A Multimodal Biometric Authentication System Using Machine Learning and Deep Learning Techniques

Ms. Hemamalini. S[1], Dr. Om Prakash Sharma[2]
School of Information Technology
SRM University, Sikkim
East- Sikkim 737101
hemamalinicse@gmail.com

Abstract

*Biometric systems relying on a single identifier, such as fingerprints or facial recognition, often struggle to deliver the accuracy and reliability required for secure personal identification. To address these challenges, we propose an advanced **multimodal biometric system** that integrates face recognition, fingerprint verification, and speaker verification. By leveraging the strengths of these diverse modalities, the system achieves superior performance, overcoming the limitations of single-biometric approaches. Preliminary results demonstrate that this integrated solution significantly enhances reliability, setting a new benchmark for biometric identification and security.*

1. Introduction

In today's interconnected digital world, the need for reliable user verification is more critical than ever, especially in scenarios like accessing multi-user computer accounts or safeguarding sensitive information. Traditionally, authentication methods have relied on **tokens**, such as ID cards ("something you have"), or **knowledge**, such as passwords ("something you know"). However, these approaches come with significant limitations—tokens can be lost, stolen, or forged, and passwords can be forgotten or compromised. Worse, they fail to distinguish between an authorized user and an imposter who has fraudulently acquired the token or knowledge. As a result, these methods often fall short of providing sufficient security for critical applications, such as access control and financial transactions.

Biometric systems, on the other hand, offer a far more secure alternative by relying on "something you are or do," such as a fingerprint, facial feature, or voice pattern. These physiological or behavioral traits are unique to each individual, making biometrics inherently more reliable and harder to compromise. Currently, nine major biometric indicators are widely used or under active research, including face, fingerprint, hand geometry, iris, voice, and signature. Each has its own strengths and weaknesses in terms of accuracy, user acceptance, and suitability for specific applications.

To address the limitations of individual biometric systems, **multimodal biometric systems** have emerged as a preferred solution. These systems combine multiple physiological or behavioral traits to enhance reliability and adaptability across various environments. For instance, in a network logon application, a user with dry or injured fingers might struggle to provide a clear fingerprint, but their face or voice could serve as effective alternatives. Similarly, in noisy environments, voice recognition may fail, while face or fingerprint recognition can still function effectively. By integrating multiple modalities, multimodal systems can provide robust identification under diverse conditions, overcoming the challenges faced by single-biometric approaches.

Research on multimodal biometric systems has been gaining traction in recent years. For example, Dieckmann et al. proposed an **abstract-level fusion scheme** known as the "2-from-3 approach," which integrates face, lip motion, and voice recognition. This method is inspired by the way humans rely on multiple cues to identify individuals. Similarly, Brunelli and Falavian introduced a **measurement-level fusion scheme**, along with a hybrid approach that combines rank-level and measurement-level methods, to enhance the accuracy and reliability of biometric systems.

Various integration strategies for multimodal biometric systems have been explored in the literature. Kittler et al. demonstrated the efficiency of a **Bayesian framework** for fusing multiple snapshots of a single biometric property. Similarly, Bigun et al. proposed a **Bayesian integration scheme** to combine different pieces of evidence for improved decision-making. Maes et al. introduced an approach to merge **biometric data** (e.g., voice) with **non-biometric data** (e.g., passwords), enhancing system robustness. Hong and Jain developed a multimodal identification system integrating **face and fingerprint recognition**, leveraging the complementary strengths of these biometrics.

