International Journal of Safety and Security Engineering
Vol. 15, No. 2, February, 2025, pp. 323-330
Journal homepage: http://iieta.org/journals/ijsse

Authentication in Liveness Detection Utilizing CNN and MobileViT Algorithm

Vera Suryani*, Fazmah Arif Yulianto, Parman Sukarno, Gian Maxmillian Firdaus

School of Computing, Telkom University, Jl. Telekomunikasi No.1, Bandung 40257, Indonesia

Corresponding Author Email: verasuryani@telkomuniversity.ac.id

https://doi.org/10.18280/ijsse.150213

Received: 9 December 2024
Revised: 2 February 2025
Accepted: 13 February 2025
Available online: 28 February 2025

Keywords:
liveness detection, CNN, MobileViT, replay attack, print attack

ABSTRACT

Liveness detection is a critical component in biometric security systems, aiming to distinguish between live and spoofed biometric samples to ensure system authentication. Image technology advancements are one of the factors that lead to attacks on liveness detection. The use of camera and mask, have made it less difficult to generate attacks that target the liveness detection system, including deep fake, replay, and print attacks. A reliable approach is required to more accurately identify these attacks. Recent advances in deep learning have shown significant promise in addressing these challenges by learning robust and adaptive features directly from raw biometric data. This paper provides an experimental research of deep learning approaches for liveness detection, focusing on Convolutional Neural Networks (CNNs) including EfficientNetV2S, EfficientNetV2M, EfficientNetV2L, and comparing with MobileViT for facial recognition in liveness detection. The datasets employed are NUAA, Synthetic, and iBeta 1. This paper examines the strengths and limitations of each method, and evaluation metrics used in the field, and highlight the latest breakthroughs in improving detection accuracy and robustness against diverse replay and print attacks. Experimental results show that EfficientNetV2S outperforms other algorithms, both in terms of accuracy and false detection rate.

1. INTRODUCTION

Liveness detection is a security mechanism in biometric systems intended to distinguish between authentic, live biometric characteristics (e.g., a living fingerprint, face, or iris) and artificial versions such as images, masks, or prosthetics. This ensures that biometric authentication systems are not effortlessly deceived by spoofing attacks, in which an impostor use counterfeit artifacts to act like an authentic user. Liveness detection is necessary for improving the reliability and security of biometrics, particularly in sensitive applications such as financial transactions, border control, and secure access, where precise identity verification is vital to avert fraud and illegal access. Integration enhances trust in biometric systems and protects against emerging threats.

Latest developments in liveness detection through deep learning have greatly improved biometric security. A review [1] performed a systematic assessment of fingerprint liveness detection techniques, emphasizing progress and persistent obstacles in mitigating presentation attacks.

In 2021, Sabaghi et al. [2] conducted an extensive assessment on deep-feature-based face anti-spoofing techniques, classifying diverse methodologies and addressing unresolved research difficulties.

Tapia et al. [3] presented a serial architecture with a modified MobileNetV2, designed to differentiate between authentic and presentation attack iris images, gaining significant success in the LivDet-Iris 2020 competition.

Koshy and Mahmood [4] implemented real-time systems that combine anisotropic diffusion with deep convolutional neural networks for face liveness detection, achieving good accuracy on the Replay-Attack and Replay-Mobile datasets.

Kuznetsov et al. [5] introduced AttackNet, a specialized convolutional neural network architecture aimed at improving biometric security via efficient liveness detection. This research highlights the crucial importance of deep learning in enhancing liveness detection across diverse biometric modalities.

The study [6] contributes to the advancement of biometric authentication and liveness detection systems that utilize facial recognition techniques. It discusses a system that utilizes algorithms such as Haar cascades and TensorFlow models to detect specific facial movements, including eye blinks, smiling, and mouth openings, through the use of multiple modules. The implementation is deployed as an API for real-time processing and achieves high accuracy in detecting these expressions, underscoring the significance of liveness detection in the protection of facial recognition systems against deceptive attacks.

The paper [7] focuses on developing robust methods for detecting spoofing attacks and ensuring liveness in face recognition systems. The paper highlights the growing concern of security breaches in biometric systems due to spoofing, where attackers use photos, videos, or 3D models to impersonate legitimate users. The author proposes the use of deep learning architecture, specifically CNNs, for detecting such spoofing attempts and ensuring the liveness of the face being recognized. The study incorporates various datasets to train and evaluate these models, demonstrating their effectiveness in distinguishing between real and spoofed faces. The paper concludes that deep learning-based approaches can significantly enhance the reliability and security of face recognition systems, making them more resilient against sophisticated spoofing methods.

