JOIV : Int. J. Inform. Visualization, 9(1) - January 2025 295-302

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

# Face recognition for Logging in Using Deep Learning for Liveness Detection on Healthcare Kiosks

Catoer Ryando [a], Riyanto Sigit [b,*], Setiawardhana [b], Bima Sena Bayu Dewantara [b]

[a] *Department of Electrical Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya, East Java, Indonesia*
[b] *Department Informatic and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya, East Java, Indonesia*
*Corresponding author: *riyanto@pens.ac.id*

***Abstract*—This study explores the enhancement of healthcare kiosks by integrating facial recognition and liveness detection technologies to address the limitations of healthcare service accessibility for a growing population. Healthcare kiosks increase efficiency, lessen the strain on conventional institutions, and promote accessibility. However, there are issues with conventional authentication methods like passwords and RFID, such as the possibility of them being lost, stolen, or hacked, which raises privacy and data security problems. Although it is more secure, face recognition is susceptible to spoofing attacks. In order to improve security, this study integrates liveness detection with face recognition. Data preparation is done using deep learning algorithms, namely FaceNet and Multi-task Cascaded Convolutional Neural Networks (MTCNN). Real-time authentication of persons is verified by the system, which provides correct identification of them. Techniques for enhancing data help the model become more accurate and robust. The system's usefulness is shown by the outcomes of the experiments. The VGG16 model outperforms alternative designs like MobileNet V2, ResNet-50, and DenseNet-121, achieving 100% accuracy in liveness detection. Face recognition and liveness detection together greatly improve security, which makes it a dependable option for real-world healthcare applications. Through the ability to differentiate between genuine and fake faces and foil spoofing efforts, facial liveness detection may boost security. This study offers insights into building biometric systems for safe and effective identity verification in the healthcare industry.**

***Keywords*— Face recognition; liveness detection; health kiosk; deep learning; computer vision.**

*Manuscript received 27 May 2024; revised 16 Jul. 2024; accepted 6 Aug. 2024. Date of publication 31 Jan. 2025.*


## I. INTRODUCTION

The constantly expanding population presents a challenge to the healthcare system due to the lack of healthcare facilities [1]. Health kiosks provide a state-of-the-art method to enhance service accessibility, lessen the load on traditional facilities, and boost productivity. These gadgets or machines are strategically positioned in places like shopping centers, workplaces, or other public venues [2]. With the use of cutting-edge technology, these kiosks enable users to monitor their own health by taking their weight, body mass index (BMI), blood pressure, and blood sugar. They may also request medical advice from medical professionals. It is easy for people to get information and evaluate their health indicators without going to a clinic or hospital [3]. The existence of these kiosks not only narrows the gap between the increasing population and the limited number of health institutions, particularly for those who reside in remote areas or have limited access to conventional health facilities [4], but it also makes health services more accessible to the community. Healthcare kiosks provide a lot of benefits but worries about user privacy and data security are also becoming more prevalent. There is a chance that personal information may leak when medical personnel or kiosk users check in to use a medical gadget or see a patient's medical history. Every authentication method used in health kiosks, including biometrics, RFID, and passwords, has disadvantages of its own. RFID is susceptible to loss or theft, and users are exposed to weak or forgotten passwords [5], [6].

With the use of face recognition, computers can recognize and confirm an individual's identification by analyzing their face characteristics [7], [8]. Many face recognition systems are currently used to reliably check faces, and liveness detection is essential for health kiosk security [9], [10]. By determining if live objects are present, liveness detection adds a crucial degree of security to face authentication. Benefits include simple access to practical healthcare services, such as reading medical history and logging in at healthcare kiosks.

This may lower the time and expense needed, improve personal data security, and lower privacy threats.

One of the most fascinating subjects in computer vision research and biometric systems is face recognition technology, which is still developing quickly [11], as evidenced by the following references: [12], [13], [14], [15]. Because deep learning can automatically extract key characteristics from facial data, it offers a major advantage in increasing face recognition accuracy in this situation.

Face recognition is thought to be more natural than other smart technologies such as voice recognition, fingerprints, and retina scans [11]. Therefore, access to information or other personal data, such as health records, is restricted to the recognized individual only when using face recognition. However, face recognition is susceptible to hacking, particularly when shielding biometric authentication systems from spoofing attempts that use printed pictures, video replays, and other similar techniques [16]. As a result, stringent controls are required to guard against data leaks in the system. Liveness detection offers an essential degree of security, a technology that emphasizes face authenticity by verifying the existence of living things. The other goals are the significant and difficult problems that establish the reliability of biometric systems. Biometric system protection against spoofing, enhanced security, and improved identification precision [17].

This study uses deep learning to analyze intricate patterns and attributes from face photos. FaceNet extracts and fully understands face characteristics via the use of deep learning architecture. The model is able to identify the unique features of a face and differentiate it from other people's faces even in different lighting situations and from different viewing angles. Preprocessing the dataset or image is necessary in deep learning in order to prepare the data for the model to process. One of the most popular pre-processing methods is (multi-task cascaded convolutional neural network) MTCNN [18].