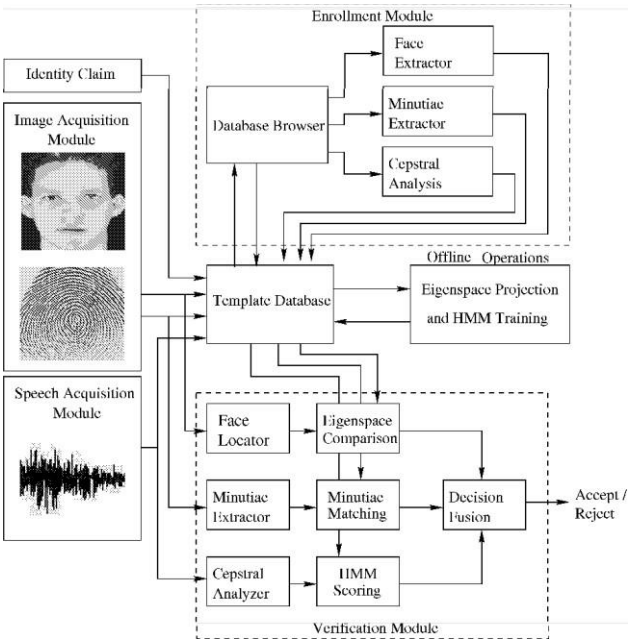


Figure 1: The block diagram of our multimodal bio—metric system.

We focus on designing a **multimodal biometric system** that combines face, fingerprint, and speech recognition for personal identification. This choice stems from the routine use of these modalities in law enforcement and commercial biometric systems, which frequently rely on fingerprint, face, or voice individually. These biometrics complement one another in their strengths: fingerprints offer exceptional verification accuracy, though they are challenging for humans to match without training. Meanwhile, face and speech are intuitive and commonly used by individuals for daily recognition tasks. Our system is tailored for **verification applications**, such as authenticating a user's identity in multiuser account systems, ensuring a seamless and secure experience.

The multimodal biometric system is composed of four primary components: **Acquisition Module**, **Template Database**, **Enrollment Module**, and **Verification Module**.

- The **Acquisition Module** is responsible for capturing biometric data, including fingerprint images, face images, and speech signals from users seeking access to the system.
- This data is stored in the **Template Database**, which acts as a secure repository containing the biometric templates of all enrolled users.
- The **Enrollment Module** handles user-related operations such as registration, deletion, updates, training, and system configuration, ensuring the system remains up-to-date and operational.
- The **Verification Module** performs the critical task of user authentication through a four-step process:

1. **Fingerprint Verification**: Matches the captured fingerprint against stored templates to compute a matching score.
2. **Face Recognition**: Compares the input face with stored face templates to generate a face matching score.
3. **Speaker Verification**: Analyzes the speech signal to produce a matching score.
4. **Decision Fusion**: Integrates the scores from all three modalities to make a final authentication decision.

This integration of multiple biometrics enhances the system's reliability, providing accurate and secure identity verification across various scenarios.

2. The Multimodal Biometric System

Each biometric modality in our multimodal system possesses unique characteristics and employs distinct matching schemes. As a result, integrating these modalities at the **decision level**—rather than at the sensor level—proves to be a more practical and effective approach for achieving accurate and reliable identification.

2. 1 Formulation

Let β\betaβ represent a biometric system, and let d1,d2,...,dNd_1, d_2, ..., d_Nd1,d2,...,dN denote the templates of the NNN users enrolled in the system, each identified by numerical labels 1,2,...,N1, 2, ..., N1,2,...,N. For simplicity, assume that each user has only one template for each biometric type stored in the system. The template for the ith user, di=(di=(dif,div,dis), consists of three components: difd_i^fdif (fingerprint), divd_i^vdiv (face), and disd_i^sdis (speech).

Let (d0,I)(d_0, I)(d0,I) represent the biometric data provided by a user attempting to access the system, where d0=(d0f,d0v,d0s) contains the measurements of the fingerprint, face, and speech modalities. The claimed identity III belongs to one of two categories:

- mt: The user is genuine, and their claim is correct.

- mf: The user is an impostor, and their claim is false.

The biometric system β\betaβ matches d0d_0d0 against the stored template did_idi for the claimed identity III to determine whether III belongs to category mt or mf.

