Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU34640

Robustness Evaluation and Feature Analysis of CNN-Based Fingerprint Recognition under Real-World Degradations

Kexin Zhang

High School Affiliated to Nanjing Normal University, Nanjing, China
Zhangkexin2027@163.com

Abstract. One of the difficulties in high-precision recognition of low-quality, moist or surgically altered fingerprints for personal identification is that it is not easy. To solve the above problem, this paper puts forward a fingerprint recognition and evaluation method based on a Convolutional Neural Network (CNN). A model was constructed and trained based on the Sokoto Coventry Fingerprint (SOCOFing) dataset, and then the other two external benchmark datasets were used to evaluate its generalization performance. The methods of experiments were to test the basic performance, robustness, cross-dataset evaluation, etc. Grad-CAM was also used to show the attention distribution at different levels of features and determine the reasons for the drop in performance. As shown in the above results, although the model has reached an accuracy of over 96% on the clean data set, this value is as low as 61% in the presence of severe data degradation and is even lower after cross-dataset testing due to domain shift. Based on the above visualisation, it can be seen that the network relies too much on local texture features and is easily damaged by noise in these features. Therefore, many kinds of data augmentation methods have been introduced to address this problem. The new model is relatively strong and can be generalized to some extent if some parts fail.

Keywords: Fingerprint Recognition, Convolutional Neural Network, Model Robustness, Data Augmentation, Cross-Dataset Generalization

1. Introduction

Fingerprint identification is now generally used to identify people because of their individual characteristics and stability. Now that it is a modern digital age, forensics have started to be used for various purposes in smartphones and the financial sector. Fingerprint recognition can simplify the complex ridge patterns of a person's fingers into digital data, and then this data is compared with one stored in a database. Its practical value is to provide a non-transferable form of identification for the public and ensure the safety of the people. The main problem is that it is still not easy to achieve high accuracy in the face of low-quality, wet or damaged fingerprints. Therefore, more elaborate calculations are required.

Due to the above reasons, the field has undergone a paradigm shift from manual feature extraction and started to use deep learning. At present, the three main types of AI-driven methods are: First of all, ResNet-50 and Siamese Networks are good at extracting hierarchical features, and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU34640

they have the advantage of a small number of training samples for verification [1, 2]. Secondly, in order to solve the problem of image degradation, Generative Adversarial Networks (GANs) and semantic segmentation models, such as DeepLabV3+, [3, 4] have been employed to reconstruct high-resolution ridges and extract features from very noisy backgrounds. Thirdly, to solve the problem of local receptive fields in conventional Convolutional Neural Networks (CNNs), a new architecture, Vision Transformer (ViT), that can obtain global dependencies has been proposed [5].

A optimized generalization fingerprint recognition workflow that can meet actual conditions is introduced in this paper. The first dataset used in this paper is SOCOFing [6]; it is very suitable for this purpose because it has a large number of artificially modified images of simulated damaged or wet fingerprints (smudging and rotation), more than 6000. In order to expand the scope of application of this paper, tests have also been carried out on the standard benchmark [7, 8] by combining the FVC2002 and forensic fingerprint datasets. A balanced model capable of real-time inference and efficient on edge devices was developed on the Edge Impulse platform. The ways this paper will do so are data augmentation to increase robustness and overfitting analysis. The following sections will introduce the details of data preprocessing, network construction, experimental results analysis, etc.

2. Methods

2.1. Dataset preparation

The three public fingerprint datasets in Figures 1, 2 and 3 are the SOCOFing dataset, the FVC2002 dataset and a forensic fingerprint dataset for this paper. The rationale for selecting the above datasets are that they are in a controlled environment and are based on real conditions.

Given that it is the primary training and baseline assessment data for the model, the SOCOFing dataset has more than 6,000 grayscale fingerprint images and includes data from 600 subjects. Each subject gave a few samples of different fingers. Both the original (unmodified) and artificially modified images are in the dataset, and they are divided into three difficulty levels: easy, medium and hard. The above altered images are in the form of real-world degradation effects, such as noise, occlusion and distortion, so they are very suitable for robustness analysis.

FVC2002 is an example set of fingerprint recognition data. It is a group of high-quality grayscale pictures obtained by many sensors, and in the general case, they are in a regular and standardized environment for model evaluation. Latent fingerprint images in forensic fingerprint datasets are often partial, noisy and low in quality. The above attributes can help know how well the model works in real forensic investigation.



