demographic factors."

The inclusion of ML techniques enables the system to learn and adapt to variations in

biometric data caused by changes in lighting, angles, or environmental conditions. These adaptive capabilities ensure that the system remains robust across diverse datasets and maintains high accuracy. Furthermore, fusion techniques will leverage ML to optimally combine iris and face features, addressing challenges such as computational overhead and scalability mentioned in [3] and [5]. By employing intelligent fusion strategies, the system can dynamically balance accuracy and efficiency, ensuring reliable performance in real-world applications.

3. Integrate AES Encryption for Securing Sensitive Data, Particularly Image Data

The third objective involves implementing the Advanced Encryption Standard (AES) to secure sensitive data generated by the biometric system, particularly image data. AES is a well-established cryptographic algorithm known for its efficiency, robustness, and compatibility with image-based data. As noted in [2], "Cipher keys are extracted dynamically from the input biometric image which can solve several numerous problems that arise in the cipher system." This system eliminates the need for traditional password-based keys by dynamically generating AES keys from the biometric features of the iris and face, enhancing both security and usability.

This objective also addresses critical concerns related to key management and security. The integration of biometric-based AES encryption reduces the risk of key leakage, as biometric traits are unique and cannot be easily replicated. Additionally, the system will focus on optimizing AES for image encryption, ensuring fast encryption and decryption processes while maintaining data integrity. By tackling limitations such as computational complexity and scalability, this objective aims to deliver a practical and efficient encryption solution suitable for real-time applications.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

6.1 System Overview

The goal of the project is to enhance the security of multimodal biometric systems through the integration of iris and face biometrics, quantization-based key generation, and AES encryption in Galois/Counter Mode (GCM). We implement machine learning (ML) techniques, including Convolutional Neural Networks (CNN) for iris feature processing and Principal Component Analysis (PCA) for face data. The system guarantees data privacy and authentication security, addressing concerns like similarity-based attacks, computational overhead, and data integrity.

The system integrates the following key components:

1. Biometric Data Acquisition (Iris and Face) - Standardize and enhance iris and face images for robust feature extraction.
2. Feature Extraction Using ML - CNN for Iris and apply Principal Component Analysis (PCA) for face dataset
3. Feature Fusion & Quantization-Based Key Generation – use extracted features and quantize them to generate a 256-bit cryptographic key.
4. AES Encryption in Galois/Counter Mode (GCM) – Securely encrypt and decrypt user-provided data
5. User Interface using Streamlit - Provide a user-friendly Streamlit-based interface for feature generation, encryption, and decryption.

The following sections explain each component and its implementation.

6.2 Techniques used

a) VGG16 (Visual Geometry Group 16)

VGG16 is a pre-trained Convolutional Neural Network (CNN) consisting of 16 layers, designed to extract hierarchical features from images. CNNs are specifically structured to automatically learn spatial hierarchies of features, progressing from basic edges to more complex textures and patterns. VGG16 employs small 3x3 convolutional filters combined with a deep architectural design to enhance performance on image recognition tasks, making

it highly effective for extracting intricate details in visual data.

b) Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while retaining the most important information. By identifying the key components— By focusing on these components, PCA simplifies complex datasets, making them easier to visualize and analyze while minimizing information loss. This method is widely used in fields like image processing, where reducing the dimensionality of image features can significantly improve computational efficiency. Additionally, PCA helps to eliminate redundancy in data by removing correlations between variables. It is particularly effective in scenarios involving high-dimensional data, such as biometric systems, where it can streamline processing without compromising accuracy.

c) Quantization

By efficiently lowering the complexity of data representation while preserving crucial information, quantization is the process of transforming continuous data into discrete bins. This system quantizes feature values into 16 different bins, each of which is represented by a binary string. These binary strings are then used to form cryptographic keys, enabling secure and efficient integration into encryption processes. This approach ensures consistency and repeatability in key generation, making it particularly suitable for biometric systems where precision and security are critical. Additionally, quantization aids in minimizing variability in feature representation, enhancing the robustness of the overall system.

d) AES-GCM (Progressed Encryption Standard – Galois/Counter Mode)

AES is a symmetric encryption calculation that guarantees information confidentiality. Galois/Counter Mode (GCM) upgrades AES by combining encryption with keenness confirmation, giving verified encryption.

6.3 System Implementation

a) Biometric Data Acquisition

Datasets:

MMU Iris Database: This dataset consists of 2000 iris images from 200 individuals under varying conditions like illumination, rotation, and scale. It provides the required iris data for feature extraction and key generation.

CelebA Dataset: This dataset includes 202,599 celebrity images annotated with 40 facial attribute labels. We use this dataset for face-based feature extraction and demonstrate the fusion of multimodal biometric data (iris and face) for enhanced key generation.

b) Iris Image Preprocessing

The MMU-Iris Database contains iris images with varying sizes and noise. To standardize these images:

To prepare the images for processing, they are first converted to grayscale to simplify computations and focus on critical iris textures while reducing data size. Next, all images are resized to a uniform size of 224x224 pixels, standardizing the input dimensions and reducing computational overhead for the CNN model. Finally, the pixel values are normalized to the range [0, 1] by dividing by 255, ensuring numerical stability and compatibility with deep learning models, which rely on normalized inputs for efficient training and inference.

```
img = cv2.imread(img_path, cv2.IMREAD_GRAYSCALE)
img_resized = cv2.resize(img, (128, 128))
img_normalized = img_resized / 255.0
```

Figure 6.1 Normalization of Iris Image

Output: A set of preprocessed grayscale iris images ready for CNN-based feature extraction.

c) Face Attribute Preprocessing

The CelebA dataset contains facial attributes stored as numeric data in a CSV file (list_attr_celeba.csv), where features such as "Smiling" or "Eyeglasses" are represented as binary values (1 or 0). The data is first loaded into a pandas DataFrame for processing. Attributes with a value of -1 are converted to 0 to standardize the binary representation. Numerical values are then standardized using StandardScaler, ensuring a zero mean and unit variance, which is essential for preparing the data for dimensionality reduction. Finally, Principal Component Analysis (PCA) is applied to reduce the dataset to 10 principal components, retaining maximum variance and simplifying the feature space for subsequent

analysis.

```
def apply_pca_to_face_data(face_data, variance_threshold=0.95):  
    # Standardize the data  
    scaler = StandardScaler()  
    standardized_face_data = scaler.fit_transform(face_data)  
  
    # Fit PCA  
    pca = PCA()  
    pca_features = pca.fit_transform(standardized_face_data)  
    explained_variance_ratio = pca.explained_variance_ratio_
```

Figure 6.2 PCA for Face Dataset

d) Feature Extraction Using ML

Feature extraction converts pre-processed biometric images into numerical representations that capture their unique patterns.

- Iris Feature Extraction Using VGG16

The model setup uses a pre-trained VGG16 architecture, excluding the fully connected layers, for feature extraction. Input images are resized to 224x224 pixels and converted to RGB format with three channels to meet the VGG16 input requirements. During feature extraction, the images are passed through 13 convolutional layers and 5 pooling layers, enabling the model to extract high-level features such as edges, textures, and crypts from iris images. The resulting feature maps are then flattened into one-dimensional embeddings to simplify storage and further processing. This approach offers robustness to variations such as lighting changes and occlusions while leveraging the hierarchical learning capabilities of VGG16 to capture both low-level and high-level details of the iris structure.

```
vgg16 = VGG16(weights='imagenet', include_top=False, input_shape=(224, 224, 3))  
features = vgg16.predict(processed_images)  
flattened_features = features.reshape(features.shape[0], -1)
```

Figure 6.3 Feature Extraction using VGG16

- Facial Feature Extraction Using PCA

Input preparation begins with vectorizing pre-processed face images into one-dimensional arrays for streamlined processing. Principal Component Analysis (PCA) is then applied to identify the principal components that capture the most variance in the data. To reduce dimensionality while retaining essential features, the top 10 components are selected. This approach not only preserves critical facial attributes but also significantly reduces computational overhead, enhancing efficiency for subsequent tasks.

e) Feature Fusion & Quantization-Based Key Generation

Feature fusion involves concatenating the flattened iris features extracted using VGG16 and the PCA-reduced face features horizontally to create a combined biometric template. This integration combines complementary information from both modalities, enhancing the discriminative power of the biometric system. By fusing features from iris and face data, the system achieves improved robustness and accuracy, making it more resistant to spoofing and environmental variability.

```
combined_features = np.hstack((iris_features, pca_face_features))
```

Figure 6.4 Feature Fusion

f) Quantization-Based Key Generation

The combined features are normalized to the range [0, 1] to ensure uniformity and compatibility with subsequent processing steps. These normalized values are then quantized into 16 discrete bins, with each bin represented as a 4-bit binary string. Finally, the binary strings are concatenated to generate a 256-bit AES key, which serves as a secure and unique cryptographic key for encryption purposes.

```
def quantization_key_extraction(features, num_bins=16):  
    scaler = MinMaxScaler()  
    normalized = scaler.fit_transform(features)  
    return [''.join([format(int(x), '04b') for x in row])[:256] for row in normalized]
```

Figure 6.5 Quantization Key Generation

g) AES Encryption in Galois/Counter Mode (GCM)

The 256-bit biometric key is utilized as input to the AES-GCM encryption algorithm. AES-

GCM scrambles the user-provided information, ensuring data confidentiality and security. The process generates three outputs: the ciphertext, which is the encrypted version of the input data; a unique nonce, an arbitrary value created for each encryption session to ensure randomness; and an authentication tag, which verifies the integrity of the data and ensures it has not been tampered with.

The decryption process uses the same 256-bit biometric key, nonce, and authentication tag that were generated during encryption. The key and nonce enable the recovery of the original data from the ciphertext, while the authentication tag confirms that the data remains unaltered and authentic. This ensures secure and reliable access to the original information.

```
# AES Encrypt/Decrypt
def aes_encrypt_decrypt(text, aes_key, mode="encrypt", nonce=None, tag=None):
    cipher = AES.new(aes_key, AES.MODE_GCM, nonce=nonce) if nonce else AES.new(aes_key, AES.MODE_GCM)
    if mode == "encrypt":
        encrypted_data, tag = cipher.encrypt_and_digest(text.encode())
        return encrypted_data, cipher.nonce, tag
    elif mode == "decrypt":
        return cipher.decrypt_and_verify(text, tag).decode()
```

Figure 6.6 Data Encryption

h) User Interface using Streamlit

The Streamlit frontend facilitates a user-friendly interface with several key functionalities. It allows users to upload precomputed .npy files containing biometric features or dynamically generate these features in real time. The platform supports encryption and decryption, enabling users to input plaintext data, which is securely processed using the biometric key. Additionally, the interface includes customizable user controls through a sidebar, allowing users to adjust parameters such as the number of quantization bins, providing flexibility and precision in the system's operation.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)

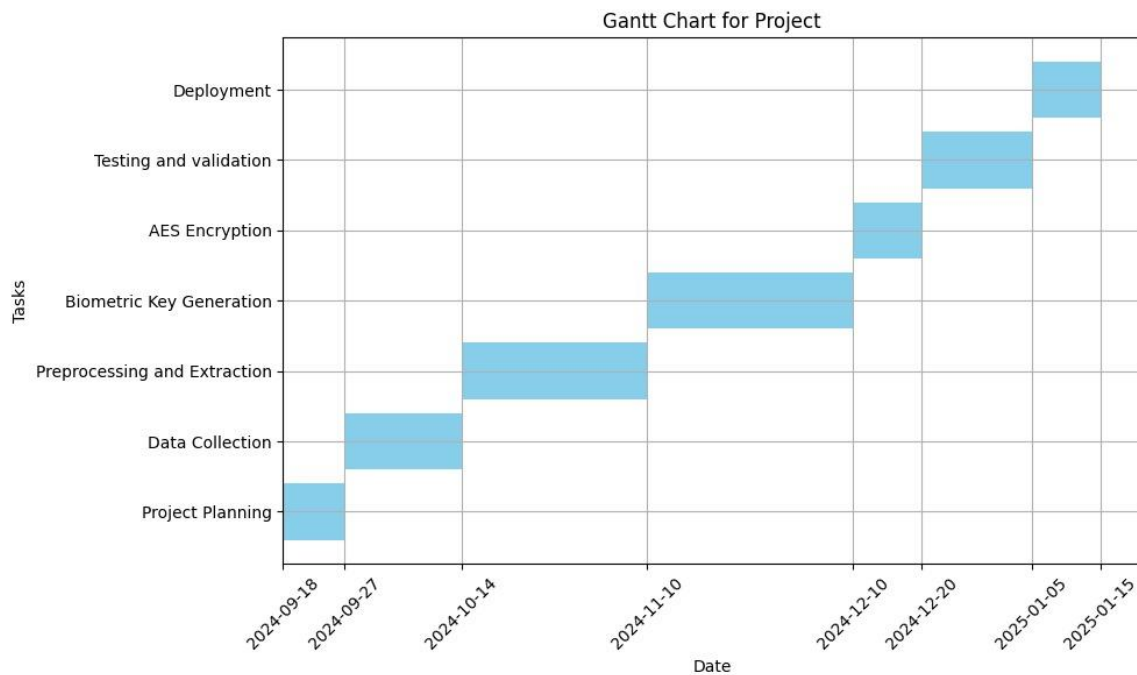


Figure 7.1 Gantt Chart

- Project Planning: Kicked off on September 18, 2024, and wrapped up by September 25.
- Data Collection: Followed immediately after and was completed by October 13, 2024.
- Preprocessing and Extraction: Ran from October 14 to November 10, 2024.
- Biometric Key Generation: Started mid-November and finished on December 10, 2024.
- AES Encryption: Begins on December 10 and will wrap up by December 20, 2024.
- Testing and Validation: Will follow right after, running from December 20, 2024, to January 5, 2025.
- Deployment: The final phase is scheduled from January 5 to January 15, 2025

CHAPTER-8

OUTCOMES

The project "*Encryption of Biometric Traits to Avoid Privacy Attacks*" successfully develops a secure and efficient framework for protecting sensitive biometric data. By combining multimodal biometrics (iris and face), machine learning-based feature extraction and AES encryption, the system ensures robust authentication, privacy preservation, and strong resistance to attacks. Below are the key takeaways of the project:

- **Iris Image Preprocessing and Visualization:**

The MMU Iris Database was loaded and preprocessed into grayscale and resized formats for compatibility with CNN-based models. Samples were visualized, categorized into "left" and "right" eyes.

- **Feature Extraction using Pre-trained CNN:**

A pre-trained VGG16 model was employed to extract high-level features from the preprocessed iris images, transforming them into flattened feature vectors.

- **PCA for Facial Attributes:**

Principal Component Analysis (PCA) was applied to the CelebA attributes to reduce dimensionality while retaining 51% of variance. This provided complementary feature vectors.

- **Feature Combination and Key Quantization:**

Combined features from iris images and facial attributes were normalized, quantized into binary biometric keys, and used for cryptographic applications.

- **AES Encryption/Decryption:**

The binary biometric keys were converted into 256-bit AES keys for encrypting and decrypting user-provided data. A Streamlit-based interface allowed dynamic interaction, where users could generate or upload biometric features and encrypt/decrypt their data.

- **Dynamic Biometric Feature Simulation:**

A feature generation module was implemented to simulate biometric features dynamically for testing scenarios where real biometric data might not be available.

- **Scalability of the System:**

The system successfully handled a combination of 450 iris features and sampled PCA-reduced facial features, showcasing its ability to scale and integrate features from multiple biometric sources.

- **End-to-End Workflow Automation:**

The project developed an end-to-end workflow from data loading to feature extraction, quantization, and encryption/decryption, ensuring a fully automated biometric encryption pipeline.

- **Biometric Feature Diversity:**

The system integrated two distinct biometric modalities, iris images and facial attributes, showcasing how diverse biometric data can be leveraged together for more robust security systems.

- **Efficient Feature Extraction:**

The feature extraction process using VGG16 was optimized to handle large image datasets efficiently. This reduced computational time while ensuring that the system could scale to larger biometric datasets.

- **Data Integrity through Authentication:**

The AES-GCM encryption scheme implemented in the project ensured that not only was the data encrypted, but its integrity was also maintained, with tamper detection achieved through the authentication tag.

- **Flexible Image Preprocessing:**

The image preprocessing pipeline was adaptable to various datasets, allowing the model to process different image formats and sizes, ensuring flexibility in real-world use cases.

