Science World Journal Vol. 21(No 2) 2026 https://dx.doi.org/10.4314/swj.v21i2.17
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Full Length Research Article

A HYBRID FRAMEWORK FOR MASK-RESILIENT FACE RECOGNITION AND ANTI-SPOOFING TO ENHANCE SECURE VOTER AUTHENTICATION

*[1]Opeyemi Lateef USMAN, [2]Khadijah Opeyemi OWODUNNI

[1]Department of Computer and Information Science, Tai Solarin Federal University of Education, Ijagun, Ogun State, Nigeria
[2]Department of Computer Science, Sikiru Adetona College of Education, Science and Technology, Omu-Ajose, Ogun State, Nigeria

*Corresponding Author Email Address: usmanol@tasued.edu.ng

ABSTRACT
Biometric authentication systems employed in Nigeria's electoral process continue to encounter significant challenges, particularly in addressing impersonation, facial occlusion, and presentation attacks. Although technologies such as the Bimodal Voter Accreditation System (BVAS) have enhanced electoral transparency and credibility, concerns regarding system reliability and susceptibility to spoofing attacks persist. To mitigate these limitations, this study proposes a hybrid framework that integrates mask-resilient face recognition with anti-spoofing mechanisms for secure voter authentication. The framework is based on a deep convolutional neural network (CNN) trained on a dataset comprising 1,478 facial images categorized into real, masked, and spoof classes. Standard preprocessing procedures, including face detection, image resizing to 128 × 128 pixels, and normalization, were applied to ensure input consistency and optimize model performance. Experimental results indicate that the proposed model achieves 97.3% classification accuracy, demonstrating its ability to effectively distinguish among real, masked, and spoof facial inputs. Furthermore, biometric evaluation metrics, namely the Attack Presentation Classification Error Rate (APCER), Bona fide Presentation Classification Error Rate (BPCER), and Average Classification Error Rate (ACER), confirm the model's effectiveness in detecting spoofing attempts while maintaining acceptable performance for legitimate users. In conclusion, the findings suggest that the proposed framework offers strong potential for real-world biometric authentication applications. It provides an efficient, integrated solution for enhancing security in electoral systems and can be extended to other security-critical domains.

Keywords: Hybrid Framework, Face Recognition, Anti-Spoofing Technique, Secure Voter Authentication, Electoral System

INTRODUCTION

Biometric technologies have become integral components of modern electoral systems, serving as critical mechanisms for verifying voters' identity and preventing impersonation. By leveraging unique physiological traits, such systems enhance the integrity and credibility of the voting process. In Nigeria, the introduction of the Bimodal Voter Accreditation System (BVAS) has significantly improved electoral transparency and trust by incorporating both fingerprint and facial recognition modalities for voter authentication (INEC 2021; 2023; 2024). Despite these advancements, several challenges continue to hinder optimal system performance. Technical limitations, including hardware and software malfunctions, have been reported during election periods. Moreover, environmental factors such as poor lighting conditions, adverse weather, and variations in image quality can significantly degrade the accuracy and reliability of biometric recognition. Empirical studies have consistently highlighted persistent issues, including biometric capture failures, environmental constraints, and operational inefficiencies, all of which undermine the robustness of these systems in real-world deployment scenarios. These limitations underscore the need for more resilient, adaptive, and context-aware biometric frameworks capable of maintaining high performance under diverse and often unpredictable electoral conditions (Abidi et al., 2026; Chingovska et al., 2020; Olayinka & Salami, 2021; Onapajo, 2021).

Facial recognition has gained widespread adoption and growing research attention due to its convenience, non-intrusive nature, and contactless operation. Unlike traditional biometric methods, it does not require physical interaction, making it particularly suitable for large-scale and real-time authentication scenarios (Jaber et al., 2022). Recent advancements in machine learning, particularly in deep learning, have significantly enhanced the performance of facial recognition systems. Convolutional Neural Networks (CNNs), in particular, have demonstrated remarkable capability in learning robust and discriminative feature representations from large-scale datasets, thereby substantially improving recognition accuracy (Schroff et al., 2015; Deng et al., 2019; Usman et al., 2025; Usman et al., 2026). Compared to earlier approaches that relied on handcrafted features, such as local descriptors and statistical methods (Ahonen et al., 2006; Jaber et al., 2022; Mercel et al., 2021; Nemavhola et al., 2025), deep learning-based models exhibit superior performance, especially in unconstrained environments characterized by variations in pose, illumination, and occlusion. These developments have positioned CNN-based facial recognition systems as a reliable and scalable solution for modern biometric authentication applications (Nowara et al., 2020; Kim et al., 2022; Yu et al., 2024).

Despite significant advancements, facial recognition systems continue to face notable limitations. One major challenge arises from occlusion, particularly when individuals wear face masks, which obscure critical facial features and degrade recognition accuracy. This issue became especially prominent during the COVID-19 pandemic, where widespread mask usage exposed the vulnerability of conventional recognition systems. Empirical studies have confirmed that occlusion of key facial regions substantially reduces performance (Deng et al., 2021; Ge et al., 2020; Zhang & Hu, 2025; Huang et al., 2023). To mitigate this limitation, recent studies have focused on mask-aware learning approaches, upper-face feature extraction, and feature fusion techniques, often validated using benchmark datasets such as the Real-World

Masked Face Recognition Dataset (RMFRD). Another significant concern is the susceptibility of facial recognition systems to presentation attacks, where adversaries attempt to deceive the system using printed images, videos, or other artifacts. To address this, face anti-spoofing techniques have been developed to distinguish between genuine and fraudulent inputs. These methods typically rely on texture analysis, motion cues, and physiological signals to enhance detection accuracy (Boulkenafet et al., 2021; Liu et al., 2020; Galbally et al., 2020; Nemavhola et al., 2025), with evaluation commonly performed on benchmark datasets such as CASIA-FASD. Therefore, while facial recognition technologies have achieved substantial progress, challenges related to occlusion and spoofing remain critical areas of ongoing research, necessitating more robust and adaptive solutions.

