Journal of Artificial Intelligence and Data Science
Vol. 1, No. 1, 2025, pp. 31-41
DOI:

Domain Generalization Performance of CNN Architectures for Face Anti-Spoofing Using a Multi-Dataset Evaluation

Angel Rica Siceliya
Program Study of Information Systems, Faculty of Engineering, Universitas Negeri Surabaya, Indonesia
e-mail address: angel.22091@mhs.unesa.ac.id (corresponding author)

Received: 1 Oktober 2025 | Revised: 10 Oktober2025 | Accepted: 20 Oktober 2025

ABSTRACT

Face recognition is widely used to improve the security level and digital attendance systems, but even still, it can susceptible to the spoofing using photos, videos or fake mask. The objective of this work is to evaluate the generalization performance of three CNN architectures (MobileNetV2, EfficientNetB0 and ResNet50) on face spoof detection by applying them into intra-dataset and cross-dataset experiments with SiW-Mv2, Replay-Attack and Paper-Attack datasets. The working of the deep-learning model is explained along with pre-processing of all datasets: firstly, splitting videos into training, validation and test sets. Secondly, frame extraction Thirdly, data augmentation for the training images; and fourthly Normalization-based preprocess_input method. The models use a fine-tuning method to train on the last 30\% of layers; maximum number of epochs is set at 30, batch size typically is {32},and hyperparameters are fixed. Experimental results are evaluated with respect to accuracy, precision, recall, F1-score, confusion matrix and average drop rate as measure of sensitivity of generalization. This experiment confirms that EfficientNetB0 has the lowest performance loss on all cross-dataset scenarios, therefore being the most stable architecture for domain variations. The adjusted model is fine-tuned by optimization of the learning rate and patience value, which performs better in testing on SiW-Mv2. The chosen model is finally tested on facial recognition based attendance systems to demonstrate practical applicability. The reported results lead to the establishment of more secure attendance systems, with higher robustness and less vulnerability to spoof attacks.

Keywords-Face Recognition; Face Anti-spoofing; Convolutional Neural Network (CNN); Cross-dataset Evaluation; EfficientNetB0; Fine-tuning; Model Generalization

I. INTRODUCTION

One of the possible biometric technologies that can be used to realize remote attendance systems is face recognition. But there are some problems which we need to face right now in face recognition systems, principal of them is their authentication. So security concerns have forced an organization where acquiring and installing a trustable attendance system is the corporate responsibility [1, 2]. In the study of Ming et al. [3], there are many ways that the Presentation Attacks (PA) may be abused by an imposter in order to cheat a face recognition system, for example, using photos, videos, 3D masks and so on.

In several recent works of anti-spoofing on face, CNN based methods have shown their growing dominance over conventional approaches for having representative feature ability and yielding higher performance [4]. Spencer et al [5] who compared deep and shallow CNN for biometric PAD and modified spoofnet architecture. It can efficiently discriminate for presentation attacks and living beings. Then based on some conditions, it can categorize into type of presentation attack. Tu et al. [6] proved that if the environment are not constant or different, this is going to cause making spoofing attacks different as well. For real-time detection systems should be adaptive to different practical situations (e.g. light conditions, background changes and camera quality/resolution distinctions) [7].

Face-recognition–based attendance systems may have the potential of such spoofing attacks in the absence of anti-spoofing detection, and false acceptance can also potentially accept fake faces for valid attendance (false acceptance) or reject real faces with genuine face authentication data (false rejection), that comprise the credibility and reliability of an attended system [8, 9].

Previous works have shown that most CNN-based face anti-spoofing methods work very well to the training set, but far do poorly on another dataset. For example, Bian et al. [10] indicated that conventional methods do not have strong generalization of cross-dataset testing and Sun et al. [11] showed that domain gap (sensors, illuminations and resolution) still bring 'feature shift', which decline the effectiveness or reducing the capability for face anti-spoofing detection.

Given these problems, in this work we will analyze and compare the generalization ability of the following three popular CNN architectures: EfficientNetB0, ResNet50, and MobileNetV2 by means of a series of intra-dataset and cross-dataset experiments over three public datasets: SiW-Mv2, Replay-Attack and Paper-Attack. The research also tries to find the architecture with minimal performance reduction as the most reliable models in operational attendance systems environments. The most accurate model in responding to the presence of the masks is then utilized in face recognition-based attendance system for increased security against spoofing attacks. Thus, the conclusions by this study offer theoretical insights on concept of generalization on CNN models as well as practical implications in terms of recommended architectures for real-world applications.

II. METHOD

The following paragraphs provide a summary of the main steps involved in the research, which span from dataset selection over model construction and training processes to evaluation strategies to assess generalization abilities of the three CNN architectures.