The paper [8] presents a novel approach to enhancing the security of face liveness detection systems on mobile devices. It addresses the vulnerability of current face recognition systems to spoofing attacks, particularly those that use static images or videos. The author introduces a method based on analyzing lip motion patterns, which are unique to live individuals and difficult to forge with conventional spoofing techniques. By leveraging motion analysis, the proposed system improves the detection of liveness while minimizing false positives and negatives. The paper demonstrates the effectiveness of this approach through experiments, showing that lip motion patterns are a reliable biometric cue that strengthens face liveness detection on mobile platforms. The study suggests that incorporating this technique can significantly enhance the security of mobile face recognition systems against spoofing attacks.

The utilization of deep learning in the liveness detection is highly prospective, as evidenced by the numerous studies that have been previously discussed. Nevertheless, the various datasets used in the study require better detection of accuracy. This study investigates several data sets to improve the detection accuracy of various attacks in liveness detection.

This paper is organized into multiple sections: the Introduction is provided in the first section, followed by Materials and Methods used in the experiment are described in the second section. The experiment's Results and Discussion are discussed in the third section, and the research summary are concluded in the conclusion.

2. MATERIAL AND METHOD

Several researchers have undertaken numerous experiments looking at video injection attacks using deep learning. Taeb and Chi [9] presented a deepfake detection framework that uses two deep learning models, Xception and MobileNet, to classify fake films from the FaceForensics++ dataset. The models, trained on modified films created by four mainstream approaches (Deepfakes, Face2Face, FaceSwap, and NeuralTextures), achieved accuracies above 90% for the majority of datasets, however performance dipped for NeuralTextures (e.g., MobileNet: 88% accuracy). A voting method that aggregates forecasts from all models increases resilience. The work emphasizes model sensitivity to certain manipulation approaches and the necessity for larger datasets and additional variables, such as inter-frame correlations, for better detection.

Elsaeidy et al. [10] presented a CNN-based system for identifying replay attacks in smart cities based on multivariate time-series data from synthetic datasets sourced from Queanbeyan, Australia. The model was tested on soil management and environmental monitoring datasets (89,566 and 178,211 instances, respectively) and achieved 99.18% accuracy (precision: 99.31%, specificity: 99.26%, sensitivity: 99.11%) on the soil dataset and 98.47% accuracy (precision: 98.56%, specificity: 98.62%, sensitivity: 98.31%) on the environmental dataset. Compared to five cutting-edge approaches, including DRN and ESNC, the suggested model outscored competition across all measures. The work underlines the necessity of modeling the time dimension in attack detection and provides a smart city benchmark dataset for future research.

Zhang et al. [11] examined defense strategies for adversarial perturbations in deep neural networks (DNNs), classifying them as perturbation detection, input modification, stochastic defense, adversarial training, and certified robustness. Adversarial training showed remarkable robustness on datasets such as CIFAR10, but it needed large processing resources. JPEG compression and denoising autoencoders like MagNet obtained great detection rates. Perturbation detection approaches, such as Mahalanobis-based confidence scoring, demonstrated up to 96% accuracy against attacks such as CW and PGD, whereas stochastic defenses based on random noise or activation pruning increased robustness while reducing accuracy loss. The paper emphasizes the trade-offs between robustness and efficiency, recommending for the use of complementary strategies to achieve optimal defense.

Kelly et al. [12] proposed using morphed photos in the training phase to make face recognition systems (FRSs) more resilient against morphing attacks. They detecting discrepancies in identification attributes suggestive of morphing by training on real and augmented photos (morphs and authentic augmented pairings) using a VGG16-based architecture and triplet loss. By reducing the Morph Accept Rate at Equal Error Rate (MAREER) from 30.20% to 20.54% and improving the Differential Equal Error Rate (D-EER) from 6.70% to 5.61% on the FRGC test set, the experiments demonstrated increased resilience. On the ASML validation set, D-EER increased from 5.86% to 4.97%, whereas MAREER decreased from 24.94% to 16.48%. Nevertheless, difficulties with generalization on pose-variation datasets, as PUT, brought to light the necessity of realistic and varied morphing datasets to improve training.

Meena and Tyagi [13] present a deep learning-based approach to picture splicing detection that combines an SVM classifier, ResNet-50 as a feature extractor, and Noiseprint preprocessing to extract noise residuals. The technique outperforms current methods with an average detection accuracy of 97.24%, according to experiments conducted on the CUISDE dataset. The algorithm demonstrated its resilience in detecting spliced photos using camera-specific noise and deep feature extraction by properly classifying 178 out of 180 fabricated images and 175 out of 183 real images. Future research proposes expanding this technique to locate spliced areas in photos and identify splicing in films.

Arora et al. [14] proposed a robust deep learning framework for detecting face spoofing attacks, such as replay attacks, 3D mask attacks, and photo attacks, that employs convolutional autoencoders for dimensionality reduction and feature extraction, followed by classification using pre-trained encoder weights and a softmax classifier. When tested on three benchmark datasets (CASIA-FASD, Replay-Attack, and 3DMAD), the framework obtained high accuracy: 99.17% for CASIA-FASD, 99.03% for Replay-Attack, and 100% for 3DMAD, with a Half Total Error Rate (HTER) of 0%. Cross-database testing yielded good results, proving the framework's robustness and generalizability for detecting spoof faces in biometric systems.

