International Journal Software Engineering and Computer Science (IJSECS)
6 (2), 2026, 752-765
Published Online August 2026 in IJSECS (http://www.journal.lembagakita.com/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v6i2.7459.

RESEARCH ARTICLE **Open Access**

Face Image Authenticity Detection for ASN Attendance Using CNN MobileNetV2: A Case Study of Tangerang City Government

Suryadi [1*], Sajarwo Anggai [2], Sudarno Wiharjo [3]

[1*,2,3] Department of Informatics Engineering, Magister Program, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.

**Corresponding author*: surya.adi13@gmail.com.

Received: June 8, 2026; Accepted: July 3, 2026; Published: August 1, 2026.

Abstract: Face recognition-based attendance systems can improve the convenience of attendance processes, but they remain vulnerable to presentation attacks, in which facial images displayed on another device are used to simulate a user's physical presence. At the Tangerang City Government, the Tangerang AYO application is used as the main attendance channel for civil servants (ASN) across government agencies (OPD), with attendance data linked to the calculation of Additional Employee Income (TPP). This study develops a face liveness detection model to distinguish between real and fake facial images using a CNN MobileNetV2 architecture with a transfer learning and two-phase fine-tuning approach. The dataset was constructed from actual ASN attendance data and consisted of real and fake images representing facial displays through a secondary device. The model was evaluated using a confusion matrix, accuracy, precision, recall, specificity, F1-score, ROC curve, and AUC. On the testing data (n=301), the model achieved an accuracy of 97.34%, precision of 96.13%, recall of 98.68%, specificity of 96.00%, F1-score of 97.39%, and AUC of 0.9934 at a threshold of 0.5. Grad-CAM analysis showed that the model produced different activation patterns between real and fake images, including contextual visual information in real images and screen-surface characteristics in fake images. A FastAPI-based REST API prototype achieved an average response time of 306.48 ms under warm conditions without a dedicated GPU, indicating the potential feasibility of the model as an additional verification layer for an ASN attendance system.

Keywords: Face Liveness Detection; MobileNetV2; CNN; Digital Attendance; ASN; Presentation Attack; Tangerang AYO.

1. Introduction

Face recognition-based attendance systems are increasingly used because they provide a fast identification process and reduce the need for manual attendance recording. However, the ability of a system to recognize facial identities does not automatically guarantee that the corresponding person is physically present in front of the camera. A face recognition system essentially matches facial characteristics captured by a camera with stored facial data. If a person's facial image is displayed on another device and can still be recognized by the system, the matching process may produce an incorrect decision. This condition represents a form of presentation attack or face spoofing, in which a facial representation is used to obtain an identification result as if the legitimate subject were physically present in front of the camera (Zawar & Chakkarwar, 2023).

This issue is directly relevant to the ASN attendance system implemented by the Tangerang City Government. The Tangerang AYO application is used as an attendance channel for ASN across government agencies (OPDs). An attendance system that relies solely on face recognition may be vulnerable to misuse when a facial image can be displayed through another device. In such a scenario, an employee who is not physically present at the designated location may submit attendance using a video recording, video call, screenshot, or other media displaying the employee's face on a secondary device. When the displayed image is directed toward the camera used for attendance, the system may interpret it as a face originating directly from the subject. This type of scenario is categorized as a presentation attack and represents a challenge in developing face biometric systems that are more resistant to spoofing.

The risks associated with such manipulation are not limited to inaccurate attendance records. Attendance data from Tangerang AYO are connected to e-Kinerja and SIMASN, which are related to employee performance management and the calculation of Additional Employee Income (TPP). Therefore, incorrectly accepting a fake image may affect the reliability of attendance data used in personnel administration. If the system accepts a fake image as real, an employee who does not perform attendance in person may still be recorded as present. Conversely, if a real image is incorrectly classified as fake, an employee who is physically present may experience an unsuccessful attendance submission and require additional verification. These conditions indicate that a face-based attendance system needs a mechanism that not only verifies identity but also assesses whether the captured facial image represents a genuine subject.

This requirement can be addressed through face liveness detection or face anti-spoofing. These approaches aim to distinguish a face originating from a real subject in front of the camera from a facial representation displayed through another medium. Deep learning methods, particularly Convolutional Neural Networks (CNNs), can be used to learn visual characteristics that distinguish between these two classes. Features such as skin texture, illumination patterns, artifacts caused by replay or recapture processes, and screen-surface characteristics can provide useful information for classification (Radad *et al.*, 2025). This approach is suitable for handling variations in presentation attacks compared with simple methods that rely only on a single characteristic, such as facial movement.

The selection of an architecture also needs to consider the conditions of practical implementation. MobileNetV2 is a CNN architecture designed with computational efficiency and model size in mind. These characteristics make MobileNetV2 relevant for implementation scenarios with limited computational resources, including local government environments (Najeebullah *et al.*, 2025). Previous research has also demonstrated that MobileNetV2 can be applied to face spoofing detection and achieve good performance on the evaluated datasets. Zawar and Chakkarwar (2023), for example, reported an accuracy of 98% for detecting photo- and video-based attacks using MobileNetV2. Najeebullah *et al.* (2025) also reported an accuracy of 91.59%, precision of 94.85%, and recall of 89% in a face spoofing detection evaluation. These findings provide a basis for considering MobileNetV2 as a relatively lightweight model with good classification capability.

Other studies have shown that presentation attacks cannot always be addressed through simple motion-based approaches. Replay attacks involving video may preserve certain facial movement patterns, making it difficult for a system to distinguish a genuine subject from a facial representation displayed on a screen (Long *et al.*, 2025). CNN-based approaches can provide additional information through texture characteristics and visual patterns contained in facial images. In Indonesia, Ardiansyah *et al.* (2025) combined nonlinear diffusion with CNN for liveness detection and achieved an accuracy of 87.76%. Meanwhile, Ahmad Abdillah *et al.* (2025) combined face recognition, liveness detection, and geolocation in a mobile attendance application and reported an accuracy of 92%. These studies indicate that liveness detection has begun to be applied to attendance systems; however, spoofing scenarios involving secondary devices require further investigation.

Based on these previous studies, there remains a research gap in the application of facial authenticity detection for ASN attendance scenarios involving presentation attacks through secondary devices. This scenario differs from the use of printed photographs or other simple attacks because the attack medium may involve video, video calls, or facial images displayed on another device. This study does not focus on comparing multiple CNN architectures. Instead, it evaluates the use of MobileNetV2 as a facial authenticity detection model in the ASN attendance environment of the Tangerang City Government. The dataset used in this study is derived from the ASN attendance environment and consists of real and fake facial images designed to represent spoofing scenarios involving a secondary device. This approach is intended to provide an assessment of model performance under conditions that are closer to the operational requirements of a face-based attendance system.