The claimed identity III falls into one of two categories:

$$I \in \begin{cases} w_1, & \text{if } F(d_0, d_I) > \epsilon \\ w_2, & \text{otherwise} \end{cases} \quad (1)$$

Where:

- w1w_1w1 represents the genuine user category (mtm_tmt), meaning the identity claim is correct.
- w2w_2w2 represents the impostor category (mfm_fmf), meaning the identity claim is false.
- F(d0,dI)F(d_0, d_I)F(d0,dI) is a function that calculates the similarity or match score between the input biometric data (d0d_0d0) and the stored template (dId_IdI).
- ϵ\epsilonϵ is a threshold value. If the similarity score F(d0,dI)F(d_0, d_I)F(d0,dI) exceeds this threshold, the claim is accepted as genuine (w1w_1w1); otherwise, it is rejected as an impostor (w2w_2w2).

This method ensures that the decision is made based on a measurable threshold, providing a systematic way to classify identities.

In biometric systems, determining the identity of a user involves matching the claimed identity against stored templates, resulting in one of four possible outcomes. These outcomes are:

1. **Outcome (i): A claimed identity in mtm_tmt is determined to be in mtm_tmt**
   - This occurs when a genuine user is correctly accepted by the system.
   - This is a *correct* decision.
2. **Outcome (ii): A claimed identity in mtm_tmt is determined to be in mfm_fmf**
   - This happens when a genuine user is mistakenly rejected by the system.
   - This is an *incorrect* decision.
3. **Outcome (iii): A claimed identity in mfm_fmf is determined to be in mfm_fmf**
   - This occurs when an impostor is correctly rejected by the system.
   - This is a *correct* decision.
4. **Outcome (iv): A claimed identity in mfm_fmf is determined to be in mtm_tmt**
   - This happens when an impostor is mistakenly accepted by the system.
   - This is an *incorrect* decision.

Correct vs. Incorrect Decisions:

- **Correct outcomes:** (i) and (iii)
  - A genuine user is accepted or an impostor is rejected.
- **Incorrect outcomes:** (ii) and (iv)
  - A genuine user is rejected, or an impostor is accepted.

Evaluating Biometric System Performance:

Due to natural variations (interclass and intraclass) in biometric indicators and system design challenges (e.g., feature extraction, decision-making), incorrect decisions are unavoidable. To quantify the system's reliability, two key metrics are used:

1. **False Acceptance Rate (FAR):**
   - The probability of outcome (iv), where an impostor is wrongly accepted.
   - A lower FAR indicates better security.
2. **False Rejection Rate (FRR):**
   - The probability of outcome (ii), where a genuine user is wrongly rejected.
   - A lower FRR ensures better user accessibility.

Ideal System:

An ideal biometric system minimizes both FAR and FRR, achieving a balance between security and accessibility. However, in practice, some trade-offs are necessary due to the inherent variability of biometric data and system limitations.

2.2 Fingerprint Verification

A **fingerprint** is characterized by a unique pattern of ridges and furrows on the surface of a fingertip. The distinctiveness of a fingerprint arises from the local ridge features and their spatial relationships. Among these features, the two most prominent **minutiae** are: (i) **ridge endings** and (ii) **ridge bifurcations**.

Fingerprint verification relies heavily on the precise comparison of these minutiae and their spatial relationships to facilitate personal identification. This process typically occurs in two key stages:

1. **Minutiae Extraction**: This stage involves the identification and extraction of minutiae from input fingerprint images. Specialized algorithms are employed to detect these key features, which are critical to the identification process.
2. **Minutiae Matching**: In this stage, the extracted minutiae are compared against a reference set to determine the degree of similarity between two fingerprint patterns. The effectiveness of this

matching process is essential for accurate and reliable fingerprint verification.

Together, these stages enable robust and efficient fingerprint-based identification, which is widely used in security systems and biometric authentication.

$$\mathcal{F}_1(\Phi_1^0, \Phi_1^I) = \frac{100U''}{p} \qquad (2)$$

- Let B0B_0B0 denote the minutiae pattern extracted from the input fingerprint image with the claimed identity.
- Let TTT denote the minutiae pattern stored in the fingerprint database (the template).
- The similarity between the input fingerprint B0B_0B0 and the template TTT is defined by a function that takes into account the minutiae in both the input and the template.

In mathematical terms:

- Let fBf_BfB represent the total number of minutiae in the input fingerprint B0B_0B0.
- Let fTf_TfT represent the total number of minutiae in the template TTT.
- Let UUU denote the total number of corresponding minutiae pairs identified between B0B_0B0 and TTT by the minutiae matching algorithm.