The previous paper [15] proposes a method to enhance the security of liveness detection by extracting human physiological components from computational ghost imaging (CGI) signals. The method achieves a 96.0% correct rate against picture and mask attacks and is resolution-independent, working even at 32×32 pixels.

Meanwhile, authors [16] address the critical issue of securing facial recognition systems against spoofing attacks, which often involve the use of photos, videos, or masks to impersonate legitimate users. The author proposes robust and reliable liveness detection models designed to accurately distinguish between real faces and spoofed ones in various environmental conditions. The paper focuses on leveraging advanced techniques, including deep learning models and feature extraction methods, to enhance the performance of liveness detection. By incorporating dynamic facial cues, such as natural facial movements, and using multimodal approaches, the proposed models aim to improve the reliability of facial recognition systems. The study demonstrates the effectiveness of these models in real-world scenarios, achieving high detection accuracy and resilience to different spoofing techniques. The paper concludes that the proposed liveness detection models offer significant improvements in securing facial recognition systems, making them more resistant to sophisticated attack methods.

According to the papers previously discussed, the most recent developments in liveness detection encompass a variety of biometric modalities, such as iris, fingerprint, and facial recognition. Methodologies used in previous researched including computational ghost imaging, convolutional neural networks, multispectral imaging, and local contrast phase descriptors. The objectives of these methods are to improve accuracy, prevent sophisticated spoofing assaults, and enhance security, thereby demonstrating substantial advancements in the field. Meanwhile, this research aims to investigate alternative deep learning techniques, specifically with regard to liveness detection for authentication.

In contrast to machine learning, deep learning has its own mechanism [17-19]. Deep learning could perform classification without requiring feature extraction; hence, the approach illustrated in Figure 1 is utilized. The specifics of each step are explained in the subsequent subsections.

2.1 Dataset and preprocessing

This research used the video dataset available on Kaggle [20], PARNEC [21] and synthetic. The dataset was processed by changing the video into an image with extension (png).

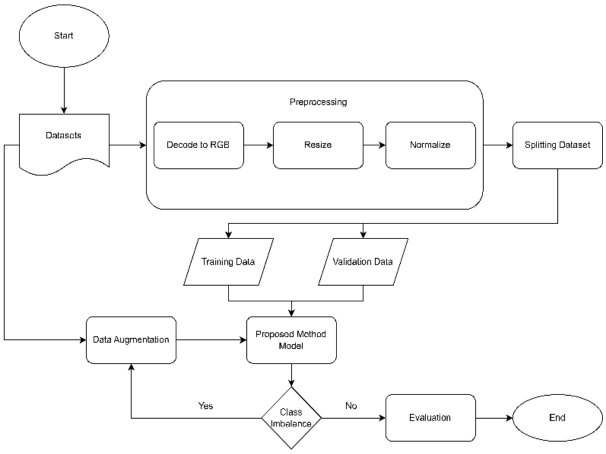


Figure 1. The flowchart of the process in detecting the fake faces using deep learning

Pre-processing was done by involving decode to RGB channel the change of image, resize with the size of 224×224 pixel and normalize. Figures 2-4 show samples of every dataset, Table 1 presents the amount of data for each dataset, meanwhile Table 2 defines the information for each dataset.

Figure 2. Sample of images in dataset [20]

Figure 3. Sample of images in dataset [21]

Figure 4. Sample of images in synthetic dataset

Table 1. Datasets faces counts

Datasets	Faces Counts
iBeta 1	3549
NUAA	9000
Synthetic	15018

2.2 Data splitting

The dataset was divided into two segments: 80% for training and 20% for testing, applicable to both augmented and non-augmented data.

2.3 Classification

In this study, a comparison among EfficientNetV2S, EfficientNetV2M, EfficientNetV2L, and MobileViT in the context of deep learning for image analysis was conducted for liveness detection.

2.3.1 EfficientNetV2

EfficientNetV2 [22] presented a new family of convolutional neural networks designed for improved training efficiency and parameter utilization. The models are developed through training-aware Neural Architecture Search (NAS), where optimizations are applied to accuracy, speed, and parameter efficiency. Techniques like progressive learning are introduced, allowing image size and regularization to be adaptively scaled during training, enabling faster convergence without sacrificing accuracy. EfficientNetV2 models are structured with Fused-MBConv layers to replace inefficient operations in early network stages, and a refined scaling strategy is employed to balance the network's complexity across layers.

The models are classified as EfficientNetV2-S, M, L, and XL, with different capacities and processing demands. EfficientNetV2-S is designed for smaller jobs, whereas EfficientNetV2-XL is intended for larger datasets. These models are intended to deliver considerable increases in training speed (up to 11 times faster) and parameter efficiency (up to 6.8 times smaller), as demonstrated on datasets such as ImageNet, CIFAR, and Flowers. Pretraining on larger datasets, such as ImageNet21k, is used to improve performance and achieve competitive accuracy with fewer computational resources. Illustration of EfficientNetV2 depicted in Figure 5.