In addition to classification performance, this study considers the requirements for practical implementation. A model with high accuracy may not be suitable for deployment if its inference time is too long or if it requires specialized hardware. Therefore, the model is evaluated using a confusion matrix, accuracy, precision, recall, specificity, F1-score, ROC curve, and AUC to obtain a broader assessment of classification performance. Grad-CAM analysis is also employed to examine the image regions that contribute to the model's predictions. Following model evaluation, the best-performing model is implemented as a

FastAPI-based REST API prototype to measure response time in an environment without dedicated GPU hardware. This process allows the study to assess not only the model's ability to distinguish real and fake images but also the initial feasibility of using the model as an additional verification layer in the attendance system. Based on the identified problem and research needs, this study has three main objectives: (1) to develop a facial authenticity detection model for the ASN attendance system using CNN MobileNetV2 to distinguish between real and fake facial images; (2) to evaluate the model's performance using a confusion matrix, accuracy, precision, recall, specificity, F1-score, ROC curve, and AUC on the Tangerang City ASN attendance dataset; and (3) to evaluate the feasibility of implementing the best-performing model as a REST API prototype for additional verification in a real-time digital ASN attendance scenario.

2. Related Work

2.1 Deep Learning-Based Face Spoofing Detection

Research on CNN-based *face anti-spoofing* has developed considerably in recent years. Zawar and Chakkarwar (2023) demonstrated that MobileNetV2 was able to detect photo- and video-based attacks in real time with an accuracy of 98% on the LCC-FASD dataset. This result indicates that a relatively lightweight CNN architecture can be used for *face spoofing* detection while maintaining good classification performance. Long *et al*. (2025) also examined the challenge of *video replay attacks*, which can pose difficulties for detection methods based on simple facial movements. Deep learning-based approaches can utilize visual and texture characteristics to distinguish genuine faces from facial representations used in spoofing attacks. In Indonesia, Ardiansyah *et al*. (2025) developed a *nonlinear diffusion* method combined with CNN for *liveness detection* and achieved an accuracy of 87.76%. Although the study is relevant to *liveness detection*, scenarios involving video calls or secondary devices were not the main focus of its evaluation. Ahmad Abdillah *et al*. (2025) combined *face recognition*, *liveness detection*, and geolocation in a mobile attendance application for Indonesian ASN and achieved an accuracy of 92%. Although the study was conducted in the context of ASN attendance, spoofing scenarios involving secondary devices were not the main focus of the evaluation. The *eye-blink detection* approach used by Sari Khatina *et al*. (2022) also demonstrates the importance of *liveness detection* mechanisms in distinguishing genuine users from facial representations that are replayed.

2.2 Architectural Efficiency for Operational Deployment

Najeebullah *et al*. (2025) compared MobileNetV2, ResNet50, and Vision Transformer (ViT) using a dataset consisting of 150,986 images for *face spoofing detection*. MobileNetV2 achieved an accuracy of 91.59%, precision of 94.85%, and recall of 89%. These results indicate that MobileNetV2 provides competitive performance for *face spoofing* detection while maintaining a relatively lightweight architecture. This characteristic is an important consideration when selecting MobileNetV2 for attendance systems operating under limited computational resources. Shinde *et al*. (2025) introduced LwFLNeT as a lightweight CNN architecture with lower model complexity and good detection performance. The study indicates that lightweight CNN models can be considered for detection systems that require computational efficiency. Meanwhile, Radad *et al*. (2025) proposed a CNN encoder based on grayscale texture features for *face spoofing* detection under varying illumination conditions. Variations in illumination should be considered in attendance systems because lighting conditions may differ between locations and operational environments.

Table 1. Summary of Related Studies

Author(s) (Year)	Method	Main Findings
Najeebullah *et al*. (2025)	MobileNetV2, ResNet50, ViT	MobileNetV2 achieved 91.59% accuracy and 94.85% precision
Radad *et al*. (2025)	CNN Encoder + grayscale texture	Detection performance was evaluated under varying illumination conditions
Shinde *et al*. (2025)	LwFLNeT (lightweight CNN)	Demonstrated good detection performance with relatively low model complexity
Ahmad Abdillah *et al*. (2025)	MobileNetV2 + geolocation	92% accuracy on Indonesian ASN attendance data
Zawar & Chakkarwar (2023)	MobileNetV2 + transfer learning	98% accuracy on the LCC-FASD dataset
Ardiansyah *et al*. (2025)	Nonlinear diffusion + CNN	87.76% accuracy; secondary-device scenarios were not the main focus of evaluation
Nugroho *et al*. (2024)	MobileNetV2 + REST API	95% accuracy [check reference and data]

Rahmawati *et al.* (2024)	MobileNetV2 + e-Kinerja integration	94% accuracy [check reference and data]

2.3 Research Position

Based on previous studies, research on *face anti-spoofing* has addressed various types of attacks, CNN architectures, and efficiency requirements for system implementation. However, research specifically examining facial authenticity detection in ASN attendance scenarios involving secondary devices remains limited, particularly in local government institutions in Indonesia. This scenario needs to be considered because facial images can be displayed through another device in the form of photographs, videos, or video calls, potentially causing a face recognition system to accept a facial representation as the actual subject. This study positions MobileNetV2 as the model used to detect facial authenticity in the ASN attendance system of the Tangerang City Government. Unlike studies that focus primarily on comparing different architectures, this study emphasizes the evaluation of an applied solution using an attendance dataset from the institution under study. The evaluation is conducted not only based on classification performance but also through model interpretation using Grad-CAM and implementation testing through a REST API. Through this approach, the study aims to assess the ability of MobileNetV2 to distinguish between *real* and *fake* facial images and to evaluate its initial feasibility as an additional verification mechanism for a digital ASN attendance system.

