### **Leveraging Multi-Feature Fusion for Fingerprint Verification Using Siamese Networks**

### **Introduction**

Fingerprint verification is a cornerstone of biometric authentication systems, celebrated for its reliability, uniqueness, and ease of use. The distinctiveness of fingerprints, formed through the intricate arrangement of dermal ridges and valleys during prenatal development, has made them a vital tool for identifying individuals [1]. Over the years, technological advancements in machine learning and image processing have significantly improved fingerprint verification systems, even in challenging scenarios involving partial or distorted fingerprints. However, the diversity of fingerprint patterns, variations in image quality, and malicious attempts to bypass these systems pose significant challenges to their robustness and security [2].

Biometric systems, which rely on unique physiological or behavioral characteristics for identification, have become an essential component of modern authentication frameworks. Among the various biometric traits, such as facial features, iris patterns, and voice recognition, fingerprint-based systems are the most widely used due to their distinctiveness, permanence, and accessibility [2]. These systems have been adopted in applications ranging from smartphone security to online payment platforms, replacing traditional authentication methods like passwords and cards that are prone to hacking and theft.

Fingerprint verification operates by matching a scanned fingerprint against a stored template to validate an individual’s identity. Matching approaches are typically classified into three categories: correlation-based, minutiae-based, and pattern-based methods. While correlation-based techniques analyze pixel-to-pixel alignment, and minutiae-based methods focus on specific ridge features, pattern-based approaches examine the global fingerprint structure, such as whorls, arches, and loops [3]. Among these, pattern-based methods have demonstrated greater robustness against image degradation and rotation, making them particularly suitable for modern verification systems.

Despite these advancements, fingerprint verification systems still face noise-related challenges, variations introduced by multiple sensors, and the limitations of individual matching algorithms. Traditional methods like SIFT (Scale-Invariant Feature Transform) effectively capture invariant local features but may fail to utilize the global structure of fingerprints. On the other hand, Convolutional Neural Networks (CNNs) excel at extracting deep hierarchical features but cannot focus on finer details crucial for fingerprint recognition.

This research bridges the gap between traditional and deep learning approaches by proposing a multi-feature fusion model that combines handcrafted and CNN-based features. The fused features are processed through a siamese network, an architecture designed for pairwise similarity learning, to enhance the accuracy and reliability of fingerprint verification. The proposed system leverages SIFT descriptors for capturing local features, CNNs for extracting global representations, and the siamese network for robust similarity measurement, addressing the limitations of existing methods and advancing the state of fingerprint verification technology.

This study not only demonstrates the effectiveness of the multi-feature fusion approach in improving fingerprint verification performance but also lays the groundwork for future research in hybrid models for biometric authentication systems.

### **Literature Review**

Deep learning has revolutionized biometric authentication by enabling automated feature extraction and robust pattern recognition. Convolutional Neural Networks (CNNs) have been particularly effective in fingerprint verification due to their ability to capture intricate patterns and hierarchical features. Wu et al. [6] demonstrated the potential of CNNs in fingerprint recognition, achieving high accuracy in categorizing fingerprint patterns. However, CNNs alone may overlook fine-grained details critical for matching.

Recent studies have explored hybrid models combining traditional feature extraction methods with deep learning architectures to enhance fingerprint verification performance. AlShehri et al. [7] highlighted the robustness of minutiae-based features when integrated with advanced machine learning techniques. Similarly, Wan et al. [10] proposed the XFinger-Net, a deep learning model capable of segmenting noisy and defective fingerprints, demonstrating improved accuracy in challenging scenarios.

Despite significant advancements, fingerprint verification systems face persistent challenges, including noise, variations across sensors, and the lack of large-scale publicly available datasets. Leung and Leung [9] addressed the issue of limited datasets by artificially expanding training samples using spatial modeling, thereby improving classification accuracy. Furthermore, studies have shown that combining local features, such as SIFT descriptors, with global features extracted by CNNs can significantly enhance performance in fingerprint verification systems [6].