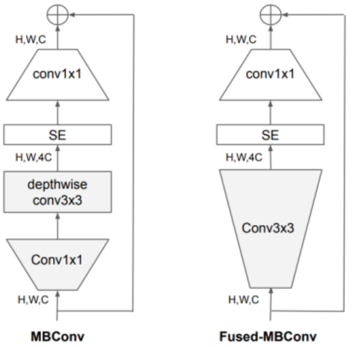


Figure 5. Illustration of the architecture of EfficientNetV2

Table 2. Datasets information

Datasets	Characteristics	Acquisition	Preprocessing
NUAA	The NUAA Imposter dataset is a publicly accessible dataset for face anti-spoofing research. It comprises authentic and synthetic facial photographs utilized for the assessment of biometric security systems. The photos were obtained via a camera in a regulated indoor setting. The collection comprises authentic facial images and counterfeit ones created by printing photographs of real faces and displaying them to the camera.	The dataset was compiled by the Pattern Recognition and Intelligent System Laboratory at Nanjing University of Aeronautics and Astronautics (NUAA). The fabricated images were generated from printed photographs of actual persons, rendering them valuable for studies in facial spoofing detection.	Adjusting photos to a uniform resolution for model training. Normalized pixel values for better model generalization. Dividing the dataset into training, and validation subsets. Data enrichment techniques, including rotation, flipping, and brightness modifications, enhance model generalization.
iBeta1	The iBeta dataset is a commonly utilized dataset for face anti-spoofing research. It comprises authentic and fabricated facial photographs, gathered under regulated conditions. The counterfeit faces were generated utilizing printed photographs and digital display assaults (e.g., projecting a facial picture on a screen). The dataset is utilized for assessing the resilience of face recognition and anti-spoofing methods.	Original facial photographs captured using iPhone and Android cameras under diverse settings. Counterfeit samples were produced utilizing printed images, digital displays, and several prevalent spoofing techniques.	Standardized picture resolutions. Implemented color normalization to provide uniformity across varying lighting conditions. Data augmented by changes such as brightness modification and random cropping. Divide into training, and validation subsets to assess model performance equitably.
Synthetic	The dataset comprises authentic and fabricated facial photos. Counterfeit images are produced by positioning printed photographs or computer screen displays in front of a camera. The collection comprises several locations, backdrops, and lighting situations. Certain counterfeit samples may display aberrations including glare, pixelation, and perspective distortions resulting from printed or screen-based presentation attacks.	Authentic photographs were obtained directly from living subjects in various environments (e.g., workplace, residence). Counterfeit photographs were produced via two primary techniques: Printed Attack: A printed image of an individual's face is positioned in front of a camera. Replay Attack: A digital image or video of an individual's face is exhibited on a mobile device or other digital display and shown to the camera. The dataset was acquired using standard cameras under diverse lighting and environmental conditions to replicate real-world situations.	All photos were scaled to a standardized resolution for model compatibility. Faces were identified and cropped utilizing a face identification technique to emphasize pertinent features. Implemented transformations such as flipping, brightness modifications, and rotation to improve model generalization. Images were partitioned into training, and validation sets to guarantee a balanced and impartial assessment.

2.4.2 MobileViT

MobileViT [23], a lightweight vision transformer designed for mobile devices, combines CNNs and transformers. MobileViT blocks combine global transformer processing with local convolutional representation learning, allowing for efficient visual interpretation while preserving spatial and patch-level order. The network uses fewer parameters and easier training procedures, making it ideal for mobile applications such as object detection and semantic segmentation. Three model variants are presented: MobileViT-XXS, XS, and S, with varying size and complexity. These models outperform classic lightweight CNNs and ViT-based approaches on benchmarks such as ImageNet-1K while being low latency and efficient on mobile devices. Illustration of the MobileViT architecture is presented in Figure 6.