As illustrated above, existing approaches to facial recognition and anti-spoofing are often developed independently, resulting in systems that either achieve high recognition accuracy with limited security or provide robust security at the expense of efficiency and real-time performance. This trade-off poses a significant challenge in electoral contexts, where both accuracy and security are essential (Adebayo & Omotayo, 2020; Akah, 2024; Li et al., 2020; Zhen & Wang, 2025). The issue is further compounded by the need to deploy solutions on resource-constrained devices, such as BVAS, which require lightweight and computationally efficient models (Ibrahim & Abdullahi, 2023; Yiaga Africa, 2023; Abidi et al., 2026; Zhang & Hu, 2025). Thus, limited research has addressed the development of unified frameworks that simultaneously satisfy these requirements. Although biometric authentication technologies have enhanced transparency in the Nigerian electoral system, they have also exposed the system to critical limitations, including reliability issues and susceptibility to fraud. Reports of authentication failures and exploitable vulnerabilities highlight weaknesses that can undermine the integrity of the voting process. These challenges underscore the urgent need for robust, secure, and efficient biometric solutions that are well-suited to real-world electoral environments, thereby motivating the present study. The main contributions of this study are summarized as follows: first, the development of an integrated framework that combines mask-resilient facial recognition with anti-spoofing mechanisms to enhance secure voter authentication; second, the design of a deep learning–based model capable of maintaining high recognition performance under facial occlusion, particularly in the presence of masks; third, the comprehensive evaluation of the proposed system using standard biometric performance metrics; and finally, the demonstration of its real-time applicability and efficiency on resource-constrained devices, such as BVAS.

Overview of Deep Convolutional Neural Networks for Image Classification and Face Recognition

Deep convolutional neural networks (CNNs) are highly effective models for image analysis, capable of automatically learning and extracting meaningful features from raw pixel data without relying on handcrafted inputs. This ability makes them particularly suitable for image classification and recognition tasks (LeCun et al., 2015; Krizhevsky et al., 2017). Advances in deep learning architectures, such as AlexNet, VGG, GoogLeNet, and ResNet, have further enhanced performance, achieving state-of-the-art results on benchmarks like ImageNet while demonstrating robustness to variations in illumination, pose, and occlusion (Schroff et al., 2015; Liu et al., 2020; Jaber et al., 2022).

In facial recognition applications, CNNs effectively capture discriminative features such as facial structure, texture, and subtle patterns, enabling accurate identification and differentiation between genuine and spoofed inputs. Their layered architecture, comprising convolutional, activation, pooling, and fully connected layers, facilitates hierarchical feature learning, which is optimized through backpropagation. To balance accuracy and efficiency, this study adopts a lightweight CNN architecture to ensure reliable performance in real-time and resource-constrained environments. Convolutional layers play a fundamental role in extracting salient features from images. They achieve this by employing learnable filters that systematically scan localized regions of the input, enabling the detection of spatial patterns and correlations among neighboring pixels. Given an input image $I(x, y, z)$, the convolutional operation produces a corresponding feature map, denoted as, I_{fas}, which captures the relevant structural and contextual information required for subsequent processing as shown in (1):

$$I_f(x,y,z) = \sum_{k=0}^{D-1} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} I(x+i, y+j, z+k) \cdot W(i,j,k) \quad (1)$$

The convolutional neural network (CNN) architecture employs learnable filter weights $W(i, j, k)$ with a kernel size of (3×3) to extract salient facial features essential for identity recognition and fake image detection. Non-linearity is introduced using the Rectified Linear Unit (ReLU), defined in (2) as:

$$\text{ReLU}(z) = \max(0, z) \quad (2)$$

This enables the model to learn complex patterns. Pooling layers, specifically max pooling, are utilized to reduce spatial dimensionality while preserving critical features. Fully connected layers then integrate the extracted features to perform high-level reasoning, culminating in a final classification layer that distinguishes between real, masked, and fake faces. To mitigate overfitting and enhance generalization, dropout regularization is applied during training (Nemavhola et al., 2025; Jaber et al., 2022; Zhang & Hu, 2025; Abidi et al., 2026; Usman et al., 2026).

CNN-based systems are widely adopted in facial recognition due to their strong capability to capture intricate facial patterns. Their application enhances security in sensitive domains such as identity verification during elections, where accurate authentication is crucial for maintaining fairness and preventing fraud. The effectiveness of CNNs lies in their ability to learn discriminative features through exposure to large datasets, enabling them to reliably distinguish genuine from manipulated facial images and verify identities against stored records.

Background of Face Recognition and Anti-Spoofing Techniques

Facial recognition systems identify individuals by extracting distinctive features from facial images and matching them against stored representations. Traditional approaches, such as Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG), achieved moderate success under controlled conditions but exhibited limited robustness to variations in illumination, pose, and occlusion. Recent advances in deep learning, particularly convolutional neural networks (CNNs), have significantly improved performance by enabling automatic learning of discriminative features from large-scale datasets. State-of-the-art models such as FaceNet and ArcFace enhance feature separability, facilitating accurate recognition even in unconstrained real-world environments (Boulkenafet et al., 2021; Liu et al., 2021; Li et al.,

Science World Journal Vol. 21(No 2) 2026 https://dx.doi.org/10.4314/swj.v21i2.17
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

2020).