The reviewed literature underscores the evolution of fingerprint verification systems, from traditional minutiae-based and pattern-based techniques to advanced deep learning-based hybrid models. While traditional approaches offer a strong foundation, the integration of deep learning and multi-feature fusion addresses the limitations posed by image quality and dataset constraints. This study builds on these advancements by proposing a novel siamese network-based multi-feature fusion approach, combining the strengths of traditional and deep learning techniques to enhance fingerprint verification accuracy and robustness.

### **Methodology**

#### **Data Collection**

Optical and capacitive sensors are commonly employed for fingerprint acquisition due to their strong performance and satisfactory accuracy. However, these sensors may face challenges in certain situations, such as when the user's finger is either unclean or dry.

The data used in this study was collected from a specific group of students at East West University, aged between 19 and 23 years. A total of 750 students participated, with each providing two sets of fingerprint data, resulting in 1,500 data entries.

Throughout the research, a strict ethical protocol was followed to ensure participant safety and maintain the integrity of the study. Prior to the data collection, participants were fully informed about the purpose, goals, and procedures of the study. They were also made aware that participation was voluntary, and their privacy and data confidentiality would be protected. Only participants who willingly gave their consent were involved in the study, and all ethical and research compliance standards set by East West University were adhered to during the process.

#### **Method**

Our approach is designed to deliver end-to-end fingerprint verification by combining the power of SIFT and CNN architectures for superior feature extraction. This hybrid methodology ensures precise and efficient identification by leveraging the strengths of both local and global features. Additionally, our preprocessing steps significantly enhance the system's ability to identify and distinguish between local and global variations, resulting in robust and accurate fingerprint verification.

#### **Model Architecture**

A fingerprint verification system typically comprises three key stages: fingerprint preprocessing, feature extraction, and matching. These stages work sequentially to ensure accurate and reliable verification. A functional block diagram illustrating the main components of the system in a step-by-step flow is presented in Fig 1.