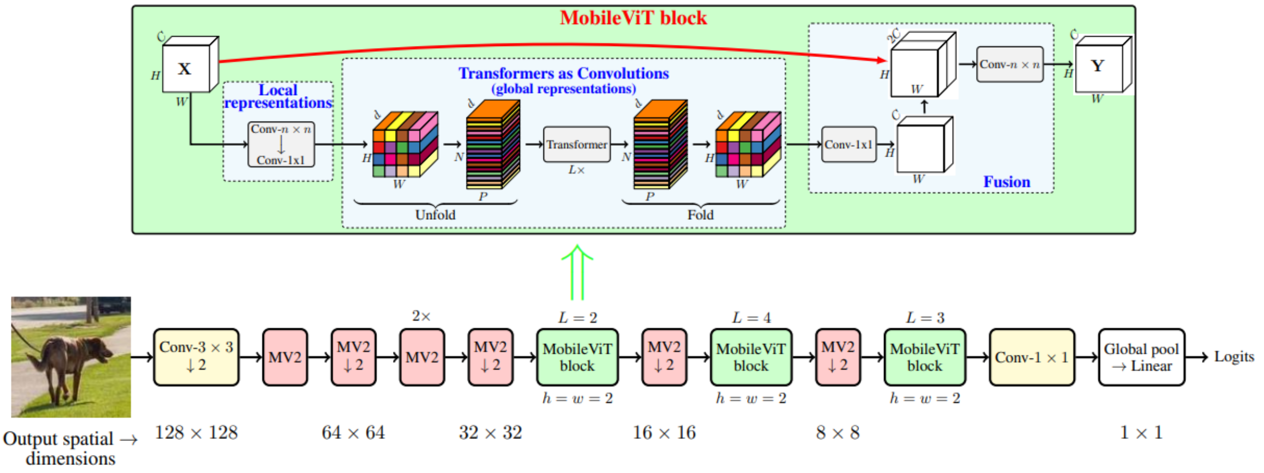


Figure 6. Illustration of the architecture of MobileViT

2.5 System performance measurement

The performance of the deep learning model was evaluated using the confusion matrix presented in Table 3. The confusion matrix comprised several metrics, including accuracy, precision, recall, F1-score, False Negative Rate (FNR), False Positive Rate (FPR), and Half Total Error Rate (HTER), as described in Eqs. (1)-(7). Four components were utilized to create the matrices: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). The matrices might benefit in predicting both the real and the fake images [24, 25].

Table 3. Confusion matrix

		Actual	
		(+)	(-)
Predicted	(+)	TP (True Positive)	FP (False Positive)
	(-)	FN (False Negative)	TN (True Negative)

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{1}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{2}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{3}$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

$$FNR = \frac{FN}{FN+TP} \tag{5}$$

$$FPR = \frac{FP}{FP+TN} \tag{6}$$

$$HTER = \frac{FNR+FPR}{2} \tag{7}$$

The model's accuracy and robustness in this paper are determined by analyzing the metrics of False Positive Rate (FPR), False Negative Rate (FNR), and Half Total Error Rate (HTER).

2.6 Data augmentation and testing scenario

Both the quantity and quality of training datasets are increased concurrently via data augmentation. This facilitates the improvement of deep learning models [26]. To enhance the diversity of data that utilize in model training, the data augmentation procedure employs ImageDataGenerator to apply a sequence of transformations to the dataset, which includes both real and fake images. This study generally entails data augmentation through several rules, including:

a. Flip: Horizontal, Vertical
b. Crop: 0% Minimum Zoom, 20% Maximum Zoom
c. Rotation: Between -15° and +15°
d. Hue: Between -15° and +15°
e. Saturation: Between -25% and +25%
f. Brightness: Between -15% and +15%
g. Exposure: Between -10% and +10%
h. Blur: Up to 2.5px
i. Noise: Up to 0.1% of pixels

3. RESULT AND DISCUSSION

Hyperparameter setting and layer configurations used for EfficientNetV2, and MobileViT depicted in Table 4 and Table 5.

As demonstrated in Tables 6-8, the classification reports for the three datasets NUAA, Synthetic, and iBeta 1 reveal that outstanding performance was attained across all models assessed. Models such as EfficientNetV2S, EfficientNetV2M, EfficientNetV2L, and MobileViT were examined, and precision, recall, F1-score, and accuracy metrics were consistently 1.00 on both the NUAA and Synthetic datasets. This means that the models were able to distinguish between "Fake" and "Real" classes without error in these datasets.

Table 4. Deep learning hyperparameters

Model	Parameter	Value
EfficientNetV2S	optimizer	Adam
	batch_size	32
	epoch	20
	learning_rate	0.001
	loss	sparse_categorical_crossentropy
	metrics	accuracy
EfficientNetV2M	optimizer	Adam
	batch_size	32

	epoch	20
	learning_rate	0.001
	loss	sparse_categorical_crossentropy
	metrics	accuracy
EfficientNetV2L	optimizer	Adam
	batch_size	32
	epoch	20
	learning_rate	0.001
	loss	sparse_categorical_crossentropy
	metrics	accuracy
MobileViT	optimizer	Adam
	batch_size	32
	epoch	20
	learning_rate	0.001
	loss	sparse_categorical_crossentropy
	metrics	accuracy

Table 5. Layer configurations

Layer	Parameter	Value
GaussianNoise	-	0.2
Conv2D	filters	16
	kernel_size	(3,3)
	padding	same
	activation	relu
	kernel_regularizer	0.1
Dropout	-	0.2
Conv2D	filters	32
	kernel_size	(3,3)
	padding	same
	activation	relu
	kernel_regularizer	0.1
Dropout	-	0.2
Dense	filters	64
	activation	relu
	kernel_regularizer	0.1
Dropout	-	0.2
Conv2D	filters	32
	activation	relu
	kernel_regularizer	0.1
Dropout	-	0.2
Dense	filters	2
	activation	softmax
	kernel_regularizer	0.1