Figure 1. Samples from forensic fingerprints dataset (picture credit: original)

DOI: 10.54254/2755-2721/2026.GU34640



Figure 2. Samples from FVC2002 fingerprints dataset (picture credit: original)



Figure 3. Samples from socofing fingerprints dataset (picture credit: original)

For the purposes of this study, the classification task is a three-class problem of fingerprint patterns: arch, loop, and whorl. All the pictures were divided according to this classification.

To ensure the fairness and regularity of the evaluation, only the original images were used, and then the SOCOFing dataset was divided into three parts: 70% for training, 15% for validation, and 15% for testing. The modified images (easy, medium and hard) were not used in the training process and were only for robustness evaluation. FVC2002 and the forensic dataset were only used as external test sets for the cross-dataset generalization test.

All pictures have been processed in the same way. First, resize the images to 96×96 pixels for uniformity and to reduce the number of data. Secondly, all the images have been converted to grayscale to reduce the number of features and lower the input dimension. Finally, to improve the stability of training and promote convergence, normalization of the pixel values was carried out. The preprocessing of the above has been completed in the Edge Impulse platform.

2.2. Edge impulse-based CNN model

Convolutional Neural Network: It was adopted as the core model for fingerprint classification, due to its strong ability to learn hierarchical spatial features ; thus, it is the primary model for fingerprint classification. Convolution filters are utilized to extract some local features that can help represent the structure of fingerprint ridges in a CNN so that local information can be obtained from it. Then a pooling layer is used to reduce the dimension of space and retain only the essential information; otherwise, it would be too large in terms of computation and overfit. Finally, a fully connected layer is used to map the extracted features to the output classes and assign one of the given categories.

Edge Impulse Platform: The model was implemented using the Edge Impulse platform, an end-to-end machine learning development environment tailored for edge artificial intelligence

DOI: 10.54254/2755-2721/2026.GU34640

applications. Edge Impulse has taken charge of all links in the data pipeline to collect, process, extract features from, train and evaluate models. One of its key advantages is its capacity to deploy optimized models on resource-constrained devices while maintaining real-time inference performance. Additionally, the platform streamlines experimentation by offering a user-friendly interface for adjusting model parameters and visualizing results.

Implementation Details: The CNN model is in the Image Classification module of Edge Impulse. The Size of the input image is 96×96 pixels and it is a single-channel grayscale image. The three output classes of the classification task are arch, loop and whorl.

To analyze the impact of training parameters, two configurations were implemented. The two sets of training parameters are as follows: In the first arrangement, a relatively large learning rate (0.0005) and a small number of epochs (10) were used; although it may converge quickly, it is also easy to overfit. Second, its small learning rate is 0.0001 and it has a relatively large number of epochs (30-50) for stable training and improved generalization.

Model performance was evaluated using multiple metrics. The main index of performance was overall classification accuracy. Confusion matrices was used to examine the performance of different classes and some common errors of misclassification. False positives and false negatives are relatively serious problems that should not be ignored in the application of biometrics for security.

2.3. Experimental design

The three stages of the experiment are shown below, and the corresponding results under various circumstances will be presented systematically.

First, the baseline performance is to be evaluated on the clean test set of the SOCOFing dataset. In the experiment, it is possible to see how well the model can recognise fingerprints in ideal circumstances under good conditions with high-quality, consistent images. The results serve as a reference for subsequent comparisons.

Secondly, some modifications were made to the fingerprint images in the SOCOFing dataset, and then reliable tests were carried out. The trained model was used to directly predict the type of damage (light, medium and severe) in the image without further training. This will help us know how sensitive the model is to noise, distortion and partial occlusion. Performance degradation across these levels provides insights into the model's limitations under realistic conditions.

Lastly, there were some cross-dataset tests to check for generalisation. The model trained on SOCOFing was then used for the FVC2002 and forensic datasets. The model is trained on new data that has a different distribution due to all kinds of acquisition circumstances, images, noise, etc. There will be a significant drop in performance, and it can be seen from this that the deployment of fingerprint recognition systems in all areas of life faces difficulties.

Collectively, these three experiments provide a comprehensive evaluation of classification accuracy, robustness, and generalization, laying the foundation for subsequent analysis and discussion.