A. Dataset

This work makes use of three public datasets, which include the most popular types of face spoofing attacks: SiW-Mv2, Replay-Attack and Paper-Attack. All the databases are split into training (70%), validation (20%) and testing (10%) as regard of data partition, where the splitting is carried out at the video level and not frame-based. This method is also important to prevent data leakage, which could cause over-estimation of the model's performance. Then all videos of each subset are extracted into frames with a resolution of 224 × 224 pixels, following the standard input size of the CNN architectures used in this research.

1) SiW-Mv2 Dataset

This dataset is one of the modern face anti-spoofing datasets, featuring variations in lighting conditions, camera types, distances, and attack categories that are more diverse compared to earlier generations of datasets. The SiW-Mv2 dataset is available through its official repository [12] and was first introduced in the publication by Guo et al. [13], which highlights a multi-domain attack design to increase the complexity of model evaluation. Tabel 1 is the total number of frames after the extraction process and Figure 1 is the types of attacks that contained in the SiW-Mv2 dataset.

TABLE 1. SIW-MV2 DATASET DISTRIBUTIONS

Dataset	Class	
	Real	***Spoof***
Train	2700	3162
Val	764	912
Test	380	486

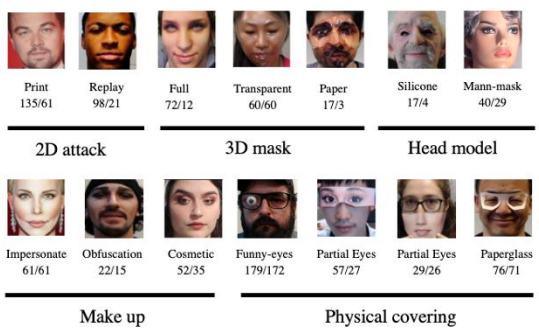

Figure. 1. Attack Types of SiW-Mv2 Dataset

2) Replay-Attack Dataset

The Replay-Attack dataset is one of the traditional benchmarks in face anti-spoofing study. This is a dataset consisting of different replay-based attacks taken with various devices and light conditions. In this study, the dataset was obtained through the Kaggle platform [14], while the complete description of its structure and characteristics was first published by Chingovska et al. [15], who introduced Replay-Attack as a standard dataset for evaluating model robustness against video replay attacks. The total number of frames obtained after the extraction process follows the train–validation–test split defined in the preprocessing stage of this research. Tabel 2 is the total number of frames after the extraction process and Figure 2 is the types of attacks that contained in the Replay-Attack dataset.

TABLE 2. REPLAY-ATTACK DATASET DISTRIBUTIONS

Set	Class	
	Real	***Spoof***
Train	1764	2259
Val	543	552
Test	280	345

Figure. 2. Attack Types of Replay-Attack Dataset

3) Paper-Attack Dataset

The Paper-Attack dataset used in this study was constructed by combining several public datasets available on the Kaggle platform, each providing variations of printed attacks or static-image–based spoofing attempts. The datasets incorporated include the Face Anti-Spoofing Dataset, Advanced Paper Attacks, IBETA Level-1 Liveness Detection, PRINT-OUT Dataset, a subset of CelebA-Spoof, and the Selfie and Video Back-Camera Dataset [16–21]. Although all datasets were obtained from Kaggle repositories, the scientific reference related to printed-photo attacks refers to the publication by Zhang et al. [22], which describes the characteristics of static-image and printed-media attacks in the context of face anti-spoofing. All sources were subsequently merged, normalized, and reprocessed to form a unified Paper-Attack dataset that aligns with the experimental standards of this research. Table 3 is the total number of frames after the extraction process and Figure 3 is the types of attacks that contained in the Paper-Attack dataset.

TABLE 3. PAPER-ATTACK DATASET DISTRIBUTIONS

Set	Class	
	Real	**Spoof**
Train	685	672
Val	189	197
Test	100	87

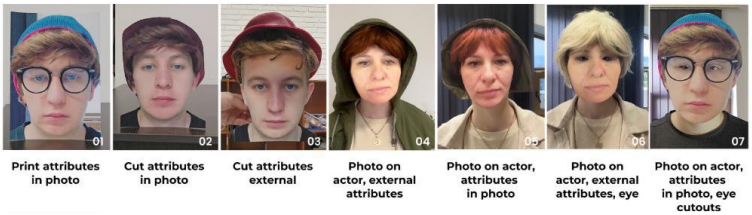

Figure. 3. Attack Types of Paper-Attack Dataset

B. Preprocessing dan Augmentasi

All extracted frames were resized to 224×224 pixels and then processed using preprocess_input according to the respective backbone architecture, which transforms the pixel range to [-1, 1]. In this study, augmentation was applied only to the training data to increase pattern diversity without introducing excessive noise. The augmentation process was implemented using tf.keras.Sequential. Three types of augmentation were applied:

1) RandomFlip: to enhance the model's robustness against variations in facial pose.
2) RandomRotation: in order to assist the model with small facial rotations..
3) RandomZoom: randomization of camera distance.