3. Methodology

3.1 Dataset

The dataset was constructed from actual ASN attendance data collected through the Tangerang AYO application, an official super app of the Tangerang City Government available on Android and iOS devices and providing attendance, e-Kinerja, and SIMASN features. The data include facial images of employees from various government agencies (OPDs) within the Tangerang City Government who performed attendance using different types of smartphones and tablets registered to their respective accounts. The dataset was classified into two classes:1. Real, referring to facial images captured directly by the camera of the primary device during the attendance process; and 2. Fake, referring to facial images originating from video recordings, video calls, screenshots, or facial images displayed on a secondary device, such as a smartphone or laptop, and directed toward the camera. The fake-image recording process was designed to simulate potential attendance fraud in operational settings. In this scenario, an employee who was not physically present could appear to perform attendance by displaying a facial image or video on another device positioned in front of the attendance camera. To represent variations in operational conditions, data collection was conducted using a combination of Android and iOS devices commonly used by ASN, with variations in screen size, resolution, and brightness settings based on typical daily usage. Image acquisition was performed in office environments under mixed illumination conditions, including indoor lighting and natural light, during working hours. These conditions were intended to approximate the illumination encountered during actual attendance activities. The dataset consisted of 2,000 images with a nearly balanced class distribution, comprising 1,001 real images and 999 fake images. The dataset was divided into training, validation, and testing subsets, as presented in Table 2.

Table 2. Dataset Distribution

Subset	Real	Fake	Total
Training	700	699	1,399
Validation	150	150	300
Testing	151	150	301
Total	1,001	999	2,000

3.2 Preprocessing and Augmentation

All images were resized to 224 × 224 pixels, corresponding to the standard input size of MobileNetV2, converted into three-channel RGB images, and normalized using the built-in preprocess_input function provided for MobileNetV2. The data pipeline was optimized using caching, shuffle(1000), and prefetch(AUTOTUNE) to improve training efficiency without intentionally altering the underlying data distribution. Data augmentation was applied only to the training set using rotation of ±15°, horizontal flipping, brightness adjustment of ±20%, and zooming of up to 10%. These augmentation ranges were selected to represent reasonable variations in facial pose and illumination during office-based attendance without substantially altering facial characteristics that could affect the distinction between real and fake images. The validation and testing sets were not augmented to maintain an objective evaluation and to reflect the distribution of the attendance images used for evaluation.

3.3 Model Architecture and Training Strategy

A MobileNetV2 model pretrained on ImageNet was used as the feature extractor with an input size of 224 × 224 × 3. A custom classifier head was added consisting of Global Average Pooling, a Dense layer with 128 neurons and ReLU activation, Dropout with a rate of 0.3, and a single-neuron output layer with sigmoid activation for binary classification between real and fake images. The sigmoid function maps the output of the final neuron to a value between 0 and 1, which is interpreted as the probability that an image belongs to the real class. In this study, a probability value of ≥0.5 was classified as real, whereas a value below 0.5 was classified as fake. Training was conducted in two phases. In Phase 1, referred to as feature extraction, all base layers of MobileNetV2 were frozen, and only the parameters of the classifier head were updated for 15 epochs using a learning rate of 1×10^{-3} and Binary Cross-Entropy as the loss function. This phase was intended to adapt the classification layer to the target task while retaining the general visual representations learned during ImageNet pretraining. In Phase 2, referred to as fine-tuning, selected upper layers of MobileNetV2 were unfrozen and retrained for an additional 10 epochs using a lower learning rate of 1×10^{-5}. This approach allowed the model to refine higher-level feature representations for characteristics associated with the ASN attendance domain, such as facial texture under office illumination and screen-surface artifacts in fake images, while preserving the lower-level features learned by the initial layers. The Adam optimizer was used in both training phases with a batch size of 32. ModelCheckpoint was used to retain the model weights corresponding to the lowest validation loss, allowing the best validation-performing model to be selected rather than relying on the weights obtained at the final training epoch.

Table 3. Model Training Configuration

Parameter	Phase 1	Phase 2
Epochs	15	+10 (25 total)
Learning Rate	1×10^{-3}	1×10^{-5}
Batch Size	32	32
Base Layers	Frozen	Selectively unfrozen
Loss Function	Binary Cross-Entropy	Binary Cross-Entropy
Optimizer	Adam	Adam

3.4 Evaluation Framework

The model was evaluated using a held-out test set consisting of 301 images. The evaluation included the following metrics:
1) Confusion Matrix, consisting of TP, TN, FP, and FN as the basis for calculating the classification metrics.
2) Accuracy, representing the proportion of correctly classified images among all test images.
3) Precision, representing the proportion of correctly predicted real images among all images predicted as real.
4) Recall, representing the model's ability to correctly identify real images.
5) Specificity, representing the model's ability to correctly reject fake images.
6) F1-score, representing the harmonic mean of precision and recall.
7) AUC-ROC, representing the model's ability to distinguish between the two classes across different classification thresholds.

In the context of ASN attendance, these metrics were interpreted in relation to the operational risks associated with the system. A False Positive (FP), in which a fake image is classified as real, represents a significant risk because it may allow attendance fraud to pass through the verification mechanism and potentially affect the integrity of TPP and employee performance records. In contrast, a False Negative (FN), in which a genuine ASN image is classified as fake, is more closely associated with user inconvenience and the potential need for additional manual verification. Therefore, precision, recall, specificity, F1-score, and AUC were considered together to evaluate the model's classification performance and its potential operational implications. Particular attention was given to the model's ability to reject fake images while maintaining an acceptable level of false rejection of genuine attendance images. Grad-CAM (*Gradient-weighted Class Activation Mapping*) was also applied as a qualitative analysis method to visualize the image regions that contributed most strongly to the model's classification output for the real and fake classes. This analysis was used to examine whether the model's predictions were associated with visually meaningful regions of the input image.

3.5 REST API Implementation

The best-performing model was implemented as a REST API prototype using FastAPI. The endpoint receives a facial image as input, performs the required preprocessing and MobileNetV2 inference, and returns a JSON response containing the predicted label (real/fake), confidence score, and recommendation status (*accepted*, *rejected*, or *review needed*). The implementation was evaluated in terms of functional correctness and response time under cold-start and warm-start conditions. Testing was performed on a device without a dedicated GPU using an Intel Core Ultra 5 processor to represent a deployment environment with limited hardware acceleration, as may occur in local government infrastructure.

4. Result and Discussion

4.1 Results

4.1.1 Model Training Results

Over 25 training epochs, the training accuracy increased from 76.70% to 97.50%, while the validation accuracy reached 96.67%. The two training phases showed different but complementary training dynamics.