However, performance remains hindered by facial occlusion, particularly due to widespread mask use, which obscures critical facial regions. To address this, mask-resilient techniques, including upper-face feature extraction, feature fusion, and mask-aware training, have been developed and are supported by specialized datasets such as RMFRD. In addition to occlusion, facial recognition systems are vulnerable to presentation (spoofing) attacks involving printed images, replayed videos, or digital displays. To counter these threats, presentation attack detection (PAD) methods leverage texture, motion, and physiological cues, with deep learning approaches demonstrating superior capability in learning discriminative spatial and temporal features (Pooshideh et al., 2024; Wang et al., 2020; Wirdiani et al., 2023). Benchmark datasets such as CASIA-FASD are commonly used for evaluation (Deng et al., 2021; Ge et al., 2020; Zhang & Hu, 2025; Galbally et al., 2020; Nemavhola et al., 2023). Despite these advancements, most systems treat recognition and anti-spoofing as separate tasks, limiting their effectiveness and increasing computational demands. Consequently, there is a need for an integrated, computationally efficient framework that simultaneously addresses mask-induced occlusion and spoofing. This study proposes a unified hybrid approach to achieve robust, secure facial authentication, particularly for applications such as electoral systems.

MATERIALS AND METHODS

Dataset Description

The dataset employed in this study comprises facial images grouped into three categories: real, masked, and spoof. It was developed by combining publicly available data with manually collected samples to reflect realistic biometric authentication scenarios. The real class includes unmasked faces representing genuine conditions; the masked class includes occluded faces typical of real-world use; and the spoof class includes presentation attack samples such as printed images and digital displays. In total, 1,478 images were utilized, including 688 real, 690 masked, and 100 spoof samples. All images were standardized to a resolution of 128×128 pixels to ensure uniformity and computational efficiency. The dataset was partitioned into training and testing sets using an 80:20 ratio, yielding approximately 1182 training and 296 testing samples to assess model generalization. Additionally, established benchmark datasets, such as RMFRD and CASIA-FASD, informed the dataset design and supported both mask-resilient recognition and anti-spoofing objectives.

Image Acquisition and Preprocessing

Facial images were acquired from established benchmark datasets, particularly CASIA-FASD, and were subjected to a uniform preprocessing pipeline to enhance model performance and maintain input consistency.

The preprocessing steps are as follows:

1. **Face Detection:** Facial regions were detected using Haar Cascade classifiers to isolate relevant facial features from background noise.
2. **Face Cropping:** Detected faces were cropped to focus on the region of interest (ROI), ensuring consistent spatial representation.
3. **Image Resizing:** All cropped images were resized to 128 × 128 pixels to match the CNN input requirements and reduce computational cost.
4. **Normalization:** Pixel values were normalized to the range [0, 1] to improve training stability and convergence.
5. **Label Encoding:** The dataset was encoded as follows:
   i. 0 → Real
   ii. 1 → Masked
   iii. 2 → Spoof

In contrast to patch-based methods, the entire facial region was used as input to preserve spatial dependencies and improve classification accuracy.

Proposed Hybrid Framework

The proposed system constitutes a hybrid deep learning framework that unifies mask-resilient face recognition and anti-spoofing detection within a single architecture. The framework consists of three main components:

1. **Face Detection Module:** Detects and extracts facial regions from input images or video streams.
2. **Feature Extraction Module:** A convolutional neural network (CNN) is used to learn hierarchical feature representations capturing both structural and texture-based facial characteristics.
3. **Classification Module:** Extracted features are passed through fully connected layers to perform multi-class classification into:
   i. Real
   ii. Masked
   iii. Spoof

This integrated architecture enables simultaneous extraction of identity-discriminative features and spoof detection, thereby improving both system security and computational efficiency. Figure 1 illustrates the simplified workflow of the proposed hybrid framework, while Algorithm 3.1 presents its stepwise procedure for integrated face recognition and presentation attack detection.

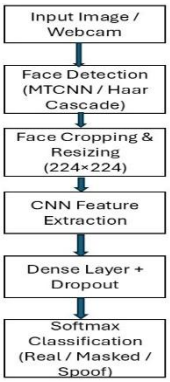

Figure 1. Proposed hybrid framework architecture.

Algorithm 1: Proposed Hybrid Face Recognition and Anti-Spoofing Framework
Input: Image frame or facial image I
Output: Classification label $L \in \{Real, Masked, Spoof\}$

1. Acquire input image I from dataset or a live webcam
2. Perform face detection using Haar Cascade
3. Extract facial region (ROI) from I
4. Resize ROI to 128 × 128 pixels
5. Normalize pixel values to range [0,1]
6. Pass preprocessed image into CNN model

Science World Journal Vol. 21(No 2) 2026 https://dx.doi.org/10.4314/swj.v21i2.17
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

7. Extract hierarchical feature representations
8. Apply fully connected layers with dropout
9. Compute Softmax probabilities for each class
10. Assign label Lbased on highest probability using Eq. (3):

$$L = \arg\max (P_{real}, P_{masked}, P_{spoof}) \quad (3)$$

11. Return classification result L

Proposed CNN Model Architecture, Training, Validation and Real-Time Implementation

The proposed model is a lightweight CNN designed to strike an optimal trade-off between classification accuracy and computational efficiency, enabling real-time deployment. By employing a streamlined architecture with reduced parameter complexity and optimized operations, the model minimizes computational overhead, memory usage, and inference latency while maintaining robust feature extraction capabilities. This design makes it particularly suitable for implementation on resource-constrained devices and edge computing environments, where efficient processing and rapid response times are critical. The model was trained using conventional computational resources, with dropout regularization employed to mitigate overfitting and enhance generalization performance. The architectural configuration of the proposed CNN is schematically illustrated in Figure 2.

To assess practical applicability, the trained model was evaluated using facial image data. The system processes input images through face detection, feature extraction, and subsequent classification using the trained CNN model. The model's lightweight architecture further demonstrates its suitability for real-time deployment in biometric authentication systems. As shown in Figure 3, which presents the real-time system output, the experimental findings indicate that the model achieves efficient inference, underscoring its potential for deployment in security-critical environments such as electoral authentication systems.