The validation and test data were not augmented in order to remain representative and unbiased.

C. Achitecture Model

We consider three Convolutional Neural Network (CNN) architectures that have demonstrated success in different visual pattern recognition tasks. We specifically choose these architectures because each belongs to a different type category lightweight, balanced and deep/high-capacity. Therefore, comparing them not only demonstrates the efficiency of individual models but also gives us more general insight into the tension between the complexity and the generalization ability for these two tasks. The first architecture is an already mentioned MobileNetV2, which has been thought for computational efficiency by using inverted residual blocks and linear bottlenecks. The architecture can also perform well on devices with low computational power, which are smartphones and real-time systems highly relevant for camera based attendance system given its need to be responsive. The second model is EfficientNetB0, the root model of the EfficientNet family and uses a compound scaling method to gradually adjust network depth, width and resolution. This method allows the model to balance well the accuracy and efficiency compared with mainstream architectures. Due to its high performance and strong feature representation capabilities, EfficientNetB0 is a strong candidate for addressing data distribution differences in cross-dataset situations. The third architecture is ResNet50, a deep network with very high capacity, which uses skip-connections to counter the problem of gradient degradation in such very deep architectures. The addition of residual connections provides a more stable gradient flow, hence the model is able to learn more complex features. ResNet50 is also a successful strong baseline in many computer vision works, thus including it is crucial to enable comparison with two more lightweight approaches.

D. Training Process

The adopted training method in this work is fine-tuning, which means that some of the pretrained backbone layers are unfrozen to adapt model weights to specific datasets like face anti-spoofing ones. The critical elements of the training process include:

1) Model Initialization

The backbone is pretrained with ImageNet weights. The lower layers of the backbone are frozen to maintain general features learned from natural image domain.

2) Training Hyperparameters

Table 4 shows the parameters that use in training process are designed to according ro the balance between learning capability and convergence stability.

TABLE 4. HYPERPARAMETER

Hyperparameter	Score
Optimizer	Adam
Learning rate	1e−3
Batch size	32
Epoch	30
Loss function	Binary Crossentropy

In the training of all models we kept the hyperparameters fixed for fair comparisons. Once we had chosen the CNN architecture with the most stable generalization, additional optimization was made, by changing some of these hyperparameters.

3) Fine-Tuning Strategy

The following fine-tuning approach was used in this study:

a). 70% of the bottom layers were frozen not to suffer catastrophic forgetting.
b). The upper layers (30%) were re-trained to adapt for texture patterns, reflections, and artifacts that are related with spoofing attacks.

Such an approach permits to use generic symbols, but can be adapted to domain-specific ones.

E. Performance Evaluation

Performance was evaluated to the how well the model can identify attacked face-spoofing patterns in both ideal and out of sample conditions. To achieve this, two evaluation scenarios were applied so that the model's generalization capability could be assessed more comprehensively.

1) Scenarios Evaluation

a) Intra-Dataset Testing

In this scenario, each model is trained and tested using the same dataset, so its performance reflects the model's ability to capture patterns specific to that domain. Table 5 presents the intra-dataset testing scenarios.

TABLE 5. INTRA-DATASET TEST SCENARIOS

No.	Train Dataset	Test Dataset
1	SiW-Mv2	SiW-Mv2
2	Replay-Attack	Replay-Attack
3	Paper-Attack	Paper-Attack

Since this study involves three models to be evaluated and each model undergoes the scenarios shown in the table above, each model will produce three intra-dataset test results. Thus, the total number of intra-dataset test results is nine.

b) Cross-Dataset Testing

This is essence of the experiment, where the model is learned on one dataset and tested on two different datasets. This case is especially critical in face anti-spoofing, where real applications are carried out under circumstances that hinder to resemble the train data. The cross-dataset testing scenarios are reported in Table 6.

TABLE 6. CROSS-DATASET TEST SCENARIOS

No.	Train Dataset	Test Dataset
1	SiW-Mv2	Replay-Attack
2	SiW-Mv2	Paper-Attack
3	Replay-Attack	SiW-Mv2
4	Replay-Attack	Paper-Attack
5	Paper-Attack	SiW-Mv2
6	Paper-Attack	Replay-Attack

All models are tested with the six scenarios as listed above, then get 18 cross-dataset testing scenarios and results in this work.

2) Evaluation Metrics

In order to offer a complete assesment of the performance of the model, we use several well-known binary classification preprocessed evaluation metrics:

a) Accuracy: the ratio of correctly predicted instances to total (all test data points)..
b) Precision: evaluates the correctness of predict spoff class by model.
c) Recall: evaluates the model's capacity to identify all true spoof cases..
d) F1-Score: a harmonic average of precision and recall which gives overall performance.
e) Confusion Matrix: describes in details commercially the number of true positives and negatives, false positive and negatives, and to analyze errors such as false acceptance or false rejection.