Furthermore, a number of deep learning architectures, including VGG, MobileNet, ResNet, and DenseNet, are used in liveness detection, which verifies that the face picture belongs to a genuine live person. Tests were conducted on the different architectures, to determine which model had the greatest accuracy in sincerity identification. Because deep learning systems can automatically extract characteristics from input data, they are substantially more accurate than previous methods. Utilizing extensive and varied datasets enhances the capacity of deep learning models to comprehend differences in face photos, leading to increased accuracy in identity verification and identification.

## II. Material and Method

This study starts with face recognition using FaceNet. FaceNet converts the facial picture into an individual vector representation, or embedding, which gives it an edge in accurate face recognition. After that, this representation is compared to a database of faces that have already undergone training, allowing for reasonably accurate identification and verification. Liveness detection comes next in the process after face recognition. Several different architectures, including VGG16, ResNet, MobileNet, and DenseNet, are used to train this liveness detection. The outcome of liveness detection is the determination of whether an individual is a spoof or real. Therefore, liveness detection and face recognition can complement each other to increase system security. By detecting liveness, a face recognition system can be more certain that the face it is identifying is a real one and not an image or fake.

### *A. Data Collection*

To create a robust anti-fake face system, the dataset should be of the highest possible quality and realism. This is due to the fact that extensive and diverse datasets are required for effective training and evaluation of deep learning models. Researchers [19], [20], and [21] have created their own datasets from various sources, including faces from printed images, faces derived from cards, faces on screens, and faces on photos and real photo faces. This is an important trend in the development of these datasets. It shows the importance of size, variety, and realism when dealing with fake facial characteristics in datasets intended to prevent counterfeiting. Considering the increasing prevalence of face locking technology on mobile devices, its significance becomes even more important. A genuine dataset covering mobile device attack scenarios is useful for building more robust and secure anti-counterfeiting systems.

Face detection, which uses algorithms to discover and identify faces in an image or video, is an essential part of face recognition systems. To do this, a number of methods like dlib, Haar, HOG, MediaPipe, and MTCNN are used. The experimental findings comparing the efficacy of several face identification systems are shown in Figure 4. This graphic illustrates how certain methods struggle to identify faces in different positions.


Fig. 1 Data collection for face recognition and liveness detection

Fig.1 displays multiple instances of data collection related to face recognition and liveness detection. The data collection process employs a mobile phone camera with the specifications listed in Table 1 for face recognition and liveness detection (labeled real in Fig. 4) and card face photos, screens, photos, and printed photos for the data collection labeled fake or spoof.

TABLE I
CAMERA SPECIFICATIONS

| Parameter | Specifications |
|---|---|
| Type | Samsung A71 |
| Resolution | 1080p/960fps |
| Focus | Fix Focus |
| Camera | 64 P |

#### *1) Preprocessing:*

To enhance image quality and streamline processing, pre-processing is a crucial step in building a solid dataset for

liveness detection and face recognition [22]. The pre-processing technique in this work is Multi-task Cascaded Convolutional Neural Network (MTCNN). Multiple face recognition applications can benefit greatly from MTCNN's resilience to noise. The rationale for the choice of MTCNN will be expounded upon in the face detection subchapter of section three.

In the preprocessing stage of the face recognition and liveness detection dataset, MTCNN is used for face detection in the source image. As can be seen in Fig. 2, the detected face region is then cropped from the source image and converted to a consistent size, such as 160×160 pixels, to be prepared for the next stage, which is augmentation on the face recognition dataset and augmentation on the liveness detection dataset.

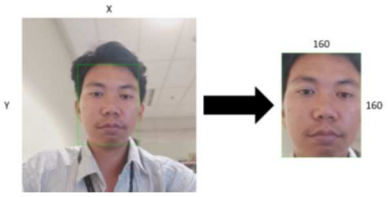


Fig. 2 Cropped image from the source image

There are two primary advantages to face recognition systems that use MTCNN for preprocessing. First off, by eliminating extraneous elements such as the backdrop, the quality of the face picture is enhanced, concentrating attention on the face and simplifying further processing. Second, by using uniformly sized face photos, the efficiency of the face detection process is expedited and improved. A high-quality facial picture that is prepared for further processing is the end product of the whole MTCNN pre-processing.

*2) Augmentation:*

One of the most important methods for raising the accuracy of contemporary picture classifiers is data augmentation [23]. By modifying photos using a variety of methods, including flipping, color space, rotation, translation, noise injection, and color space, the quantity and diversity of data are expanded [24]. Better deep learning models are produced as a consequence of these modifications, which enhance the training dataset. Since data augmentation increases data adequacy and variety and prevents overfitting in artificial neural networks [25], it is an essential component of training deep learning [26] models using picture data. This process of augmentation may be critical to the effectiveness of applying deep learning models to picture data.

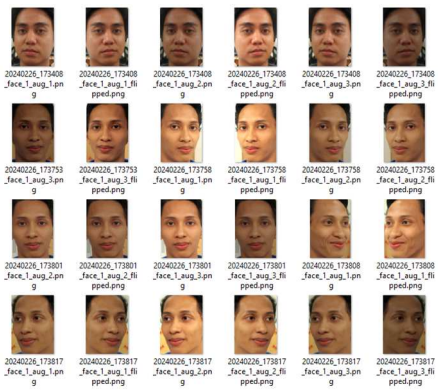

Fig. 3 Augmentation

The purpose of data augmentation is to increase the training dataset's variety and quality for face recognition and liveness detection. The dataset that has been enhanced by augmentation is shown in Figure 3, where three augmented photographs are created from one source photo. The photo is initially augmented by flipping it horizontally. Next, the brightness is adjusted (0.1–0.6). Finally, the Elastic Transform is used with randomly selected alpha and sigma values (0.5–1.5 for alpha and 0.05-0.15 for sigma). This enhanced dataset may be utilized for facial recognition and liveness detection training and validation.