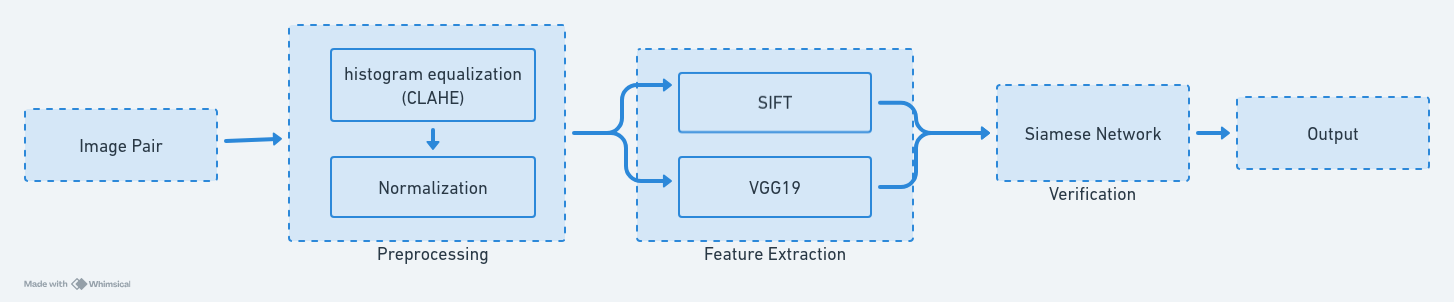


Figure 1. Block Diagram

##### **3.1 Pre-processing**

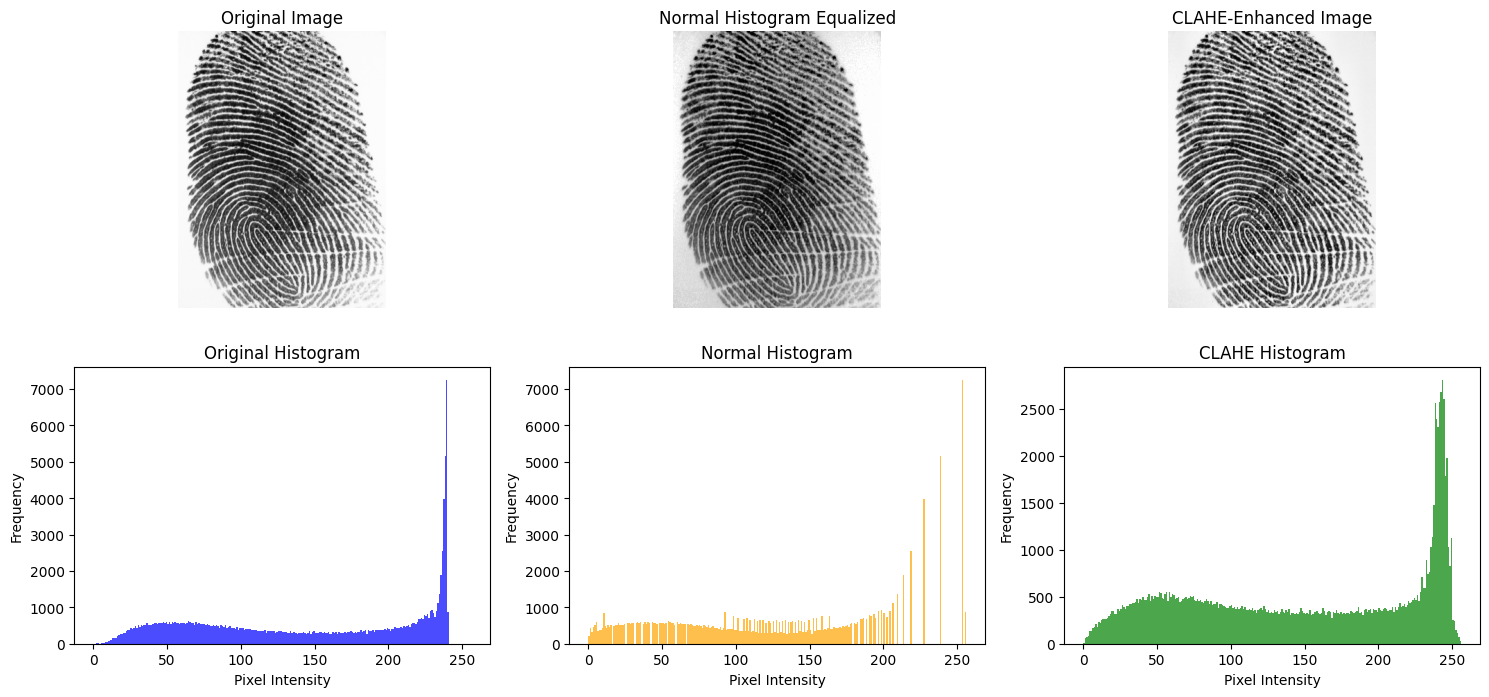
Fingerprint images acquired from sensors or other media may not always be well defined (i.e., are not assured with perfect quality), due to elements of noise that corrupt the clarity of the ridge structures. Consequently, fingerprint enhancement methods are usually applied to reduce the noise present in the image and enlighten the clarity and continuity of ridge and valley structures. This greatly helps in better extraction of potentially relevant features and finding stable and robust feature matches. In this work, as an initial pre-processing step, an adaptive local–global technique for contrast image enhancement based on local histogram equalization is carried out to increase the contrast between ridges and valleys and to connect the false broken ridges due to an excessive or inadequate amount of ink shown in fig 2.