All experiments were conducted in Python utilizing TensorFlow and Keras frameworks. Model training was performed in a GPU-enabled environment to expedite computational processes. The technical configuration is consistent with standard image classification pipelines and reflects the resource constraints commonly emphasized in energy-efficient modeling studies (Usman & Muniyandi, 2020; Usman et al., 2021).

Model Architecture comprises:

i. Conv2D (32 filters, 3×3) + ReLUii. MaxPooling (2×2)
iii. Conv2D (64 filters, 3×3) + ReLU
iv. MaxPooling (2×2)
v. Conv2D (128 filters, 3×3) + ReLU
vi. MaxPooling (2×2)
vii. Flatten
viii. Dense (128 neurons) + ReLU
ix. Dropout (0.5)
x. Dense (3 neurons, Softmax output)

Training Configuration involves:

i. Optimizer: Adam
ii. Learning rate: 0.0001
iii. Loss function: Sparse categorical cross-entropy
iv. Batch size: 32
v. Epochs: 10
vi. Regularization: Dropout (0.5)

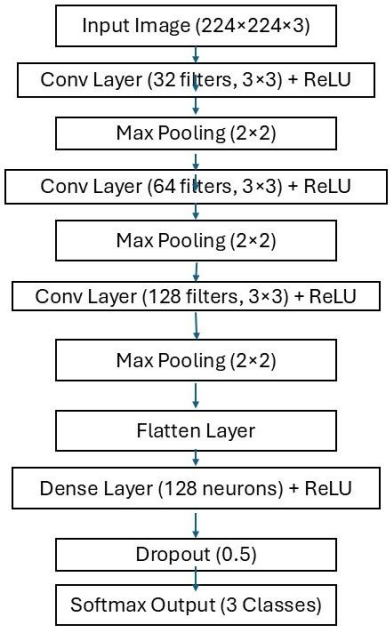

Figure 2. Sketch of the proposed CNN architecture.

Model Evaluation Metrics

The performance of the proposed system was assessed using both classification and biometric security metrics. The classification aspect utilizes standard evaluation measures, whereas the biometric metrics offer a more comprehensive assessment of recognition accuracy and resilience to spoofing attacks. These metrics are derived from the confusion matrix presented in Figure 3 and are defined as follows:

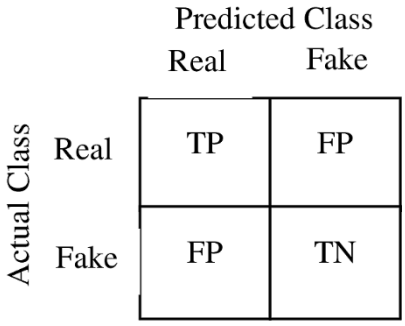

Figure 3. Confusion matrix.

1. **Accuracy:** Accuracy quantifies the overall correctness of the proposed model by measuring the proportion of correctly classified instances relative to the total number of predictions made. It provides a general indication of the model's predictive performance across all classes. The metric is formally defined in Eq. (3) as follows:

$$Accurcy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2. **Precision:** Precision evaluates the extent to which instances predicted as positive are truly positive, thereby reflecting the model's ability to minimize false positive predictions. The metric is formally defined in Eq. (4) as follows:

Science World Journal Vol. 21(No 2) 2026 https://dx.doi.org/10.4314/swj.v21i2.17
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3. **Recall:** Recall quantifies the proportion of actual positive instances that are correctly identified by the model, thereby indicating its effectiveness in capturing true positives while minimizing false negatives. It is formally defined in Eq. (5) as follows:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

4. **F1-score:** The F1-score provides a balanced measure of a model's performance by integrating both precision and recall into a single metric, thereby capturing the trade-off between false positives and false negatives. The F1-score is computed as the harmonic mean of precision and recall and is formally defined in Eq. (6) as follows:

$$F1 - Score = \frac{2(Precision \times Recall)}{Precision+Recall} \quad (6)$$

5. **APCER:** The Attack Presentation Classification Error Rate (APCER) quantifies the frequency with which spoofing attempts (presentation attacks) are incorrectly classified as genuine samples, thereby reflecting the system's susceptibility to false acceptance of attacks. This metric is critical for assessing the robustness of biometric systems against adversarial manipulation and presentation-based threats. APCER is formally defined in Eq. (7) as follows:

$$APCER = \frac{FP}{FP+TN} \quad (7)$$

6. **BPCER:** The Bona Fide Presentation Classification Error Rate (BPCER) measures the proportion of genuine (bona fide) facial presentations that are incorrectly classified as attacks, thereby indicating the system's tendency to falsely reject legitimate users. BPCER is formally defined in Eq. (8) as follows:

$$BPCER = \frac{FN}{FN+TP} \quad (8)$$

7. ACER: The Average Classification Error Rate (ACER) represents the mean of the error rates associated with both attack and bona fide presentations, thereby providing a unified measure of overall system performance in biometric security evaluation. ACER is formally defined in Eq. (9) as follows:

$$ACER = \frac{APCER+BPCER}{2} \quad (9)$$

An example of a typical real-time system output is shown in Figure 4, which illustrates a comparison of genuine, masked, and spoof facial instances.

Figure 4. Typical real-time system output.

RESULTS AND DISCUSSION

Classification Performance

The performance of the proposed hybrid framework was assessed using standard classification metrics, including accuracy, precision, recall, and F1-score. The model attained an overall classification accuracy of 97.3% on the test dataset, indicating its effectiveness in distinguishing among real, masked, and spoof facial inputs. A more detailed class-wise evaluation is presented in Table 1, which summarizes the corresponding performance metrics for each category. The results reveal that the model performs best on real-face samples, while it shows relatively lower performance on masked and spoof classes. This discrepancy is likely due to facial occlusion and the visual similarity between masked and spoofed inputs, which complicate feature discrimination.