3. Results and discussion

3.1. Baseline performance and feature learning behavior

To know how efficient the new fingerprint recognition system is, it is possible to take different models and provide the performance reference for a comparison. Based on the above experiments, it

DOI: 10.54254/2755-2721/2026.GU34640

can be concluded that there is a clear correlation between hyperparameter tuning and model performance. Model A was trained for 10 epochs at a learning rate of 0.0005, and at the beginning, it had reached a accuracy of 94.8%. Further optimization process was carried out to reduce the learning rate to 0.0001 and extend the number of training epochs to 30 for Model B. The above arrangement has an accuracy of 96.2%. Based on the above results, it can be seen that a relatively small learning rate and a relatively large number of training iterations help the model better converge and obtain fine-grained ridge features that are suitable for fingerprint identification.

This configuration yielded a superior performance, and then the CNN model has reached a peak classification accuracy of 96.2% on the clean SOCOFing test set. Therefore, it can be concluded that the model has learned the discriminant spatial features of fingerprint ridge patterns, such as local structure, ridge flow, bifurcation and curvature.

In addition to the overall accuracy, the confusion matrix (Figure 4) shows some subtle class-dependent behavior. Loop Patterns have been classified with relatively higher accuracy; Arch Patterns exhibit a higher misclassification rate. This discrepancy can be attributed to the inherent structural complexity of the classes. Loop and whorl patterns contain more distinctive directional flows and curvature variations, which are easier for convolutional filters to capture. In contrast, arch patterns are relatively simple and lack strong directional features, making them more prone to confusion with other classes.

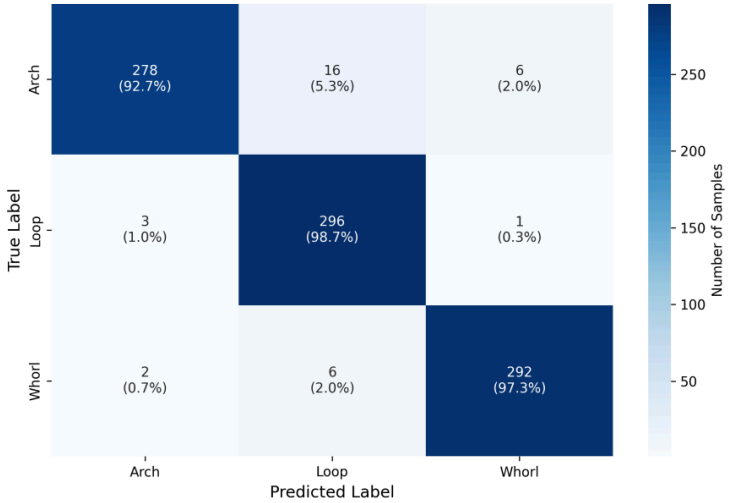

Figure 4. Confusion matrix (clean test set) (picture credit: original)

Confusion Matrix of the CNN Model on the Clean SOCOFing Test Set. Most of the samples were classified correctly, and only a few were confusing for arch and loop patterns.

In terms of representation learning, it can be seen that CNNs mainly rely on local texture gradients and curvature cues and do not consider the entire global structure. It is adequate under clean environment, but when the local conditions are degraded, this reliance becomes a limiting factor.

3.2. Robustness degradation: sensitivity to local feature corruption

Table 1: Performance Degradation of the CNN Model with Reduced Quality of Fingerprint Images. There will be a relatively large drop in the medium and hard situations.

DOI: 10.54254/2755-2721/2026.GU34640

Table 1. Performance degradation of the CNN model with reduced quality of fingerprint images. There will be a relatively large drop in the medium and hard situations

Condition	Accuracy(%)	Relative Drop(%)
Clean	96.2	-
Easy	77.5	-18.7
Medium	75.3	-20.9
Hard	61.7	-34.5

As shown in the results of the robustness test in Table 1, with an increase in the degree of image degradation, the corresponding recognition accuracy gradually drops from 96.2% for a clear image to 61.7% for a serious one. The decline in signal strength is not only a loss of signal quality but also a defect in the feature-extraction limitation of the model.

Under degraded conditions, noise and distortion will severely damage the detailed information of the ridges that are employed by CNNs for classification. Convolutional Layers operated on local receptive field, and if there is any damage to the local pattern, such as a broken ridge, a smudge or partial occlusion, it will directly affect the features. Therefore, the learned filters failed be employed reliably leading to unstable predictions.