The similarity function between B0B_0B0 and TTT is typically a measure of how many minutiae points from the input fingerprint match with those in the template. The more minutiae pairs (denoted by UUU) that match between the two, the higher the similarity score.

The exact mathematical form of the similarity function could vary, but a general formulation might look like this:

$$\text{Similarity}(B_0, T) = \frac{U}{\sqrt{f_B \cdot f_T}} \qquad (3)$$

Where:

- U is the number of corresponding minutiae pairs found between the input and template.
- fB is the number of minutiae in the input fingerprint.
- fT is the number of minutiae in the stored template.

This formula normalizes the number of matched minutiae by the total number of minutiae in both the input and the template, which helps ensure that the similarity score is not biased by the absolute number of minutiae in each fingerprint but rather by how well the input and template match.

2.3 Face Recognition

In the context of personal identification, **face recognition** refers to the process of identifying or verifying a person based on their facial features, typically using a controlled, full-frontal portrait image. There are two main tasks involved in face recognition:

1. **Face Location**: This task involves detecting the presence of a face in the input image and determining its location.
2. **Face Recognition**: Once a face is located, this task compares the located face with stored templates to determine the identity of the person.

Numerous face recognition techniques have been proposed in the literature, many of which have demonstrated impressive performance. In our system, we employ the **Eigenface approach** for face recognition, as outlined in [9].

The Eigenface-based face recognition method consists of two main stages:

1. Training Stage

In this stage, a set of **orthonormal images** is computed. These images capture the most significant variations in the training facial images and define a **lower-dimensional subspace** (also known as the **eigenspace**) that best represents the distribution of facial features. The training images are then projected into this eigenspace, generating a lower-dimensional representation of each facial image.

2. Operational Stage

During this stage, when a facial image is detected, it is projected into the same eigenspace as the training images. The similarity between the input facial image and the stored templates is then calculated within this eigenspace.

Let ϕ0\phi_0ϕ0 denote the representation of the input face image, which is associated with the claimed identity, and let ϕ2\phi_2ϕ2 represent the stored template in the system. The **similarity function** between ϕ0\phi_0ϕ0 and ϕ2\phi_2ϕ2 is defined as:

$$\mathcal{F}_2(\Phi_2^0, \Phi_2^I) \quad -\|\Phi_2^I - \Phi_2^0\|, \qquad (4)$$

where || • || denotes the £ *norm.*

2.4 Speaker Verification

Anatomical variations that naturally occur among different individuals, along with differences in their learned speaking habits, manifest as variations in the acoustic properties of the speech signal. By analyzing and identifying these differences, it is possible to discriminate among speakers [3].

In our system, we have implemented a **text-dependent speaker recognition** method, which utilizes a **left-to-right Hidden Markov Model (HMM)** of the **10th order Linear Prediction Coefficients (LPC)** derived from the **cepstrum** to perform speaker verification [3].

$$\mathcal{F}_3(\Phi_3^0, \Phi_3^I) \qquad \log\{p(s(1:L)|\Phi_3^I)\} \tag{5}$$

The biometric system uses conditional probabilities, denoted as **p** and **q**, to characterize its decision-making process, particularly in face recognition and speaker verification. These probabilities are essential for determining the likelihood of accurate or inaccurate classifications. Although alternative algorithms for face recognition and speaker verification are available, the system employs specific methods based on the resources accessible in the laboratory.

The primary focus of this paper is not on optimizing the performance of individual biometric modalities but rather on demonstrating the improvement in overall system accuracy through the integration of multiple biometric indicators, such as fingerprint, face, and speech. By combining these modalities, the system capitalizes on the strengths of each to enhance its reliability and accuracy. This approach highlights the significant advantage of multimodal biometric systems in overcoming the limitations of single-modality systems, emphasizing the value of integration over individual optimization.