Table 1. Classification Performance Metrics

Class	Precision	Recall	F1-score
Real	0.98	0.96	0.97
Masked	0.99	0.97	0.98
Spoof	0.86	1.00	0.92

Confusion Matrix Analysis

To further examine classification performance, the confusion matrix is presented in Figure 5. The matrix exhibits strong diagonal dominance, indicating that most samples are correctly classified across all categories. The model demonstrates a high true positive rate for real-face classification, reflecting robust performance under non-occluded conditions. Misclassifications are minimal, occurring in a small number of cases observed between the masked and spoof classes, suggesting that the model experiences slight confusion between real and masked classes. However, there are no spoof samples classified as genuine, as demonstrated by the zero APCER value, indicating strong resistance to presentation attacks. Therefore, the confusion matrix confirms that the proposed framework has learned strong, effective discriminative representations, as evidenced by the high classification accuracy across all classes. The absence of spoof misclassification further demonstrates the model's robustness in detecting presentation attacks. While only minimal errors are observed, primarily between visually similar classes, the overall performance indicates that the system achieves a high level of reliability and effectiveness for face classification and anti-spoofing tasks.

Science World Journal Vol. 21(No 2) 2026 https://dx.doi.org/10.4314/swj.v21i2.17
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

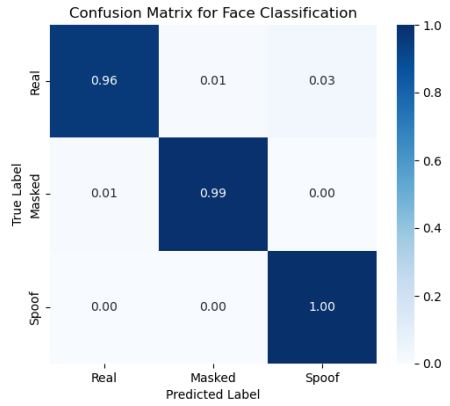

Figure 5. Confusion matrix of classification results for real, masked, and spoof classes.

Biometric Security Evaluation

In addition to conventional classification metrics, biometric-specific evaluation measures were employed to assess the system's robustness to spoofing attacks. The corresponding results, summarized in Table 2, provide further insight into the system's security-usability balance. The relatively low Average Classification Error Rate (ACER) of 0.02 indicates that the proposed framework achieves an effective compromise between spoof resistance and genuine user accessibility. Furthermore, the Attack Presentation Classification Error Rate (APCER) of 0.00 indicates that the system can reject all spoofing attempts, reinforcing its strong security performance. In parallel, the Bona fide Presentation Classification Error Rate (BPCER) of 0.03 reflects an acceptable level of accuracy in recognizing legitimate users, ensuring that usability is not significantly compromised. Collectively, these metrics highlight the model's potential for deployment in real-world biometric authentication systems, where both security robustness and user convenience are critical considerations.

Table 2. Biometric evaluation metrics.

Metric	Value
APCER	0.00
BPCER	0.03
ACER	0.02

Training and Validation Performance

The training and validation performance of the model are depicted in Figures 6(a) and (b). The learning curves exhibit a consistent trend of convergence, with both training and validation accuracy improving progressively across successive epochs. The relatively narrow gap between the training and validation metrics suggests limited overfitting and indicates strong generalization, despite the dataset's constrained size. Moreover, the stability observed in the validation curve suggests that the model effectively captures underlying feature representations rather than memorizing the training data. This behavior reflects the effectiveness of the applied regularization strategies and architectural design in promoting robust learning, thereby enhancing the model's reliability for deployment in real-world scenarios.

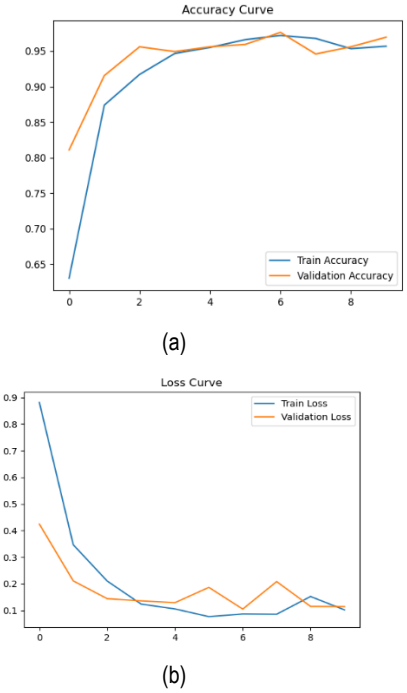

Figure 6. Training and validation performance curve. (a) Training curve (b) validation accuracy and loss curve.

Comparative Analysis with the State-of-the-Art and Computational Performance

To assess the effectiveness of the proposed framework, this study presents both a literature-based comparison with prior work and an experimental comparison with baseline models evaluated on the study dataset. Table 3 presents the literature-based comparison with selected state-of-the-art methods. The results show that George et al. (2021) reported an accuracy of 68.0% using a Multi-task CNN with Depth Estimation, while Yu et al. (2023) reported 94.0% using a Vision Transformer-based detection method. In comparison, the proposed framework achieved an accuracy of 97.3%. However, the results in Table 3 are presented only for contextual comparison, as the previous studies were conducted with different datasets, preprocessing procedures, and experimental settings. Therefore, to ensure a fairer comparison, baseline models were implemented and evaluated in Python using the same study dataset, preprocessing procedure, train-test split, and evaluation metrics as the proposed hybrid framework.