After applying augmentation techniques to each of the eight original photographs for each label obtained from the face recognition data collecting procedure, a total of 32 images were obtained for each label. As shown in Figure 5, the face recognition dataset had 192 photos in total following the augmentation process. The same augmentation method was used for the liveness detection procedure, yielding three enhanced photographs for every source photo. This method yielded 364 images from the 91 original face shots, while the augmented fake face data, which started off at 768 images, ended up with 3072 images.

Preprocessing and augmentation are crucial when it comes to data collection. Preprocessing and augmentation contribute to consistent dataset sizes and increase the quantity and variety of data. This makes the trained model more adaptable to different environments and disruptions and improves its recognition and differentiation capabilities between real and fake faces.

*B. Deep Learning*

"Deep Learning" is a branch of artificial intelligence that employs multi-layered artificial neural networks for data processing and decision making. By using these methods, computer systems are able to learn from vast amounts of data and produce more intricate and abstract representations of the incoming data. Convolutional neural networks (CNN), a kind of neural network essential to deep learning, are used in this procedure to recognize different face traits by obtaining their visual representation [27]. The first step in this process is data collection, which starts with face detection using MTCNN, as explained in the section on preprocessing. After the face is detected, augmentation is performed on the face to enrich the dataset. The augmented dataset is then ready to be used as input to train the deep learning architecture to liveness detection. ResNet and VGG are two popular deep learning architectures. Using deep learning for liveness detection in face recognition makes the system more secure and offers a more dependable and efficient way to authenticate users.

Deep learning is used in this study to liveness detection. Every architecture in deep learning differs significantly from the others. VGG16 has three fully connected layers, thirteen convolution layers, three 3x3 convolution layers, and pooling. MobileNet V2, which is effective for devices with constrained resources, employs inverted residuals with bottleneck layers and depth wise separable convolutions. ResNet-50 allows for extremely deep network training without disappearing gradients by introducing residual blocks with shortcut connections in 50 layers. DenseNet-121 enhances information flow and parameter usage efficiency by connecting every layer to every prior layer across 121 layers. This implies that

multiple methodologies are used in the construction of each deep learning architecture to maximise processing power, efficiency, and performance according to the specific needs and limitations of the targeted application.

With Google's FaceNet face recognition in 2015, deep learning for face recognition performed well. Through the use of the Inception-ResNet network architecture, this system has completely changed the field of face recognition technology [28]. FaceNet's ability to map faces to Euclidean space makes it simple to recognize and categorise faces [29]. During training, the system optimises triplet loss and embedding using convolutional networks. VGG represents faces as 512 dimensional vectors [28] in order to facilitate a more precise and effective recognition procedure.

FaceNet's primary advantage is its capacity to handle low-quality pictures, which comes in handy for things like distant shots or security cameras. Furthermore, FaceNet is very versatile since it can identify faces even when they are obscured by accessories [30]. FaceNet is used in several industries, including security and entertainment. The technology is used in security for access control and identity verification [22], [31]. Enhances face recognition and liveness detection through deep learning. Thus, detection of liveness, which confirms that the detected face is a real face and not just an image or video, makes face recognition safer.

## III. Result and Discussion

### *A. Face Detection*

Face detection is a vital component in systems for face recognition and liveness detection, where algorithms are tasked with locating and identifying faces within an image or video. Techniques such as dlib, Haar, HOG, MediaPipe, FaceNet-MTCNN and MTCNN are employed to accomplish this objective. Figure 4 presents experimental comparisons of the efficacy of these face detection methods. This illustration reveals that certain techniques struggle to detect faces in diverse poses.

From Figure 4, it is evident that various face detection techniques successfully identify faces in different poses. In Figure 4(a) with the forward-facing pose, all face detection techniques accurately detect the face. In Figure 4(b), where the face is turned to the left at approximately 1.5 meters from the camera, the techniques that successfully detect the face are MTCNN and FaceNet MTCNN. Meanwhile, in Figure 4(c), with the face tilted downward, face detection techniques such as MediaPipe, MTCNN, and FaceNet MTCNN effectively recognize the face. Thus, in various poses, techniques like MTCNN and FaceNet MTCNN prove to be effective in detecting faces.

Two efficient face detection techniques are MTCNN (Multi-task Cascaded Convolutional Networks) and FaceNet MTCNN. MTCNN employs a cascaded convolutional neural network with three stages: Proposal Network (P-Net), Refine Network (R-Net), and Output Network (O-Net) to generate bounding box candidates, R-Net to filter the candidates, and O-Net to refine the bounding box and facial landmarks. MTCNN is robust in a variety of situations, including changing lighting and angle. In contrast, FaceNet MTCNN combines the FaceNet model for extracting facial features with the MTCNN algorithm for initial face detection. Because of its exceptional facial feature extraction capacity and excellent recognition accuracy, this model is very appropriate for face detection.