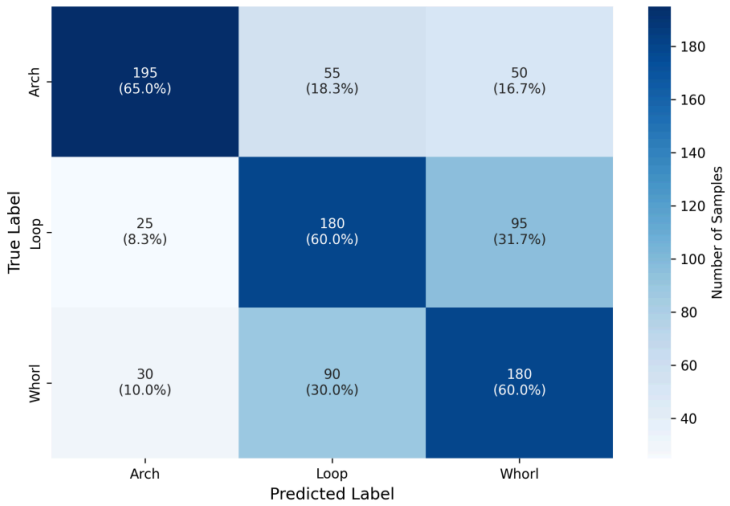

Figure 5. Confusion matrix after hard degradation (picture credit: original)

Figure 5 is the confusion matrix under hard degradation and it further supports the observation. There is a relatively significant increase in the misclassifications of loop and whorl patterns. It can be seen that both classes have the similar local ridge structure, and the difference between them is often reflected in the global flow pattern of the ridge. If there are damage to the local features, the model will be unable to infer the corresponding whole-part relationship and thus will work ambiguously.

Therefore, it highlights that standard CNNs are not very capable at capturing long-range dependencies. Unlike transformer-based architectures, which explicitly capture global context, CNNs are inherently biased toward local feature extraction. Therefore, the robustness is not suitable for situations that need to understand the whole of global structure, but constrained in scenarios.

DOI: 10.54254/2755-2721/2026.GU34640

3.3. Cross-dataset generalization: domain shift and distribution mismatch

Cross-dataset evaluation shows that there is a relatively large drop in performance for the model on FVC2002 (71.4%) and the forensic dataset (63.8%). This result reflects a classic domain shift problem, where the training and testing data distributions differ substantially.

Rationales for the loss in performance are:

1) Sensor variability: Different datasets can be obtained from different fingerprint sensors, and thus there will be variations in resolution, contrast and noise.

2) Acquisition Conditions: Typically, the controlled dataset (SOCOFing) is relatively clean and well-aligned for fingerprints whereas forensic datasets include partial and distorted prints.

3) Data distribution bias: he model learns dataset-specific features that do not generalize well to unseen data.

From a machine learning perspective, this indicates that the model has learned a narrow feature space representation, optimized for the SOCOFing distribution but not invariant to external variations. The sharp performance drop on the forensic dataset further emphasizes the gap between laboratory conditions and real-world applications.

3.4. Overfitting and optimization dynamics

The training dynamics reveal clear evidence of overfitting in models trained with higher learning rates and fewer epochs. Specifically, a divergence between training and validation accuracy is observed. It indicates that the model memorizes training data rather than learning generalizable features.

At the same time, the configuration with a lower learning rate and more epochs demonstrates smoother convergence and reduced overfitting. It can be seen that a relatively small learning rate allows the optimizer to explore the loss landscape more gradually, so it avoids sharp minima and thus performs poorly on new data.

Even in the optimised case, some degree of overfitting persists. Therefore, the reason for the small result is not only due to the hyperparameter setting but also reflects the limitations of the intrinsic capacity of the model and dataset diversity.

4. Feature explanation and model improvement

4.1. Feature-level explanation via visualization overfitting and optimization dynamics

To better understand the underlying causes of performance degradation, feature-level visualization is conducted to analyze how the CNN model processes fingerprint images under different conditions.

4.1.1. Methodology

Gradient-weighted Class Activation Mapping (Grad-CAM) is employed to show which parts of the input image have been used by the model to make a certain prediction (Figure 6). The three typical cases to be analyzed are as follows: 1) Correctly classified clean images, 2) Correctly classified degraded images, and 3) Misclassified images in the degraded condition.