The results in Table 4 present the experimental and computational comparison between the proposed hybrid framework and the implemented baseline models using the same study dataset. The comparison was performed using accuracy, precision, recall, F1-score, inference time per image, and model size. This ensures that the reported values were obtained under the same experimental conditions rather than directly copied from previous studies. As shown in Table 4, the proposed hybrid framework achieved an accuracy of 97.3%, precision of 97.8%, recall of 97.3%, and F1-score of 97.4%. This result demonstrates that the proposed framework achieved strong and competitive classification performance across the three classes: real, masked, and spoof. The model outperformed the Simple CNN and ResNet50 baselines

and achieved comparable accuracy to MobileNetV2, while achieving higher precision and F1-score. Although VGG16 achieved the highest accuracy of 98.3%, it recorded a higher inference time of 0.061537 sec/image compared with 0.007003 sec/image for the proposed framework. In terms of computational performance, inference time and model size were used to evaluate deployment suitability. Inference time refers to the average time required by the trained model to classify a single image, while model size refers to the storage size of the saved trained model file. The proposed framework achieved an inference time of 0.007003 sec/image and a model size of 37.9 MB, indicating a strong balance between classification performance and computational efficiency. This makes the framework suitable for near-real-time deployment in security-sensitive authentication systems such as the Bimodal Voter Accreditation System (BVAS).

Table 3. Literature-based comparison with existing studies

Authors (Year)	Proposed Method	Accuracy
George et al. (2021).	Multi-task CNN + Depth Estimation	68.0%
Yu et al. (2023).	Vision Transformer-based Detection	94.0%
Usman & Owodunni (2026).	Proposed framework	97.3%

Table 4. Experimental and computational comparison of baseline models using the study dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Inference Time per Image (sec)
Simple CNN Baseline	96.0	96.0	96.0	95.9	0.005180
MobileNetV2 Baseline	97.3	97.3	97.3	97.3	0.031810
ResNet50 Baseline	67.6	64.5	67.6	65.2	0.051556
VGG16 Baseline	98.3	98.4	98.3	98.3	0.061537
Proposed Hybrid Framework	97.3	97.8	97.3	97.4	0.007003

The proposed hybrid framework achieved an accuracy of 97.3%, precision of 97.8%, recall of 97.3%, and F1-score of 97.4%. These results show that the proposed framework achieved strong and competitive classification performance across the three classes: real, masked, and spoof. Compared with the Simple CNN baseline, which achieved 96.0% accuracy, the proposed framework improved classification accuracy while maintaining a relatively low inference time. MobileNetV2 also achieved an accuracy of 97.3%, matching that of the proposed framework. However, the proposed hybrid framework recorded a higher precision value of 97.8% compared with 97.3% for MobileNetV2, and a slightly higher F1-score of 97.4% compared with 97.3%. In addition, the proposed framework had a faster inference time of 0.007003 sec/image, compared with 0.031810 sec/image for MobileNetV2. This indicates that the proposed framework provides better classification balance and faster prediction speed. ResNet50 recorded the lowest performance among the evaluated models, with an accuracy of 67.6%, precision of 64.5%, recall of 67.57%, and F1-score of 65.2%. The weak performance of ResNet50 may be due to poor adaptation of the pre-trained feature extractor to the specific characteristics of the study dataset, especially the spoof class. This suggests that deeper architectures do not always guarantee improved performance when applied to a dataset with specific class characteristics and limited sample distribution. VGG16 achieved the highest accuracy of 98.3%, with precision, recall, and F1-score values of 98.4%, 98.3%, and 98.3%, respectively. However, VGG16 also recorded a higher inference time of 0.061537 sec/image, which is significantly slower than the proposed framework. Although VGG16 achieved the highest classification accuracy, its slower inference speed may limit its suitability for time-sensitive authentication systems that require rapid decision-making.

Overall, the proposed hybrid framework provides a strong balance between classification performance and computational efficiency. While VGG16 achieved slightly higher accuracy, the proposed framework maintained competitive performance with a much lower inference time. This makes the proposed model more suitable for near real-time deployment in security-sensitive applications such as biometric authentication and BVAS-related identity verification, where both prediction accuracy and response speed are important.

Table 5. Computational performance.

Metric	Definition	Value	Interpretation
Inference time	Average time required by the trained model to classify one input frame/image	0.15 sec/frame	Indicates near real-time authentication capability
Model size	Storage size of the saved trained model file	37.87 MB	Indicates that the model is lightweight enough for practical deployment

The computational performance of the proposed framework was evaluated using inference time and model size. Inference time refers to the average time required by the trained model to classify a single input frame or image, while model size refers to the storage size of the saved trained model file. As shown in Table 5, the proposed framework achieved an inference time of 0.15 sec/frame, indicating that it can support near real-time authentication. The saved model size was 37.87 MB, suggesting that the framework has a relatively lightweight storage requirement and is suitable for practical deployment in resource-constrained and security-sensitive environments such as BVAS.

DISCUSSION OF FINDINGS

The experimental findings demonstrate that the proposed hybrid framework effectively addresses the dual challenges of facial occlusion and presentation attacks. By integrating mask-resilient feature learning with anti-spoofing mechanisms into a unified architecture, the system can simultaneously perform identity classification and security verification. The model exhibits strong performance in recognizing genuine facial inputs, while showing comparatively moderate performance in distinguishing masked and spoofed samples. This limitation can be attributed to feature similarity between classes, constraints in dataset size and diversity, and variations in image quality. Although the overall accuracy of 97.3% is modest relative to large-scale deep learning models, it is

important to emphasize that the proposed framework operates under challenging real-world conditions, such as occlusion and spoofing, while maintaining computational efficiency suitable for embedded and resource-constrained systems. Furthermore, the model demonstrates practical viability for real-time deployment, reinforcing its applicability in dynamic operational environments. However, factors such as illumination variability, camera quality, and subject motion were found to influence prediction stability and consistency.