Figure 2: Fingerprint with inadequate amount of ink

###### **3.1.1. Histogram Equalization**

Histogram equalization is a process that aims to distribute the gray levels in the image so that they are equally distributed across their range. This process efficiently resets the brightness value of every pixel based on the image histogram and intends to extend the pixel value distribution to increase the perceptional information. In this work, contrast limited adaptive histogram equalization (CLAHE) is employed. While traditional histogram equalization works on the entire image, the CLAHE operates locally on the image in small regions called tiles. CLAHE increases the contrast of small tiles and incorporates the neighboring tiles in the image using bilinear interpolation which removes the boundaries that are artificially caused. Additionally, the ’clip limit’ factor is applied to avoid excessively saturating the image, specifically the inhomogeneous areas that have high peaks in the histogram of certain image tiles due to many pixels that fall within the same range of gray level [[1](https://www.mdpi.com/2076-3417/12/4/2028#B16-applsci-12-02028)0]. [Fig](https://www.mdpi.com/2076-3417/12/4/2028#fig_body_display_applsci-12-02028-f004) 3 shows the histogram-equalized results of a fingerprint image where improvements in average intensity and contrast are apparent. Furthermore, the histogram distributed over the entire intensity increases the contrast and the average intensity level in the equalized image histogram is higher (lighter) than the original intensity level [[1](https://www.mdpi.com/2076-3417/12/4/2028#B17-applsci-12-02028)1].

Figure 3: Comparison of original, normal histogram equalized and CLAHE-Enhanced image

###### **3.1.2. Normalization**

Normalization is a process of adjusting the pixel intensity values range; it is also called contrast stretching. This process is an easy and significant pre-processing step to enhance the quality of the image by eliminating noise from the image. The image normalization consists of changing the intensity of each pixel, hence which means changing the entire image to some of the pre-defined values. Normalization maintains the clarity and contrast of the ridges and valley structure and it is a pixel-wise operation [[1](https://www.mdpi.com/2076-3417/12/4/2028#B18-applsci-12-02028)7]. The normalized image 𝑁(𝑖,𝑗) is defined in equation (1)

*(1)*

so that new intensities of the pixel for the normalized image would mostly be between −1 and 1, making the subsequent calculations easier.

##### **3.2 Feature Extraction**

Effective and reliable feature extraction is the most crucial and substantial issue to the final accuracy of the feature matching process, which heavily depends on the perfection of the previous pre-processing steps [13]. In this work, an optimized combination of the CNN and SIFT algorithm is presented, which is potentially expected to generate a set of most represented features with high repeatability and excellent matching properties.

###### **3.2.1 SIFT**

The Scale-Invariant Feature Transform (SIFT) algorithm is a widely used technique for detecting and describing local features in images. It is highly robust to variations in scale and rotation, making it effective for various computer vision applications. The process begins with scale-space construction, where a Difference of Gaussians (DoG) is computed to identify key points at different scales using the equation shown in equation (2).

*(2)*

Here G(x,y,σ) represents the Gaussian function, σ is the scale, k is a constant multiplier, and I(x,y) is the input image. After detecting key points, keypoint localization refines them by eliminating unstable points to ensure high repeatability. To achieve rotation invariance, orientation assignment is performed by analyzing local gradient directions, computed in equation (3).

*(3)*

Here L(x,y) is the Gaussian-smoothed image. SIFT focuses on extracting local features, which are highly robust to variations in scale, rotation, and noise, making it an excellent choice for detailed local analysis shown in fig 4.

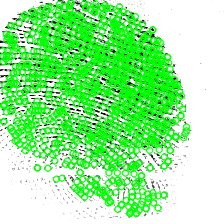
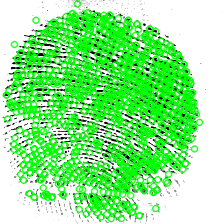


Figure 4: Local features extracted by SIFT

###### **3.2.2 CNN (VGG19)**

The Convolutional Neural Network (CNN) used in this work is based on the VGG19 architecture, which is renowned for its deep structure and ability to extract hierarchical features from images. VGG19 consists of 19 layers, including convolutional layers, pooling layers, and fully connected layers, and is pre-trained on large-scale datasets such as ImageNet. The feature extraction process involves layer-wise feature extraction. Lower layers capture simple patterns like edges and textures, while deeper layers identify more complex features such as shapes and objects. The feature map output F can be expressed as equation (4).