3) Generalization Measurement Using Average Drop Rate

Besides the common ones, special attention is made to the Average Drop Rate that in our case quantifies how much does the performance drop when moving from intra-dataset to cross-dataset. The drop rate is the value obtained by substituting into (1):

$$Drop\ Rate = \frac{|A_{intra} - A_{cross}|}{A_{intra}} \times 100\% \quad (1)$$

Models with lower drop estimates may be more robust to real-world jitter, and might be used in operational deployments.

III. PROPOSED METHOD

This section describes the proposed approach used in the study to evaluate and determine the CNN architecture with the most stable generalization capability for face spoofing detection. The proposed method focuses on comparing the performance of three CNN models through cross-dataset testing scenarios, which are designed to simulate real operational conditions of face-recognition-based attendance systems.

A. Proposed Method Workflow

The research approach comprises several integrated steps, from data collection to implementing the best-performing model. In general, the proposed method is shown in Figure 4.

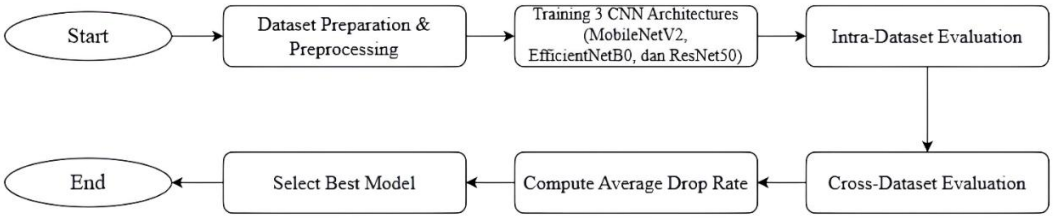

Figure. 4. Proposed Method

The pipeline starts with the preparation of datasets and preprocessing where all videos are split into training, validation, and testing sets using a 70 to 20 to 10 percentage ratio at the video level in order to avoid data leakage. The frames in each video are then resized to 224 × 224 (as opposed to the original 256x256 resolution for technical reasons) and then augmented and normalized to add variability in data and obtain standard input quality. After pre-processing the image data, three convolutional neural network models: MobileNetV2, EfficientNetB0 and ResNet50 are trained with a fine-tuning method proposed with homogeneous hyperparameters to ensure the unbiased and fair comparison. Following training, each model is intra-dataset evaluated on the same dataset as where it's training samples originate, serving as a baseline performance reference for controlled settings. To evaluate the generalization ability, we also perform cross-dataset evaluation by testing on 2 extra datasets which have different characteristics compared to the training set so that how robust our method is against domain shift can be analyzed. The average drop rate is the performance decline level across data-sets, which is used as the primary index for measuring cross-domain stability. Finally, the model that has the lowest average drop rate (shows least degradation with respect to image source) is chosen as most stable architecture. This method is preferred in practice for an attendance system because of its better robustness and stable performance under different data conditions.

B. Generalization Measurement Approach

This work takes a generalization measurement perspective based on Average Drop Rate as we believe accuracy is not always enough to completely understand how well a model generalizes with respect to its stability when faced with distributional shifts. The adoption of Average Drop Rate enables the analysis to recognize how sensitive the model on variations, including lighting condition and device differences, as well as measuring how much the model is dependent on certain patterns exists in training dataset. Furthermore, this measure represents a less biased and more general evaluation of robustness than assessing the cross-dataset performance in isolation. Therefore, the Average Drop Rate is the main characteristic of the method we propose and it is used to look for the best model, i.e., that with lower value of this rate and a small variance.

IV. RESULT AND DISCUSSION

This section discusses the experimental results obtained from training and testing three CNN architectures, there are MobileNetV2, EfficientNetB0, and ResNet50, across two main scenarios: intra-dataset testing and cross-dataset testing. The primary objective of this analysis is to assess the generalization capability of the models, namely how well their performance remains stable when evaluated on domains that differ from the training data. Generalization is a critical factor in the context of implementing face-recognition-based attendance systems, as real-world operational conditions are highly variable and cannot be fully represented by a single dataset.

A. Intra-Dataset Testing Results

In the intra-dataset testing, all models achieved high accuracy across the three datasets (Table 7). This behavior coincides with results already observed in prior works where CNN reply very well to testing data which derives from the same sample distribution used for training. This indicates the models are able to learn important texture, reflection and visual pattern features in the training data.

TABLE 7. INTRA-DATASET TEST RESULTS

Model	Dataset	Accuracy	Precision	Recall	F1-Score
EfficientNetB0	SiW-Mv2	0.95	0.95	0.95	0.95
	Replay-Attack	0,94	0,94	0,94	0,94
	Paper-Attack	0,66	0,67	0,67	0,66
ResNet50	SiW-Mv2	0,94	0,94	0,94	0,94
	Replay-Attack	0,98	0,98	0,98	0,98
	Paper-Attack	0,72	0,72	0,72	0,72
MobileNetV2	SiW-Mv2	0,86	0,86	0,86	0,86
	Replay-Attack	0,98	0,98	0,98	0,98
	Paper-Attack	0,78	0,78	0,77	0,77

B. Cross-Dataset Testing Results

In order to test the transferability of these models and their ability to remain generalizable in face of new domains that were not present in the training data, cross-dataset evaluation was performed.