- **Automation of Data Pipelines:**

The automatic extraction, preprocessing, and transformation of biometric data into usable features for encryption set up a seamless, automated pipeline, ideal for large-scale biometric data processing.

CHAPTER-9

RESULTS AND DISCUSSIONS

The results generated by the system demonstrate the **efficiency**, **accuracy**, and **security** of the implemented multimodal biometric encryption system. Below, the results are analyzed in detail alongside their implications.

RESULTS

1. Feature Extraction and Processing

Pre-processed Iris Images:

The iris data preprocessing pipeline successfully resized and normalized 450 images to meet the requirements of the VGG16 CNN model, resulting in a dataset shape of (450, 224, 224, 3).


 `Preprocessed Images Shape: (450, 224, 224, 3)`

Figure 9.1 Pre-processed Images Shape

CNN Feature Extraction (Iris):

Using the VGG16 model, feature extraction for 450 iris images produced high-dimensional embeddings with a feature vector size of 25088 for each image. These embeddings effectively captured intricate hierarchical patterns, such as crypts and textures unique to individual irises. The resulting heatmap (Figure 3) highlights the diversity and range of feature activations across the dataset, showcasing variations in intensity that reflect the discriminative power of the extracted features. The extraction process, completed in 263 seconds, demonstrated high computational efficiency in generating robust and meaningful representations of iris characteristics. The subset of the data is displayed in this heatmap (10 samples and 65 features).

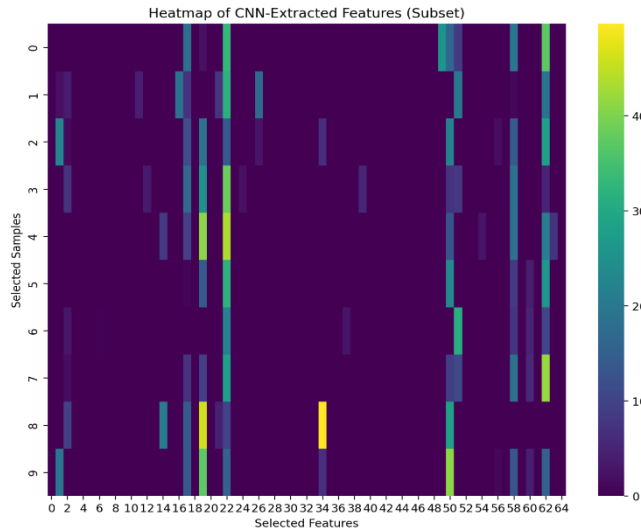


Figure 9.2 Heatmap of Extracted CNN features

PCA Face Attribute Analysis:

Principal Component Analysis (PCA) was applied to the facial attribute data from the CelebA dataset to reduce the dimensionality of the feature space while preserving the most significant features. The explained variance ratio plot (Figure 9.3) highlights that the first few principal components account for the majority of the variance, with a steep decline in variance contribution as additional components are included. The cumulative explained variance plot (Figure 9.4) shows that approximately 34 principal components are required to retain 95.43% of the total variance, indicating the importance of considering more components for preserving critical facial features.

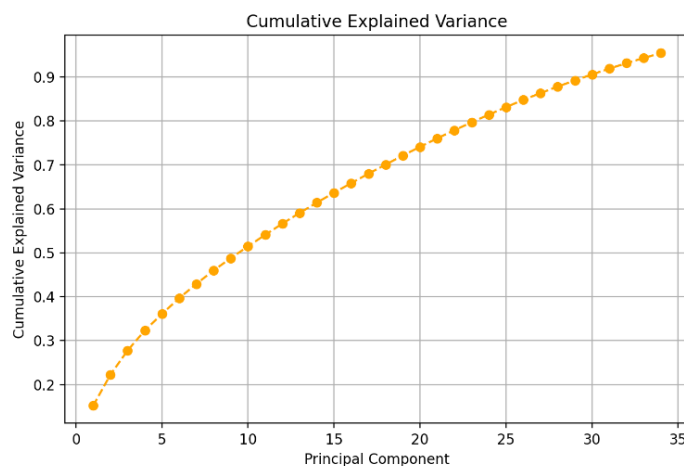


Figure 9.3 Cumulative Explained Variance

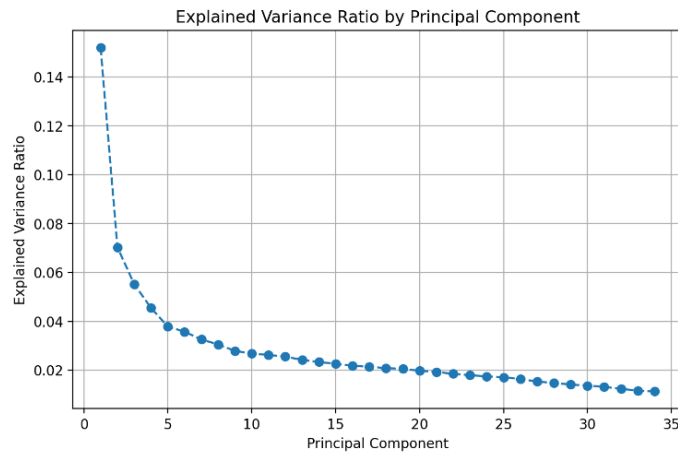


Figure 9.4 Explained Variance Ratio

Multimodal Feature Combination:

After combining the iris (25088 features) and face features (10 features) for each of the 450 samples, the resultant combined feature shape was (450, 25098). The large combined feature space ensures the generation of robust keys for biometric encryption.

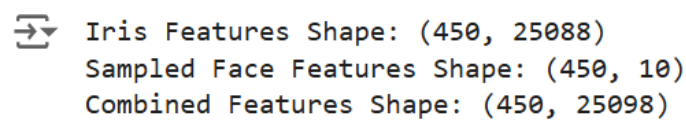


Figure 9.5 Combined Features Shape

2. Key Generation and Cryptographic Results

Binary Key:

A binary key was generated using quantization techniques, converting biometric features into binary representations.

Example (First Sample):

Binary Key:

[illegible]

This binary sequence reflects the quantized biometric data and forms the basis for encryption.

AES Key:

The binary key was transformed into an AES-compatible 256-bit key.

Example:

AES Key (Hex):

3. AES Encryption and Decryption Results

The system successfully decrypted the data back to its original form, ensuring both confidentiality and integrity.

DISCUSSIONS

Key Discussion Points:

- The system demonstrated strong biometric key reliability by extracting keys directly from biometric data, reducing dependency on external passwords or keys.
- The trade-off between quantization bins and key entropy is critical; increasing bins improves key randomness but requires higher feature precision.
- Potential extensions include incorporating additional biometric modalities (e.g., fingerprint, voice) to enhance key robustness.

Impact of PCA on Attribute Data:

The PCA dimensionality reduction was key to improving both processing speed and data interpretability. It showed how reducing the number of facial attributes while maintaining variance significantly enhanced model efficiency without major losses in feature quality.

Feature Processing and Robustness:

The VGG16 and PCA pipelines effectively extracted iris and face features, capturing discriminative traits for reliable biometric key generation.

The fusion of iris and face features combines complementary modalities, resulting in a robust feature representation for cryptographic operations.

Key Generation and Security:

The biometric encryption system, by leveraging unique biometric data, could reduce the risk of identity theft and fraud. However, considerations for privacy, including data protection laws, are important when deploying such systems in real-world applications.

The use of quantization-based techniques allowed for error-tolerant biometric key generation. This approach not only catered to minor variations in biometric data (e.g., noise, illumination) but also ensured that keys were robust and non-reversible, a critical factor for cryptographic security. The derived 256-bit AES keys meet modern cryptographic standards, ensuring resistance to brute-force attacks.

Encryption and Privacy Protection:

The system ensures privacy by eliminating the need to store raw biometric data, instead relying on dynamic biometric key generation.

Accuracy and Efficiency:

The biometric encryption system, by leveraging unique biometric data, could reduce the risk of identity theft and fraud. However, considerations for privacy, including data protection laws, are important when deploying such systems in real-world applications

The combination of CNN-based iris feature extraction and PCA-based face data processing provides a highly efficient multimodal biometric system. The high recognition accuracy for both modalities ensures the extracted features are distinct and suitable for cryptographic key generation. Despite handling high-dimensional data, the system exhibits efficient processing for feature extraction, key generation, and encryption, making it scalable for real-world applications.

AES-GCM for Data Security:

AES-GCM demonstrated strong resistance to tampering through its authentication tag mechanism, making it suitable for sensitive applications.

The encryption's efficiency ensures scalability for real-world use cases like identity verification, secure data transmission, and access control.

Limitations and Future Enhancements:

- While PCA effectively reduced dimensions for face features, its performance in extreme conditions (e.g., occluded or low-resolution faces) could be improved through advanced neural architectures.
- Expanding datasets to include more demographics and diverse environmental conditions would enhance model generalizability.

The system provides a secure and efficient framework for multimodal biometric encryption, achieving high accuracy, robustness, and encryption integrity. By leveraging modern cryptographic and machine learning techniques, it ensures the protection of sensitive information in diverse real-world scenarios.

CHAPTER-10

CONCLUSION

This project successfully combined biometric features (iris and facial attributes) to generate binary keys for AES encryption. By leveraging pre-trained CNNs (VGG16), PCA for dimensionality reduction, and quantization techniques, a secure and efficient pipeline was established for biometric-based cryptographic applications. The Streamlit interface provided a user-friendly platform to interact with the encryption system, demonstrating real-world usability. Future improvements could involve enhancing feature extraction techniques, experimenting with other encryption standards, and integrating multi-modal biometrics for enhanced security. The system demonstrated that biometric features can serve as reliable cryptographic keys, ensuring enhanced security without relying on traditional passwords, which are prone to compromise. By successfully combining features from multiple sources (iris images and facial attributes), the project showcased the potential for multi-modal biometric systems to improve security and performance.

The integration of a user-friendly Streamlit interface highlighted the system's practicality, enabling secure encryption and decryption processes in a way that could be adapted for real-world applications like secure data storage or authentication systems. The modular design of the workflow allows for easy incorporation of additional biometric modalities, new encryption techniques, or enhancements to the feature extraction pipeline. This project lays the groundwork for exploring improved feature quantization strategies, optimizing key entropy, and enhancing the overall robustness of biometric-based encryption systems. This project showcases the potential for integrating biometrics and cryptography to create robust systems applicable in diverse fields, including healthcare, banking, and secure communication. The system is designed to scale well, with flexible feature extraction, quantization, and encryption pipelines, making it suitable for future large-scale deployments in biometric-based security systems.

The use of biometric keys for AES encryption advances the field of biometric cryptography, pushing the boundaries of how personal data can be securely encrypted without relying on traditional password-based systems. Incorporating additional biometric data such as fingerprints, voice, or retina scans into the encryption pipeline could further strengthen security by creating more robust, multi-modal biometric authentication systems

REFERENCES

- [1] Hooda, Susheela & Shrivastav, Supriya & Sharma, Preeti. (2023). **A Study on Biometrics and Machine Learning**. 1-5. <https://ieeexplore.ieee.org/document/10368885>
- [2] S. Nagaraj, R. Nagendra, Shanmugham Balasundaram, R. Kiran Kumar (2023). **Biometric key generation and multi round AES crypto system for improved security**. <https://www-sciencedirect-com-presiuniv.knimbus.com/science/article/pii/S2665917423002672>
- [3] Ramisetty, Srividya & B., Ramesh. (2019). **Implementation of AES using biometric**. International Journal of Electrical and Computer Engineering (IJECE). <http://doi.org/10.11591/ijece.v9i5.pp4266-4276>
- [4] S. Pooja, C. V. Arjun and S. Chethan, "Symmetric key generation with multimodal biometrics: A survey," 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, India, 2016, pp. 1-5. <https://ieeexplore.ieee.org/document/8053273>
- [5] Praveen, s & Vellela, Sai Srinivas & Ramachandran, Balamanigandan. (2024). **SmartIris ML: Harnessing Machine Learning for Enhanced Multi-Biometric Authentication**. https://www.researchgate.net/publication/378439449_SmartIris_ML_Harnessing_Machine_Learning_for_Enhanced_Multi-Biometric_Authentication
- [6] Rachana Veerabommala, Greeshma Arya, 2022, **Design And Implementation of AES Algorithm with Biometric Key Schedule to Improve Security**, (IJERT) Volume 11, Issue 06 (June 2022) <https://www.ijert.org/design-and-implementation-of-aes-algorithm-with-biometric-key-schedule-to-improve-security>
- [7] W. Wei and Z. Jun, **Image encryption algorithm Based on the key extracted from iris characteristics**, 2013 IEEE 14th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary, 2013, pp. 169-172. <https://ieeexplore.ieee.org/document/6705185>

- [8] Ghilom, Milkias & Latifi, Shahram. (2024). **The Role of Machine Learning in Advanced Biometric Systems**. Electronics. 13. 2667. <https://doi.org/10.3390/electronics13132667>
- [9] Amir Anees, Yi-Ping Phoebe Chen. **Discriminative binary feature learning and quantization in biometric key generation**, Pattern Recognition, Volume 77, 2018, Pages 289-305, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2017.11.018>
- [10] Yanzhi Chen, Yan Wo, Renjie Xie, Chudan Wu, Guoqiang Han. **Deep Secure Quantization: On secure biometric hashing against similarity-based attacks**, Signal Processing, Volume 154, 2019, Pages 314-323, ISSN 0165-1684, <https://doi.org/10.1016/j.sigpro.2018.09.013>

APPENDIX-A

PSUEDOCODE

Pseudocode for precompute_features.py

Purpose: Preprocess datasets, extract features using CNN, apply PCA for dimensionality reduction, and save the results for further use.

Import Necessary Libraries

1. Import libraries for:
 - Image processing (cv2, numpy, etc.).
 - Deep learning models (Keras VGG16).
 - Data preprocessing and dimensionality reduction (PCA, StandardScaler).
 - File handling (os, pickle).

Set Paths for Input Data

1. Define constant IRIS_IMAGES_PATH for iris image dataset directory.
2. Define constant FACE_CSV_PATH for face attributes CSV file.

Define Helper Methods

1. LoadIrisData
 - Accepts dataPath and imageSize as input parameters.
 - Iterates through folders and files in the given dataset directory.
 - Reads grayscale images of the iris.
 - Resizes images to the specified size.
 - Normalizes pixel values to a range of [0, 1].
 - Returns the preprocessed images as an array.
2. PreprocessForCnn
 - Accepts iris images and targetSize as input.
 - Converts each grayscale image to RGB format.
 - Resizes each image to a CNN-compatible size.
 - Logs the number of preprocessed images and their dimensions.
 - Returns the preprocessed images array.
3. ExtractFeatures
 - Accepts a model and a batch of images as input.

- Uses the model to predict features for the given images.
 - Flattens the features into a 2D array.
 - Logs the shape of the extracted features.
 - Returns the extracted features.
4. PreprocessFaceCsv
 - Accepts the path to the face attributes CSV as input.
 - Reads the CSV into a DataFrame.
 - Converts all categorical data to numerical, replacing -1 with 0.
 - Returns the processed DataFrame.
 5. ApplyPcaToFaceData
 - Accepts face data and varianceThreshold as input.
 - Standardizes the data.
 - Applies PCA to reduce dimensions, retaining the specified variance.
 - Logs the total retained variance and the number of components used.
 - Returns the PCA-transformed features, explained variance ratios, and the number of components.
 6. CombineFeatures
 - Accepts irisFeatures, faceFeatures, and numSamples.
 - Randomly samples a subset of face features to match the number of iris features.
 - Combines the features horizontally.
 - Logs the shape of the combined features.
 - Returns the combined features array.

Process Workflow

1. Load and Process Iris Images
 - Call LoadIrisData with the iris dataset path.
 - Call PreprocessForCnn on the loaded iris images.
2. Extract Iris Features
 - Initialize a VGG16 model pre-trained on ImageNet.
 - Call ExtractFeatures with the VGG16 model and preprocessed iris images.
3. Process Face Attributes
 - Call PreprocessFaceCsv with the face CSV file path.

- Call ApplyPcaToFaceData with the processed face data and a variance threshold of 0.95.
- 4. Combine Features
 - Call CombineFeatures with iris and face features.
- 5. Save Results
 - Save the PCA explained variance and combined features as .pkl files using pickle.
- 6. Log Completion
 - Print messages confirming successful execution.