*(4)*

Here Wl,i represents the filter weights, Fl−1,i is the input feature map, and bl is the bias term. By utilizing a pre-trained VGG19 model, domain-specific features are extracted without requiring extensive re-training.

**Feature Fusion**: The high-level semantic features extracted by CNN complement the local invariant features detected by SIFT, leading to a more comprehensive feature representation. CNN focuses on extracting global features that capture the overall semantic information of the image, providing a broader contextual understanding. The combination of SIFT and CNN (VGG19) leverages the strengths of both approaches, with SIFT excelling in local feature extraction and CNN providing robust global feature representation. This synergy results in robust and highly descriptive feature sets that improve the accuracy and reliability of feature matching.

##### **3.3 Verification**

The verification stage is the final and most critical step in the fingerprint verification system, where the extracted features are used to determine if two fingerprint images belong to the same individual. In this work, the combined features obtained from SIFT and the VGG19-based CNN are utilized as input to a Siamese network, designed specifically for one-shot learning and similarity comparison tasks.

###### **3.3.1 Siamese Network Architecture**

A Siamese network is a neural network architecture that consists of two identical sub-networks, sharing the same weights and parameters. These sub-networks process two input feature sets in parallel and output a similarity score, which quantifies the likelihood that the two inputs belong to the same source. The architecture of the siamese network used in this work is shown in fig 5.

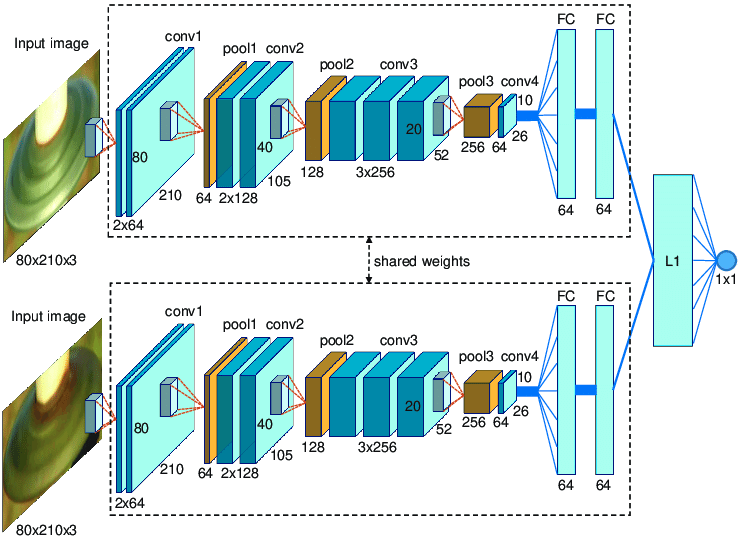


Figure 5: Siamese network architecture

**Input Layers:** The combined feature vectors of SIFT and CNN for the two fingerprints being compared.

**Shared Layers:** A series of dense layers with shared weights to process the input feature vectors and extract high-level embeddings.

**Distance Metric Layer:** Computes the absolute difference between the feature embeddings generated by the two sub-networks. This step ensures symmetry and invariance to input order.

**Output Layer:** A sigmoid activation function outputs a similarity score ranging from 0 to 1, where a score closer to 1 indicates a high likelihood of the fingerprints being from the same source, while a score closer to 0 suggests otherwise.

###### **3.3.2 Training the Siamese Network**

To train the Siamese network, a labeled dataset containing pairs of fingerprints is used. Each pair is assigned a label of 1 if the fingerprints are from the same individual and 0 otherwise. The training process involves:

**Loss Function:** The contrastive loss function is employed to minimize the distance between embeddings of matching pairs while maximizing the distance for non-matching pairs is shown in *equation (5).*

*(5)*

Here Y is the label (1 for a match, 0 otherwise), D is the Euclidean distance between the two embeddings, and m is the margin parameter that defines the minimum distance between non-matching pairs.