1) Trained Model for SiW-Mv2

The Cross-dataset test results for each trained model obtained across datasets are showed in Table 8.

TABLE 8. CROSS-DATASET TEST RESULTS WHEN TRAIN IN SIW-MV2

Model	SiW-Mv2 (Intra)	Replay-Attack (Cross)	Paper-Attack (Cross)	Drop Rate (Replay-Attack)	Drop Rate (Paper-Attack)	Avg Drop Rate
EfficientNetB0	0,95	0,83	0,82	12,63%	13,68%	13,16%
MobileNetV2	0,94	0,78	0,53	17,02%	43,62%	30,32%
ResNet50	0,86	0,78	0,55	9,30%	36,05%	22,67%

According to the findings, it can be noticed that EfficientNetB0 is the most stable model with the average drop rate of 13.16%, which also means that this architecture can capture rather general spoofing patterns. MobileNetV2 was also the one with prominent performance drop to handle the Paper-Attack in comparison to other models. ResNet50 On the other end, ResNet50 presented a fair performance, obtaining less drop rate at Replay-Attack which it was already quite big on against Paper-Attack. In a nutshell, training on a more diverse dataset, SiW-Mv2 in our case results into models that generalise better against domain shifts.

2) Trained Model for Replay-Attack

The Cross-dataset test results is presented in Table 9 for each model trained with Replay-Attack dataset.

TABLE 9. CROSS-DATASET TEST RESULTS WHEN TRAIN IN REPLAY-ATTACK

Model	Replay-Attack (Intra)	SiW-Mv2 (Cross)	Paper-Attack (Cross)	Drop Rate (SiW-Mv2)	Drop Rate (Paper-Attack)	Avg Drop Rate
EfficientNetB0	0,94	0,63	0,88	32,98%	6,38%	19,68%
MobileNetV2	0,98	0,56	0,47	42,86%	52,04%	47,45%
ResNet50	0,98	0,52	0,76	46,94%	22,45%	34,69%

As can be seen from the table 9, when models are trained on the Replay-Attack dataset, all the architectures suffered significant performance drop on SiW-Mv2 and Paper-Attack. EfficientNetB0 has the lowest drop rate on average of 19.68% comparatively to MobileNetV2 and ResNet50. The poor performance on SiW-Mv2 shows that Replay-Attack does not cover enough diversity of conditions and types of attack for the model to generalize well. MobileNetV2 exhibited the most reduction in performance, with a loss rate of about 50%, and ResNet50 achieved relatively acceptable results but still showed slippage. These results emphasise the fact that over-simplistic datasets can propagate to a low generalisation model.

3) Trained Model for Paper-Attack

The Cross-dataset test results for each models trained on the Paper-Attack dataset from Table 10.

TABLE 10. CROSS-DATASET TEST RESULTS WHEN TRAIN IN PAPER-ATTACK

Model	Paper-Attack (Intra)	SiW-Mv2 (Cross)	Replay-Attack (Cross)	Drop Rate (SiW-Mv2)	Drop Rate (Replay-Attack)	Avg Drop Rate
EfficientNetB0	0,66	0,49	0,49	25,76%	25,76%	25,76%
MobileNetV2	0,72	0,58	0,61	19,44%	15,28%	17,36%
ResNet50	0,78	0,46	0,52	41,03%	33,33%	37,18%

The overall generalization performance of training on the Paper-Attack dataset is the weakest according to these results. Meanwhile, MobileNetV2 remained themost robust architecture with an average drop rate of 17.36%, although it was marginally worse than EfficientNetB0. The significant performance drop on SiW-Mv2 for EfficientNetB0 and ResNet50 implies that the features of printed-based spoofing attacks are not generalized enough to describe more dynamic attacks like replay, makeup, or mask. Overall, these results suggest that training on a static dataset like Paper-Attack is less effective for learning representations which can generalize to a broader variety of real-world spoofing variations.

C. Average Drop Rate Results

And to fairly compare the stability of the architectures, we also computed its Average Drop Rate among all cross-dataset scenarios. This average drop in accuracy from intra-dataset to cross-dataset evaluations is a direct measure of generalization. Table 11 is how the Average Drop Rate looks like for each model.