Pseudocode for main.py

Purpose: Provide an interactive Streamlit app for demonstrating biometric-based AES encryption, file encryption, and data visualization.

Import Necessary Libraries

1. Import libraries for:
 - Data handling (pandas, numpy).
 - Visualization (matplotlib, seaborn).
 - Encryption (Cryptodome. Cipher.AES).
 - Streamlit functionality.

Set Paths for Input Data

1. Define constant paths for:
 - IRIS_IMAGES_PATH for iris dataset.
 - FACE_CSV_PATH for face attributes CSV.

Define Helper Methods

1. LoadPrecomputedFeatures
 - Reads and returns the combined features from a .pkl file.
2. LoadExplainedVariance
 - Reads and returns the explained variance ratios from a .pkl file.
3. LoadFaceAttributes
 - Reads and preprocesses the face attributes CSV:
 - Converts categorical attributes to numerical by replacing -1 with 0.

4. LoadRandomIrisImages
 - Reads random iris images from the dataset.
 - Resizes them to the specified size.
 - Returns a list of image arrays.
5. QuantizationKeyExtraction
 - Normalizes biometric features to [0, 1].
 - Divides the range into bins.
 - Converts features into binary keys based on bin indices.
6. BinaryToBytes
 - Truncates a binary string to the required key length.
 - Converts the truncated string into a byte object.
7. AesEncryptDecrypt
 - Performs AES encryption or decryption using the specified mode (encrypt or decrypt).
 - For encryption: Encrypts the text and generates a tag and nonce.
 - For decryption: Decrypts the text and verifies its authenticity.

Streamlit Application Layout

1. **Main Application**
 - Display a title and description for the app.
 - Provide a sidebar with options:
 - Generate Biometric Key Dynamically.
 - Encrypt a File.
 - Visualize Data.
2. **Option: Generate Biometric Key**
 - Load precomputed features.
 - Extract binary keys using QuantizationKeyExtraction.
 - Display the first generated key.
 - Allow the user to input text and encrypt it using the binary key.
 - Allow the user to decrypt the encrypted text.
3. **Option: Encrypt a File**
 - Allow the user to upload a file.
 - Encrypt the file using the biometric key.

- Provide an option to download the encrypted file.

4. Option: Visualize Data

- Provide options to visualize:
 - Raw iris images.
 - Face attribute statistics.
 - PCA explained variance.
 - CNN feature heatmap.

5. Visualization

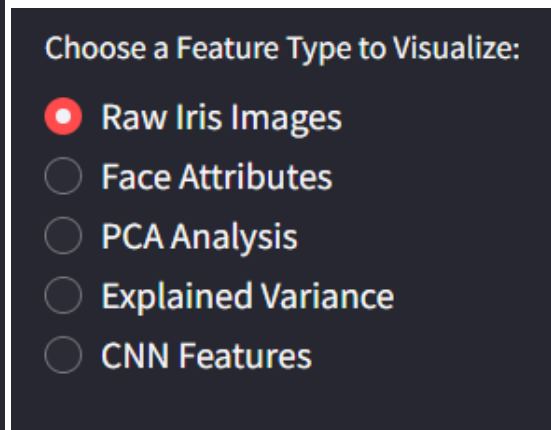
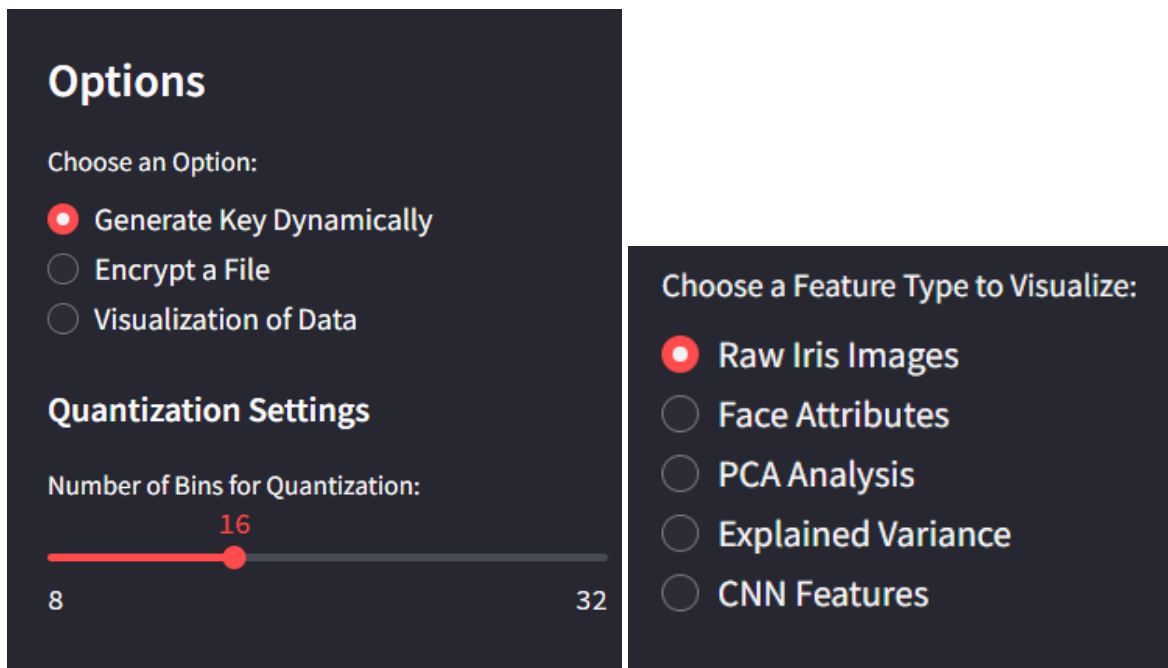
- Dynamically generate plots and display them based on the selected option.

6. Footer


- Display footer text with copyright information.

APPENDIX-B

SCREENSHOTS



Biometric Key-Based AES Encryption

Encrypt Your Data Using AES-GCM and Biometric Key 

Biometric Feature Selection

Generate Biometric Features

Biometric Features Generated Dynamically!

Biometric Key Ready for Encryption!

Encrypt Your Data

Enter Data to Encrypt

Sensitive Information

Encrypt Data

Data Encrypted Successfully!

Encrypted Data (Hex): 3a8f0f0b6850bb3c220813cd32674ab2fbb043b477

Decrypt Your Data

Decrypt Data

Data Decrypted Successfully!

Decrypted Data: Sensitive Information

A Streamlit app to demonstrate AES encryption using biometric keys.

File Encryption

Upload a File to Encrypt



Drag and drop file here

Limit 200MB per file • TXT, CSV, JSON

Browse files

© 2024 Karen Rena C. All rights reserved.



Data Visualization for Biometric Features

Explore raw and processed biometric features from iris and face datasets.

Randomly Selected Iris Images



Iris 1



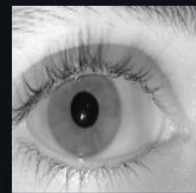
Iris 2



Iris 3



Iris 4



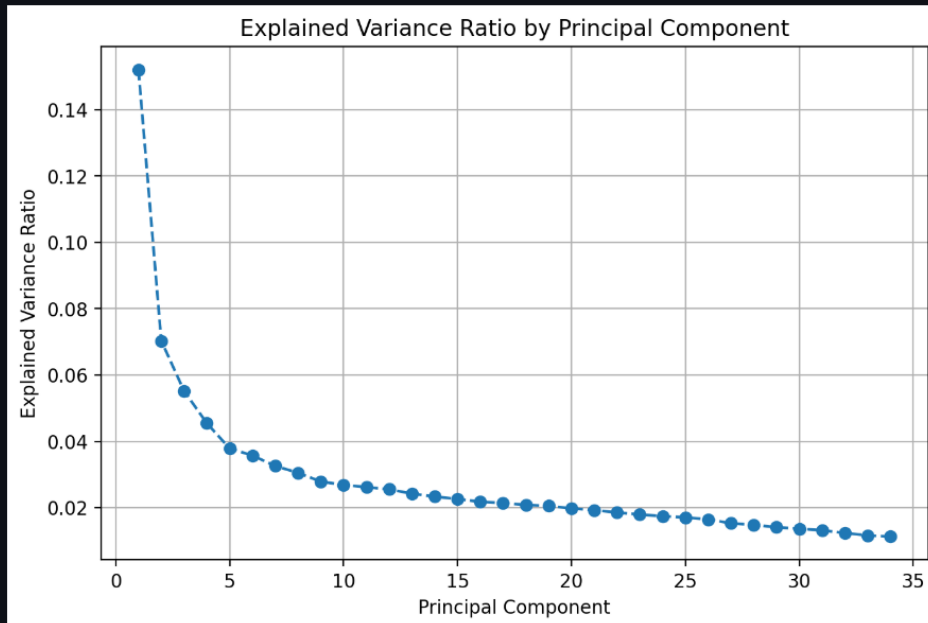
Iris 5



Data Visualization for Biometric Features

Explore raw and processed biometric features from iris and face datasets.

Scree Plot of PCA Explained Variance



Data Visualization for Biometric Features

Explore raw and processed biometric features from iris and face datasets.

Heatmap of Extracted CNN Features

Number of Samples to Visualize:

10

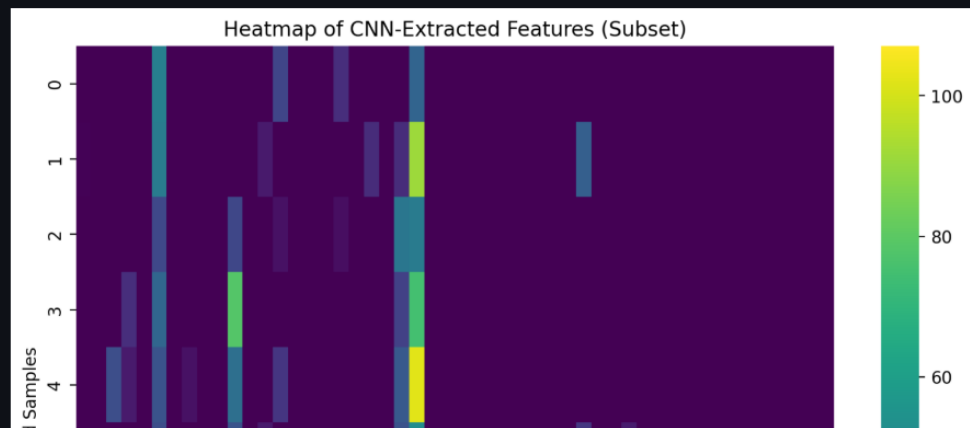
1

450

Number of Features to Visualize:

1

25098



APPENDIX-C

ENCLOSURES





International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal Since 2013)



CERTIFICATE OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

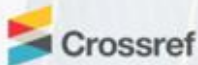
PAYAMAN S SURAJ

**UG Student, Dept. of Computer Science and Engineering, Presidency
University, Bengaluru, Karnataka, India**

In Recognition of Publication of the Paper Entitled

**“Encryption of Biometric Traits for Privacy Attacks using
AES Encryption”**

in IJIRCCE, Volume 13, Issue 1, January 2025



e-ISSN: 2320-9801
p-ISSN: 2320-9798



www.ijircce.com ijircce@gmail.com

Himansu_Sekhar_Rout_report_check_for_similarity

ORIGINALITY REPORT

11 %	8 %	8 %	5 %
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Presidency University Student Paper	5 %
2	Susheela Hooda, Supriya Shrivastav, Preeti Sharma. "A Study on Biometrics and Machine Learning", 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), 2023 Publication	1 %
3	S. Nagaraju, R. Nagendra, Shanmugham Balasundaram, R. Kiran Kumar. "Biometric key generation and multi round AES crypto system for improved security", Measurement: Sensors, 2023 Publication	1 %
4	www.researchgate.net Internet Source	1 %
5	www.ijert.org Internet Source	<1 %

Mapping project with the Sustainable Development Goals (SDGs).



The project worked carried out here is mapped to the following SDGs:

Goal 9: Industry, Innovation, and Infrastructure

The project introduces a novel approach by combining biometric authentication with advanced cryptographic techniques (e.g., AES encryption) and machine learning for feature extraction. This kind of innovation strengthens secure digital infrastructures, which are critical in modern industries such as banking, healthcare, and transportation.

Biometric systems are at the heart of smart technologies that underpin innovative industries, especially in areas like identity management, secure communication, and fraud prevention. By addressing challenges like privacy risks and computational overhead, this project helps improve the efficiency and resilience of infrastructure that supports industries. For example, biometric-based secure access systems can be implemented in critical infrastructures like airports and government facilities.

Goal 11: Sustainable Cities and Communities

Biometric systems secured with this project's approach can play a key role in making cities safer by enabling secure access control in public spaces, transportation hubs, and private

infrastructure. By ensuring privacy-preserving authentication, the system can facilitate smart city initiatives, where secure identity verification is essential for digital governance, public service delivery, and community safety. Multimodal biometrics (iris and facial features) as used in this project can be employed in smart community solutions, such as secure voting systems, public resource management, and crime prevention, ensuring more resilient and inclusive communities.

Goal 16: Peace, Justice, and Strong Institutions

The project directly addresses the privacy risks associated with biometric systems, which are increasingly being used in governance (e.g., voter identification systems) and justice systems (e.g., forensic identification). The use of robust encryption to protect biometric data ensures that sensitive personal information is not misused or exploited, contributing to accountability and trust in institutions. Biometric-based systems secured through encryption help in reducing fraud and identity theft, which are critical for ensuring justice and fair access to resources (e.g., welfare schemes, subsidies).

Goal 17: Partnerships for the Goals

Biometric security is a critical area where cross-sector partnerships (governments, private companies, and academic institutions) are necessary to develop and deploy innovative solutions. This project demonstrates the potential to foster such partnerships by offering a robust, scalable framework. The project's application in real-world contexts (e.g., secure data transmission, access control systems) can lead to collaborations with tech companies, security agencies, and public institutions to advance digital infrastructure and global security initiatives. By emphasizing privacy and data security, your project contributes to building trust in digital solutions, a key enabler for global partnerships. For example, this technology could be shared between countries to build a more interconnected and secure digital ecosystem for identity management.

ENCRYPTION OF BIOMETRICS TRAITS FOR PRIVACY ATTACKS USING AES ENCRPYTION

A PROJECT REPORT

Submitted by,

Ms. Karen Rena C -20211CSD0169

Ms. Samprity Singha -20211CSD0044

Mr. Pavaman S Suraj -20211CSD0126

Under the guidance of,

Prof. Himansu Sekhar Rout

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

At



PRESIDENCY UNIVERSITY

BENGALURU

DECEMBER 2024

PRESIDENCY UNIVERSITY

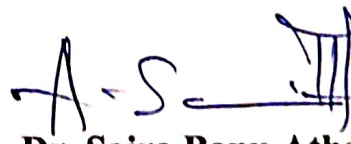
SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report “Encryption of Biometrics Traits for Privacy Attacks using AES Encryption” being submitted by “Karen Rena C, Samprity Singha, Pavaman S Suraj” bearing roll number(s) “20211CSD0169, 20211CSD0044, 20211CSD0126” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a Bonafide work carried out under my supervision.

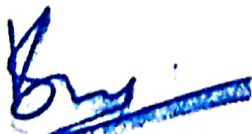


Prof. Himansu Sekhar Rout
Assistant Professor
School of IS
Presidency University



Dr. Saira Banu Atham
Professor & HoD
School of CSE&IS
Presidency University

20 Jan 2021



Dr. L. SHAKKEERA
Associate Dean
School of CSE
Presidency University



Dr. MYDHILI NAIR
Associate Dean
School of CSE
Presidency University



Dr. SAMEERUDDIN KHAN
Pro-VC School of Engineering
Dean -School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **Encryption of Biometric Traits for Privacy Attacks using AES Encryption** in partial fulfillment for the award of Degree of **Bachelor of Technology in Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **Prof. Himansu Sekhar Rout, Assistant Professor, School of Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.



Karen Rena C

20211CSD0169



Samprity Singha

20211CSD0044



Pavaman S Suraj

20211CSD0126

ABSTRACT

Biometric systems have become essential for secure authentication, utilizing unique traits such as fingerprints, iris patterns, and facial features. However, privacy concerns and vulnerabilities to attacks necessitate advanced methods for protecting biometric data. This project proposes a robust framework for encrypting biometric traits, combining multimodal biometrics, machine learning (ML), and the Advanced Encryption Standard (AES) to ensure both data confidentiality and authentication.