In summary, the results indicate that the proposed framework achieves a balanced trade-off among accuracy, security, and computational efficiency. This balance underscores its potential as a reliable solution for secure voter authentication in electoral systems, particularly in scenarios requiring real-time processing and robust spoof resistance. Specifically, the proposed hybrid framework effectively combines convolutional neural networks with anti-spoofing mechanisms to facilitate secure and reliable facial authentication. The system demonstrates the ability to accurately distinguish among real, masked, and spoofed inputs while maintaining computational efficiency suitable for real-time applications. The findings indicate that the model successfully learns discriminative facial representations even in the presence of occlusion, and biometric evaluation metrics further validate its robustness against presentation attacks. Although certain limitations persist in distinguishing masked from spoofed samples, the framework establishes a solid foundation for developing secure and efficient biometric authentication systems. systems.

CONCLUSION

This study presented a hybrid framework that integrates mask-resilient face recognition with anti-spoofing mechanisms to enable secure voter authentication in Nigeria's electoral system. The proposed approach employs a convolutional neural network to simultaneously perform identity recognition and presentation attack detection in a unified architecture, thereby addressing key limitations of conventional biometric systems. Experimental results indicate that the model achieves an overall classification accuracy of 97.3%, with strong performance in recognizing genuine facial inputs and comparatively moderate effectiveness in distinguishing masked and spoofed presentations. The incorporation of biometric evaluation metrics, such as APCER, BPCER, and ACER, further substantiates the framework's ability to maintain a balanced trade-off between security and usability.

The model demonstrates real-time inference capability and is well-suited for deployment in resource-constrained environments, such as polling units, thereby confirming its practical applicability. Its ability to perform real-time face detection and classification is particularly significant for mitigating challenges related to impersonation and identity fraud in Nigeria's electoral process. Despite these promising outcomes, certain limitations were identified. The relatively small dataset size and limited diversity constrain the model's capacity for generalization. Additionally, environmental variations, such as illumination conditions and camera quality, as well as the similarity of features between masked and spoof samples, contribute to classification inaccuracies. Future research will focus on enhancing the robustness and scalability of the system through several directions, including: (i) expanding the dataset with more diverse and large-scale real-world samples; (ii) incorporating transfer learning using advanced architectures such as ResNet and MobileNet; (iii) integrating temporal feature analysis to improve video-based spoof detection; and (iv) optimizing the model for deployment on low-resource devices, including BVAS platforms. In conclusion, the integration of mask-resilient face recognition with anti-spoofing techniques offers a viable, scalable solution for secure biometric authentication. The proposed framework contributes to the advancement of reliable and fraud-resistant electoral systems. It has potential for extension to other security-critical domains, such as border control, financial services, and access control.

REFERENCES

Abidi, S. M. H., Hassan, S. A., Razza, S. M. & Beliatis, M. J. (2026). Advances in Face Recognition: A Comprehensive Review of Feature Extraction and Dataset Evaluation. *Electronics (MDPI), 15*(338), 1–27. https://doi.org/10.3390/electronics15020338

Adebayo, A. A., & Omotayo, F. O. (2020). Biometric technology and election credibility in Nigeria. *African Journal of Political Science and International Relations, 14*(2), 45–57.

Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(12), 2037–2041. https://doi.org/10.1109/TPAMI.2006.244

Akah, A. U. (2024). Elections administration and the bimodal voter accreditation system (BVAS) in Nigeria's 2023 presidential election. *Journal of African Elections, 23*(1), 1–21.

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2021). Face spoofing detection using color texture analysis. *IEEE Transactions on Information Forensics and Security, 15*, 181–196. https://doi.org/10.1109/TIFS.2019.2924642

Chingovska, I., Anjos, A., & Marcel, S. (2020). Biometrics evaluation under spoofing attacks. *IEEE Transactions on Information Forensics and Security, 15*, 215–230. https://doi.org/10.1109/TIFS.2019.2925650

Deng, J., Guo, J., & Tao, X. (2021). Understanding the impact of facial occlusion on deep face recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications, 17*(2), 1–22. https://doi.org/10.1145/3433296

Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4690–4699. https://doi.org/10.1109/CVPR.2019.00482

Galbally, J., Gómez-Barrero, M., & Fierrez, J. (2020). Vulnerabilities of face recognition systems to spoofing attacks. *IEEE Access, 8*, 142023–142040. https://doi.org/10.1109/ACCESS.2020.3013922

Ge, Y., Zhang, C., & Gao, W. (2020). Masked face recognition with feature fusion and angular margin loss. *Neurocomputing, 410*, 350–364. https://doi.org/10.1016/j.neucom.2020.07.128

George, A., & Marcel, S. (2021). Deep pixel-wise binary supervision for face presentation attack detection. Pattern Recognition Letters, 147, 110–117. https://doi.org/10.1016/j.patrec.2021.03.017

Huang, R., Zhao, Q., & Yang, J. (2023). Face recognition under occlusion with generative mask augmentation. *Neural Networks, 156*, 478–490.

https://doi.org/10.1016/j.neunet.2022.12.011

Ibrahim, M., & Abdullahi, A. (2023). Electoral integrity and biometric challenges in Nigeria's general elections. *Journal of African Electoral Studies*, 12(1), 45–60

INEC. (2021). *Election technologies handbook*. Independent National Electoral Commission

INEC. (2023). *Manual for election officials: 2023 general elections*. Independent National Electoral Commission

INEC. (2024). *Report of the 2023 general election*. Independent National Electoral Commission

Jaber, A. G., Muniyandi, R. C., Usman, O. L., & Singh, H. K. R. (2022). A hybrid method of enhancing the accuracy of a facial recognition system using a Gabor filter and a stacked sparse autoencoder deep neural network. *Applied Sciences (MDPI)*, *12*(21), 1–20. https://doi.org/10.3390/app122111052

Kim, S.-H., Jeon, S.-M., Lee, E. C. (2022). Face biometric spoof detection method using a remote photoplethysmography signal. *Sensors, 22*(8), 3070. https://doi.org/10.3390/s22083070

Kollreider, K., Fronthaler, H., & Bigun, J. (2021). Non-intrusive liveness detection by motion analysis. *Image and Vision Computing, 106*, 104089. https://doi.org/10.1016/j.imavis.2020.104089

Krizhevsky, A., Sutskever, I., Hinton, G. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communication ACM*, 84–90.