Table 6. Classification report on NUAA dataset

Model	Class	Precision	Recall	F1-Score	Accuracy
EfficientNetV2S	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	
EfficientNetV2M	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	
EfficientNetV2L	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	
MobileViT	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	

Table 7. Classification report on synthetic dataset

Model	Class	Precision	Recall	F1-Score	Accuracy
EfficientNetV2S	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	
EfficientNetV2M	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	
EfficientNetV2L	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	
MobileViT	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	

Table 8. Classification report on iBeta 1 dataset

Model	Class	Precision	Recall	F1-Score	Accuracy
EfficientNetV2S	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	
EfficientNetV2M	Fake	0.99	1.00	0.99	1.00
	Real	1.00	0.99	0.99	
EfficientNetV2L	Fake	1.00	1.00	1.00	1.00
	Real	1.00	1.00	1.00	
MobileViT	Fake	0.99	1.00	1.00	1.00
	Real	1.00	0.99	1.00	

However, in the iBeta 1 dataset, there were minor differences in performance between EfficientNetV2M and MobileVit. While both models correctly classified the "Fake" class (accuracy and recall of 1.00), the recall for the "Real" class decreased to 0.99, resulting in an F1-score of 0.99 for this class. These fluctuations indicate that there were few misclassifications in the iBeta 1 dataset for the "Real" class, although overall accuracy for all models stayed at 1.00. The results show a modest change in model resilience based on dataset features.

Table 9. Error metrics on NUAA dataset

Model	FPR	FNR	HTER
EfficientNetV2S	0.0006	0	0.0003
EfficientNetV2M	0.0027	0.0006	0.0017
EfficientNetV2L	0.0013	0.0006	0.0010
MobileViT	0.0027	0	0.0013

Table 10. Error metrics on synthetic dataset

Model	FPR	FNR	HTER
EfficientNetV2S	0	0	0
EfficientNetV2M	0	0	0
EfficientNetV2L	0.0022	0	0.0011
MobileViT	0	0	0

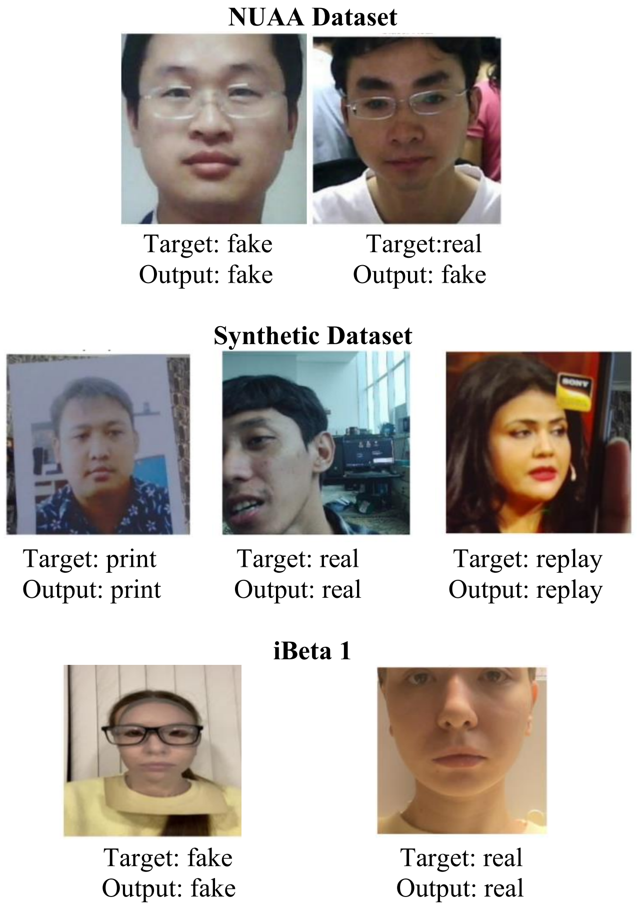
Table 11. Error metrics on iBeta 1 dataset

Model	FPR	FNR	HTER
EfficientNetV2S	0	0	0
EfficientNetV2M	0.0029	0.0081	0.0055
EfficientNetV2L	0	0	0
MobileViT	0.0029	0.0054	0.0042

Table 12. Training duration for every dataset

Dataset	Model	Training Duration (Seconds)
NUAA	EfficientNetV2S	1537.53 seconds
	EfficientNetV2M	2012.85 seconds
	EfficientNetV2L	6814.88 seconds
	MobileViT	158.48 seconds
Synthetic Dataset	EfficientNetV2S	771.93 seconds
	EfficientNetV2M	1664.67 seconds
	EfficientNetV2L	2899.20 seconds
	MobileViT	67.27 seconds
iBeta1	EfficientNetV2S	506.27 seconds
	EfficientNetV2M	749.57 seconds
	EfficientNetV2L	2456.14 seconds
	MobileViT	46.95 seconds

Examples of detection results from each dataset depicted in Figure 7.