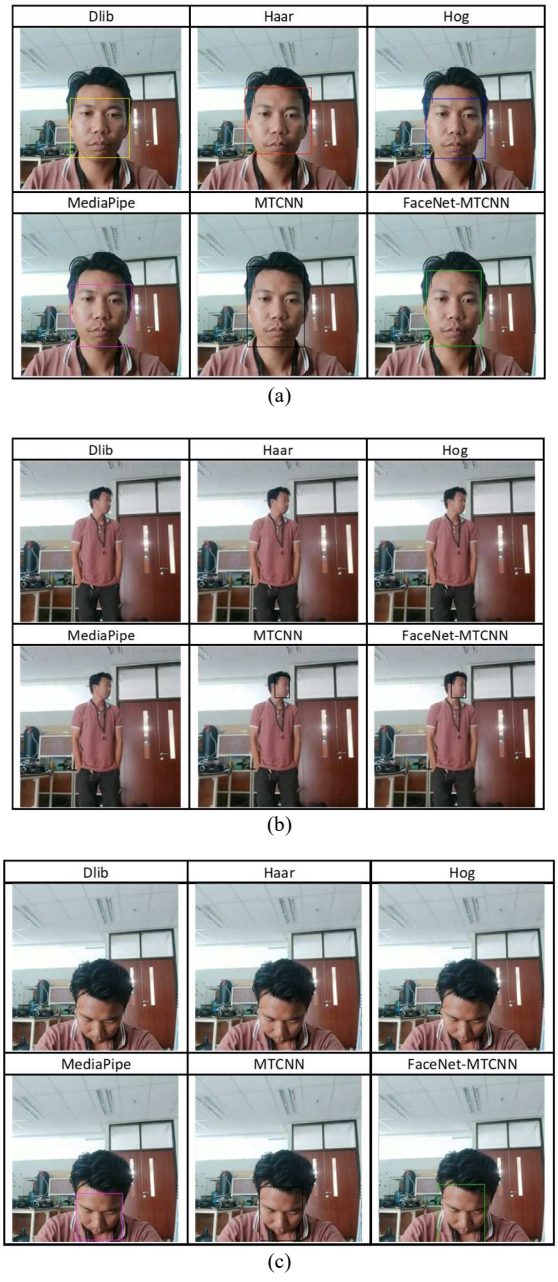


Fig. 4 Face detection (a) face forward, (b) face distance from the camera is about 1.5 meters, (c) face downwards

FaceNet maps facial photos into a feature space so that faces may be compared using deep convolutional networks. Additional methods like MediaPipe, HOG, and Haar Cascade are also beneficial. MediaPipe tracks and detects faces quickly and effectively using a multi-stage pipeline structure. Using edges and gradients in the picture, HOG (Histogram of Oriented Gradients) recognizes faces, while Haar Cascade employs a sequence of cascade classifiers trained using the AdaBoost technique. On the other hand, MTCNN consistently produced the best results while addressing different face detection problems. It performs better than the other solutions in terms of face detection consistency and quality.

## B. *Training Face Recognition*

As mentioned in the data collection subchapter in section II, the dataset used to train the face recognition model using the FaceNet architecture comprises annotated photos of six distinct people. Every person has eight unique photos. Three augmentation rounds were applied to each original picture in order to increase the dataset's size, yielding a total of 32 photos each label. Fig. 5 confusion matrix displays the findings.

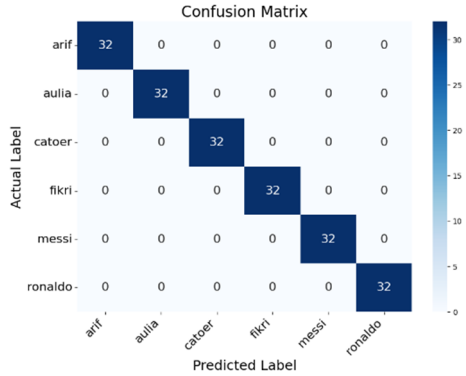


Fig. 5 Confusion matrix uses FaceNet for facial recognition

TABLE II
PERFORMANCE EVALUATION OF FACENET

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Arif | 1.00 | 1.00 | 1.00 | 32 |
| Aulia | 1.00 | 1.00 | 1.00 | 32 |
| Catoer | 1.00 | 1.00 | 1.00 | 32 |
| Fikri | 1.00 | 1.00 | 1.00 | 32 |
| Messi | 1.00 | 1.00 | 1.00 | 32 |
| Ronaldo | 1.00 | 1.00 | 1.00 | 32 |
| Accuracy | | | 1.00 | 192 |
| Macro Avg | 1.00 | 1.00 | 1.00 | 192 |
| Weighted Avg | 1.00 | 1.00 | 1.00 | 192 |

FaceNet as a technique in face recognition. The data in Fig. 5 and Table 2 demonstrate FaceNet ability to provide high-quality face embedding. In Fig. 5, the confusion matrix displays perfect predictions with no errors, showing that the FaceNet model not only works well on the dataset used but also has remarkable accuracy in understanding and classifying photos. This reliability is additionally supported by the accuracy, recall, f1-score, and support statistics shown in Table 2. By using these techniques, it is confirmed that the model successfully classified all six subjects, with a precision, recall, and f1-score of 1.00.