The system integrates iris and face biometrics, generating cryptographic keys through ML-driven feature extraction techniques from Pre-trained CNN models like VGG. These features are used to create a biometric key using Quantization which utilizes 16 bins, then converted to 256 byte key used in AES encryption for securing image data. Optimized AES implementations, including non-linear S-Box designs and Galois/Counter Mode, further enhance security and performance. Multimodal biometrics improve accuracy and resilience against spoofing attacks, addressing limitations of unimodal systems.

Research demonstrates that biometric-based key generation ensures unique and secure cryptographic keys while eliminating the need for traditional passwords. Machine learning enhances feature extraction and multimodal fusion, achieving high recognition accuracy. Combined with AES, this approach provides efficient, robust encryption resistant to brute force and spoofing attacks.

This framework addresses critical challenges like template instability, privacy risks, and computational overhead, making it ideal for real-world applications such as secure identity verification, data transmission, and access control. The integration of multimodal biometrics, ML, and AES represents a transformative step towards scalable, secure, and privacy-preserving authentication systems for the digital age.

ENCRYPTION OF BIOMETRICS TRAITS FOR PRIVACY ATTACKS USING AES ENCRPYTION

A PROJECT REPORT

Submitted by,

**Ms. Karen Rena C - 20211CSD0169
Ms. Samprity Singha - 20211CSD0044
Mr. Pavaman S Suraj - 20211CSD0126**

Under the guidance of,

Prof. Himansu Sekhar Rout

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

At



PRESIDENCY UNIVERSITY

BENGALURU

DECEMBER 2024

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report “**Encryption of Biometrics Traits for Privacy Attacks**” being submitted by “Karen Rena C, Samprity Singha, Pavaman S Suraj” bearing roll number(s) “20211CSD0169, 20211CSD0044, 20211CSD0126” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a Bonafide work carried out under my supervision.

Prof. Himansu Sekhar Rout
Assistant Professor
School of IS
Presidency University

Dr. Saira Banu Atham
Professor & HoD
School of CSE
Presidency University

Dr. L. SHAKKEERA
Associate Dean
School of CSE
Presidency University

Dr. MYDHILI NAIR
Associate Dean
School of CSE
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-VC School of Engineering
Dean -School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **Encryption of Biometric Traits for Privacy Attacks** in partial fulfillment for the award of Degree of **Bachelor of Technology in Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **Prof. Himansu Sekhar Rout, Assistant Professor, School of Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Student Name(s)	Roll No.(s)	Signature(s)
Karen Rena C	20211CSD0169	
Samprity Singha	20211CSD0044	
Pavaman S Suraj	20211CSD0126	

ABSTRACT

Biometric systems have become essential for secure authentication, utilizing unique traits such as fingerprints, iris patterns, and facial features. However, privacy concerns and vulnerabilities to attacks necessitate advanced methods for protecting biometric data. This project proposes a robust framework for encrypting biometric traits, combining multimodal biometrics, machine learning (ML), and the Advanced Encryption Standard (AES) to ensure both data confidentiality and authentication.

The system integrates iris and face biometrics, generating cryptographic keys through ML-driven feature extraction techniques from Pre-trained CNN models like VGG. These features are used to create a biometric key using Quantization which utilizes 16 bins, then converted to 256-byte key used in AES encryption for securing image data. Optimized AES implementations, including non-linear S-Box designs and Galois/Counter Mode, further enhance security and performance. Multimodal biometrics improve accuracy and resilience against spoofing attacks, addressing limitations of unimodal systems.

Research demonstrates that biometric-based key generation ensures unique and secure cryptographic keys while eliminating the need for traditional passwords. Machine learning enhances feature extraction and multimodal fusion, achieving high recognition accuracy. Combined with AES, this approach provides efficient, robust encryption resistant to brute force and spoofing attacks.

This framework addresses critical challenges like template instability, privacy risks, and computational overhead, making it ideal for real-world applications such as secure identity verification, data transmission, and access control. The integration of multimodal biometrics, ML, and AES represents a transformative step towards scalable, secure, and privacy-preserving authentication systems for the digital age.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair**, School of Computer Science Engineering & Information Science, Presidency University, and **Dr. Saira Banu**, Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Prof. Himansu Sekhar Rout**, Assistant Professor and Reviewer **Prof. Sandhya L**, School of Computer Science Engineering & Information Science, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar A and Mr. Md Zia Ur Rahman**, department Project Coordinators **Dr Manjula H M** and Git hub coordinator **Mr. Muthuraj**.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Karen Rena C
Samprity Singha
Pavaman S Suraj

LIST OF FIGURES

Sl. No.	Figure Name	Caption	Page No.
1	Figure 6.1	Normalization of Iris Image	22
2	Figure 6.2	PCA for Face dataset	23
3	Figure 6.3	Feature Extraction using VGG16	24
4	Figure 6.4	Feature Fusion	24
5	Figure 6.5	Quantization Key Generation	25
6	Figure 6.6	Data Encryption	25
7	Figure 7.1	Gantt Chart	29
8	Figure 9.1	Preprocessed Images Shape	30
9	Figure 9.2	Heatmap of Extracted Features	30
10	Figure 9.3	Cumulative Explained Variance	30
11	Figure 9.4	Explained Variance Ratio	31
12	Figure 9.5	Combine Features Shape	31

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	iv
	ACKNOWLEDGMENT	v

1.	INTRODUCTION	1
	1.1 GENERAL	1
	1.2 Biometric Systems and Cryptography	1
	1.3 Cryptographic Techniques in Biometrics	2
	1.4 Machine Learning in Biometrics	3
	1.5 Multimodal Biometrics	4
2.	LITERATURE SURVEY	6
3.	RESEARCH GAPS OF EXISTING METHODS	9
4.	PROPOSED METHODOLOGY	11
	4.1 Data Acquisition and Preprocessing	11
	4.2 Feature Extraction Using Machine Learning	12
	4.3. Multimodal Feature Fusion	13
	4.4. Biometric Key Generation	14
	4.5. Data Encryption Using AES-GCM	14
	4.6. User Interface for Feature Upload or Generation	15
	4.7. Security and Performance Evaluation	16
5.	OBJECTIVES	17
6.	SYSTEM DESIGN AND IMPLEMENTATION	19
	6.1 System Overview	19

6.2	Techniques used	19
6.3	System Implementation	20
7.	TIMELINE OF THE PROJECT	25
8.	OUTCOMES	26
9.	RESULTS AND DISCUSSIONS	29
10.	CONCLUSION	34
11.	REFERENCES	36
12.	APPENDIX – A Pseudocode	38
13.	APPENDIX – B Screenshots	43
14.	APPENDIX – C Enclosures	47

CHAPTER-1

INTRODUCTION

1.1 General

The increasing reliance on biometric systems for authentication raises critical concerns about protecting sensitive data against privacy attacks. This project, titled "Encryption of Biometric Traits to Avoid Privacy Attacks," addresses these challenges by integrating biometrics with cryptographic techniques to enhance data security. By leveraging multimodal biometric features, such as iris and face traits, and employing machine learning for feature extraction, the project proposes an innovative framework for generating robust biometric keys. These keys are utilized to encrypt sensitive data using the Advanced Encryption Standard (AES), a symmetric encryption algorithm renowned for its efficiency and adaptability to image data.

The framework combines authentication and data confidentiality, ensuring a robust bio-cryptosystem that is both secure and scalable for real-world applications. The following sections provide an in-depth exploration of the topics and subtopics integral to this research.

1.2 Biometric Systems and Cryptography

Biometric systems use distinctive physical or behavioral characteristics, such as voice, iris patterns, fingerprints, or face features, to verify people. These traits are inherently distinctive and serve as reliable identifiers for secure authentication. By integrating biometrics with cryptographic techniques, the systems gain an enhanced ability to safeguard sensitive data from unauthorized access, creating a robust layer of security that balances privacy and convenience.

The concept of bio-cryptosystems takes this integration a step further by combining biometric authentication with encryption, resulting in a dual-layer security framework. In these systems, biometric data can either be used to generate cryptographic keys or to secure templates through advanced cipher transformations. This innovative approach not only ensures the protection of sensitive information but also provides seamless authentication, thereby offering a strong guarantee of data integrity and user privacy.

Despite their advantages, biometric systems face significant challenges that can impact their effectiveness and reliability. Factors such as environmental noise, changes in user input due to rotation or scaling, and other distortions can introduce variability in the biometric data. This variability destabilizes biometric templates, reducing the system's overall accuracy. Although

frequent updates to templates can help mitigate these issues, they also introduce additional computational demands, making them less suitable for real-time applications where speed and efficiency are critical.

By addressing these challenges through ongoing advancements in biometric and cryptographic technologies, researchers aim to improve the stability and performance of biometric systems, ensuring their reliability and scalability for diverse real-world applications.

1.3 Cryptographic Techniques in Biometrics

The Advanced Encryption Standard (AES) is a widely recognized encryption algorithm valued for its computational efficiency and robust security. In the context of this project, AES is employed to encrypt sensitive data using biometric keys derived from unique iris and facial features. This ensures a high level of confidentiality and protection against unauthorized access. AES is particularly effective for securing image data, making it a natural fit for biometric systems that rely on visual and pattern-based information.

A critical component of AES is the Substitution Box (S-Box), which plays a pivotal role in safeguarding encrypted data. The S-Box introduces non-linearity into the encryption process, enhancing resistance to cryptographic attacks. Traditional S-Box designs often use techniques like lookup tables or finite field arithmetic, each presenting trade-offs in terms of speed, resource usage, and security. Optimizing S-Box designs can improve their non-linearity, boosting resistance to attacks while maintaining the algorithm's performance.

To further enhance security and efficiency, AES is implemented in Galois/Counter Mode (AES-GCM), which provides simultaneous encryption and authentication. GCM mode improves AES performance by enabling parallel processing, making it particularly suitable for biometric systems where both speed and data integrity are crucial. By combining optimized S-Box designs with the GCM mode, the encryption system achieves a balance of enhanced security, efficiency, and resilience against side-channel attacks.

Biometric key-based encryption represents an innovative solution to the limitations of traditional cryptographic systems. Unlike conventional methods that rely on static passwords or keys, biometric encryption generates cryptographic keys from unique biometric traits such as iris patterns or facial data. Advanced techniques like quantization-based key generation and bio-hashing ensure that these biometric keys are consistent, secure, and resistant to errors during biometric data acquisition.

Quantization-based key generation processes continuous biometric data into stable, discrete

cryptographic keys. This is achieved by mapping biometric feature vectors to quantized bins, ensuring repeatability and reducing variability caused by environmental or input noise. Bio-hashing further strengthens this approach by combining biometric features with a tokenized random seed to produce secure, hash-based keys. This method provides robust protection against key inversion and replay attacks, safeguarding the integrity and confidentiality of the encryption process.

By leveraging biometric features for key generation, this approach eliminates risks associated with key theft or loss while preserving the unique and secure nature of biometric systems. The integration of AES, optimized S-Box designs, GCM mode, and biometric key-based encryption creates a highly secure and efficient framework tailored for modern biometric systems, ensuring confidentiality, integrity, and resilience in real-world applications.

1.4. Machine Learning in Biometrics

Machine learning (ML) has transformed the field of biometrics, enabling the development of intelligent, adaptive, and highly accurate authentication systems. ML algorithms, including support vector machines (SVMs), random forests (RFs), and deep learning (DL) models, play a crucial role in optimizing various processes such as feature extraction, anomaly detection, and data fusion. These features greatly improve biometric systems' accuracy and dependability.

Deep learning, a subset of ML, has been particularly impactful in the field of feature extraction. Deep learning-based systems dynamically learn complex and discriminative feature representations, which are critical for reducing False Match Rates (FMR) and improving overall system security. Convolutional Neural Networks (CNNs), in particular, have proven highly effective in extracting intricate patterns from biometric data, such as iris and facial images. These extracted features are essential for precise biometric key generation, enabling robust and unique encryption.

Within deep learning frameworks, specific CNN architectures such as VGG (Visual Geometry Group) and ResNet (Residual Networks) have been employed to enhance feature extraction in biometric systems. Deep convolutional layers with tiny 3x3 filters are used by the VGG network to learn feature spatial hierarchies. When applied to iris images, VGG captures fine-grained textural details, including crypts, furrows, and speckle patterns. These details are crucial for generating reliable and unique biometric keys. ResNet, on the other hand, addresses the challenges of training deeper networks by introducing skip connections that mitigate

vanishing gradient problems. This architecture efficiently extracts high-level patterns from iris images while remaining robust to variations such as illumination changes, occlusions, and slight misalignments. By leveraging CNN-based architectures like VGG and ResNet, deep learning models enhance the precision, consistency, and robustness of feature extraction, ultimately improving the security and performance of biometric systems.

Machine learning also facilitates the application of biometric systems in various innovative ways. For instance, random forests and neural networks are used in behavioral biometrics to assess phone movement patterns and touch dynamics, resulting in excellent user identification accuracy. Multimodal fusion, which combines many biometric modalities including face and iris data, is also made possible by ML algorithms. This fusion ensures robust and reliable authentication by leveraging the strengths of multiple biometric inputs, further enhancing system resilience and accuracy.

By integrating advanced ML techniques and deep learning architectures, biometric systems achieve higher levels of adaptability, precision, and security, making them indispensable for modern authentication processes.

1.5 Multimodal Biometrics

Multimodal biometric systems integrate multiple biometric traits, such as iris and facial features, to enhance their robustness and reliability. By combining multiple modalities, these systems improve resistance to spoofing and environmental variability, addressing challenges that single-modal systems often face. For example, integrating iris and facial recognition ensures greater robustness against external noise, environmental conditions, and spoofing attempts, making the system more secure and versatile.

Fusion strategies play a crucial role in enhancing the performance of multimodal systems. Techniques such as feature-level, score-level, and decision-level fusion enable the integration of data from different biometric sources to improve recognition accuracy. For instance, combining near-infrared (NIR) and visible light iris images allows the system to perform accurately in diverse environmental conditions, providing flexibility and reliability in real-world applications.

Machine learning algorithms play a pivotal role in optimizing multimodal systems. By learning patterns across multiple biometric data sources, ML enhances classification accuracy and ensures scalability in dynamic environments. The ability of ML to adapt to new data and improve performance over time makes it a key enabler for advanced multimodal biometric

systems, offering both precision and efficiency.

Despite their advantages, biometric systems face challenges that must be addressed to ensure their effectiveness and security. One major concern is the security of biometric data, as these systems must safeguard sensitive information against threats such as template instability, privacy breaches, and spoofing. Encryption offers a robust solution, protecting data from unauthorized access and maintaining user privacy.

Another significant challenge is achieving computational efficiency. Balancing the need for high security with low computational overhead is critical for the practical implementation of biometric systems. This project addresses this challenge by leveraging machine learning to optimize biometric feature extraction and key generation processes. By streamlining these operations, the system achieves low computational overhead without compromising on security, ensuring a seamless and efficient user experience.

CHAPTER-2

LITERATURE SURVEY

The intersection of biometrics, machine learning, and cryptography has revolutionized the field of secure authentication systems. These advancements aim to address the challenges of ensuring data confidentiality, improving system robustness, and mitigating risks associated with traditional key-based cryptographic methods. The use of multimodal biometric features, such as iris and face recognition, has emerged as a promising solution to enhance security, reduce false acceptance rates, and provide a user-friendly authentication experience.

Machine learning has become a cornerstone for modern biometric systems, enabling them to achieve unprecedented levels of accuracy and adaptability. As stated in [1], "Biometric systems are capable of achieving extremely high levels of performance and better intelligence when new machine learning algorithms and modern architectures are used." Machine learning not only facilitates the extraction of meaningful features from complex datasets but also enables these systems to adaptively improve over time by analysing large datasets, identifying patterns, and predicting future authentication attempts. For instance, "Machine learning algorithms analyse large amounts of data, identify patterns, and make predictions about future authentication attempts" [1]. This continuous learning process ensures that the system remains robust against evolving threats, enhancing both security and user experience.