**Optimization:** The Adam optimizer is used to update network weights based on the computed gradients, ensuring efficient convergence.

**Validation:** A validation set is used to monitor performance and prevent overfitting during training.

### **Results and Discussion**

The Siamese network is employed for the verification process by analyzing two fingerprint feature sets and generating a similarity score. The decision on whether the fingerprints belong to the same individual is determined by a predefined threshold.

**Match (Same Individual):** If similarity score(S) ≥ threshold(T), it means the fingerprints are similar enough to be classified as belonging to the same individual. If S < T, they are considered different.

**Mismatch (Different Individuals):** If S<T, the fingerprints are classified as belonging to different individuals.

To evaluate the effectiveness of the proposed hybrid feature extraction method, which combines SIFT and CNN, its performance is compared against systems utilizing only CNN for feature extraction.

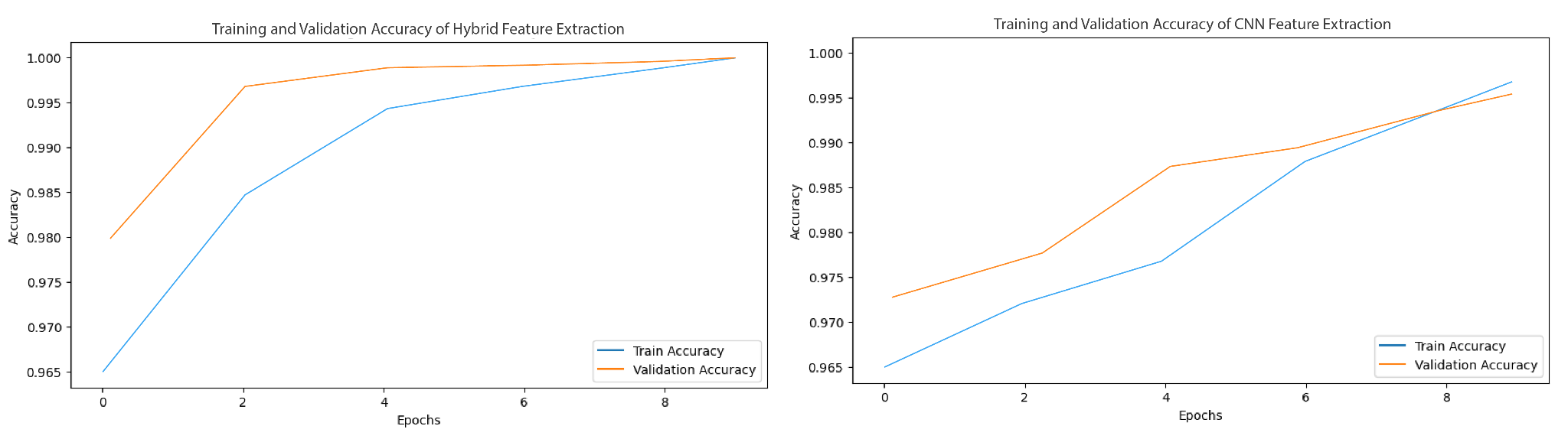


Figure 6: Comparison of training and validation accuracy of hybrid feature extraction and CNN feature extraction

Fig 6 compares training and validation accuracy of Hybrid Feature Extraction and CNN Feature Extraction for fingerprint verification. The hybrid approach, incorporating SIFT feature extraction, enables faster training, requiring less computational power and converging in fewer epochs. It achieves near-perfect accuracy quickly with better generalization, while the CNN-based method improves more gradually and shows slight overfitting.