In conclusion, the FaceNet model shows very high effectiveness in the face recognition task, as evidenced by the F1 score of 1.00 or in other words, 100% accuracy. Due to the high level of reliability, this model is very useful for applications that require accurate face recognition.

## C. *Liveness Detection Training*

The dataset for liveness detection has undergone preprocessing and augmentation to improve the quality and functionality of the model, as described in section II under data collection. These procedures are essential to ensure that the model performs at its best. The performance of the model is significantly affected by the volume of data, the augmentation strategy used, and the preprocessing. The dataset consists of 3435 images used for real presence detection training. This dataset is divided into 70% for training (2402 images) and 30% for validation (1033 images). Of the 2402 training images, there are 2150 images labeled as "fake or spoof" and 252 images labeled as "real". For validation, there are 922 images labeled "fake" and 112 images labeled "real", which will be used as input for training the deep learning model.

TABLE III
LIVENESS DETECTION ACCURACY TRAINING RESULTS

| CNN Structure (Accuracy % ) | | | | |
|---|---|---|---|---|
| **Epoch Value** | **VGG16** | **MobileNet V2** | **ResNet 50** | **DenseNet 121** |
| 100 | 100 % | 95.00 % | 90.00 % | 90.00 % |

The accuracy of the several deep learning architectures in the classification problem is shown in Table 3. After 100 training epochs, the VGG16 model achieved 100% accuracy, making it the best-performing model. This suggests that VGG16 can effectively identify patterns in the training data to differentiate between the two classes. MobileNet V2, ResNet50, and DenseNet 121 exhibit lesser accuracy (75%, 66.67%, and 58.33%, respectively) compared to other deep learning architectures. This implies that VGG16 performs better than the other models in this particular classification occupation.

The VGG16 model performs very well in model training, as shown in Fig. 6 and Fig. 7. Fig. 6 illustrates the high precision, recall, and f1-score values in Table 4 that demonstrate the model's excellent accuracy. The confusion matrix, shown in Fig. 7, indicates that the model can accurately categorize pictures into true positive and true negative categories. In summary, the VGG16 model demonstrated efficacy in the task of picture categorization.

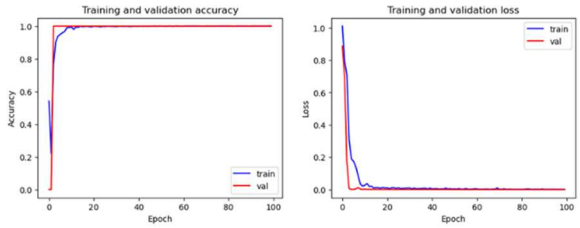


Fig. 6 Accuracy and loss with vgg16 architecture

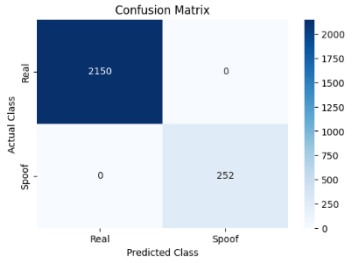


Fig. 7 Confusion matrix with vgg16 architecture

TABLE IV
PERFORMANCE EVALUATION OF VGG16 ARCHITECTURE MODEL

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Real | 1.00 | 1.00 | 1.00 | 2150 |
| Spoof | 1.00 | 1.00 | 1.00 | 252 |
| Accuracy | | | 1.00 | 2402 |
| Macro Avg | 1.00 | 1.00 | 1.00 | 2402 |
| Weighted Avg | 1.00 | 1.00 | 1.00 | 2402 |

Based on Table 4, the VGG16 model shows outstanding performance in training the model to identify "real" and "fake" images. This is proven by flawless accuracy, recall, and f1-score values (1.00) for both classes. The total accuracy of the model likewise hit 1.00, meaning that the model was able to categorize all photos (2402 pieces) properly. These findings demonstrate that the VGG16 model is quite good at differentiating between actual (real) pictures and false (spoof) ones.

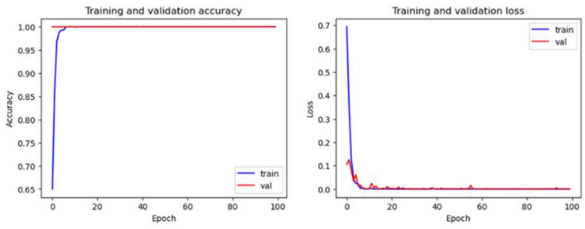


Fig. 8 Accuracy and loss with mobilenet v2 architecture

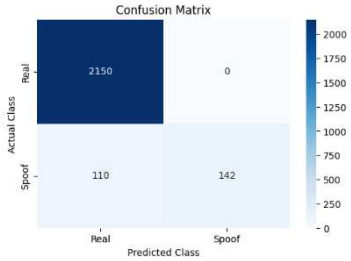


Fig. 9 Confusion matrix with mobilenet v2 architecture

TABLE V
PERFORMANCE EVALUATION OF MOBILENET V2 ARCHITECTURE MODEL

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Real | 0.95 | 1.00 | 0.98 | 2150 |
| Spoof | 1.00 | 0.56 | 0.72 | 252 |
| Accuracy | | | 0.95 | 2402 |
| Macro Avg | 0.98 | 0.78 | 0.85 | 2402 |
| Weighted Avg | 0.96 | 0.95 | 0.95 | 2402 |

The MobileNet V2 model demonstrates good performance in model training, as demonstrated in Fig. 8 and Table 5. The high, but not perfect, accuracy, recall and f1-score values illustrate the model's ability to accurately categorize pictures. The confusion matrix in Fig. 9 indicates an overall accuracy of 95% in differentiating between real positives and true negatives.