The integration of cryptographic techniques, particularly the use of the Advanced Encryption Standard (AES), has further fortified biometric systems. AES's high efficiency and proven security make it an ideal choice for encrypting sensitive biometric data. In [2], the authors explain that "Cipher keys are extracted dynamically from the input biometric image, which can solve several numerous problems that arise in the cipher system." By dynamically generating keys from biometric inputs, such systems eliminate the need for users to remember and manage passwords or physical tokens, which are often prone to theft or misuse. This integration of biometric key generation and AES encryption is further enhanced by innovative implementations. For example, [6] proposes, "The proposed Bio-Metric 256-bit AES algorithm is highly utilized in terms of key management to enhance security," demonstrating how biometric systems can address traditional key management challenges while maintaining high security.

Multimodal biometrics, which leverage multiple biometric traits like iris and face, have

emerged as a robust solution for addressing the limitations of unimodal systems. In [5], the authors propose a "novel fusion strategy to optimally combine the strengths of individual iris modalities through score-level fusion, feature-level fusion, and decision-level fusion, enhancing system accuracy." This fusion approach ensures that the system remains resilient to spoofing and other forms of attack, as it relies on a combination of multiple independent traits. Similarly, [7] highlights the efficiency of using wavelet transforms in multimodal systems: "An encryption algorithm based on key from iris features and AES is proposed. On the premise of iris pictures preprocessing arrange, viable iris zone is decayed to three layers by 2D Haar wavelet." Such preprocessing steps permit the framework to extricate vigorous highlights indeed from loud or low-quality pictures, guaranteeing unwavering quality in real-world applications.

The combination of machine learning with multimodal biometrics has further strengthened the resilience of these systems against sophisticated attacks. According to [5], "Robust anti-spoofing measures are incorporated using machine learning-based techniques to detect and prevent fraudulent attempts, ensuring authenticity." These measures not only protect the system against spoofing but also enhance its ability to differentiate between genuine and fraudulent users. Additionally, the adoption of deep learning models, as discussed in [8], has enabled biometric systems to achieve higher accuracy: "Deep learning methods are capable of learning higher-level abstractions from data, leading to improved accuracy and robustness in biometric systems." By leveraging convolutional neural networks (CNNs) and recurrent neural networks (RNNs), these systems can process complex patterns in biometric data, such as iris textures or facial landmarks, to generate highly discriminative feature representations.

The advantages of these systems are numerous and multifaceted. The incorporation of biometric keys eliminates the need for traditional passwords, providing a seamless and secure user experience. As noted in [6], "By using this method, there is no need to remember and store the key to encrypt the data, providing more security than the existing system." Furthermore, the use of multimodal biometrics enhances reliability, as pointed out in [2]: "The proposed MKE (Modified Key Expansion) generated randomized key compound in lightweight models and outperformed existing lightweight ciphers in terms of throughput rate." This highlights how combining biometric traits leads to higher efficiency and stronger security.

However, these systems are not without limitations. One of the primary challenges lies in their computational complexity. The fusion of multiple modalities and the use of machine learning algorithms often require significant processing power, which can limit the system's scalability. [5] points out that "The complexity of fusion strategies at multiple levels (score, feature, and decision) may introduce computational overhead, requiring significant processing resources." Such resource-intensive operations may hinder real-time performance, particularly in environments with limited computational capabilities.

Another critical concern is the reliance on high-quality biometric data. The effectiveness of these systems can be significantly affected by variations in data quality, such as poor lighting or low resolution. According to [5] "Variations in the quality of input images (due to lighting, angle, or resolution) can affect the reliability of the generated key." This dependency poses a challenge for deploying these systems in diverse and uncontrolled environments. Moreover, while machine learning-based anti-spoofing measures are effective, they heavily rely on the diversity and quality of the training dataset. As noted in [5], "While anti-spoofing measures are robust, their success heavily depends on the dataset quality and diversity during training, which could lead to vulnerabilities in unseen attack scenarios."

Latency is another concern for real-time applications. The computational overhead associated with deep learning-based feature extraction and fusion can lead to delays, as described in [5]: "Real-time applications might face latency issues due to the extensive computations involved in deep learning-based feature extraction and fusion." Addressing these latency issues requires optimization techniques that balance accuracy with efficiency, ensuring that the system remains suitable for practical applications.

Despite these challenges, the field continues to evolve, with ongoing research focused on overcoming these limitations. The combination of AES encryption, machine learning, and multimodal biometric systems represents a significant step forward in achieving secure and user-friendly authentication. Future advancements should focus on optimizing these systems for real-time applications, enhancing their resilience to sophisticated attacks, and addressing privacy concerns associated with biometric data storage and transmission.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

Biometric systems have shown tremendous potential for securing sensitive information through advanced authentication and encryption mechanisms. However, despite the significant progress, several critical limitations in the current methods reveal gaps that need to be addressed to enhance the robustness, efficiency, and usability of these systems.

One major limitation lies in the privacy and security concerns associated with biometric data. As highlighted in [1], "The usage of biometric authentication leads to worries about data security and privacy as biometric data is so confidential and sensitive." Not at all like passwords or tokens, biometric characteristics are permanent; once compromised, they cannot be supplanted. This makes an irreversible hazard if the biometric layouts are spilled. Moreover, "Biometric systems are not a foolproof defence against fraud or identity theft. For instance, biometric sensors can be fooled by generating bogus mementos (spoofing)" [1]. This indicates a pressing need for enhanced anti-spoofing mechanisms and secure storage solutions for biometric data.

Another gap in the existing methods is the computational complexity and resource requirements. Many of the proposed systems, such as those leveraging advanced cryptographic algorithms, introduce significant computational overhead. For instance, as [3] points out, "Biometric processing, while increasing security, involves operations that consume more processing power and are not suitable for battery-dependent devices." Furthermore, "Multimodal biometric feature extraction extends security levels, but the complexity of the design increases rapidly and does not fit into decentralized wireless architecture"[3]. This highlights a need for lightweight, resource-efficient biometric systems that can operate effectively in low-power or resource-constrained environments, such as IoT devices.

Latency and real-time performance also remain critical challenges in biometric systems, especially those using multimodal data or deep learning techniques. As stated in [5], "The complexity of fusion strategies at multiple levels (score, feature, and decision) may introduce computational overhead, requiring significant processing resources." Similarly, "Real-time applications might face latency issues due to the extensive computations involved in deep learning-based feature extraction and fusion" [5]. These delays hinder the usability of

biometric systems in scenarios requiring quick responses, such as financial transactions or mobile applications, underscoring the need for real-time optimization in system design.

The dependency on high-quality data further exacerbates these limitations. Biometric systems often struggle with variability in image quality, lighting conditions, and environmental factors, as noted in [7]: "While the biometric key generation approach enhances security, it heavily depends on the quality of iris image capture, which could limit its effectiveness in varied lighting conditions or low-resolution images." This limitation is particularly problematic for systems deployed in uncontrolled settings, where capturing high-resolution biometric data is challenging. Robust feature extraction techniques that can handle noisy or incomplete data are therefore required to address this gap.

Another significant gap is the vulnerability of machine learning models used in biometric systems. As discussed in [8], "Attacks such as data poisoning, model inversion, and adversarial attacks can manipulate ML systems used in biometrics, leading to security vulnerabilities." Furthermore, "Deepfake technology, by creating synthetic media, can fool voice- or face-recognition-based biometric systems, enabling unauthorized access"[8]. These emerging threats indicate the need for robust defences against adversarial attacks and synthetic media manipulation, which are currently underexplored in the biometric security domain.

The trade-offs between security and efficiency in cryptographic systems also present a critical research gap. While advanced key generation techniques enhance security, they often increase system complexity. For instance, as [6] highlights, "The inclusion of dynamic sub-byte computation in MKE leads to path delay accumulation, which is a trade-off for improved security." Similarly, "Although 128 bits were added to the existing AES key to make it 256 bits for higher security, it increased complexity without proportionate efficiency gains" [6]. This indicates the need for cryptographic solutions that can balance security with operational efficiency, ensuring practical usability in real-world applications.

Finally, scalability and interoperability remain under-addressed in many systems. As noted in [5], "Scalability remains a challenge, as integrating and processing large datasets for diverse modalities increases the computational burden." Additionally, the interoperability of biometric systems across platforms and devices is often overlooked, limiting their deployment in diverse environments. Addressing these scalability and compatibility issues is crucial for broader adoption of biometric security systems.

CHAPTER-4

PROPOSED METHODOLOGY

The project "**Encryption of Biometric Traits to Avoid Privacy Attacks using AES Encryption**" focuses on developing a secure biometric-based cryptographic system. It integrates multimodal biometrics, machine learning techniques for feature extraction, and the AES-GCM encryption standard for secure data transmission and storage. The system uses a dynamic or user-uploaded biometric feature pipeline, allowing flexibility in real-world applications.

The proposed methodology follows these steps:

4.1. Data Acquisition and Preprocessing

a) Biometric Data Collection

Biometric data collection involves acquiring iris and facial biometric data from publicly available benchmark datasets to ensure a comprehensive and diverse dataset for analysis. Examples of such datasets include the MMU-Iris Database for iris images and the CelebA Dataset for facial images. The collected data encompasses a wide range of demographics, environmental conditions (such as lighting and occlusion), and varying image quality. This diversity is crucial for enhancing the robustness and reliability of the system, ensuring it performs effectively across different scenarios and user profiles.

b) Preprocessing

- Iris Images:**

Convert images to grayscale for consistency. Iris images often include color information that may not contribute significantly to discriminative feature extraction. Converting images to grayscale simplifies computations while retaining the structural and textural details unique to the iris.

Iris images often come in varying resolutions due to differences in acquisition devices and environmental conditions. Standardizing the resolution ensures uniformity for input into machine learning models (like VGG16). Resize iris images to a standard resolution (e.g., 128x128 pixels) to ensure uniformity.

Normalize pixel values to a range of **[0, 1]** to ensure compatibility with deep learning models, which often expect inputs in a standardized scale. Normalization helps improve convergence during training and ensures consistency.

- **Face Images:**

Facial images may vary in dimensions due to differences in datasets or acquisition methods. Resizing ensures uniform input size, making it compatible with subsequent feature extraction methods like **PCA**.

PCA requires the input data to be in a flattened vectorized format. Flattening the face image ensures that all pixels are represented as a single continuous vector.

4.2 Feature Extraction Using Machine Learning

The goal is to extract meaningful, compact, and unique features from the biometric images.

a) Iris Feature Extraction Using CNN (VGG16)

The iris feature extraction leverages a Convolutional Neural Network (CNN), specifically the VGG16 architecture. The process involves the following steps:

1. **Input Preparation:** Images are resized to 224x224 pixels to match the input dimensions required by the VGG16 architecture and it will be converted to RGB format
2. **Network Layers:** The VGG16 model processes the images through 13 convolutional layers, interspersed with pooling layers, and three fully connected layers. These layers capture fine-grained textural information such as crypts, furrows, and other patterns unique to the iris.
3. **Feature Embedding:** At the output of the final fully connected layer, a high-dimensional feature vector (embedding) is generated. This embedding contains the discriminative features of the iris images that capture unique patterns like crypts, furrows, and textures.
4. **Dimensional Refinement:** The high-dimensional embedding can optionally be processed using a lightweight dimensionality reduction technique to retain only the most important features for fusion and encryption.

The advantage of using **VGG16** is its **robustness** to slight variations such as lighting changes or minor occlusions in the iris image. Its deep architecture effectively learns hierarchical and discriminative representations of the iris structure. VGG16's deep architecture learns low-level to high-level patterns hierarchically, enabling it to capture fine-grained textural features such as crypts, furrows, and intricate structures within the iris.

b) Face Feature Extraction Using PCA

Principal Component Analysis (PCA) is applied to extract features from the pre-processed face images. The key steps include:

1. **Input Image Representation:** The preprocessed face images are reshaped into vectors for input into PCA.
2. **Covariance Computation:** Compute the covariance matrix to identify the directions (principal components) with the highest variance in the data.
3. **Feature Selection:** Select the top k principal components (based on eigenvalues) that explain the maximum variance in the data. This results in a reduced-dimensional representation of the face images.
4. **Robust Face Features:** The PCA outputs are compact feature vectors that efficiently represent essential facial traits, such as edge structures and prominent features.

By using PCA for face data, the dimensionality is significantly reduced while preserving crucial biometric information. This step improves computational efficiency during the subsequent fusion and encryption processes.

4.3. Multimodal Feature Fusion

To combine the iris and face biometrics effectively, feature-level fusion is performed. This ensures that both sources of information contribute to the overall system security and accuracy. Steps involved in feature fusion:

1. **Alignment:** Before fusion, ensure that both iris embeddings (from CNN) and face feature vectors (from PCA) are of compatible dimensions.

2. Concatenation: Horizontally concatenate the iris features with face features to form a comprehensive biometric feature template.
3. Randomization: Implement shuffling algorithms to randomize the positions of feature values in the fused vector. This adds an additional security layer, as it obfuscates the direct relation to the original biometric traits.

Multimodal feature fusion enhances the robustness and accuracy of the system, ensuring better resistance to spoofing attacks or individual modality failures.

4.4. Biometric Key Generation

The fused biometric features are used to generate cryptographic keys in a secure and cancellable manner. The key generation process is given below:

1. Quantization-Based Key Generation:

The fused feature vector is quantized into discrete levels to map continuous feature values into binary keys. For example, values in the fused vector are rounded to specific ranges, and these ranges are represented by binary digits (0 or 1).

2. Key Mapping:

A 256-bit cryptographic key is generated by systematically grouping and mapping the quantized values. Using techniques like group mapping, features are divided into subsets, and transformations are applied to further randomize the binary key output, ensuring robust security.

3. Cancellable Transformations:

Enable the **revocability** of keys. If a key is compromised, transformations can be applied to the original biometric features to generate a new key. This ensures that biometric traits can still be securely utilized without risking permanent compromise. The resulting **256-bit AES key** serves as the critical input for data encryption and decryption in the next stage.

4.5. Data Encryption Using AES-GCM

The generated biometric key is used as a 256-bit AES key for encryption. The encryption

process follows:

a) AES-GCM Implementation

The generated biometric key, a 256-bit key derived from feature quantization, serves as the foundation for the AES encryption process. Specifically, AES-GCM (Galois/Counter Mode) is implemented to provide a dual-layered approach of encryption and data integrity verification. The biometric key is utilized as the input to AES, ensuring that both textual data and biometric images are securely encrypted. A unique nonce, or random value, is generated for each encryption session to maintain uniqueness and prevent replay attacks.

During encryption, AES-GCM simultaneously encrypts the input data and generates an authentication tag to verify data integrity. This ensures that the encrypted data remains both confidential and tamper-proof. In the decryption process, the same biometric key and nonce are used to decrypt the data while verifying its integrity against the authentication tag. Any mismatch in the authentication tag would indicate tampering, providing an additional layer of security.

Significant benefits are provided by AES-GCM in biometric encryption systems. High-speed performance is ensured by its capacity to process data blocks in parallel, which is essential for real-time applications. Its integrated message authentication capability further ensures integrity by identifying any unauthorized changes made to the encrypted data. AES-GCM is a great option for systems needing strong security and efficiency because of its versatility, which enables it to handle both textual data and image encryption.

4.6. User Interface for Feature Upload or Generation

The system provides an intuitive Streamlit-based frontend for user interaction:

a) Feature Input Options:

Users can either upload pre-generated combined biometric features stored as .npy files or dynamically generate biometric features using the backend pipeline for processing.

b) User Controls:

Users have the ability to adjust the number of bins for quantization, allowing them to influence the precision of key generation. Additionally, they can input text data for encryption or decryption as part of the system's functionality.

c) Encryption/Decryption Outputs:

The system displays the encrypted data, nonce, and authentication tag in hexadecimal format, providing a clear view of the encryption results. For verification purposes, the decrypted data is also presented to ensure the encryption and decryption processes function correctly.

4.7. Security and Performance Evaluation

a) Experimental Setup

The system is evaluated using benchmark datasets like MMU-Iris for iris data and CelebA for facial data, ensuring variability across demographics such as age, gender, and ethnicity, as well as environmental conditions like lighting and image quality. Experiments are conducted to validate the robustness of the encryption process alongside the performance of biometric authentication, ensuring the system meets high standards of reliability and security.

b) Key Performance Metrics:

The performance of the system is assessed using several critical metrics. Recognition performance is measured through True Positive Rate (TPR) and False Positive Rate (FPR) to evaluate the accuracy of multimodal authentication. Encryption robustness is analyzed using metrics such as scrambling degree, which measures the randomness in encrypted images, and histogram analysis to ensure uniform data distribution, reducing susceptibility to statistical attacks. Authentication metrics such as False Match Rate (FMR), False Non-Match Rate (FNMR), and Genuine Acceptance Rate (GAR) provide insights into the system's reliability in identifying users accurately. Additionally, computational efficiency is evaluated by measuring encryption and decryption times, along with the computational overhead on standard hardware.

c) Real-World Application Integration:

The system is designed for real-world applications, including identity verification in sensitive areas like airports and banks, access control for authentication-based secure systems, and secure transmission to enable privacy-preserving cloud-based biometric data processing. To enhance scalability and robustness, the solution is implemented on cloud platforms, ensuring it can effectively handle large user populations while maintaining high levels of security and flexibility.