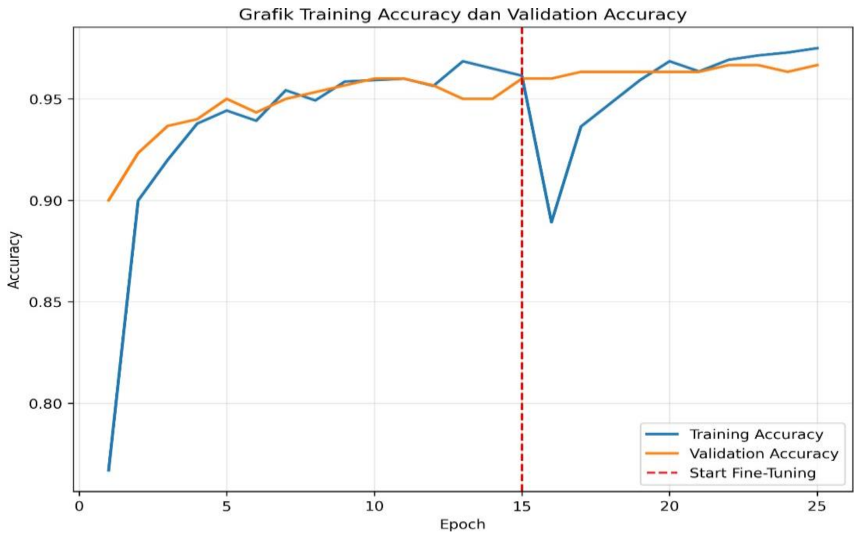

Figure 1. Training and Validation Accuracy.

During Phase 1 (epochs 1–15), the accuracy and loss curves converged relatively steadily, indicating that the classifier head was able to adapt to the feature representations provided by the pretrained MobileNetV2 model. At the beginning of Phase 2 (epoch 16), a temporary increase in training loss was observed after selected upper layers of MobileNetV2 were unfrozen. This change is consistent with the adjustment of previously frozen pretrained weights during the initial stage of fine-tuning. At the same point, the validation loss increased only moderately to approximately 0.12, suggesting that the feature representations obtained during Phase 1 remained relatively stable during the initial fine-tuning process. During the final stage of Phase 2 (epochs 17–25), the training loss continued to decrease and reached 0.0736, while the validation loss remained within approximately 0.10–0.12. This pattern suggests a mild tendency toward overfitting. ModelCheckpoint retained the model state from epoch 24, which achieved the lowest validation loss of 0.1043 and a validation accuracy of 96.67%, rather than using the weights from the final epoch.

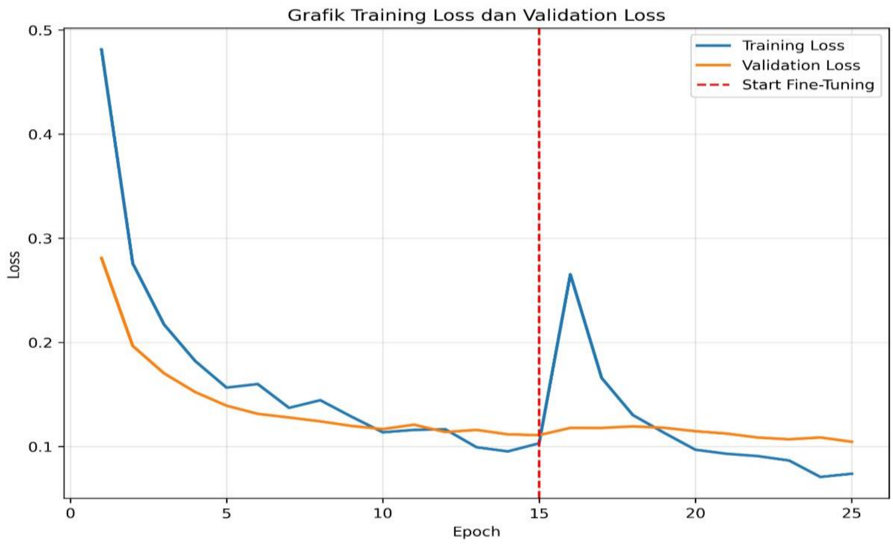

Figure 2. Training and Validation Loss.

4.1.2 Classification Performance Evaluation

The confusion matrix on the test set (n = 301) produced TN = 144, TP = 149, FP = 6, and FN = 2. Overall, the distribution indicates strong classification performance, with correct predictions substantially outnumbering incorrect predictions. However, the errors were not completely symmetrical because the number of FP cases was higher than the number of FN cases. This indicates that a small number of fake images were classified as real, although the overall number of classification errors remained low. The higher number of FP than FN may be associated with the visual characteristics of some fake images. In particular, screen-based facial representations may not always exhibit strong and consistent display artifacts under different capture conditions. In some cases, faces displayed through video calls, replayed recordings, or secondary devices may appear visually similar to genuine facial images, particularly when screen reflections, pixel patterns, or color distortions are not clearly visible. However, this interpretation should be considered a possible explanation rather than a direct causal finding unless the misclassified samples were examined individually. From an operational perspective, the FP and FN results have different implications for ASN attendance. An FN occurs when a genuine facial image is classified as fake, which may result in user inconvenience or the need for additional verification. In contrast, an FP occurs when a fake image is classified as real and therefore represents a more direct risk to the integrity of the attendance verification process. The observed FP = 6 and FN = 2 therefore indicate that further mitigation may be considered at the implementation stage, such as threshold calibration, manual review for uncertain cases, or additional verification mechanisms.

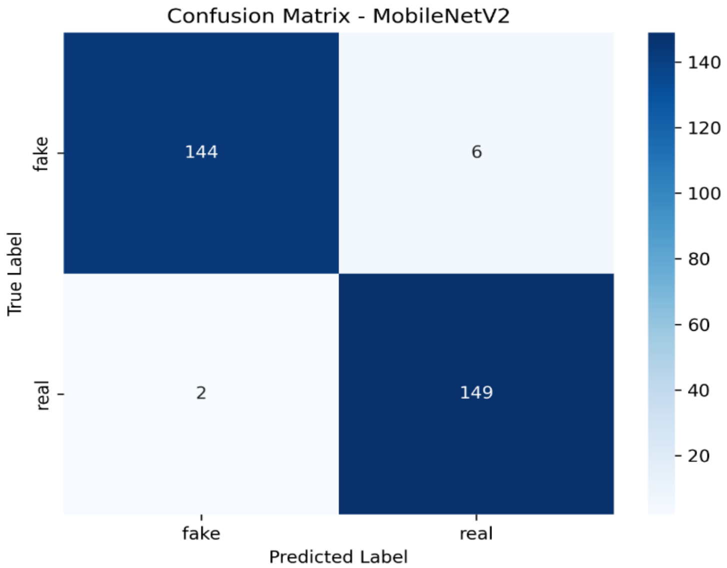

Figure 3. Confusion Matrix.