These findings demonstrate that the MobileNet V2 model is extremely successful and promising as a dependable tool for the categorization of "real" and "fake" photos, , although there is still room for further development.

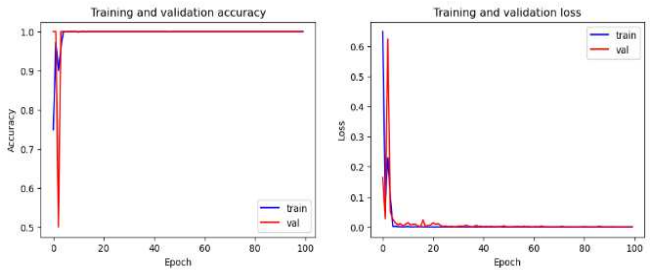


Fig. 10 Accuracy and loss with resnet-50 architecture

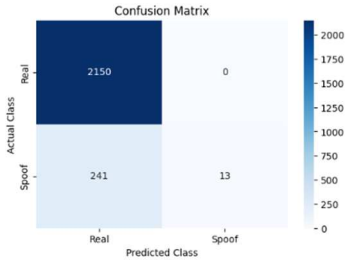


Fig. 11 Confusion matrix with resnet-50 architecture

TABLE VI
PERFORMANCE EVALUATION OF RESNET-50 ARCHITECTURE MODEL

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Real | 0.90 | 1.00 | 0.95 | 2150 |
| Spoof | 1.00 | 0.04 | 0.08 | 252 |
| Accuracy | | | 0.90 | 2402 |
| Macro Avg | 0.95 | 0.52 | 0.52 | 2402 |
| Weighted Avg | 0.91 | 0.90 | 0.86 | 2402 |

The training graph using ResNet-50 architecture is shown in Fig. 10. The graph indicates a steady decline in loss and a rise in accuracy. At the end, the model's accuracy was 90%. The confusion matrix shown in Fig. 11 indicates that the majority of the actual data were properly identified by the model. However, the model struggles to detect fake data, particularly when the recall is poor. The spoof class's accuracy was flawless, but the poor recall also resulted in a low F1-score value. Table 6 illustrates that the model's overall accuracy is 90%, with varying precision, recall, and F1-score values for the two classes.

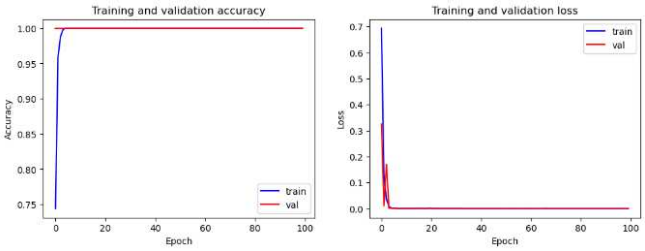


Fig. 12 Accuracy and loss with densenet-121 architecture

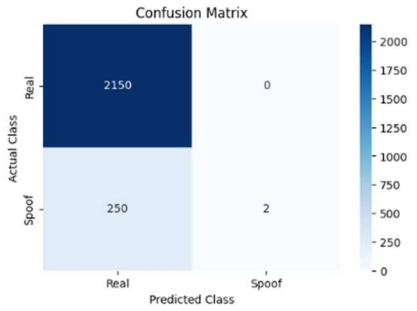


Fig. 13 Confusion matrix with densenet-121 architecture

TABLE VII
PERFORMANCE EVALUATION OF DENSENET-121 ARCHITECTURE MODEL

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Real | 0.95 | 1.00 | 0.95 | 2150 |
| Spoof | 1.00 | 0.01 | 0.02 | 252 |
| Accuracy | | | 0.90 | 2402 |
| Macro Avg | 0.95 | 0.50 | 0.48 | 2402 |
| Weighted Avg | 0.91 | 0.90 | 0.85 | 2402 |

Fig. 12 shows the performance of the DenseNet-121 model with 90% accuracy in training the model. Despite its high accuracy, the model has low performance in identifying fake data. The confusion matrix in Fig. 13 shows that the model successfully classified 2150 original data correctly (TP) and only misclassified 0 original data as fake data (FP). However, the model only correctly classified 250 fake data (TN) and 2 fake data that were misclassified as original data (FN). The low recall and F1-score values for the fake data class in Table 7 indicate the poor performance of the model in classifying this class. Improvements are needed in identifying fake data to improve the overall performance of the model.

### *D. Testing*

Very excellent accuracy was achieved in liveness detection using deep learning training with VGG16 and in face recognition using the model or training results with FaceNet as detailed in section III results and discussion subsection b training face recognition. Liveness Detection in Section c provides a description of these findings. As shown in Fig. 16, tests were carried out directly on a camera using the FaceNet model for face recognition and the VGG16 model for liveness detection.