NUAA Dataset

Target: fake
Output: fake

Target:real
Output: fake

Synthetic Dataset

Target: print
Output: print

Target: real
Output: real

Target: replay
Output: replay

iBeta 1

Target: fake
Output: fake

Target: real
Output: real

Figure 7. Output from every datasets

Tables 9-12 depict the performance comparison of three EfficientNetV2 models (versions S, M, and L) towards MobileViT across three datasets: NUAA, Synthetic, and iBeta 1.

EfficientNetV2S consistently outperforms other models over all datasets. It achieves outstanding outcomes (zero FPR, FNR, and HTER) on the Synthetic and iBeta datasets, while having a minimal error rate on the NUAA dataset (HTER = 0.0003). The "S" version demonstrates exceptional proficiency in differentiating between authentic and counterfeit samples in both synthetic and real-world contexts.

EfficientNetV2S in this experiment has shown superiority as the optimal approach due to its lightweight structure, fast inference, and robust feature extraction capabilities. It is particularly beneficial for liveness detection, considered as a real-time application where efficiency and accuracy are essential.

In contrast, EfficientNetV2M showed the highest error rate, particularly on the iBeta dataset, where its HTER is 0.0055, related to its high FNR of 0.0081. This indicates that EfficientNetV2M struggles with generalization relative to other models.

MobileViT performs well, exhibiting low error rates on the iBeta and NUAA datasets and no mistakes on the Synthetic dataset. But compared to EfficientNetV2S and EfficientNetV2L, MobileViT is less reliable. The EfficientNetV2L variation demonstrates commendable performance, attaining an optimal score on the iBeta dataset and exhibiting a reduced HTER on NUAA compared to EfficientNetV2M; nevertheless, its efficacy on the Synthetic data is considerably hindered by its minimal FPR.

Regarding datasets, the Synthetic dataset has little challenges for all models, as most of them attain 100% accuracy. Nonetheless, the iBeta and NUAA datasets demonstrate greater complexity, particularly for EfficientNetV2M and MobileViT, signifying heightened challenges in managing real-world data. The results are verified by the confusion matrix values obtained from all algorithms throughout the experiment. The confusion matrix values can be seen in the Appendix section. Also, the results underscore the substantial robustness of EfficientNetV2S and EfficientNetV2L, suggesting that forthcoming advancements in EfficientNetV2M and MobileViT could bridge the performance disparity.

4. CONCLUSIONS

This study investigates liveness detection through deep learning techniques. The employed deep learning algorithm is CNN, utilizing EfficientNetV2S, EfficientNetV2M, EfficientNetV2L, and MobileViT models, applied to the NUAA, Synthetic, and iBeta 1 datasets. The experimental results, as indicated by the parameters of the confusion matrix, error metric, and training duration, indicate that EfficientNetV2S is superior to other approaches among the three datasets. This is not surprising, considering that the advantages of the EfficientNetV2S algorithm are small model size, fast Inference Speed, and ease of Deployment Feasibility.

New datasets with various attack classes could be added for future research enhancements, in contrast to those used in this study. This aims to boost the detection of precision of new attacks that remain undetectable by this research.

ACKNOWLEDGMENT

The research was supported by the Ministry of Education, Culture, Research and Technology of Indonesia (Grant No.: 043/SP2H/ RT-MONO/LL4/2024).

REFERENCES

[1] Ametefe, D.S., Sarnin, S.S., Ali, D.M., Zaheer, M.Z. (2022). Fingerprint liveness detection schemes: A review on presentation attack. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 10(2): 217-240. https://doi.org/10.1080/21681163.2021.2012826

[2] Sabaghi, A., Oghbaie, M., Hashemifard, K., Akbari, M. (2021). Deep learning meets liveness detection: Recent advancements and challenges. arXiv preprint arXiv:2112.14796. http://arxiv.org/abs/2112.14796

[3] Tapia, J.E., Gonzalez, S., Busch, C. (2021). Iris liveness detection using a cascade of dedicated deep learning networks. IEEE Transactions on Information Forensics and Security, 17: 42-52. https://doi.org/10.1109/TIFS.2021.3132582

[4] Koshy, R., Mahmood, A. (2020). Enhanced deep learning architectures for face liveness detection for static and video sequences. Entropy, 22(10): 1186. https://doi.org/10.3390/e22101186

[5] Kuznetsov, O., Zakharov, D., Frontoni, E., Maranesi, A. (2024). AttackNet: Enhancing biometric security via

tailored convolutional neural network architectures for liveness detection. Computers & Security, 141: 103828. https://doi.org/10.1016/j.cose.2024.103828

[6] Jie, O.Z., Ming, L.T., Wee, T.C. (2023). Biometric authentication based on liveness detection using face landmarks and deep learning model. JOIV: International Journal on Informatics Visualization, 7(3-2): 1057-1065. https://doi.org/10.30630/joiv.7.3-2.2330

[7] Priyadarsini, M.J.P., Ramya, K., Parlakota, S., Tadi, N.K.R., Jabeena, A., Rajini, G. (2023). Face anti-spoofing and liveness detection using deep learning architectures. Journal of Engineering Science and Technology, 18: 217-227.