Table 4. Classification Metrics of the MobileNetV2 Model on the Test Set

Metric	Value
Accuracy	97.34%
Precision	96.13%
Recall	98.68%
Specificity	96.00%
F1-score	97.39%
AUC	0.9934

Table 4 shows that the model achieved an accuracy of 97.34%, precision of 96.13%, recall of 98.68%, specificity of 96.00%, F1-score of 97.39%, and AUC of 0.9934 on the test set. The accuracy obtained in this study was higher than the 91.59% accuracy reported by Najeebullah *et al.* (2025). However, the two results were obtained using different datasets and experimental settings, so the values should not be interpreted as a direct benchmark comparison. The AUC of 0.9934 indicates very strong discrimination between the real and fake classes across different classification thresholds. This result suggests that the model's classification performance was not limited to the selected threshold of 0.5 and that the predicted scores provided strong separation between the two classes on the test set.

4.1.3 Threshold Analysis

The predicted probability distribution on the test set showed a relatively clear separation between the two classes. Most fake images were concentrated at low predicted probabilities, whereas real images were concentrated at high predicted probabilities. This distribution indicates that the 0.5 threshold provided a clear separation between the two classes under the test conditions. A small number of samples were located near the decision threshold. These samples may be associated with less favorable image conditions, such as low resolution, partial facial occlusion, or strong illumination contrast, which can make facial texture and screen artifacts less distinguishable. Such samples tend to produce prediction probabilities closer to 0.5, indicating lower model confidence in one of the two classes. Correctly classified samples generally showed probabilities farther from the 0.5 decision threshold, either toward 0 for fake images or toward 1 for real images. In contrast, borderline samples were closer to the decision boundary and were therefore more sensitive to changes in the classification threshold. This indicates that prediction probabilities can provide additional information for designing the system's decision process rather than being used only to produce the final class label. For operational deployment, this distribution provides a basis for considering a gray zone around the main classification threshold. Images within this zone could be directed to manual review or an additional verification mechanism instead of being immediately accepted or rejected. Such an approach may help reduce the risk of false acceptance while limiting unnecessary rejection of genuine users. However, the appropriate operational threshold should be determined through additional threshold analysis and should consider the institution's tolerance for false acceptance, the capacity for manual review, and the requirements of the attendance process.

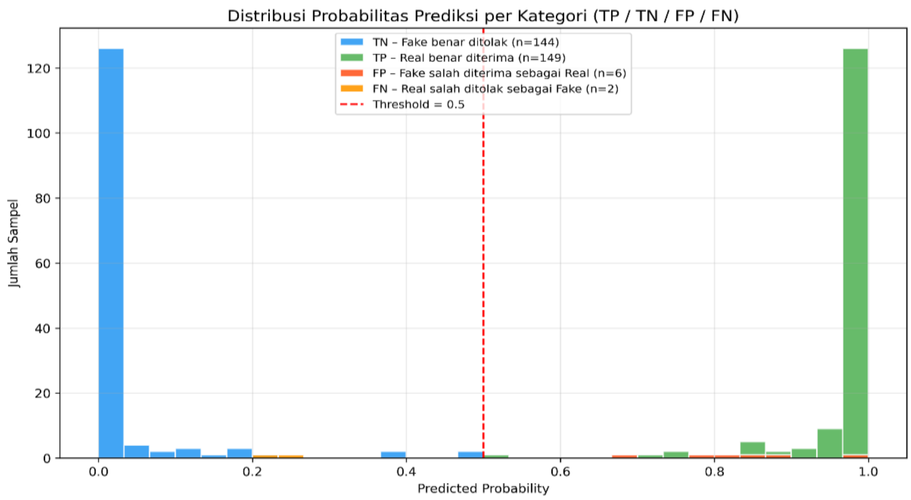

Figure 4. Prediction Probability Distribution.

4.1.4 Grad-CAM Interpretability Analysis

Grad-CAM was used to provide qualitative information about the image regions contributing to the MobileNetV2 predictions. The visualization helps examine whether the model's decisions are associated with visually relevant regions of the input images.

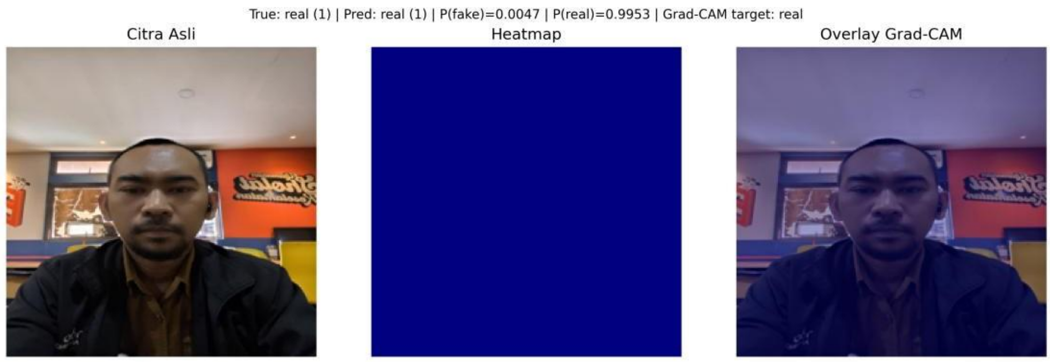

Figure 5. Grad-CAM Visualization of a Real Image.

In the real-image example, the activation was concentrated around peripheral facial regions and parts of the surrounding context, including areas affected by natural illumination and spatial depth. This pattern suggests that the model used information beyond the central facial region when producing its prediction. The observed pattern is consistent with the use of visual texture and contextual information in CNN-based face anti-spoofing approaches, although the Grad-CAM visualization alone cannot establish that these features are the sole basis of the model's decision.

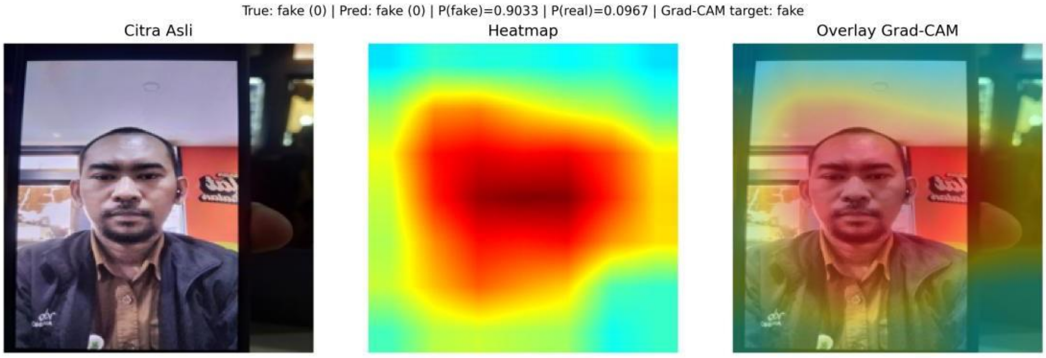

Figure 6. Grad-CAM Visualization of a Fake Image.