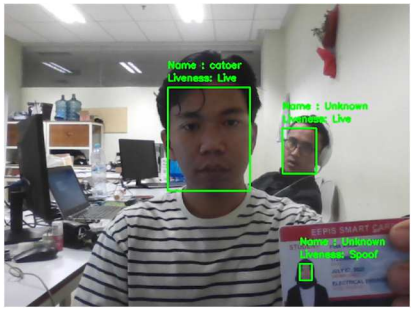


Fig. 16 Testing of facial recognition and liveness detection models

It is evident from the test results that three faces were identified by face detection. As seen in Table 2, the face detected with the name "Catoer" belongs to an individual in the dataset who has either been trained or registered in the face recognition process. The other two faces, however, are identified as "Unknown" as they have not yet been added to the dataset and have not been used in FaceNet deep learning training process.

Direct testing is carried by employing camera inputs. Two faces were identified as "live" in the liveness detection process since the tests verified that the faces were genuine. As can be seen in Fig. 1 of the "Fake" dataset, one face was identified as spoof as it was based on a picture on a card. This demonstrates the model's high degree of accuracy in differentiating between actual and fake faces, providing assurance that the system may be used in real applications to increase security and dependability.

The capacity of the model to differentiate between faces that are created from photos or recordings and genuine faces is shown in this live test. By confirming that the detected face is authentic and not a fake, liveness detection using VGG16 enhances the security of face recognition in the FaceNet model. The objective of this real time assessment, which employs data from cameras positioned between 30 and 60 centimeters apart, is to verify the precision and dependability of the system in authentic scenarios. The goal of this distance limitation is to minimize the loss of significant data while preserving the face's picture quality. The picture quality is preserved at a certain distance, thus allowing facial recognition and liveness detection to be as accurate as possible.

## IV. Conclusion

Studies on face recognition and liveness detection have been carried out using different models and approaches. The findings demonstrate high accuracy rates of various methods, including FaceNet for face recognition and MTCNN and FaceNet for face detection. With 100% accuracy, the VGG16 model had the best performance in authenticity detection training. While it had a somewhat reduced accuracy rate, MobileNet V2 nevertheless functioned well. The ResNet-50 and DenseNet-121 models, on the other hand, recognize real data with a fair accuracy rate but struggle to detect fraudulent data. These findings show that liveness detection in face recognition can distinguish real faces from fake faces and prevent face forgery. Moreover, liveness detection can improve the reliability of face recognition, especially in real-time applications.

Thus, by using a variety of approaches and models, this publication offers significant insights into the development of face recognition and liveness detection systems. The findings of this study may be used as a basis for future advancements in facie recognition and security.

## Acknowledgment

We thank the Academic Director of Vocational Higher Education and Electronics Engineering Polytechnic Institute of Surabaya through the Directorate of Research and Community Service. We also would like to extend our sincere thanks to the researchers at the Signal, Vision, and Graphics Laboratory, Electronics Engineering Polytechnic Institute of Surabaya.

## References

[1] R. Sigit, Z. Arief, dan M. Mobed Bachtiar, "Development of Healthcare Kiosk for Checking Heart Health," *Emit. Int. J. Eng. Technol.*, vol. 3, no. 2, 2015.

[2] M. Letafat-nejad, P. Ebrahimi, M. Maleki, dan A. Aryankhesal, "Utilization of integrated health kiosks: A systematic review," *Med. J. Islam. Repub. Iran*, no. August, 2020, doi: 10.47176/mjiri.34.114.

[3] I. D. Maramba, R. Jones, D. Austin, K. Edwards, E. Meinert, dan A. Chatterjee, "The Role of Health Kiosks: Scoping Review," *JMIR Med. Informatics*, vol. 10, no. 3, 2022, doi: 10.2196/26511.

[4] W. Cheng *et al.*, "Evaluation of a village-based digital health kiosks program: A protocol for a cluster randomized clinical trial," *Digit. Heal.*, vol. 8, no. October, 2022, doi: 10.1177/20552076221129100.

[5] S. Rahmawati, P. W. Ciptadi, dan R. H. Hardyanto, "Sistem Smart Class untuk Presensi Mahasiswa dan Akses Pintu Kelas Berbasis RFID," *Semin. Nas. Din. Inform.*, hal. 185–189, 2021.

[6] M. Sucianto, C. I. Gosal, dan E. A. Lisangan, "Perancangan Prototipe Sistem Kelola Gudang Menggunakan RFID Berbasis Android," *Konstelasi Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 2, hal. 366–375, 2022, doi: 10.24002/konstelasi.v2i2.5611.

[7] L. Li, X. Mu, S. Li, dan H. Peng, "A Review of Face Recognition Technology," *IEEE Access*, vol. 8, hal. 139110–139120, 2020, doi: 10.1109/access.2020.3011028.

[8] C. F. Rosa, R. Amelia, dan F. Mulyasih, "Home Door Security System with Face Recognition Based on Internet of Things," *Pap. Knowl. . Towar. a Media Hist. Doc.*, no. 16040007, hal. 73, 2019.