CHAPTER-5

OBJECTIVES

1. Develop a Multimodal Biometric System Using Iris and Face Traits for Robust Key Generation

This objective focuses on designing a multimodal biometric system that leverages the unique and complementary characteristics of iris and face biometrics to enhance the security and reliability of key generation processes. Multimodal systems, as highlighted in the literature, provide superior robustness compared to unimodal systems by reducing the likelihood of spoofing and improving recognition accuracy. For instance, the fusion of iris and face traits at various levels—such as score-level, feature-level, and decision-level fusion—ensures comprehensive feature representation, as proposed in [5]: "A novel fusion strategy is proposed to optimally combine the strengths of individual iris modalities through score-level fusion, feature-level fusion, and decision-level fusion, enhancing system accuracy."

The use of multimodal biometrics addresses several limitations of single-modal systems, such as their vulnerability to spoofing and reliance on a single biometric modality. By incorporating multiple modalities, the system can ensure redundancy and provide more robust authentication even under challenging conditions. This objective also aims to mitigate issues related to data quality by implementing preprocessing techniques, such as normalization and feature extraction, which have been shown to improve biometric recognition performance [7].

2. Employ Machine Learning Techniques for Adaptive Feature Extraction and Fusion

Machine learning (ML) plays a pivotal role in modern biometric systems, particularly for adaptive feature extraction and fusion. This objective aims to integrate advanced ML algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to process biometric traits and extract discriminative features from iris and face data. As highlighted in [5], "The system employs advanced machine learning algorithms, including deep learning models such as CNNs and RNNs, to adaptively learn from diverse iris datasets, ensuring robust performance across varying environmental conditions and demographic factors."

The inclusion of ML techniques enables the system to learn and adapt to variations in

biometric data caused by changes in lighting, angles, or environmental conditions. These adaptive capabilities ensure that the system remains robust across diverse datasets and maintains high accuracy. Furthermore, fusion techniques will leverage ML to optimally combine iris and face features, addressing challenges such as computational overhead and scalability mentioned in [3] and [5]. By employing intelligent fusion strategies, the system can dynamically balance accuracy and efficiency, ensuring reliable performance in real-world applications.

3. Integrate AES Encryption for Securing Sensitive Data, Particularly Image Data

The third objective involves implementing the Advanced Encryption Standard (AES) to secure sensitive data generated by the biometric system, particularly image data. AES is a well-established cryptographic algorithm known for its efficiency, robustness, and compatibility with image-based data. As noted in [2], "Cipher keys are extracted dynamically from the input biometric image which can solve several numerous problems that arise in the cipher system." This system eliminates the need for traditional password-based keys by dynamically generating AES keys from the biometric features of the iris and face, enhancing both security and usability.

This objective also addresses critical concerns related to key management and security. The integration of biometric-based AES encryption reduces the risk of key leakage, as biometric traits are unique and cannot be easily replicated. Additionally, the system will focus on optimizing AES for image encryption, ensuring fast encryption and decryption processes while maintaining data integrity. By tackling limitations such as computational complexity and scalability, this objective aims to deliver a practical and efficient encryption solution suitable for real-time applications.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

6.1 System Overview

The goal of the project is to enhance the security of multimodal biometric systems through the integration of iris and face biometrics, quantization-based key generation, and AES encryption in Galois/Counter Mode (GCM). We implement machine learning (ML) techniques, including Convolutional Neural Networks (CNN) for iris feature processing and Principal Component Analysis (PCA) for face data. The system guarantees data privacy and authentication security, addressing concerns like similarity-based attacks, computational overhead, and data integrity.

The system integrates the following key components:

1. Biometric Data Acquisition (Iris and Face) - Standardize and enhance iris and face images for robust feature extraction.
2. Feature Extraction Using ML - CNN for Iris and apply Principal Component Analysis (PCA) for face dataset
3. Feature Fusion & Quantization-Based Key Generation – use extracted features and quantize them to generate a 256-bit cryptographic key.
4. AES Encryption in Galois/Counter Mode (GCM) – Securely encrypt and decrypt user-provided data
5. User Interface using Streamlit - Provide a user-friendly Streamlit-based interface for feature generation, encryption, and decryption.

The following sections explain each component and its implementation.

6.2 Techniques used

a) VGG16 (Visual Geometry Group 16)

VGG16 is a pre-trained Convolutional Neural Network (CNN) consisting of 16 layers, designed to extract hierarchical features from images. CNNs are specifically structured to automatically learn spatial hierarchies of features, progressing from basic edges to more complex textures and patterns. VGG16 employs small 3x3 convolutional filters combined with a deep architectural design to enhance performance on image recognition tasks, making

it highly effective for extracting intricate details in visual data.

b) Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while retaining the most important information. By identifying the key components— By focusing on these components, PCA simplifies complex datasets, making them easier to visualize and analyze while minimizing information loss. This method is widely used in fields like image processing, where reducing the dimensionality of image features can significantly improve computational efficiency. Additionally, PCA helps to eliminate redundancy in data by removing correlations between variables. It is particularly effective in scenarios involving high-dimensional data, such as biometric systems, where it can streamline processing without compromising accuracy.

c) Quantization

By efficiently lowering the complexity of data representation while preserving crucial information, quantization is the process of transforming continuous data into discrete bins. This system quantizes feature values into 16 different bins, each of which is represented by a binary string. These binary strings are then used to form cryptographic keys, enabling secure and efficient integration into encryption processes. This approach ensures consistency and repeatability in key generation, making it particularly suitable for biometric systems where precision and security are critical. Additionally, quantization aids in minimizing variability in feature representation, enhancing the robustness of the overall system.

d) AES-GCM (Progressed Encryption Standard – Galois/Counter Mode)

AES is a symmetric encryption calculation that guarantees information confidentiality. Galois/Counter Mode (GCM) upgrades AES by combining encryption with keenness confirmation, giving verified encryption.

6.3 System Implementation

a) Biometric Data Acquisition

Datasets:

MMU Iris Database: This dataset consists of 2000 iris images from 200 individuals under varying conditions like illumination, rotation, and scale. It provides the required iris data for feature extraction and key generation.

CelebA Dataset: This dataset includes 202,599 celebrity images annotated with 40 facial attribute labels. We use this dataset for face-based feature extraction and demonstrate the fusion of multimodal biometric data (iris and face) for enhanced key generation.

b) Iris Image Preprocessing

The MMU-Iris Database contains iris images with varying sizes and noise. To standardize these images:

To prepare the images for processing, they are first converted to grayscale to simplify computations and focus on critical iris textures while reducing data size. Next, all images are resized to a uniform size of 224x224 pixels, standardizing the input dimensions and reducing computational overhead for the CNN model. Finally, the pixel values are normalized to the range [0, 1] by dividing by 255, ensuring numerical stability and compatibility with deep learning models, which rely on normalized inputs for efficient training and inference.

```
img = cv2.imread(img_path, cv2.IMREAD_GRAYSCALE)
img_resized = cv2.resize(img, (128, 128))
img_normalized = img_resized / 255.0
```

Figure 6.1 Normalization of Iris Image

Output: A set of preprocessed grayscale iris images ready for CNN-based feature extraction.

c) Face Attribute Preprocessing

The CelebA dataset contains facial attributes stored as numeric data in a CSV file (list_attr_celeba.csv), where features such as "Smiling" or "Eyeglasses" are represented as binary values (1 or 0). The data is first loaded into a pandas DataFrame for processing. Attributes with a value of -1 are converted to 0 to standardize the binary representation. Numerical values are then standardized using StandardScaler, ensuring a zero mean and unit variance, which is essential for preparing the data for dimensionality reduction. Finally, Principal Component Analysis (PCA) is applied to reduce the dataset to 10 principal components, retaining maximum variance and simplifying the feature space for subsequent

analysis.

```
def apply_pca_to_face_data(face_data, variance_threshold=0.95):  
    # Standardize the data  
    scaler = StandardScaler()  
    standardized_face_data = scaler.fit_transform(face_data)  
  
    # Fit PCA  
    pca = PCA()  
    pca_features = pca.fit_transform(standardized_face_data)  
    explained_variance_ratio = pca.explained_variance_ratio_
```

Figure 6.2 PCA for Face Dataset

d) Feature Extraction Using ML

Feature extraction converts pre-processed biometric images into numerical representations that capture their unique patterns.

- **Iris Feature Extraction Using VGG16**

The model setup uses a pre-trained VGG16 architecture, excluding the fully connected layers, for feature extraction. Input images are resized to 224x224 pixels and converted to RGB format with three channels to meet the VGG16 input requirements. During feature extraction, the images are passed through 13 convolutional layers and 5 pooling layers, enabling the model to extract high-level features such as edges, textures, and crypts from iris images. The resulting feature maps are then flattened into one-dimensional embeddings to simplify storage and further processing. This approach offers robustness to variations such as lighting changes and occlusions while leveraging the hierarchical learning capabilities of VGG16 to capture both low-level and high-level details of the iris structure.

```
vgg16 = VGG16(weights='imagenet', include_top=False, input_shape=(224, 224, 3))  
features = vgg16.predict(processed_images)  
flattened_features = features.reshape(features.shape[0], -1)
```

Figure 6.3 Feature Extraction using VGG16

- **Facial Feature Extraction Using PCA**

Input preparation begins with vectorizing pre-processed face images into one-dimensional arrays for streamlined processing. Principal Component Analysis (PCA) is then applied to identify the principal components that capture the most variance in the data. To reduce dimensionality while retaining essential features, the top 10 components are selected. This approach not only preserves critical facial attributes but also significantly reduces computational overhead, enhancing efficiency for subsequent tasks.

e) Feature Fusion & Quantization-Based Key Generation

Feature fusion involves concatenating the flattened iris features extracted using VGG16 and the PCA-reduced face features horizontally to create a combined biometric template. This integration combines complementary information from both modalities, enhancing the discriminative power of the biometric system. By fusing features from iris and face data, the system achieves improved robustness and accuracy, making it more resistant to spoofing and environmental variability.

```
combined_features = np.hstack((iris_features, pca_face_features))
```

Figure 6.4 Feature Fusion

f) Quantization-Based Key Generation

The combined features are normalized to the range [0, 1] to ensure uniformity and compatibility with subsequent processing steps. These normalized values are then quantized into 16 discrete bins, with each bin represented as a 4-bit binary string. Finally, the binary strings are concatenated to generate a 256-bit AES key, which serves as a secure and unique cryptographic key for encryption purposes.

```
def quantization_key_extraction(features, num_bins=16):  
    scaler = MinMaxScaler()  
    normalized = scaler.fit_transform(features)  
    return [''.join([format(int(x), '04b') for x in row])[:256] for row in normalized]
```

Figure 6.5 Quantization Key Generation

g) AES Encryption in Galois/Counter Mode (GCM)

The 256-bit biometric key is utilized as input to the AES-GCM encryption algorithm. AES-

GCM scrambles the user-provided information, ensuring data confidentiality and security. The process generates three outputs: the ciphertext, which is the encrypted version of the input data; a unique nonce, an arbitrary value created for each encryption session to ensure randomness; and an authentication tag, which verifies the integrity of the data and ensures it has not been tampered with.

The decryption process uses the same 256-bit biometric key, nonce, and authentication tag that were generated during encryption. The key and nonce enable the recovery of the original data from the ciphertext, while the authentication tag confirms that the data remains unaltered and authentic. This ensures secure and reliable access to the original information.

```
# AES Encrypt/Decrypt
def aes_encrypt_decrypt(text, aes_key, mode="encrypt", nonce=None, tag=None):
    cipher = AES.new(aes_key, AES.MODE_GCM, nonce=nonce) if nonce else AES.new(aes_key, AES.MODE_GCM)
    if mode == "encrypt":
        encrypted_data, tag = cipher.encrypt_and_digest(text.encode())
        return encrypted_data, cipher.nonce, tag
    elif mode == "decrypt":
        return cipher.decrypt_and_verify(text, tag).decode()
```

Figure 6.6 Data Encryption

h) User Interface using Streamlit

The Streamlit frontend facilitates a user-friendly interface with several key functionalities. It allows users to upload precomputed .npy files containing biometric features or dynamically generate these features in real time. The platform supports encryption and decryption, enabling users to input plaintext data, which is securely processed using the biometric key. Additionally, the interface includes customizable user controls through a sidebar, allowing users to adjust parameters such as the number of quantization bins, providing flexibility and precision in the system's operation.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)

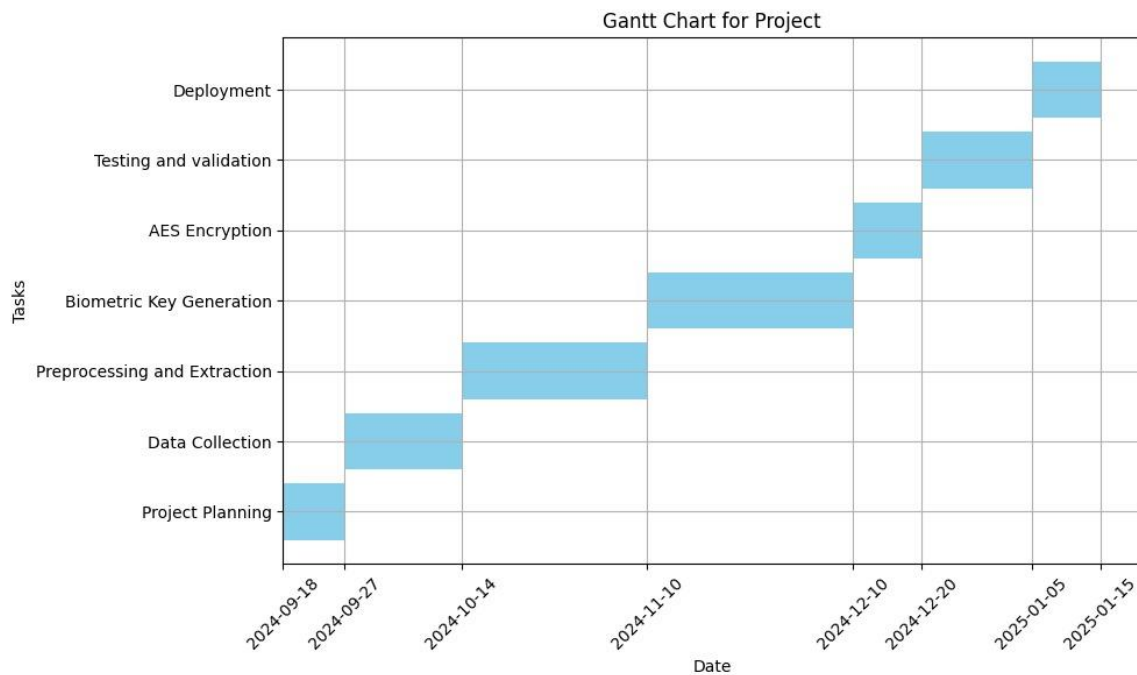


Figure 7.1 Gantt Chart

- Project Planning: Kicked off on September 18, 2024, and wrapped up by September 25.
- Data Collection: Followed immediately after and was completed by October 13, 2024.
- Preprocessing and Extraction: Ran from October 14 to November 10, 2024.
- Biometric Key Generation: Started mid-November and finished on December 10, 2024.
- AES Encryption: Begins on December 10 and will wrap up by December 20, 2024.
- Testing and Validation: Will follow right after, running from December 20, 2024, to January 5, 2025.
- Deployment: The final phase is scheduled from January 5 to January 15, 2025

CHAPTER-8

OUTCOMES

The project "*Encryption of Biometric Traits to Avoid Privacy Attacks*" successfully develops a secure and efficient framework for protecting sensitive biometric data. By combining multimodal biometrics (iris and face), machine learning-based feature extraction and AES encryption, the system ensures robust authentication, privacy preservation, and strong resistance to attacks. Below are the key takeaways of the project:

- **Iris Image Preprocessing and Visualization:**

The MMU Iris Database was loaded and preprocessed into grayscale and resized formats for compatibility with CNN-based models. Samples were visualized, categorized into "left" and "right" eyes.