In the fake-image example, the activation was concentrated around central facial regions, including the forehead, eyes, cheeks, and nose. These areas may contain visual characteristics associated with facial representations displayed on screens, such as relatively uniform texture, pixel patterns, or color gradients. The difference between the activation patterns in the real and fake examples provides qualitative evidence that different image regions contributed to the two predictions. These visualizations provide an additional perspective on model behavior. However, because Grad-CAM is a qualitative interpretability method, the results should not be interpreted as definitive evidence that the model relies exclusively on physical screen artifacts or genuine skin microtexture. Further analysis using multiple samples would be required to establish whether the observed patterns are consistent across the dataset.

4.1.5 REST API Implementation Feasibility

Functional testing confirmed that the REST API prototype was able to receive a facial image, perform the required preprocessing and MobileNetV2 inference, and return a JSON response containing the predicted label, confidence score, and recommendation status. An example using a real image produced a confidence score of 0.9997, with an accepted decision and a review_needed value of false. This indicates that the API can return a direct decision for cases with a high prediction score.

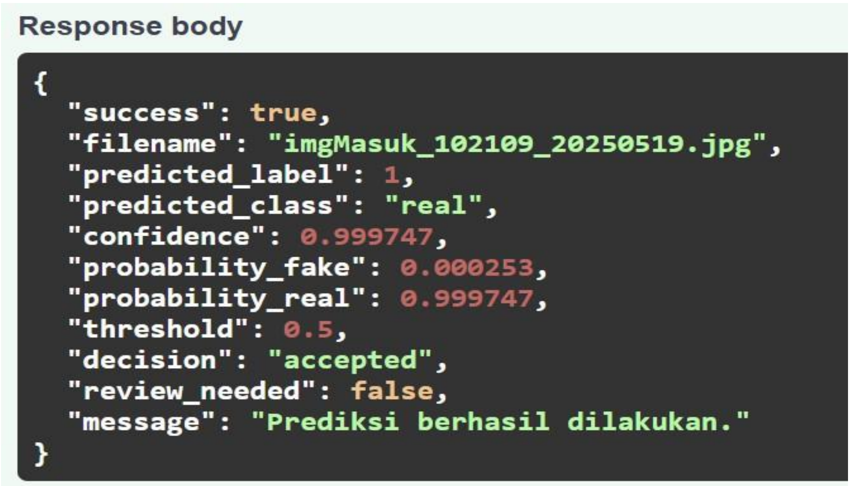

Figure 7. REST API JSON Response.

Table 5. REST API Response Time Sampling Results

Sample	Initial Connection	TTFB (ms)	Total Response (ms)	Condition
R-01	316.10 ms	567.56 ms	889.21 ms	Cold start
R-02	–	531.36 ms	545.65 ms	Warm
R-03	–	182.83 ms	200.35 ms	Warm
R-04	–	167.72 ms	173.45 ms	Warm
R-05	306.71 ms	988.53 ms	1,300.20 ms	Cold start
Warm Average	–	294.04 ms	306.48 s	–

The response-time measurements are summarized in Table 5. Among the five initial samples, the two cold-start requests produced relatively high total response times of 889.21 ms and 1,300.20 ms. In contrast, the three warm requests ranged from 173.45 to 545.65 ms, with an average total response time of 306.48 ms. Additional testing involving 20 requests showed a similar pattern, with cold-start requests producing higher response times and warm requests generally showing lower latency. The distribution indicates that service initialization contributes substantially to response-time variation. After the service is initialized, the REST API operates with lower and more consistent response times on the tested device without a dedicated GPU.

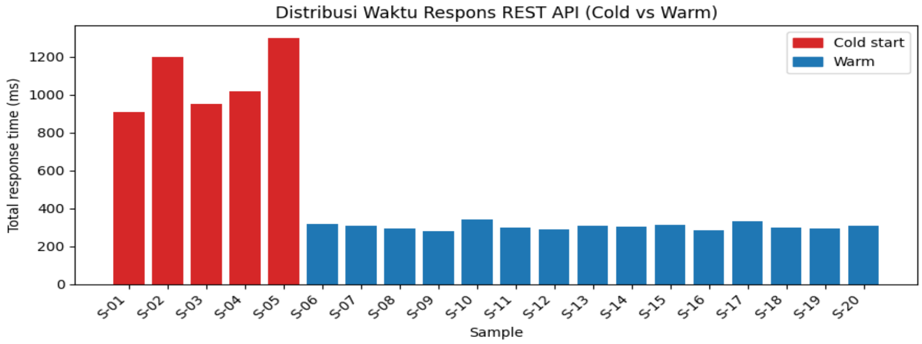

Figure 8. REST API Response Time Distribution.

The observed response-time results suggest that MobileNetV2 can be considered for use as an additional verification component in an ASN attendance system with limited computational resources. However, the measured API response time alone does not establish complete real-time or end-to-end operational feasibility. Additional testing involving simultaneous operation with the existing face-matching system, concurrent requests, network conditions, and actual deployment infrastructure would be required before drawing a final conclusion about production deployment.

4.2 Discussion

The results demonstrate that MobileNetV2 can distinguish between real and fake facial images in the ASN attendance scenario with strong classification performance. The model achieved an accuracy of 97.34%, precision of 96.13%, recall of 98.68%, specificity of 96.00%, F1-score of 97.39%, and AUC of 0.9934 on the test set. These results indicate that the model was able to identify genuine facial images while rejecting most images generated through secondary-device presentation attacks. The high recall indicates that most real images were correctly recognized, while the specificity of 96.00% shows that the majority of fake images were successfully rejected. The classification results are particularly relevant to the attendance scenario because the two types of classification errors have different operational consequences. The test set produced six false positives and two false negatives. A false positive occurs when a fake image is classified as real, allowing a potentially fraudulent attendance attempt to pass the verification stage. In contrast, a false negative occurs when a genuine image is classified as fake, which may require the employee to repeat the verification process or undergo additional manual checking. Therefore, although the overall error rate was low, the presence of six false positives indicates that the model should not be considered a complete solution for preventing attendance fraud. Instead, the model is more appropriately positioned as an additional verification layer that can reduce the likelihood of spoofing attacks.