[9] M. Saini, "Liveness Detection for Face Recognition in Biometrics: A Review," *IOSR J. Comput. Eng.*, vol. 02, no. 02, hal. 31–36, 2016, doi:10.9790/0661-15010020231-36.

[10] S. Liu, Y. Song, M. Zhang, J. Zhao, S. Yang, dan K. Hou, "An identity authentication method combining liveness detection and face recognition," *Sensors (Switzerland)*, vol. 19, no. 21, 2019, doi:10.3390/s19214733.

[11] T. Hussain, D. Hussain, I. Hussain, H. AlSalman, S. Hussain, S.S. Ullah, dan S. Al-Hadhrami, "[Retracted] Internet of Things with Deep Learning-Based Face Recognition Approach for Authentication in Control Medical Systems," *Computational and Mathematical Methods in Medicine*, 2022.

[12] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," Pattern Recognit., vol. 137, 2023, doi: 10.1016/j.patcog.2023.

[13] A. Setiawan, R. Sigit, dan R. Rokhana, "Face Recognition Using Convolution Neural Network Method with Discrete Cosine Transform Image for Login System," *International Journal on Informatics Visualization*, vol. 7, no. 2, 2023.

[14] A.T. Aditya, R. Sigit, dan B.S.B. Dewantara, "Face Recognition Using Deep Learning as User Login on Healthcare Kiosk," in *ICITEE 2022 - Proceedings of the 14th International Conference on Information Technology and Electrical Engineering*, 2022.

[15] A. Setiawan, R. Sigit, dan R. Rokhana, "Implementation of Face Recognition Using Discrete Cosine Transform on Convolutional Neural Networks," in *IES 2022 - 2022 International Electronics Symposium: Energy Development for Climate Change Solution and Clean Energy Transition, Proceeding*, 2022.

[16] S. Khairnar, S. Gite, K. Kotecha, dan S. D. Thepade, "Face Liveness Detection Using Artificial Intelligence Techniques: A Systematic Literature Review and Future Directions," *Big Data Cogn. Comput.*, vol. 7, no. 1, 2023, doi: 10.3390/bdcc7010037.

[17] M. Basurah, W. Swastika, dan O. H. Kelana, "Implementation of Face Recognition and Liveness Detection System Using Tensorflow.Js," *J. Inform. Polinema*, vol. 9, no. 4, hal. 509–516, 2023, doi:10.33795/jip.v9i4.1332.

[18] J. Xiang dan G. Zhu, "Joint face detection and facial expression recognition with MTCNN," *Proc. - 2017 4th Int. Conf. Inf. Sci. Control Eng. ICISCE 2017*, hal. 424–427, 2017, doi:10.1109/icisce.2017.95.

[19] S. Chakraborty dan D. Das, "An Overview of Face Liveness Detection," *Int. J. Inf. Theory*, vol. 3, no. 2, hal. 11–25, 2014, doi:10.5121/ijit.2014.3202.

[20] X. Wang, K. Wang, and S. Lian, "A Survey on Face Data Augmentation," no. 1, 2019, doi: 10.1007/s00521-020-04748-3.

[21] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, dan S. Z. Li, "A face antispoofing database with diverse attacks," *Proc. - 2012 5th IAPR Int. Conf. Biometrics, ICB 2012*, hal. 26–31, 2012, doi:10.1109/ICB.2012.6199754.

[22] A. F. S. Moura, S. S. L. Pereira, M. W. L. Moreira, dan J. J. P. C. Rodrigues, "Video Monitoring System using Facial Recognition: A FaceNet-based Approach," *Proc. - IEEE Glob. Commun. Conf. GLOBECOM*, vol. 2020-January, 2020, doi:10.1109/globecom42002.2020.9348216.

[23] S. Banuchitra, "A Comprehensive Survey of Content Based Image Retrieval Techniques," *International Journal of Engineering and Computer Science*, Aug. 2016, doi: 10.18535/ijecs/v5i8.26.

[24] E. Ghorbel and F. Ghorbel, "3D Face Data Augmentation Based on Gravitational Shape Morphing for Intra-Class Richness," *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*, pp. 1294–1299, 2024, doi: 10.5220/0012466700003636.

[25] S. Yang, W. Xiao, M. Zhang, S. Guo, dan J. Zhao, "Image Data Augmentation for Deep Learning : A Survey," 2020.

[26] E.D. Cubuk et al., "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 113-123.

[27] M. Wang dan W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215-244, 2021.

[28] F. Schroff, D. Kalenichenko, dan J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June-2015, p.815–823, 2015, doi: 10.1109/CVPR.2015.7298682.

[29] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, dan S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, hal. 5962–5979, 2022, doi: 10.1109/tpami.2021.3087709.

[30] I. William, D. R. Ignatius Moses Setiadi, E. H. Rachmawanto, H. A. Santoso, dan C. A. Sari, "Face Recognition using FaceNet (Survey, Performance Test, and Comparison)," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, no. October, 2019, doi:10.1109/ICIC47613.2019.8985786.

[31] L. Q. Vu, P. T. Trieu, dan H. S. Nguyen, "Implementation of FaceNet and support vector machine in a real-time web-based timekeeping application," *IAES Int. J. Artif. Intell.*, vol. 11, no. 1, hal. 388–396, 2022, doi: 10.11591/ijai.v11.i1.pp388-396.