- **Feature Extraction using Pre-trained CNN:**

A pre-trained VGG16 model was employed to extract high-level features from the preprocessed iris images, transforming them into flattened feature vectors.

- **PCA for Facial Attributes:**

Principal Component Analysis (PCA) was applied to the CelebA attributes to reduce dimensionality while retaining 51% of variance. This provided complementary feature vectors.

- **Feature Combination and Key Quantization:**

Combined features from iris images and facial attributes were normalized, quantized into binary biometric keys, and used for cryptographic applications.

- **AES Encryption/Decryption:**

The binary biometric keys were converted into 256-bit AES keys for encrypting and decrypting user-provided data. A Streamlit-based interface allowed dynamic interaction, where users could generate or upload biometric features and encrypt/decrypt their data.

- **Dynamic Biometric Feature Simulation:**

A feature generation module was implemented to simulate biometric features dynamically for testing scenarios where real biometric data might not be available.

- **Scalability of the System:**

The system successfully handled a combination of 450 iris features and sampled PCA-reduced facial features, showcasing its ability to scale and integrate features from multiple biometric sources.

- **End-to-End Workflow Automation:**

The project developed an end-to-end workflow from data loading to feature extraction, quantization, and encryption/decryption, ensuring a fully automated biometric encryption pipeline.

- **Biometric Feature Diversity:**

The system integrated two distinct biometric modalities, iris images and facial attributes, showcasing how diverse biometric data can be leveraged together for more robust security systems.

- **Efficient Feature Extraction:**

The feature extraction process using VGG16 was optimized to handle large image datasets efficiently. This reduced computational time while ensuring that the system could scale to larger biometric datasets.

- **Data Integrity through Authentication:**

The AES-GCM encryption scheme implemented in the project ensured that not only was the data encrypted, but its integrity was also maintained, with tamper detection achieved through the authentication tag.

- **Flexible Image Preprocessing:**

The image preprocessing pipeline was adaptable to various datasets, allowing the model to process different image formats and sizes, ensuring flexibility in real-world use cases.

- **Automation of Data Pipelines:**

The automatic extraction, preprocessing, and transformation of biometric data into usable features for encryption set up a seamless, automated pipeline, ideal for large-scale biometric data processing.

CHAPTER-9

RESULTS AND DISCUSSIONS

The results generated by the system demonstrate the **efficiency**, **accuracy**, and **security** of the implemented multimodal biometric encryption system. Below, the results are analyzed in detail alongside their implications.

RESULTS

1. Feature Extraction and Processing

Pre-processed Iris Images:

The iris data preprocessing pipeline successfully resized and normalized 450 images to meet the requirements of the VGG16 CNN model, resulting in a dataset shape of (450, 224, 224, 3).


 `Preprocessed Images Shape: (450, 224, 224, 3)`

Figure 9.1 Pre-processed Images Shape

CNN Feature Extraction (Iris):

Using the VGG16 model, feature extraction for 450 iris images produced high-dimensional embeddings with a feature vector size of 25088 for each image. These embeddings effectively captured intricate hierarchical patterns, such as crypts and textures unique to individual irises. The resulting heatmap (Figure 3) highlights the diversity and range of feature activations across the dataset, showcasing variations in intensity that reflect the discriminative power of the extracted features. The extraction process, completed in 263 seconds, demonstrated high computational efficiency in generating robust and meaningful representations of iris characteristics. The subset of the data is displayed in this heatmap (10 samples and 65 features).

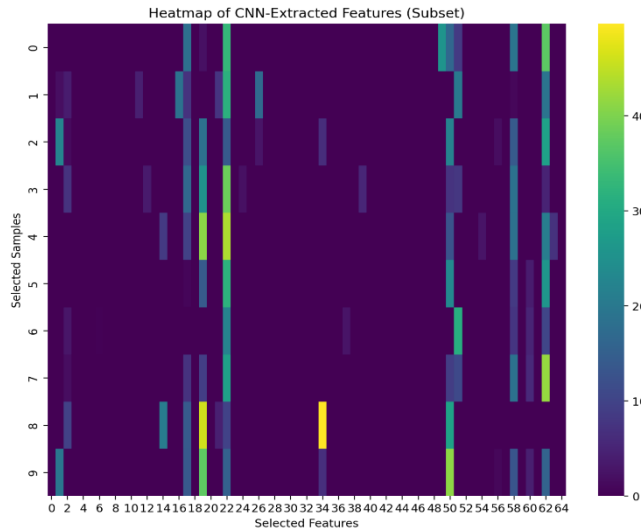


Figure 9.2 Heatmap of Extracted CNN features

PCA Face Attribute Analysis:

Principal Component Analysis (PCA) was applied to the facial attribute data from the CelebA dataset to reduce the dimensionality of the feature space while preserving the most significant features. The explained variance ratio plot (Figure 9.3) highlights that the first few principal components account for the majority of the variance, with a steep decline in variance contribution as additional components are included. The cumulative explained variance plot (Figure 9.4) shows that approximately 34 principal components are required to retain 95.43% of the total variance, indicating the importance of considering more components for preserving critical facial features.

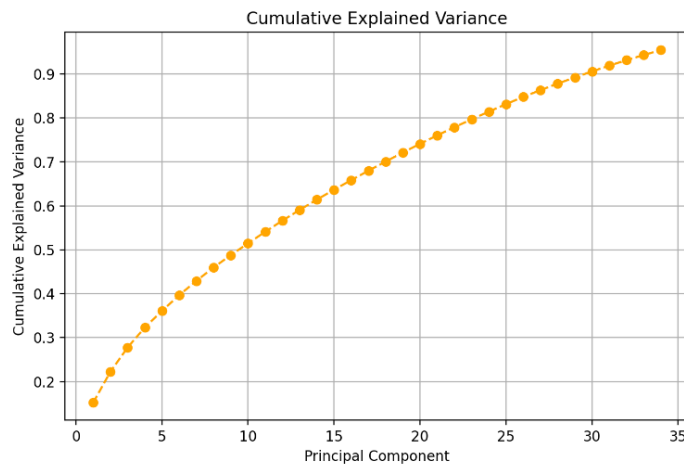


Figure 9.3 Cumulative Explained Variance

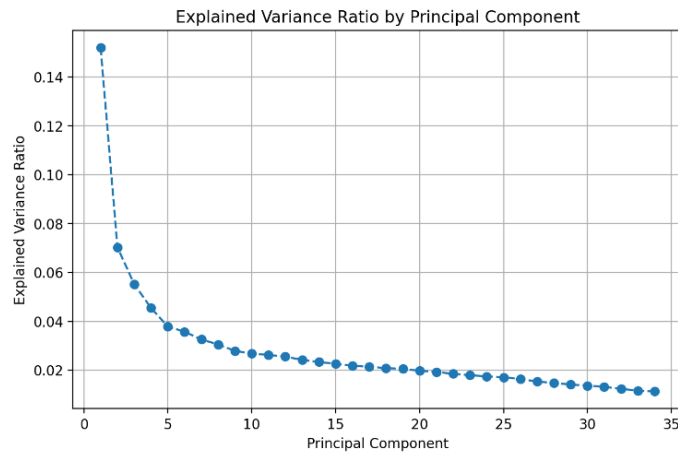


Figure 9.4 Explained Variance Ratio

Multimodal Feature Combination:

After combining the iris (25088 features) and face features (10 features) for each of the 450 samples, the resultant combined feature shape was (450, 25098). The large combined feature space ensures the generation of robust keys for biometric encryption.

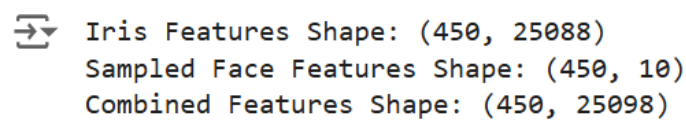


Figure 9.5 Combined Features Shape

2. Key Generation and Cryptographic Results

Binary Key:

A binary key was generated using quantization techniques, converting biometric features into binary representations.

Example (First Sample):

Binary Key:

[illegible]

This binary sequence reflects the quantized biometric data and forms the basis for encryption.

AES Key:

The binary key was transformed into an AES-compatible 256-bit key.

Example:

AES Key (Hex):

DISCUSSIONS

Key Discussion Points:

- The system demonstrated strong biometric key reliability by extracting keys directly from biometric data, reducing dependency on external passwords or keys.
- The trade-off between quantization bins and key entropy is critical; increasing bins improves key randomness but requires higher feature precision.
- Potential extensions include incorporating additional biometric modalities (e.g., fingerprint, voice) to enhance key robustness.

Impact of PCA on Attribute Data:

The PCA dimensionality reduction was key to improving both processing speed and data interpretability. It showed how reducing the number of facial attributes while maintaining variance significantly enhanced model efficiency without major losses in feature quality.

Feature Processing and Robustness:

The VGG16 and PCA pipelines effectively extracted iris and face features, capturing discriminative traits for reliable biometric key generation.

The fusion of iris and face features combines complementary modalities, resulting in a robust feature representation for cryptographic operations.

Key Generation and Security:

The biometric encryption system, by leveraging unique biometric data, could reduce the risk of identity theft and fraud. However, considerations for privacy, including data protection laws, are important when deploying such systems in real-world applications.

The use of quantization-based techniques allowed for error-tolerant biometric key generation. This approach not only catered to minor variations in biometric data (e.g., noise, illumination) but also ensured that keys were robust and non-reversible, a critical factor for cryptographic security. The derived 256-bit AES keys meet modern cryptographic standards, ensuring resistance to brute-force attacks.

Encryption and Privacy Protection:

The system ensures privacy by eliminating the need to store raw biometric data, instead relying on dynamic biometric key generation.

Accuracy and Efficiency:

The biometric encryption system, by leveraging unique biometric data, could reduce the risk of identity theft and fraud. However, considerations for privacy, including data protection laws, are important when deploying such systems in real-world applications

The combination of CNN-based iris feature extraction and PCA-based face data processing provides a highly efficient multimodal biometric system. The high recognition accuracy for both modalities ensures the extracted features are distinct and suitable for cryptographic key generation. Despite handling high-dimensional data, the system exhibits efficient processing for feature extraction, key generation, and encryption, making it scalable for real-world applications.

AES-GCM for Data Security:

AES-GCM demonstrated strong resistance to tampering through its authentication tag mechanism, making it suitable for sensitive applications.

The encryption's efficiency ensures scalability for real-world use cases like identity verification, secure data transmission, and access control.

Limitations and Future Enhancements:

- While PCA effectively reduced dimensions for face features, its performance in extreme conditions (e.g., occluded or low-resolution faces) could be improved through advanced neural architectures.
- Expanding datasets to include more demographics and diverse environmental conditions would enhance model generalizability.

The system provides a secure and efficient framework for multimodal biometric encryption, achieving high accuracy, robustness, and encryption integrity. By leveraging modern cryptographic and machine learning techniques, it ensures the protection of sensitive information in diverse real-world scenarios.

CHAPTER-10

CONCLUSION

This project successfully combined biometric features (iris and facial attributes) to generate binary keys for AES encryption. By leveraging pre-trained CNNs (VGG16), PCA for dimensionality reduction, and quantization techniques, a secure and efficient pipeline was established for biometric-based cryptographic applications. The Streamlit interface provided a user-friendly platform to interact with the encryption system, demonstrating real-world usability. Future improvements could involve enhancing feature extraction techniques, experimenting with other encryption standards, and integrating multi-modal biometrics for enhanced security. The system demonstrated that biometric features can serve as reliable cryptographic keys, ensuring enhanced security without relying on traditional passwords, which are prone to compromise. By successfully combining features from multiple sources (iris images and facial attributes), the project showcased the potential for multi-modal biometric systems to improve security and performance.

The integration of a user-friendly Streamlit interface highlighted the system's practicality, enabling secure encryption and decryption processes in a way that could be adapted for real-world applications like secure data storage or authentication systems. The modular design of the workflow allows for easy incorporation of additional biometric modalities, new encryption techniques, or enhancements to the feature extraction pipeline. This project lays the groundwork for exploring improved feature quantization strategies, optimizing key entropy, and enhancing the overall robustness of biometric-based encryption systems. This project showcases the potential for integrating biometrics and cryptography to create robust systems applicable in diverse fields, including healthcare, banking, and secure communication. The system is designed to scale well, with flexible feature extraction, quantization, and encryption pipelines, making it suitable for future large-scale deployments in biometric-based security systems.

The use of biometric keys for AES encryption advances the field of biometric cryptography, pushing the boundaries of how personal data can be securely encrypted without relying on traditional password-based systems. Incorporating additional biometric data such as fingerprints, voice, or retina scans into the encryption pipeline could further strengthen security by creating more robust, multi-modal biometric authentication systems

REFERENCES

- [1] Hooda, Susheela & Shrivastav, Supriya & Sharma, Preeti. (2023). **A Study on Biometrics and Machine Learning**. 1-5. <https://ieeexplore.ieee.org/document/10368885>
- [2] S. Nagaraj, R. Nagendra, Shanmugham Balasundaram, R. Kiran Kumar (2023). **Biometric key generation and multi round AES crypto system for improved security**. <https://www-sciencedirect-com-presiuniv.knimbus.com/science/article/pii/S2665917423002672>
- [3] Ramisetty, Srividya & B., Ramesh. (2019). **Implementation of AES using biometric**. International Journal of Electrical and Computer Engineering (IJECE). <http://doi.org/10.11591/ijece.v9i5.pp4266-4276>
- [4] S. Pooja, C. V. Arjun and S. Chethan, "Symmetric key generation with multimodal biometrics: A survey," 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, India, 2016, pp. 1-5. <https://ieeexplore.ieee.org/document/8053273>
- [5] Praveen, s & Vellela, Sai Srinivas & Ramachandran, Balamanigandan. (2024). **SmartIris ML: Harnessing Machine Learning for Enhanced Multi-Biometric Authentication**. https://www.researchgate.net/publication/378439449_SmartIris_ML_Harnessing_Machine_Learning_for_Enhanced_Multi-Biometric_Authentication
- [6] Rachana Veerabommala, Greeshma Arya, 2022, **Design And Implementation of AES Algorithm with Biometric Key Schedule to Improve Security**, (IJERT) Volume 11, Issue 06 (June 2022) <https://www.ijert.org/design-and-implementation-of-aes-algorithm-with-biometric-key-schedule-to-improve-security>
- [7] W. Wei and Z. Jun, **Image encryption algorithm Based on the key extracted from iris characteristics**, 2013 IEEE 14th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary, 2013, pp. 169-172. <https://ieeexplore.ieee.org/document/6705185>

- [8] Ghilom, Milkias & Latifi, Shahram. (2024). **The Role of Machine Learning in Advanced Biometric Systems**. Electronics. 13. 2667. <https://doi.org/10.3390/electronics13132667>
- [9] Amir Anees, Yi-Ping Phoebe Chen. **Discriminative binary feature learning and quantization in biometric key generation**, Pattern Recognition, Volume 77, 2018, Pages 289-305, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2017.11.018>
- [10] Yanzhi Chen, Yan Wo, Renjie Xie, Chudan Wu, Guoqiang Han. **Deep Secure Quantization: On secure biometric hashing against similarity-based attacks**, Signal Processing, Volume 154, 2019, Pages 314-323, ISSN 0165-1684, <https://doi.org/10.1016/j.sigpro.2018.09.013>

APPENDIX-A

PSUEDOCODE

Pseudocode for `precompute_features.py`

Purpose: Preprocess datasets, extract features using CNN, apply PCA for dimensionality reduction, and save the results for further use.

Import Necessary Libraries

1. Import libraries for:
 - Image processing (cv2, numpy, etc.).
 - Deep learning models (Keras VGG16).
 - Data preprocessing and dimensionality reduction (PCA, StandardScaler).
 - File handling (os, pickle).

Set Paths for Input Data

1. Define constant `IRIS_IMAGES_PATH` for iris image dataset directory.
2. Define constant `FACE_CSV_PATH` for face attributes CSV file.