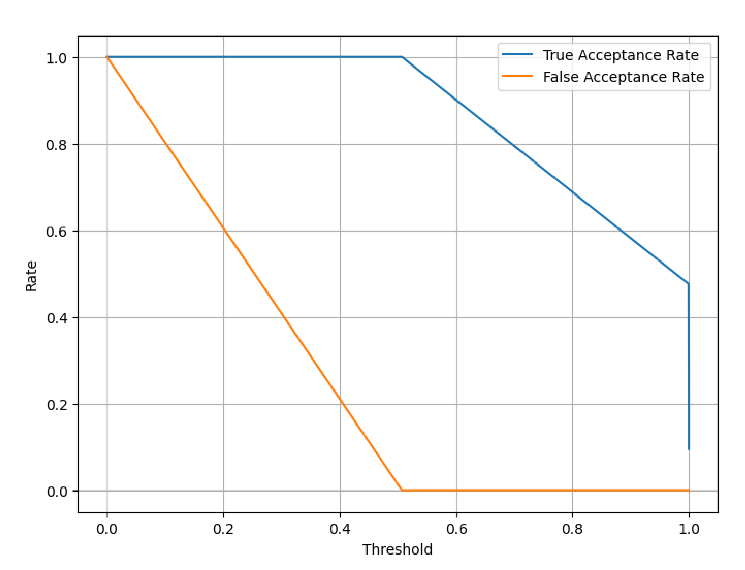
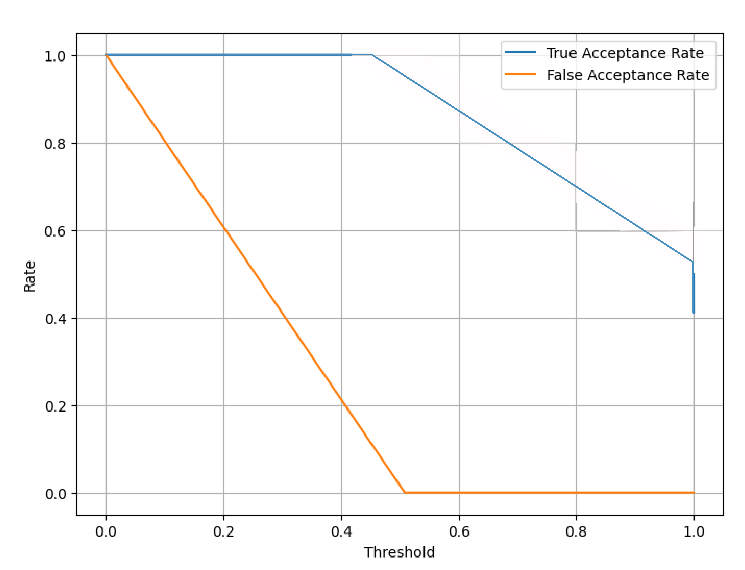


Figure 7: Comparison of True Acceptance Rate and False Acceptance Rate of (a) CNN Feature Extraction and (b) SIFT-CNN Feature Extraction

The comparative analysis of SIFT-CNN hybrid feature extraction in Fig 7(a) and CNN-based feature extraction in fig 7(b) reveals that the hybrid approach offers superior performance in biometric verification. The SIFT-CNN model achieves a True Acceptance Rate (TAR) of approximately 98.5% at a threshold of 0.6, while its False Acceptance Rate (FAR) drops to nearly 0% at the same threshold. In contrast, the CNN-based method achieves a TAR of around 96.8% at the same threshold, but its FAR remains slightly higher at approximately 0.5%. Additionally, the Equal Error Rate (EER) for the SIFT-CNN approach is around 1.2%, whereas the CNN-based method has an EER of approximately 2.5%, indicating that the hybrid model achieves better overall classification accuracy. This suggests that integrating SIFT with CNN enhances feature representation, leading to improved discrimination between genuine and imposter samples. In contrast, the CNN-based feature extraction, though effective, shows a slightly higher EER and a slower decline in FAR, making it marginally less reliable in reducing misclassification errors.

#### **Discussion**

These results highlight the complementary strengths of SIFT and CNN in feature extraction. SIFT Excels at extracting stable local features, such as ridge endings and bifurcations, which are invariant to transformations like scaling and rotation. On the other hand, CNN captures high-level global features that provide a broader contextual understanding of the fingerprint image. By combining these methods, the hybrid approach mitigates the limitations of each individual technique. For instance, SIFT alone struggles with complex image variations, while CNN alone overlooks critical local details. The synergy achieved through this hybrid approach ensures a balance between local detail and global context, leading to improved matching accuracy and robustness. These findings highlight the advantages of hybrid feature extraction, particularly in applications where high precision and security are critical.