The higher number of false positives may be related to the visual similarity between some fake samples and genuine facial images. A face displayed through a secondary device does not necessarily produce strong screen artifacts in every capture condition. Variations in screen brightness, image quality, viewing angle, illumination, and camera characteristics can reduce the visibility of pixel patterns, reflections, or other characteristics associated with a displayed image. As a result, some fake samples may contain visual patterns that are sufficiently similar to genuine images to be classified as real. However, this explanation should be interpreted as a possible factor rather than a confirmed cause because the study does not provide a separate analysis of the characteristics of each misclassified sample. The performance obtained in this study is also consistent with previous research showing the potential of lightweight CNN architectures for face anti-spoofing. Zawar and Chakkarwar (2023) reported an accuracy of 98% using MobileNetV2 on the LCC-FASD dataset, demonstrating that MobileNetV2 can achieve high performance while maintaining a relatively lightweight architecture. Najeebullah *et al.* (2025) also reported an accuracy of 91.59% for MobileNetV2 in a comparative evaluation involving MobileNetV2, ResNet50, and ViT. The 97.34% accuracy obtained in this study is higher than the accuracy reported by Najeebullah *et al.* (2025). Nevertheless, these results should not be interpreted as a direct superiority claim because the datasets, attack types, data distributions, and experimental settings were different. The results instead support the suitability of MobileNetV2 as a candidate architecture for the specific attendance environment investigated in this study.

The high AUC value of 0.9934 provides additional evidence of the model's ability to separate the real and fake classes across different classification thresholds. This is important because the classification decision in the current implementation uses a threshold of 0.5. The strong AUC indicates that the model maintains good class discrimination beyond this single operating point. However, an AUC close to 1.0 does not automatically indicate that a threshold of 0.5 is optimal for operational deployment. In an attendance system, the selection of an operational threshold should consider the relative consequences of false acceptance and false rejection. Because accepting a fake image may have a greater impact on attendance integrity than rejecting a genuine image, threshold calibration can be considered in future implementation.

The prediction probability distribution also supports the use of confidence information as an additional decision signal. Most test samples were located relatively far from the 0.5 decision boundary, while a smaller number of samples were located closer to the threshold. Samples near the boundary represent cases in which the model has lower confidence in its classification. In an operational system, such cases could be assigned to a review category instead of being automatically accepted or rejected. This approach is particularly relevant for attendance systems because it allows the system to treat high-confidence cases automatically while directing uncertain cases to an additional verification process. However, the appropriate range for such a review zone should be determined through additional threshold analysis rather than selected arbitrarily.

The Grad-CAM analysis provides qualitative information about the visual regions contributing to the model's predictions. In the real-image example, activation appeared around facial peripheral regions and parts of the surrounding context, whereas the fake-image example showed stronger activation around central facial regions such as the forehead, eyes, cheeks, and nose. These differences suggest that the model may use different visual characteristics when distinguishing between the two classes. For fake images, the activated regions may contain characteristics associated with facial representations displayed on a secondary screen, including relatively uniform textures or display-related visual patterns. However, Grad-CAM should be interpreted as a qualitative explanation rather than definitive evidence of the features learned by the model. The visualization does not establish that MobileNetV2 exclusively relies on skin texture or screen artifacts, and analysis of a larger number of correctly and incorrectly classified samples would be needed to establish whether the observed activation patterns are consistent.

The REST API results further indicate that the proposed model has potential for practical deployment as an additional verification component. The API successfully processed facial images and returned the prediction label, confidence score, and recommendation status in JSON format. The warm-start tests produced an average total response time of 306.48 ms, while cold-start requests required substantially longer processing times. This difference indicates that service initialization contributes to the overall response time. Once the service is initialized, the observed response time suggests that MobileNetV2 can operate with relatively low latency on a device without a dedicated GPU. This is relevant for local government environments where deployment infrastructure may not always include specialized hardware for deep learning inference.

Despite these results, the REST API evaluation should not be interpreted as evidence of complete production readiness. The response-time experiment was conducted on a specific hardware configuration and under a limited testing scenario. The study also did not evaluate concurrent requests, network variability, end-to-end integration with the existing face-matching process, or long-term service stability. Therefore, the current REST API implementation should be regarded as a prototype demonstrating technical feasibility rather than a fully validated production deployment. The findings indicate that MobileNetV2 provides a reasonable balance between classification performance and computational efficiency for face anti-spoofing in an ASN attendance scenario. The model achieved strong test-set performance while requiring an architecture that is comparatively lightweight for deployment. More importantly, the study addresses a practical spoofing scenario involving facial images presented through secondary devices, which differs from a conventional face recognition problem because the system must determine not only whose face is present but also whether the captured facial representation corresponds to a genuine physical presence. The combination of classification, probability-based decision making, Grad-CAM analysis, and REST API implementation therefore provides a practical basis for developing an additional verification layer for digital ASN attendance systems. Further evaluation with subject-independent data splitting, broader attack variations, threshold calibration, larger-scale testing, and integration with the existing attendance infrastructure would be necessary to establish the robustness of the proposed approach under real operational conditions.

5. Conclusion and Recommendations

This study demonstrates that a MobileNetV2-based face authenticity detection model using transfer learning and two-phase fine-tuning can achieve strong classification performance on a domain-specific ASN attendance dataset. On the 301-image test set, the model misclassified eight images, consisting of six false positives and two false negatives, and achieved an accuracy of 97.34%, an F1-score of 97.39%, and an AUC of 0.9934. These results indicate that the proposed model can effectively distinguish between genuine facial images and facial images presented through secondary devices under the evaluated conditions. The distribution of classification errors, with more false positives than false negatives, indicates that some fake images were still classified as real. This finding is relevant to ASN attendance systems because false acceptance can allow unauthorized attendance submissions and potentially affect the integrity of attendance and TPP-related data. Therefore, although the model demonstrated strong overall performance, the remaining false-positive cases indicate that additional verification mechanisms may still be required for operational deployment. The Grad-CAM analysis provides qualitative evidence regarding the image regions that contributed to the model's predictions. Different activation patterns were observed between the real and fake examples, particularly around facial regions and surrounding visual context. However, these visualizations should be interpreted as qualitative evidence and should not be considered definitive proof that the model relies exclusively on physically meaningful features.