TABLE 11. AVERAGE DROP RATE RESULTS

Model	Avg Drop Rate (All Scenarios)
EfficientNetB0	19,53%
MobileNetV2	31,71%
ResNet50	31,52%

Table 11 shows that EfficientNetB0 has the lowest Average Drop Rate of 19.53% such that this model is comparably stable when evaluated on other domains. MobileNetV2 and ResNet50 present dropout ratio values of 31% or more, which implies them to be quite sensitive for data distribution. The results imply that the compound scaling and architecture structure of EfficientNet possess an advantage in learning more global spoofing patterns, so that the model is less reliant on specific characteristics of the training dataset. Thus, EfficientNetB0 is chosen as the most suitable candidate for further fine-tuning and deployment in the attendance systems of this paper.

D. Fine-Tuning of Selected Model (EfficientNetB0) for Attendance Systems Adoption.

According to the final testing on 5 datasets overall, EfficientNetB0 was chosen as it has the most stable cross-dataset generalization. For its applicability to a face-recognition-based attendance systems, further tuning was carried out by using

SiW-Mv2 as the main dataset. We use this dataset because of its diversity and representative nature, so that the final model is more robust across different spoofing attacks. The optimization was carried out with some procedures of trial-and-errors where certain hyperparameters were adjusted improving stability and accuracy. Table 12 displays the modified hyperparameter values and their corresponding RMSEs.

TABLE 12. MODEL EFFICIENTNETB0 OPTIMIZATION RESULTS

Experiment	Drop Out	Patience	Trainable layers (%)	Learning Rate	Epochs	Accuracy
1	0,3	3	30%	1e-3	30	0,95
2	0,3	3	30%	1e-5	30	0,95
3	0,3	5	30%	1e-5	30	0.97

The optimization process of the EfficientNetB0 network was conducted through three experiments by exploring some important hyperparameters considered in detail to achieve high stability and accuracy on SiW-Mv2 dataset. For the first experiment, we trained the model with a rate of dropout 0.3 with EarlyStopping patience equals to 3 along 30% trainable layers using learning rate of 1e-3 and needed to train for 30 epochs, achieving an accuracy of 0.95. The second experiment was the same configuration as the first simulation, only with a smaller learning rate 1e-5 and yielding an accuracy of 0.95 as well. Its reduction improved stability of training, but did not noticeably increase accuracy. The third experiment showed a performance leap and we used the learning rate as 1e-5 with the patience now as 5 compared to 3. The accuracy is brought up to 0.97 by giving a more time for the model to finetune.

These results demonstrate that a higher patience value has a positive impact on the optimization process, while the combination of a 0.3 dropout rate, 30% trainable layers, and a low learning rate of 1e-5 proves to be the most optimal configuration for producing a stable final EfficientNetB0 model ready for integration into an anti-spoofing attendance systems. Figure 5 is the visualization of the confusion matrix and classification report for the optimized EfficientNetB0 model.

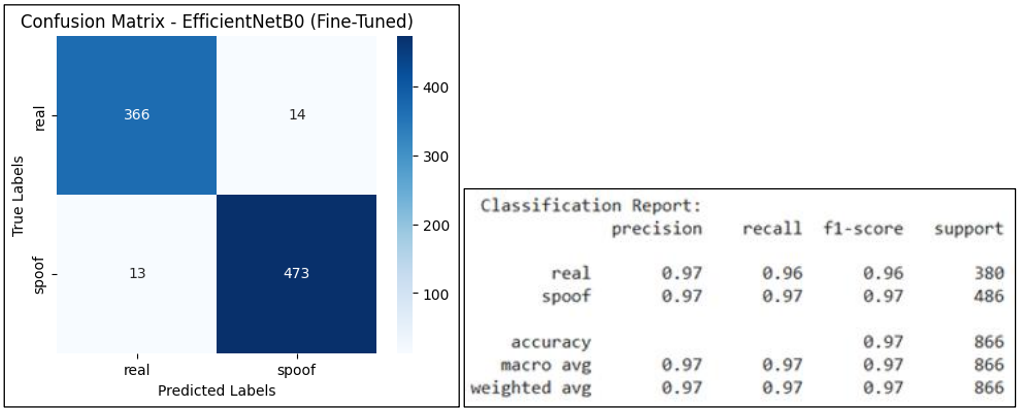

Figure. 5. Confussion Matrix & Classification Report

E. Implementation of the Model in the Attendance Systems

Finally, the fine-tuned EfficientNetB0 model was incorporated into this study's face recognition–based attendance application. This decision merges the face detection module, the face recognition module and the anti-spoofing detection one working with camera on a device. Figure 6 shows the system works well under normal operating conditions where registered users' faces are successfully detected and classified as real with a green bounding-box appears around the user along with the name label of the recognized user. On the other hand (Figure 7), if a user is trying to take attendance from an image or video of the displayed photo or video on any remote device, then the system may alert for suspicious action and mark them as spoofing attacks through red bounding boxes and indicating that it has detected with status message. Attendance attempts are logged in the attendance history (Figure 8) with both status (Present or Spoof Detected) and time-stamp details. These implementation results indicate that the EfficientNetB0 model works responsively and accurately in real application scenario improving the security of attendance systems by mitigating risks of manipulation with fake faces.