DOI: 10.54254/2755-2721/2026.GU34640



Figure 6. Grad-CAM visualisation of model attention in different scenarios. Red areas are very influential in the classification decision (picture credit: original)

4.1.2. Analysis

As shown by the visualisation results, the attention pattern of the model has changed:

Clean Image: The model focuses on the continuous ridge structure and curvature area regions which are highly discriminative for classification.

Degraded but correctly classified images: The attention is more scattered, but it is still partially aligned with the ridge pattern; thus, the feature confidence has dropped.

Misclassified Images: The model often attends to irrelevant regions, such as background noise or broken ridge fragments, rather than coherent fingerprint structures.

This behavior suggests that the CNN primarily relies on local texture features, such as ridge continuity and edge gradients. When these features are disrupted by noise or distortion, the model fails to extract meaningful representations, leading to incorrect predictions.

The rationale can be attributed to the local receptive field of CNN is biased in terms of theory. CNNs do not have an inherent mechanism to address the problem of long-range spatial correlation; thus, if local feature information is corrupted, it can make the model vulnerable.

4.2. Controlled robustness analysis via augmentation

4.2.1. Methodology

Artificial Noise and Distortion are added to the test images at different degrees of degradation. The following are the transformations [9]: Gaussian noise, Gaussian blur, Rotation perturbation.

According to the experiment results, the better the integrity of the data, the higher its recognition accuracy will be; that is, it is resistant to external environmental disturbances. At a baseline of 0% noise, the model achieves a peak accuracy of 96.2%; but when controlled noise was added to simulate the signal degradation of real life, the accuracy of the model reflects a steady decline, and the following results were obtained: at a noise level of 10%, it was 85.7%; at 20%, it was 71.5%; and ultimately, at a noise level of 30%, it fell to 54.8%.

The noise level is gradually increased to evaluate performance sensitivity (Figure 7).

DOI: 10.54254/2755-2721/2026.GU34640

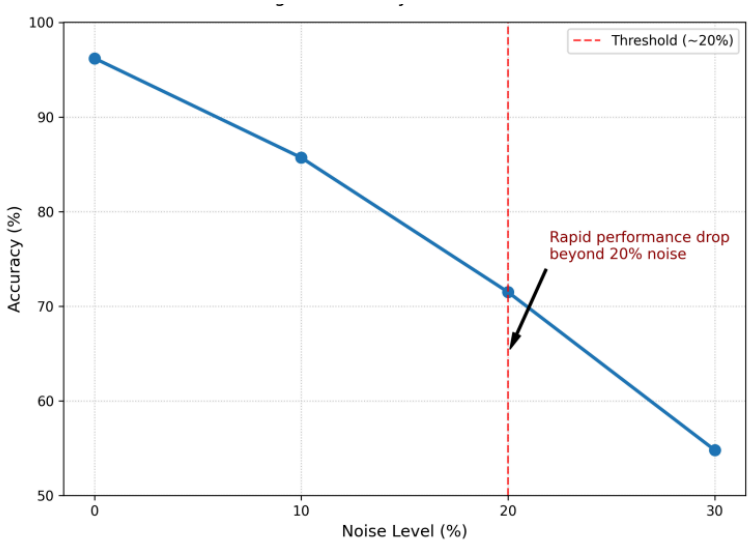

Figure 7. Accuracy vs. noise level (picture credit: original)

4.2.2. Analysis

As the level of noise increases, so does a decrease in performance; it is monotonic. It is to be noted that the model has a threshold effect and declines sharply after a certain amount of noise (around 20%). Therefore, the model has limited tolerance to feature modification.

This controlled experiment confirms that the degradation observed in Section 3 is not dataset-specific but is fundamentally linked to the model's sensitivity to local feature corruption.

4.3. Improvement via data augmentation

4.3.1. Methodology

The augmentations of the training data are: random rotation, addition of Gaussian noise, blurring and horizontal flipping. Augmentations are to expose the model to a wider range of variations and help it learn more robust and generalizable features [10].