The REST API evaluation also indicates that the proposed model has potential as an additional verification component. The API successfully performed image preprocessing and inference and returned the predicted class, confidence score, and recommendation status. The average response time under warm-start conditions was 306.48 ms on the tested device without a dedicated GPU. This result suggests that MobileNetV2 can be considered for deployment in environments with limited computational resources. Nevertheless, the response-time evaluation represents a prototype-level assessment and does not by itself establish complete production readiness. The dataset contains actual ASN attendance images collected across OPDs within the Tangerang City Government environment, providing relatively broad representation within a single local government context. However, the findings cannot yet be assumed to generalize to other government institutions. Differences in devices, image-capture conditions, lighting environments, and spoofing practices may affect model performance in other settings. Independent validation using data from other institutions is therefore necessary.

Future research should expand the spoofing scenarios evaluated by the model, including deepfake-based attacks, generative-AI facial manipulation, and three-dimensional mask attacks. These additional attack types would provide a broader assessment of the model's ability to handle increasingly sophisticated presentation attacks. Further studies should also evaluate the model using data from other local government institutions with different device types, employee populations, and environmental conditions. Subject-independent and cross-institutional testing would be particularly useful for determining whether the model has learned general anti-spoofing characteristics rather than patterns specific to the current dataset. A comparative evaluation involving architectures such as EfficientNet and Vision Transformer should also be conducted under identical experimental conditions. Such comparisons could provide a stronger basis for determining whether MobileNetV2 offers the most appropriate balance between classification performance and computational efficiency for the intended deployment environment. For implementation, the proposed model should initially be deployed as an additional verification layer alongside the existing face-matching process rather than immediately replacing the current mechanism. Threshold calibration and a manual-review mechanism can also be considered to handle uncertain predictions and reduce the risk of false acceptance. Before full deployment, additional testing should evaluate concurrent requests, network conditions, end-to-end processing time, and long-term service stability. Future implementations may consider combining facial authenticity detection with other independent signals, such as geolocation or application-use behavior. A multimodal verification approach could provide additional protection because an attack that bypasses one verification layer may still be detected by another independent signal. Such an approach should, however, be evaluated carefully in terms of accuracy, privacy, computational requirements, and operational feasibility.

References

Ahmad, F. A., & Pratiwi, N. (2025). Implementasi Pengenalan Wajah, Deteksi Kehadiran, dan Geolokasi menggunakan TensorFlow Lite dan Google ML Kit pada Aplikasi Absensi Mobile. *SISTEMASI: Jurnal Sistem Informasi*, *14*(1), 172-186.

Ardiansyah, A. C., Yulian, H., Simon, S., Gunawan, G., & Winardi, S. (2025). Penerapan *liveness detection* dengan metode nonlinear diffusion dan convolutional neural network. *METHODIKA: Jurnal Teknik Informatika dan Sistem Informasi, 11*(1), 1–7. https://doi.org/10.46880/mtk.v11i1.3691

Arnab, A., Dehghani, M., Heigold, G., Sun, C., Lučić, M., & Schmid, C. (2021, October). Vivit: A video vision transformer. In *2021 IEEE/CVF international conference on computer vision (ICCV)* (pp. 6816-6826). IEEE. https://doi.org/10.1109/ICCV48922.2021.00676

George, A., & Marcel, S. (2019, June). Deep pixel-wise binary supervision for face presentation attack detection. In *2019 international conference on biometrics (ICB)* (pp. 1-8). IEEE.

Goodfellow, I. (2016). Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. *Artificial intelligence*, *17*(1-3), 185-203.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications.* http://arxiv.org/abs/1704.04861

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Long, X., Zhang, J., & Shan, S. (2025). Confidence aware learning for reliable face anti-spoofing. *IEEE Transactions on Information Forensics and Security, 20*(8), 5083–5093. https://doi.org/10.1109/TIFS.2025.3564878

Mittal, S., & Patel, S. (2023). Age invariant face recognition techniques: A survey on the recent developments, challenges and potential future directions. *International Journal of Engineering Trends and Technology, 71*(5), 435–460. https://doi.org/10.14445/22315381/IJETT-V71I5P243

Najeebullah, Salman, M., & Swati, Z. N. K. (2025). Face spoofing detection using deep learning. *arXiv preprint arXiv:2503.19223*. https://doi.org/10.48550/arXiv.2503.19223

Nurhakim, B., Rifai, A., Kurnia, D. A., Sudrajat, D., & Supriatna, U. (2025). Smart attendance tracking system employing deep learning for face anti-spoofing protection. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), 10*(3), 496–505. https://doi.org/10.33480/jitk.v10i3.5992

Radad, M., Enab, A. E., Elagooz, S. S., El-Fishawy, N. A., & El-Rashidy, M. A. (2025). Face anti-spoofing detection based on novel encoder convolutional neural network and texture's grayscale structural information. *International Journal of Computational Intelligence Systems, 18*(1). https://doi.org/10.1007/s44196-025-00757-z

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018, June). Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 4510-4520). Ieee.

Sari Khatina, Jasmir, Arvita, & Yulia. (2022). Perancangan sistem absensi facial recognition menggunakan CNN dan liveness detector pada BPR Central Dana Mandiri. *Jurnal Informatika dan Rekayasa Komputer (JAKAKOM), 2*(1), 70–80. https://doi.org/10.33998/jakakom.2022.2.1.63

Shinde, S. R., Bongale, A. M., Dharrao, D., & Thepade, S. D. (2025). An enhanced lightweight face liveness detection method using deep convolutional neural network. *MethodsX, 14*, 103229. https://doi.org/10.1016/j.mex.2025.103229

Simard, P., Bottou, L., Haffner, P., & LeCun, Y. (1998). Boxlets: a fast convolution algorithm for signal processing and neural networks. *Advances in neural information processing systems*, *11*.

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition.* http://arxiv.org/abs/1409.1556

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PmLR.

Xie, Z., Li, J., & Shi, H. (2019). A face recognition method based on CNN. *Journal of Physics: Conference Series, 1395*(1), 012006. https://doi.org/10.1088/1742-6596/1395/1/012006

Xing, H., Tan, S. Y., Qamar, F., & Jiao, Y. (2025). Face anti-spoofing based on deep learning: A comprehensive survey. *Applied Sciences, 15*(12), 6891. https://doi.org/10.3390/app15126891

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PmLR.

Zawar, R., & Chakkarwar, V. (2023). *Real-time face liveness detection and face anti-spoofing using deep learning.* Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-196-8_47