1) The System when the Sttendance is Successful

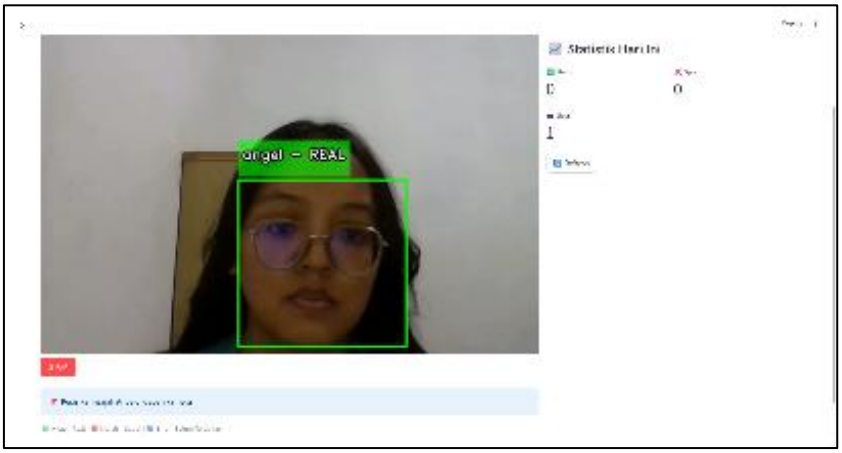

Figure. 6. Attendance Systems Detecting Real Face

2) The System when the Attendance is fails (Spoofing Detected)

Figure. 7. Attendance Systems Detecting Fake Face/Spoof

3) Attendance History

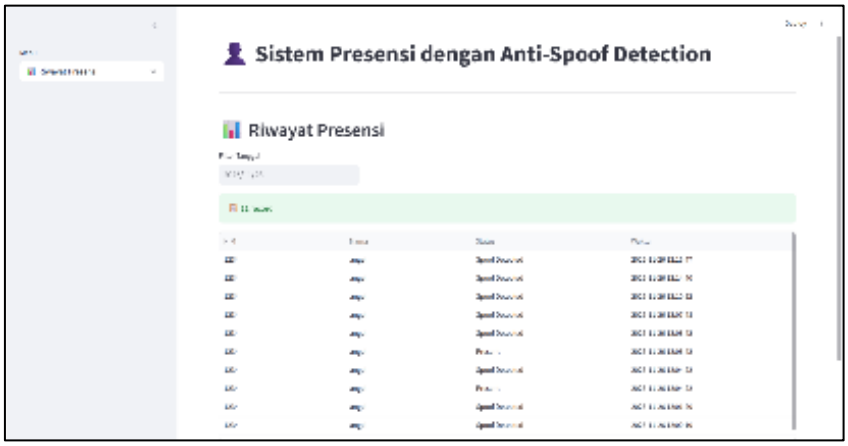

Figure. 8. Attendance History

V. CONCLUSION

This study evaluates the generalization capability of three CNN architectures—MobileNetV2, EfficientNetB0, and ResNet50, through a series of intra-dataset and cross-dataset tests using the SiW-Mv2, Replay-Attack, and Paper-Attack datasets. The experimental results show that all three models experience performance degradation when tested on different domains, confirming that face spoofing detection is highly sensitive to variations in data distribution. Among the three architectures, EfficientNetB0 consistently achieves the lowest average drop rate in two out of three scenarios, allowing it to be identified as the model with the most stable generalization capability.

The fine-tuning step of EfficientNetB0 The selected base model, EfficientNetB0 was further tuned with different hyperparameter settings applied to the SiW-Mv2 dataset. This optimization, especially the increased value of EarlyStopping patience reveals a better accuracy performance (up to 0.97) and thus making the model more deployable on real time attendance systems. The implementation of the model in the face recognition–based attendance application demonstrated that the system can reliably distinguish between genuine and spoofed faces, responding effectively under normal usage conditions as well as during simple spoofing attempts.

So, this paper helps to understand how CNN generalizes in face spoofing detection, and the findings show that combining a strong generalizable model can improve the security of digital attendance systems. Possible future works include generalization/adaptation to domains other than the available training sets, inclusion of more diverse fine tune tasks and operational scenarios that are representative with respect to complexity demand challenge against a wider variety of spoofs.

REFERENCE

[1] A. Anshari, S. A. Hirtranusi, D. I. Sensuse, Kautsarina, and R. R. Suryono, "Face Recognition for Identification and Verification in Attendance System: A Systematic Review," in 10th IEEE International Conference on Communication, Networks and Satellite, Comnetsat 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Jul. 2021, pp. 316–323. doi: 10.1109/COMNETSAT53002.2021.9530817.