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.

Li, X., Wang, Y., & Zhao, G. (2020). Remote photoplethysmography-based face anti-spoofing. *IEEE Transactions on Information Forensics and Security, 15*, 2408–2420. https://doi.org/10.1109/TIFS.2020.2978582

Liu, Y., Jourabloo, A., & Liu, X. (2020). Learning deep models for face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(4), 930–945. https://doi.org/10.1109/TPAMI.2018.2881879

Marcel, S., Nixon, M. S., Fierrez, J., & Evans, N. (2021). *Handbook of biometric anti-spoofing* (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-54159-1

Nemavhola, A., Viriri, S. & Chibaya, C. (2025). A scoping review of literature on deep learning techniques for face recognition. *Human Behavior and Emerging Technologies (Wiley), 2025*(5979728), 1–14. https://doi.org/10.1155/hbe2/5979728

Nowara, E. M., Sabharwal, A., & McDuff, D. (2020). Impact of motion and illumination on rPPG. *IEEE Transactions on Biomedical Engineering, 67*(8), 2340–2352. https://doi.org/10.1109/TBME.2019.2955240

Olayinka, O., & Salami, A. (2021). Challenges of biometric voter authentication in Nigeria. *African Journal of Governance and Development*, 10(2), 89–104

Onapajo, H. (2021). Why biometric voter authentication fails in Africa. *Journal of Modern African Studies, 59*(4), 537–556. https://doi.org/10.1017/S0022278X21000417

Pooshideh, M., Mohammadi, A., & Marcel, S. (2024). Presentation attack detection: A systematic review. *ACM Computing Surveys, 56*(7), Article 164. https://doi.org/10.1145/3687264

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 815–823. https://doi.org/10.1109/CVPR.2015.7298682

Usman, O. L., Adeusi, O. O., Kareem, M. A., Owoade, A. A. & Muniyandi, R. C. (2026). Quantitative study on the impact of an EfficientNet-based deep transfer learning model for pneumonia detection with explainable artificial intelligence using chest radiographs. *Zamfara International Journal of Science, Technology, Education and Mathematics. 3*(1), 45–56. https://doi.org/10.64348/zije.2026344

Usman, O. L., Muniyandi, R. C., Omar, K., Mohamad, M., Owoade, A. A., & Kareem, M. A. (2025). HoRNS-CNN Model: An energy-efficient fully homomorphic residue number system convolutional neural network model for privacy-preserving classification of dyslexia neural-biomarkers. *Brain Informatics, 12*(11), 1–28. https://doi.org/10.1186/s40708-025-00256-z

Usman, O. L. & Muniyandi, R. C. (2020). CryptoDL: Predicting Dyslexia Biomarkers from Encrypted Neuroimaging Dataset Using Energy-Efficient Residue Number System and Deep Convolutional Neural Network. Symmetry MDPI-Basel, 12(836), 1–24. https://doi.org/10.3390/sym12050836

Usman, O. L., Muniyandi, R. C., Omar, K., & Mohamad, M. (2021). Advanced machine learning methods for dyslexia biomarker detection: A review of implementation details and challenges. IEEE Access, 9, 36879–36894. https://doi.org/10.1109/ACCESS.2021.3062709

Wang, Z., Deng, W., & Hu, J. (2020). Masked face recognition dataset and benchmark. *arXiv preprint arXiv:2003.09093*. https://arxiv.org/abs/2003.09093

Wirdiani, A., Putra, K. G. D., Sudarma, M., Hartati, R. S., Lofiana, S. A. (2023). Real-time face recognition system using a deep learning method. *Lontar Komputer, 14*(1), 62-70. DOI: 10.24843/LKJITI. 2023. v14.i01.p06

Yiaga Africa. (2023). *Watching the vote: Final report on Nigeria's 2023 general elections*. https://yiaga.org

Yu, Z., Cai, R., Li, Z., Yang, W., Shi, J. & Kot, A. C. (2024). Benchmarking joint face spoofing and forgery detection with visual and physiological cues. *IEEE Transactions on Dependable and Secure Computing, 21*(5), 4327-4342. https://doi.org/10.1109/TDSC.2024.3352049

Yu, Z., Qin, Y., Li, X., Zhao, C., Lei, Z., & Li, S. Z. (2023). Learning deep models for face anti-spoofing: Binary or auxiliary supervision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2), 1234–1248. https://doi.org/10.1109/TPAMI.2022.3141234

Zhang, J. & Hu, N. (2025). Accuracy and robustness evaluation of deep learning algorithms in facial recognition systems. *Systems and Soft Computing, 7*(200252), 1–10. https://doi.org/10.1016/j.sasc.2025.200252

Zhang, K., Liu, Y., Wang, H., & Li, S. Z. (2023). Hybrid CNN–Transformer framework for face anti-spoofing. *IEEE Transactions on Information Forensics and Security, 18*, 1123–1137. https://doi.org/10.1109/TIFS.2022.3223127

Zhen, C. & Wang, Y. (2025). Design of a face recognition system based on deep learning. *MLNN 2025: 2025 International Conference on Machine Learning and Neural Networks*. https://dl.acm.org/doi/10.1145/3747227.3747243