2.5 Decision Fusion

The final decision made by our system integrates the outputs of the fingerprint verification module, face recognition module, and speaker verification module. If the output of each module is only a categorical label, such as xtx_txt (indicating the claimed identity is true) or ¬xt\neg x_t¬xt (indicating the claimed identity is not true), without any associated confidence value, the integration can only be performed at an abstract level. In this case, a majority rule may be employed to reach a conclusion [12][12][12].

However, if the output of each module is a similarity score, a more accurate decision can be achieved at a rank or measurement level by accumulating the confidence associated with each individual decision. Let AtA_tAt, NtN_tNt, and NsN_sNs represent random variables indicating the similarity (or dissimilarity) between an input and a template for fingerprint verification, face recognition, and speaker verification, respectively.

Let pijp_{ij}pij (where j=1,2,…,nj = 1, 2, \dots, nj=1,2,…,n and i=1,2i = 1, 2i=1,2) denote the class-conditional probability density functions of AtA_tAt, NtN_tNt, and NsN_sNs. Assuming AtA_tAt, NtN_tNt, and NsN_sNs are statistically independent, the joint class conditional probability density function of AtA_tAt, NtN_tNt, and NsN_sNs can be expressed as:

$$p(X_1, X_2, X_3|w_i) = \prod_{\mathrm{v}='}^{3} p_j(X_j|w_i), \quad i = 1, 2. \tag{6}$$

Depending on the application requirements for verification accuracy, various statistical decision theory frameworks can be employed. In biometrics, performance requirements are typically defined in terms of the False Acceptance Rate (FAR). In such cases, the decision fusion scheme should establish a decision boundary that meets the specified FAR while minimizing the False Rejection Rate (FRR).

$$(X_1^0, X_2^0, X_3^0) \in \begin{matrix} \mathrm{mt}, & \text{if } \frac{p_1(X_1^0, X_2^0, X_3^0|w_1)}{p_2(X_1^0, X_2^0, X_3^0|w_2)} > \lambda \\ w_2, & otherwise, \end{matrix} \tag{7}$$

where A is the minimum value that satisfies the following condition:

$$\begin{matrix} \lambda \\ \mathrm{c_0} \end{matrix} \quad \begin{matrix} \frac{p_1(X_1, X_2, X_3|w_1)}{p_2(X_1, X_2, X_3|w_2)} \, and \\ \int_{R_1} p_2(X_1, X_2, X_3|w_2) dX_1 dX_2 dX_3. \end{matrix} \tag{8}$$

3. Performance Evaluation

The performance benchmark assesses the capabil ity o1' the system at the point-of-identification, which depends heavily on how the system is used, whether the users are willing to cooperate, etc. A test which simulates the operating environment is needed to as- sess the performance benchmark of an implemented system. We have evaluated the pertormance of our multimodal biometric system on a small set o1' data which is acquired in a laboratory environment.

3.1 Databases

A training database consisting of fingerprints, face images, and speech samples from 50 users was collected. For each user, 10 fingerprint images (totaling 500 images), 9 face speech samples (totaling 600 samples) were acquired. The fingerprint images were captured using an optical fingerprint scanner, with the condition that the fingers be

placed approximately at the center of the scanner and oriented within a 90° angle. The face images were captured using a Panasonic video camera under normal indoor lighting conditions.

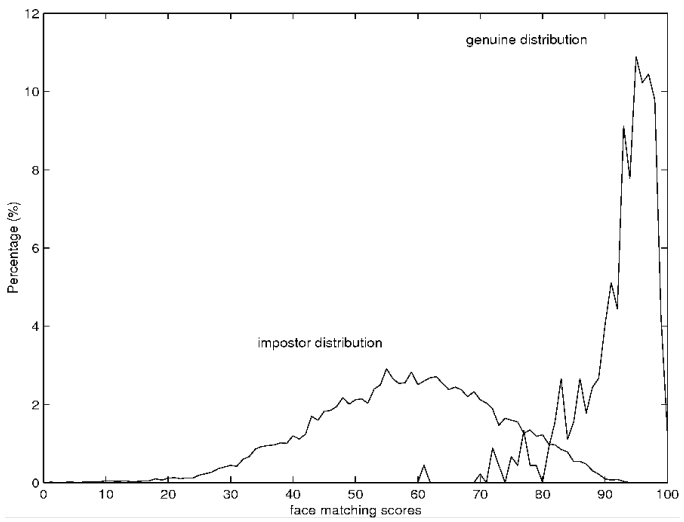
Figure 2: Fingerprint, Face and Speech samples.