[2] P. Anthony, B. Ay, and G. Aydin, "A review of face anti-spoofing methods for face recognition systems," in 2021 International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Aug. 2021. doi: 10.1109/INISTA52262.2021.9548404.

[3] Z. Ming, M. Visani, M. M. Luqman, and J. C. Burie, "A survey on anti-spoofing methods for facial recognition with rgb cameras of generic consumer devices," J Imaging, vol. 6, no. 12, Dec. 2020, doi: 10.3390/jimaging6120139.

[4] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep Learning for Face Anti-Spoofing: A Survey," IEEE Trans Pattern Anal Mach Intell, vol. 45, no. 5, pp. 5609–5631, May 2023, doi: 10.1109/TPAMI.2022.3215850.

[5] J. Spencer et al., "Presentation Attack Detection using Convolutional Neural Networks and Local Binary Patterns," 2023. doi: 10.48550/arXiv.2312.00041.

[6] X. Tu, H. Zhang, M. Xie, Y. Luo, Y. Zhang, and Z. Ma, "Deep Transfer Across Domains for Face Anti-spoofing," Dec. 2019, doi: 10.1117/1.JEI.28.4.043001.

[7] H. Xing, S. Y. Tan, F. Qamar, and Y. Jiao, "Face Anti-Spoofing Based on Deep Learning: A Comprehensive Survey," Applied Sciences (Switzerland), vol. 15, no. 12, Jun. 2025, doi: 10.3390/app15126891.

[8] N. Surantha and B. Sugijakko, "Lightweight face recognition-based portable attendance system with liveness detection," Internet of Things (Netherlands), vol. 25, Apr. 2024, doi: 10.1016/j.iot.2024.101089.

[9] D. Saraswat, P. Bhattacharya, T. Shah, R. Satani, and S. Tanwar, "Anti-spoofing-enabled Contactless Attendance Monitoring System in the COVID-19 Pandemic," in Procedia Computer Science, Elsevier B.V., 2022, pp. 1506–1515. doi: 10.1016/j.procs.2023.01.129.

[10] Y. Bian, P. Zhang, J. Wang, C. Wang, and S. Pu, "Learning Multiple Explainable and Generalizable Cues for Face Anti-spoofing," Feb. 2022, doi: 10.48550/arXiv.2202.10187.

[11] Y. Sun, Y. Liu, X. Liu, Y. Li, and W.-S. Chu, "Rethinking Domain Generalization for Face Anti-spoofing: Separability and Alignment," Mar. 2023, doi: 10.48550/arXiv.2303.13662.

[12] Computer Vision Lab, Michigan State University, "SiW-Mv2 Dataset." [Online]. Available: https://cvlab.cse.msu.edu/siwm-v2-dataset.html.

[13] X. Guo, Y. Liu, A. Jain, and X. Liu, "Multi-domain Learning for Updating Face Anti-spoofing Models," Apr. 2023, doi: 10.48550/arXiv.2208.11148.

[14] TrainingDataPro, "Real vs Fake Anti-Spoofing Video Classification Dataset." [Online]. Available: https://www.kaggle.com/datasets/trainingdatapro/real-vs-fake-anti-spoofing-video-classification

[15] I. Chingovska, A. Anjos, and S. Marcel, "On the Effectiveness of Local Binary Patterns in Face Anti-spoofing," Proceedings of the 11th International Conference of the Biometrics Special Interes Grou, 2012. [Online]. Available: http://www.idiap.ch/dataset/replayattack.

[16] AxonData, "Face Anti-Spoofing Dataset." [Online]. Available: https://www.kaggle.com/datasets/axondata/face-anti-spoofing-dataset.

[17] AxonData, "Advanced Paper Attacks Dataset." [Online]. Available: https://www.kaggle.com/datasets/axondata/face-anti-spoofing-advanced-paper-attacks.

[18] TrainingDataPro, "IBETA Level-1 Liveness Detection Dataset Part 1." [Online]. Available: https://www.kaggle.com/datasets/trainingdatapro/ibeta-level-1-liveness-detection-dataset-part-1.

[19] TrainingDataPro, "Printout Dataset." [Online]. Available: https://www.kaggle.com/datasets/trainingdatapro/printout.

[20] TrainingDataPro, "CelebA Spoof Dataset." [Online]. Available: https://www.kaggle.com/datasets/trainingdatapro/celeba-spoof-dataset.

[21] Tapakah68, "Selfie and Video on Back Camera Dataset." [Online]. Available: https://www.kaggle.com/datasets/tapakah68/selfie-and-video-on-back-camera

[22] Y. Zhang, Z. Yin, L Y. Li, G. Yin. J. Yan, J. Shao, Z. Liu, "CelebA-Spoof: Large-Scale Face Anti-Spoofing Dataset With Rich Annotations," in *Proc. European Conference on Computer Vision (ECCV)*, pp. 70–85, 2020, doi: 10.1007/978-3-030-58610-2_5.