#### **Conclusion**

The experimental results demonstrate that the hybrid SIFT-CNN feature extraction method, coupled with a Siamese network for verification, delivers superior performance compared to single-method approaches. This improvement is particularly evident in scenarios involving noisy or low-quality fingerprint images, making the hybrid approach a promising solution for robust and accurate fingerprint verification.

#### **Reference**

1. Praseetha, V.M., Bayezeed, Saad and Vadivel, S.. "Secure Fingerprint Authentication Using Deep Learning and Minutiae Verification" *Journal of Intelligent Systems*, vol. 29, no. 1, 2020, pp. 1379-1387.<https://doi.org/10.1515/jisys-2018-0289>
2. Bakheet, S.; Al-Hamadi, A.; Youssef, R. A Fingerprint-Based Verification Framework Using Harris and SURF Feature Detection Algorithms. *Appl. Sci.* 2022, *12*, 2028. https://doi.org/10.3390/app12042028
3. M. Nur-A-Alam, M. Ahsan, M. A. Based, J. Haider, and M. Kowalski, "An intelligent system for automatic fingerprint identification using feature fusion by Gabor filter and deep learning," *Computers and Electrical Engineering*, vol. 95, p. 107387, 2021. doi: 10.1016/j.compeleceng.2021.107387.
4. Cherrat EM, Alaoui R, Bouzahir H. Convolutional neural networks approach for multimodal biometric identification system using the fusion of fingerprint, finger-vein and face images. PeerJ Comput Sci. 2020 Jan 6;6:e248. doi: 10.7717/peerj-cs.248. PMID: 33816900; PMCID: PMC7924518.
5. Abernethy, M., Rai, S.M. (2014). Cooperative Feature Level Data Fusion for Authentication Using Neural Networks. In: Loo, C.K., Yap, K.S., Wong, K.W., Teoh, A., Huang, K. (eds) Neural Information Processing. ICONIP 2014. Lecture Notes in Computer Science, vol 8834. Springer, Cham. <https://doi.org/10.1007/978-3-319-12637-1_72>
6. AlShehri, Helala. “Cross-Sensor Fingerprint Matching Method Based on Orientation, Gradient, and Gabor-HoG Descriptors With Score Level Fusion.” IEEE Access (2018): n. pag. Print.
7. AlShehri, H.; Hussain, M.; AboAlSamh, H.; AlZuair, M. A Large-Scale Study of Fingerprint Matching Systems for Sensor Interoperability Problem. *Sensors* 2018, *18*, 1008. https://doi.org/10.3390/s18041008
8. Wang Y, Shi D, Zhou W. Convolutional Neural Network Approach Based on Multimodal Biometric System with Fusion of Face and Finger Vein Features. Sensors (Basel). 2022 Aug 12;22(16):6039. doi: 10.3390/s22166039. PMID: 36015799; PMCID: PMC9412820.
9. Kumar, Naresh, and Parag Verma. "Fingerprint image enhancement and minutia matching." *International Journal of Engineering Sciences & Emerging Technologies* 2.2 (2012): 37-42.
10. Sepasian, M., W. Balachandran, and C. Mares. "Image enhancement for fingerprint minutiae-based algorithms using CLAHE, standard deviation analysis and sliding neighborhood." *Proceedings of the World congress on Engineering and Computer Science*. 2008.
11. Bhargava, Neeraj, Anchal Kumawat, and Ritu Bhargava. "Fingerprint matching of normalized image based on Euclidean distance." *International Journal of Computer Applications* 120.24 (2015): 20-23.
12. Geteneh, Ashenafi. *Designing fingerprint based verification system using image processing: the case of national examination*. Diss. 2019.