[8] Zhou, M., Wang, Q., Li, Q., Zhou, W., Yang, J., Shen, C. (2024). Securing face liveness detection on mobile devices using unforgeable lip motion patterns. IEEE Transactions on Mobile Computing, 23(10): 9772-9788. https://doi.org/10.1109/TMC.2024.3367781

[9] Taeb, M., Chi, H. (2022). Comparison of deepfake detection techniques through deep learning. Journal of Cybersecurity and Privacy, 2(1): 89-106. https://doi.org/10.3390/jcp2010007

[10] Elsaeidy, A.A., Jagannath, N., Sanchis, A.G., Jamalipour, A., Munasinghe, K.S. (2020). Replay attack detection in smart cities using deep learning. IEEE Access, 8: 137825-137837. https://doi.org/10.1109/ACCESS.2020.3012411

[11] Zhang, X., Zheng, X., Mao, W. (2021). Adversarial perturbation defense on deep neural networks. ACM Computing Surveys (CSUR), 54(8): 1-36. https://doi.org/10.1145/3465397

[12] Kelly, U.M., Veldhuis, R.N., Spreeuwers, L. (2020). Improving deep-learning-based face recognition to increase robustness against morphing attacks. In 9th International Conference on Signal, Image Processing and Pattern Recognition, SPPR 2020. Academy and Industry Research Collaboration Center (AIRCC), pp. 1-12. https://doi.org/10.5121/csit.2020.101901

[13] Meena, K.B., Tyagi, V. (2021). A deep learning based method for image splicing detection. Journal of Physics: Conference Series, 1714(1): 012038. https://doi.org/10.1088/1742-6596/1714/1/012038

[14] Arora, S., Bhatia, M.P.S., Mittal, V. (2022). A robust framework for spoofing detection in faces using deep learning. The Visual Computer, 38(7): 2461-2472. https://doi.org/10.1007/s00371-021-02123-4

[15] Guan, Q., Deng, H., Liang, W., Ni, M., Gao, X., Ma, M., Zhong, X., Gong, X. (2023). Resolution-independent liveness detection via computational ghost imaging. Applied Physics Letters, 123(2). https://doi.org/10.1063/5.0155365

[16] Anjum, H., Arshad, U., Ali, R.H., Abideen, Z.U., Shah, M.H., Khan, T.A., Ijaz, A.Z., Siddique, A.B., Imad, M. (2023). Robust and reliable liveness detection models for facial recognition systems. In 2023 International Conference on Frontiers of Information Technology (FIT), pp. 292-297. https://doi.org/10.1109/FIT60620.2023.00060

[17] Jondri, J., Rizal, A. (2020). Classification of premature ventricular contraction (PVC) based on ECG signal using convolutional neural network. Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 8(3): 494-499. https://doi.org/10.11591/ijeei.v8i3.1530

[18] Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies, 32(1): e4150. https://doi.org/10.1002/ett.4150

[19] Helm, J.M., Swiergosz, A.M., Haeberle, H.S., Karnuta, J.M., Schaffer, J.L., Krebs, V.E., Spitzer, A.I., Ramkumar, P.N. (2020). Machine learning and artificial intelligence: Definitions, applications, and future directions. Current Reviews in Musculoskeletal Medicine, 13: 69-76. https://doi.org/10.1007/s12178-020-09600-8

[20] Kaggle. (2023). iBeta Level 1 Liveness Detection Dataset-Part 1. https://www.kaggle.com/datasets/trainingdatapro/ibeta-level-1-liveness-detection-dataset-part-1, accessed on Nov. 17, 2024.

[21] Tan, X., Li, Y., Liu, J., Jiang, L. (2010). Face liveness detection from a single image with sparse low rank bilinear discriminative model. In Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI 11, pp. 504-517. https://doi.org/10.1007/978-3-642-15567-3_37

[22] Tan, M., Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In International Conference on Machine Learning, pp. 10096-10106. https://doi.org/10.48550/arXiv.2104.00298

[23] Mehta, S., Rastegari, M. (2021). Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. arXiv preprint arXiv:2110.02178. https://doi.org/10.48550/arXiv.2110.02178

[24] Hasnain, M., Pasha, M.F., Ghani, I., Imran, M., Alzahrani, M.Y., Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. IEEE Access, 8: 90847-90861. https://doi.org/10.1109/ACCESS.2020.2994222

[25] Luque, A., Carrasco, A., Martín, A., de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91: 216-231. https://doi.org/10.1016/j.patcog.2019.02.023

[26] Shorten, C., Khoshgoftaar, T.M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1): 1-48. https://doi.org/10.1186/s40537-019-0197-0