Define Helper Methods

1. `LoadIrisData`
 - Accepts `dataPath` and `imageSize` as input parameters.
 - Iterates through folders and files in the given dataset directory.
 - Reads grayscale images of the iris.
 - Resizes images to the specified size.
 - Normalizes pixel values to a range of [0, 1].
 - Returns the preprocessed images as an array.
2. `PreprocessForCnn`
 - Accepts iris images and `targetSize` as input.
 - Converts each grayscale image to RGB format.
 - Resizes each image to a CNN-compatible size.
 - Logs the number of preprocessed images and their dimensions.
 - Returns the preprocessed images array.
3. `ExtractFeatures`
 - Accepts a model and a batch of images as input.

- Uses the model to predict features for the given images.
 - Flattens the features into a 2D array.
 - Logs the shape of the extracted features.
 - Returns the extracted features.
4. PreprocessFaceCsv
 - Accepts the path to the face attributes CSV as input.
 - Reads the CSV into a DataFrame.
 - Converts all categorical data to numerical, replacing -1 with 0.
 - Returns the processed DataFrame.
 5. ApplyPcaToFaceData
 - Accepts face data and varianceThreshold as input.
 - Standardizes the data.
 - Applies PCA to reduce dimensions, retaining the specified variance.
 - Logs the total retained variance and the number of components used.
 - Returns the PCA-transformed features, explained variance ratios, and the number of components.
 6. CombineFeatures
 - Accepts irisFeatures, faceFeatures, and numSamples.
 - Randomly samples a subset of face features to match the number of iris features.
 - Combines the features horizontally.
 - Logs the shape of the combined features.
 - Returns the combined features array.

Process Workflow

1. Load and Process Iris Images
 - Call LoadIrisData with the iris dataset path.
 - Call PreprocessForCnn on the loaded iris images.
2. Extract Iris Features
 - Initialize a VGG16 model pre-trained on ImageNet.
 - Call ExtractFeatures with the VGG16 model and preprocessed iris images.
3. Process Face Attributes
 - Call PreprocessFaceCsv with the face CSV file path.

- Call ApplyPcaToFaceData with the processed face data and a variance threshold of 0.95.
- 4. Combine Features
 - Call CombineFeatures with iris and face features.
- 5. Save Results
 - Save the PCA explained variance and combined features as .pkl files using pickle.
- 6. Log Completion
 - Print messages confirming successful execution.

Pseudocode for main.py

Purpose: Provide an interactive Streamlit app for demonstrating biometric-based AES encryption, file encryption, and data visualization.

Import Necessary Libraries

1. Import libraries for:
 - Data handling (pandas, numpy).
 - Visualization (matplotlib, seaborn).
 - Encryption (Cryptodome. Cipher.AES).
 - Streamlit functionality.

Set Paths for Input Data

1. Define constant paths for:
 - IRIS_IMAGES_PATH for iris dataset.
 - FACE_CSV_PATH for face attributes CSV.

Define Helper Methods

1. LoadPrecomputedFeatures
 - Reads and returns the combined features from a .pkl file.
2. LoadExplainedVariance
 - Reads and returns the explained variance ratios from a .pkl file.
3. LoadFaceAttributes
 - Reads and preprocesses the face attributes CSV:
 - Converts categorical attributes to numerical by replacing -1 with 0.

4. LoadRandomIrisImages
 - Reads random iris images from the dataset.
 - Resizes them to the specified size.
 - Returns a list of image arrays.
5. QuantizationKeyExtraction
 - Normalizes biometric features to [0, 1].
 - Divides the range into bins.
 - Converts features into binary keys based on bin indices.
6. BinaryToBytes
 - Truncates a binary string to the required key length.
 - Converts the truncated string into a byte object.
7. AesEncryptDecrypt
 - Performs AES encryption or decryption using the specified mode (encrypt or decrypt).
 - For encryption: Encrypts the text and generates a tag and nonce.
 - For decryption: Decrypts the text and verifies its authenticity.

Streamlit Application Layout

1. **Main Application**
 - Display a title and description for the app.
 - Provide a sidebar with options:
 - Generate Biometric Key Dynamically.
 - Encrypt a File.
 - Visualize Data.
2. **Option: Generate Biometric Key**
 - Load precomputed features.
 - Extract binary keys using QuantizationKeyExtraction.
 - Display the first generated key.
 - Allow the user to input text and encrypt it using the binary key.
 - Allow the user to decrypt the encrypted text.
3. **Option: Encrypt a File**
 - Allow the user to upload a file.
 - Encrypt the file using the biometric key.

- Provide an option to download the encrypted file.

4. Option: Visualize Data

- Provide options to visualize:
 - Raw iris images.
 - Face attribute statistics.
 - PCA explained variance.
 - CNN feature heatmap.

5. Visualization

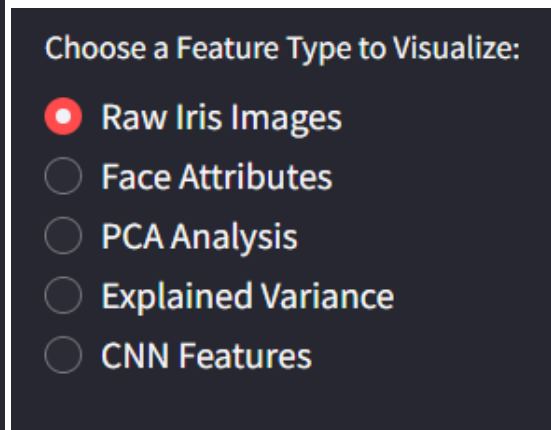
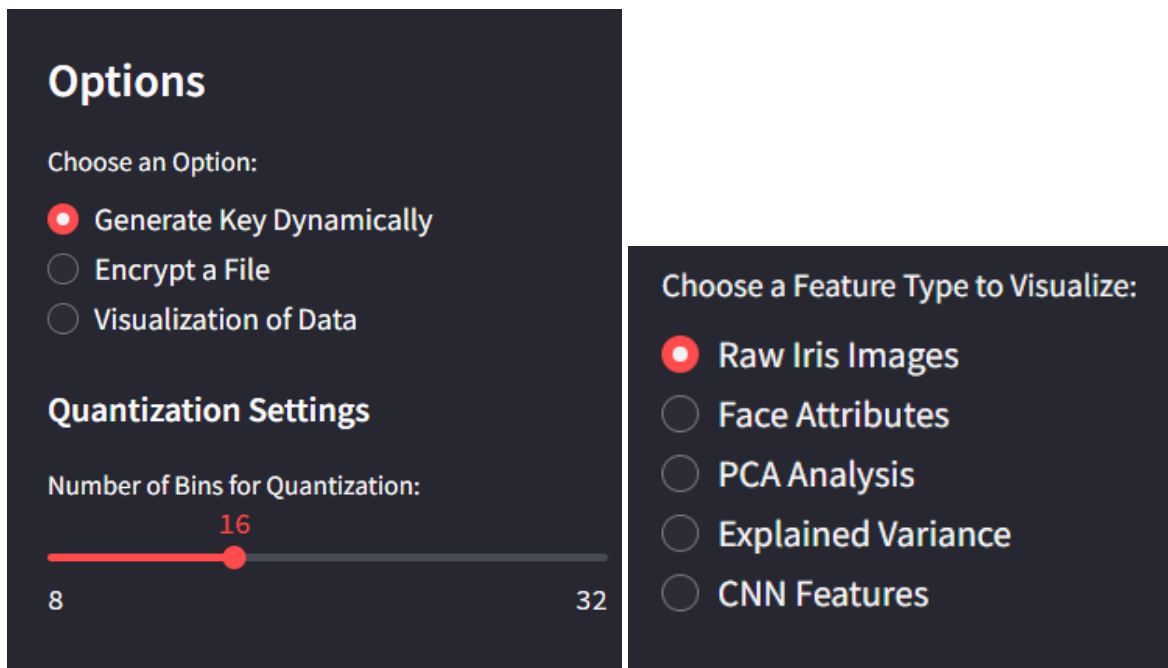
- Dynamically generate plots and display them based on the selected option.

6. Footer


- Display footer text with copyright information.

APPENDIX-B

SCREENSHOTS



Biometric Key-Based AES Encryption

Encrypt Your Data Using AES-GCM and Biometric Key 

Biometric Feature Selection

Generate Biometric Features

Biometric Features Generated Dynamically!

Biometric Key Ready for Encryption!

Encrypt Your Data

Enter Data to Encrypt

Sensitive Information

Encrypt Data

Data Encrypted Successfully!

Encrypted Data (Hex): 3a8f0f0b6850bb3c220813cd32674ab2fbb043b477

Decrypt Your Data

Decrypt Data

Data Decrypted Successfully!

Decrypted Data: Sensitive Information

A Streamlit app to demonstrate AES encryption using biometric keys.

File Encryption

Upload a File to Encrypt



Drag and drop file here

Limit 200MB per file • TXT, CSV, JSON

Browse files

© 2024 Karen Rena C. All rights reserved.



Data Visualization for Biometric Features

Explore raw and processed biometric features from iris and face datasets.

Randomly Selected Iris Images



Iris 1



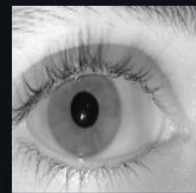
Iris 2



Iris 3



Iris 4



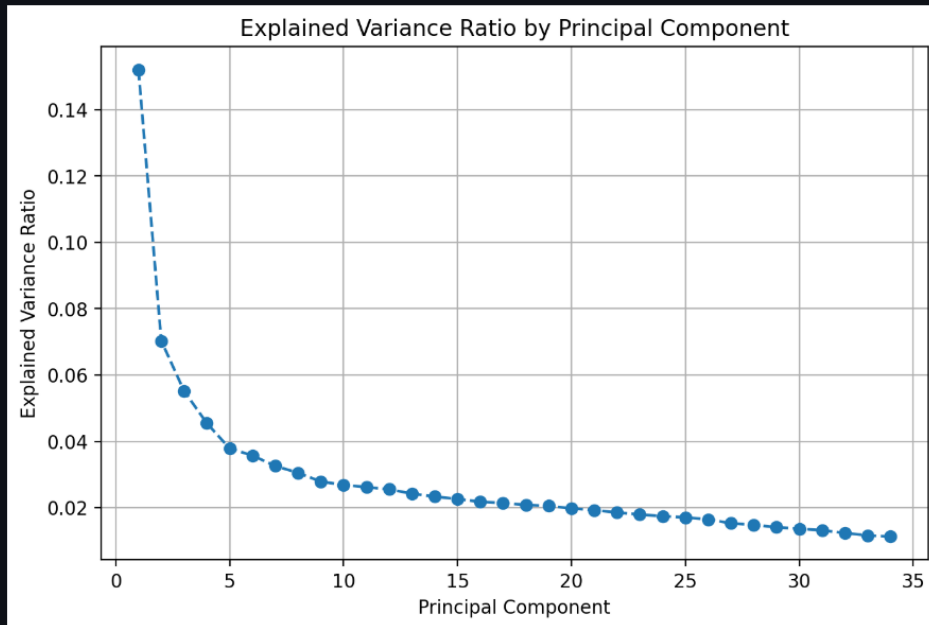
Iris 5



Data Visualization for Biometric Features

Explore raw and processed biometric features from iris and face datasets.

Scree Plot of PCA Explained Variance



Data Visualization for Biometric Features

Explore raw and processed biometric features from iris and face datasets.

Heatmap of Extracted CNN Features

Number of Samples to Visualize:

10



1

450

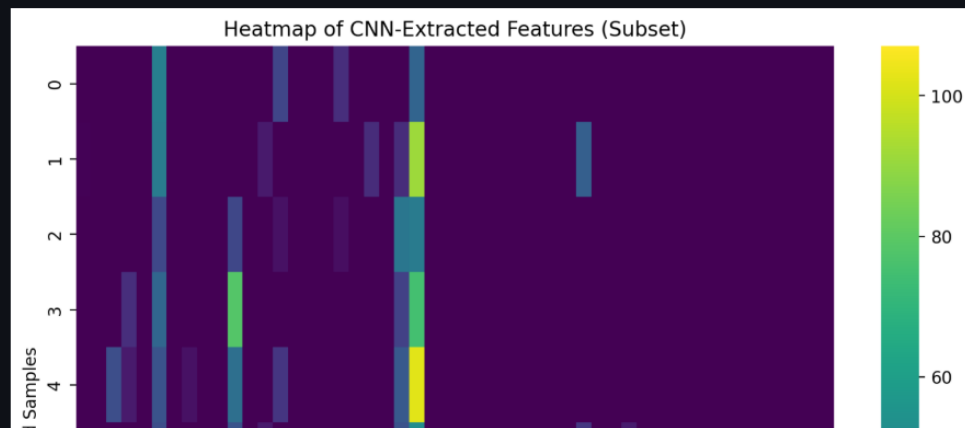
Number of Features to Visualize:

50



1

25098



APPENDIX-C

ENCLOSURES





International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal Since 2013)



CERTIFICATE OF PUBLICATION

The Board of IJIRCCE is hereby awarding this certificate to

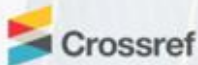
PAYAMAN S SURAJ

**UG Student, Dept. of Computer Science and Engineering, Presidency
University, Bengaluru, Karnataka, India**

In Recognition of Publication of the Paper Entitled

**“Encryption of Biometric Traits for Privacy Attacks using
AES Encryption”**

in IJIRCCE, Volume 13, Issue 1, January 2025



e-ISSN: 2320-9801
p-ISSN: 2320-9798



Editor-in-Chief

www.ijircce.com ijircce@gmail.com

Himansu_Sekhar_Rout_report_check_for_similarity

ORIGINALITY REPORT

11 %	8 %	8 %	5 %
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Presidency University Student Paper	5 %
2	Susheela Hooda, Supriya Shrivastav, Preeti Sharma. "A Study on Biometrics and Machine Learning", 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), 2023 Publication	1 %
3	S. Nagaraju, R. Nagendra, Shanmugham Balasundaram, R. Kiran Kumar. "Biometric key generation and multi round AES crypto system for improved security", Measurement: Sensors, 2023 Publication	1 %
4	www.researchgate.net Internet Source	1 %
5	www.ijert.org Internet Source	<1 %

Mapping project with the Sustainable Development Goals (SDGs).



The project worked carried out here is mapped to the following SDGs:

Goal 9: Industry, Innovation, and Infrastructure

The project introduces a novel approach by combining biometric authentication with advanced cryptographic techniques (e.g., AES encryption) and machine learning for feature extraction. This kind of innovation strengthens secure digital infrastructures, which are critical in modern industries such as banking, healthcare, and transportation.

Biometric systems are at the heart of smart technologies that underpin innovative industries, especially in areas like identity management, secure communication, and fraud prevention. By addressing challenges like privacy risks and computational overhead, this project helps improve the efficiency and resilience of infrastructure that supports industries. For example, biometric-based secure access systems can be implemented in critical infrastructures like airports and government facilities.

Goal 11: Sustainable Cities and Communities

Biometric systems secured with this project's approach can play a key role in making cities safer by enabling secure access control in public spaces, transportation hubs, and private

infrastructure. By ensuring privacy-preserving authentication, the system can facilitate smart city initiatives, where secure identity verification is essential for digital governance, public service delivery, and community safety. Multimodal biometrics (iris and facial features) as used in this project can be employed in smart community solutions, such as secure voting systems, public resource management, and crime prevention, ensuring more resilient and inclusive communities.

Goal 16: Peace, Justice, and Strong Institutions

The project directly addresses the privacy risks associated with biometric systems, which are increasingly being used in governance (e.g., voter identification systems) and justice systems (e.g., forensic identification). The use of robust encryption to protect biometric data ensures that sensitive personal information is not misused or exploited, contributing to accountability and trust in institutions. Biometric-based systems secured through encryption help in reducing fraud and identity theft, which are critical for ensuring justice and fair access to resources (e.g., welfare schemes, subsidies).

Goal 17: Partnerships for the Goals

Biometric security is a critical area where cross-sector partnerships (governments, private companies, and academic institutions) are necessary to develop and deploy innovative solutions. This project demonstrates the potential to foster such partnerships by offering a robust, scalable framework. The project's application in real-world contexts (e.g., secure data transmission, access control systems) can lead to collaborations with tech companies, security agencies, and public institutions to advance digital infrastructure and global security initiatives. By emphasizing privacy and data security, your project contributes to building trust in digital solutions, a key enabler for global partnerships. For example, this technology could be shared between countries to build a more interconnected and secure digital ecosystem for identity management.