Table 2. Comparison of performance before and after augmentation

Model	Clean (%)	Hard (%)	FVC2002 (%)
Baseline CNN	96.2	61.7	71.4
+Augmentation	95.4	70.3	78.6

DOI: 10.54254/2755-2721/2026.GU34640

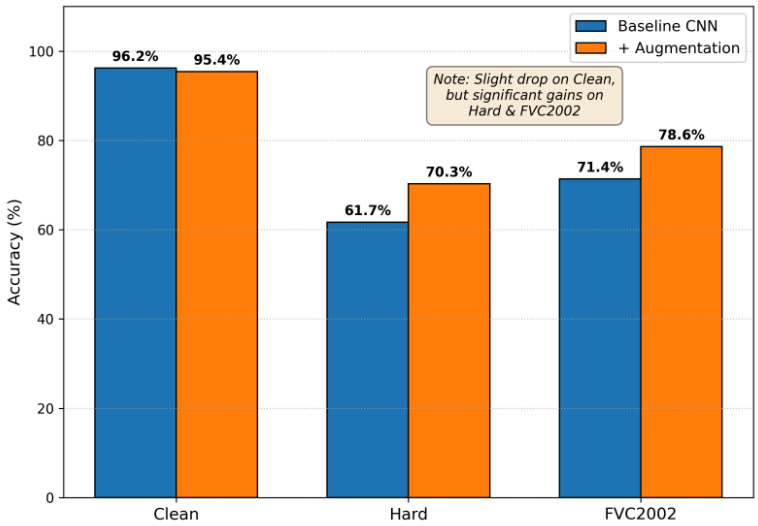

Figure 8. Comparison of performance before and after improvement (picture credit: original)

4.3.2. Analysis

Due to the increase in training complexity, it can be expected that there will be a slight drop in accuracy on clean data for the augmented model (Table 2, Figure 8). It also addresses the improvement in the degraded state and the cross-dataset test.

The above improvement can be explained by the fact that data augmentation has increased the size of the training data, thus reducing overfitting and helping the model learn more generalized features. Add some noise and distortion to the training set so that the model is not too sensitive to the fluctuations of the test data.

Augmentation can help learn features that are invariant to changes and noise from the perspective of representation learning. This directly addresses the problems mentioned in the previous sections.

5. Conclusion

When the quality of the fingerprint is poor, it is noisy or distorted, the recognition accuracy will be low, so this paper proposed an optimized fingerprint recognition workflow to address the critical bottleneck of maintaining high accuracy when dealing with low-quality, noisy, or altered fingerprints. This paper developed a Convolutional Neural Network using the Edge Impulse platform and systematically evaluated its performance, robustness, and generalization across the Sokoto Coventry Fingerprint dataset and two external benchmarks. Experimental results showed that the baseline model achieved over ninety-six percent accuracy on clean data but experienced significant degradation under severe noise and domain shifts due to its over-reliance on local texture features. By implementing targeted data augmentation strategies, the robustness and cross-dataset generalization of the model were substantially improved. A limitation of the current methodology is the inherent bias of standard convolutional architectures, which lack the mechanism to model long-range spatial dependencies when local ridge features are completely corrupted. In the future, the further study plans to explore advanced architectures, such as Vision Transformers, to explicitly capture global contextual relationships and further enhance the reliability of biometric security systems in unpredictable real-world environments.

DOI: 10.54254/2755-2721/2026.GU34640

References

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.
[2] Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1994). Signature verification using a "Siamese" time delay neural network. Advances in Neural Information Processing Systems, 737–744.
[3] Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5967–5976.
[4] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations.
[5] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European Conference on Computer Vision, 833–851.
[6] Shehu, Y. I., Ruiz-Garcia, A., Palade, V., & James, A. (2018). Sokoto coventry fingerprint dataset. arXiv.
[7] Maio, D., Maltoni, D., Cappelli, R., Wayman, J. L., & Jain, A. K. (2002). FVC2002: Second fingerprint verification competition. Proceedings of the 16th International Conference on Pattern Recognition, 3, 811–814.
[8] Imfeld, R., Verhelst, M., & Van Berkel, K. (2022). Edge impulse: An MLOps platform for tiny machine learning. arXiv.
[9] Sajjad, M., et al. (2025). Fingerprint matching for noisy and distorted patterns using a siamese network with ResNet50 and multihead attention. IEEE Access , vol. 13, pp. 1-15.
[10] Al-Qizwini, M., et al. (2025). A hybrid deep learning approach for secure biometric authentication using fingerprint data. Computers , vol. 14, no. 5, p. 178.