For the Training set, data was collected from 50 individuals during the second round of data collection. Each participant contributed 15 fingerprint images (375 total), 15 face images (375 total), and 15 speech samples (375 total). The data collection process spanned three sessions over a period of approximately two weeks.

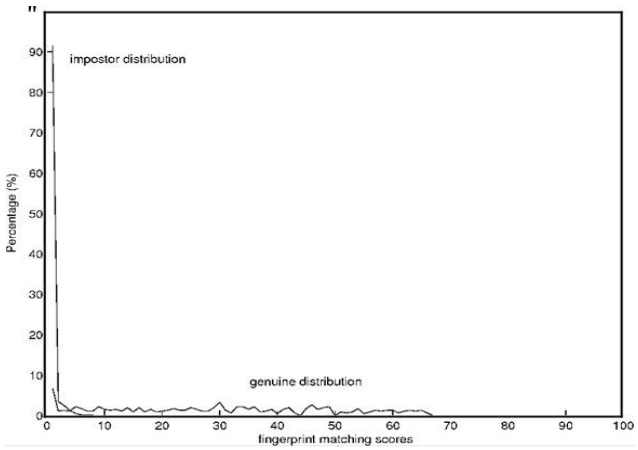
3.2 Benchmarks

The genuine and impostor distributions for each biometric indicator cannot be precisely defined using a known statistical model; instead, they must be estimated from empirical data. In our testing process, an "all-against-all" verification test was conducted on the training database to generate the genuine and impostor distributions. Each distribution was discretized into 100 bins, and the results are depicted in Figure 3. Using these estimated distributions, the decision boundary that meets a specified False Acceptance Rate (FAR) is determined based on the Neyman-Pearson rule, as described in Section 2.

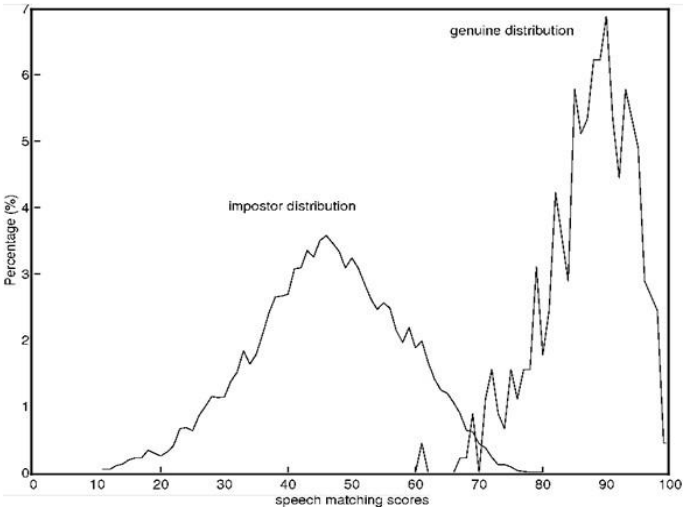
Since the pre-specified FAR for a biometric system is typically very low (less than 0.001), demonstrating compliance with this performance specification requires a large number of representative samples. However, obtaining such a dataset is both costly and time-consuming. Given that our test database contains only a limited number of samples, we adopted a practical workaround. In this approach, a different fingerprint, face, and speech sample are combined to form a probe during each test iteration. While this method addresses the issue of insufficient sample size, it may inadvertently lead to an overestimation of system performance.



(a)



(b)



(c)

Figure 3: Genuine and impostor distributions of face recognition, fingerprint verification, and speaker verification; matching scores are normalized to [0, 100].

However, since the three biometric indicators are independent, this approach appears reasonable. In our test, a total of 36,796 impostor probes and 358 genuine probes were generated and evaluated. The Receiver Operating Characteristic (ROC) curve for decision fusion, along with the ROC curves for face recognition, fingerprint verification, and speaker verification, is shown in Figure 4. In these curves, the authentic acceptance rate (the percentage of genuine individuals accepted, i.e., 1 − False Rejection Rate or FRR) is plotted against the False Acceptance Rate (FAR). From these results, it is evident that integrating fingerprint, face, and speech biometrics significantly improves verification performance.

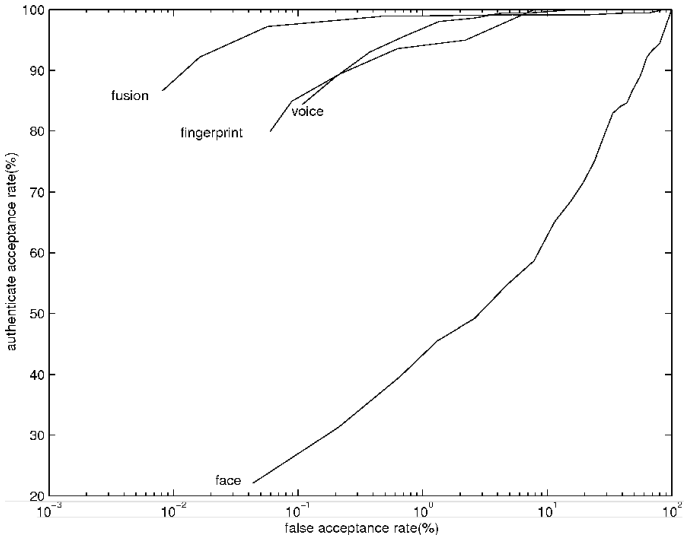


Figure 4: Receiver Operating Curves using Neyman-Pearson rule.

4. Summary and Conclusions

A multimodal biometric technique, which combines multiple biometric modalities for personal identification, effectively addresses the limitations of individual biometric systems. We developed a multimodal biometric system that integrates face recognition, fingerprint verification, and speaker verification to achieve accurate identification. To evaluate the efficiency of this integrated approach, we conducted experiments simulating real-world operating conditions using a small dataset collected in a laboratory environment. The results demonstrated that our system performs exceptionally well. However, further testing on a larger dataset in real-world scenarios is necessary to validate its performance comprehensively

References

[1] Bigun, E. S., Bigun, J., Duc, B., & Fischer, S. (1997, March). Expert conciliation for multimodal person authentication systems by Bayesian statistics. *Proceedings of the International Conference on Audio Video-Based Personal Authentication*, 327–334. Crans-Montana, Switzerland.

[2] Brunelli, R., & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(10), 1042–1052.

[3] J. Campbell. Speaker recognition: A tutorial. *Proceedings oJ the IEEE,* 85(9):1437 1462, September 1997.

[4] R. Chellappa, C. Wilson, and A. Sirohey. Human and machine recognition of' laces: A survey. *Proceedings IEEE,* 83(5):705 740, 1995.

[5] U. Dieckmann, P. Plankensteiner, and T. Wagner. Sesam: A biometric person identification system using sensor fusion. *Pattern* fiecopnition *Letters,* 18(9):827 833, 1997.

[6] L. Hong and A. Jain. Integrating faces and fingerprints for personal identification. In *Proc. 3rd Asian Oonference on Computer Vision,* pages 16—23, Hong Kong, China, 1998.

[7] A. Jain, R. Bolle, and S. Pankanti. *Biometrics: Personal Identi fication in Networked Soci*ety. Kluwer Academic Publishers, Boston, 1998.

8] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. PAMI,* 12(1):103 108, 1990.

9] J. Kittler, Y. Li, J. Matas, and M. U. Sanchez. Combining evidence in multimodal personal identity recognition systems. In *Proc. 1st* Int. *Conf. on Audio Video-Based Personal Authenti*cation, pages 327 334, Crans-Montana, Switzerland, March 1997.

10]. A. Zuev and S. K. Ivanov. The voting as a wayto increase t e ecision re ia i ity. n roc. oun*datious of In Jormatio n/Decision Fusion with Applications to Engineering Problems,* pages 206 210, Washington, D. C., August 1996.