Face Detection Based on Anti-Spoofing with FaceNet Method for Filtering Contract Cheating in Online Exam

Erik Iman Heri Ujianto[1,*], I Gede Susrama Mas Diyasa[2], Achmad Junaidi[3], Ryan Reynickha Fatullah[4], Wahyu Melinda Permanasari[5], Allan Ruhui Fatmah Sari[6]

[1]*Master of Information Technology Program, University of Technology Yogyakarta, Indonesia.*

[2,3,4,5,6]*Faculty of Computer Science, University of Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia*

(Received: September 3, 2025; Revised: November 10, 2025; Accepted: January 15, 2026; Available online: February 17, 2026)

Abstract

This study develops a reliable face-based verification system for online examinations by integrating a face recognition model with a blink detection mechanism to minimize the risk of identity fraud, also known as "contract cheating," and static image manipulation. "Contract cheating" refers to the practice where students hire others to complete their exams or assignments, compromising academic integrity. The growing reliance on online exams has raised concerns about the credibility of facial verification, as conventional methods are often vulnerable to spoofing attempts. To address this issue, the proposed system combines FaceNet, a deep learning model for identity recognition, with Dlib's eye blink detection to provide a stronger layer of protection. The system was evaluated using 5-fold and 10-fold K-fold cross-validation, and additional testing assessed the impact of different video frame rates on performance. The results show that the system performs effectively in identifying legitimate users and detecting spoofing. FaceNet achieved an accuracy of 96.67 percent, outperforming DeepFace, which showed poorer results in precision, recall, and F1 score for some participants. Both models were evaluated on the same dataset, consisting of 150 images. The preprocessing pipeline, including face detection using MTCNN, cropping, and resizing, was applied consistently to both models to ensure a fair comparison of their performance. The system also demonstrated adaptability, achieving correct classifications at both 15 and 30 frames per second. Anti-spoofing tests based on the eye blink detection system detected all real faces, while static images were classified as spoofing. These results confirm that combining face recognition with liveness detection enhances the security of online examination platforms. The findings demonstrate the system's potential to reduce contract cheating and impersonation fraud, making online examinations more credible. Future work may focus on implementing adaptive thresholding for blink detection and integrating multimodal verification techniques to improve robustness across diverse real-world environments.

Keywords: Face Recognition, FaceNet, Dlib, Blink Detection, Spoofing, Online Exam Validity, Identity Verification, DeepFace

1. Introduction

The advent of deep learning has significantly enhanced the accuracy and reliability of systems, making face recognition one of the most prominent areas of research within the broader field of computer vision [1]. Deep learning has enabled the integration of face recognition technology into various sectors, providing real-time identification capabilities that are crucial for improving security [2], [3]. One of the sectors that have benefited from this advancement is education, where automated facial recognition systems have been implemented to monitor online examinations, ensuring integrity and consistency [4], [5].

While online exams provide flexibility for students, they also increase the risk of cheating, particularly through methods such as "contract cheating" or "exam jockeying" [6]. This type of cheating undermines the integrity of education, especially in the context of online exams, which historically have lacked strong vigilance and reliable identity verification. The effectiveness of tracking and verification systems in place for these exams is often inadequate, leaving them vulnerable to cheating [7]. Several prior studies have explored identity authentication using biometric technology. For instance, Salsabila et al. [8] developed an IoT-based security system using ESP32-CAM for facial recognition to

***Corresponding author: Erik Iman Heri Ujianto (dr.eih.ujianto@gmail.com)**
DOI: https://doi.org/10.47738/jads.v7i1.1167

open doors automatically. While effective in well-lit environments, this system becomes unreliable in low-light conditions and is susceptible to spoofing attacks, including vulnerabilities to low-resolution images and videos. This highlights the limitations of current facial recognition technologies, which fail to address more sophisticated face-ID fraud attempts.

Similarly, deep learning-based gesture recognition shows promising results in helping the deaf community with an accuracy of 99.25%. However, the system's resource-intensive nature makes it difficult to deploy in real-time on low-resource devices, presenting a challenge for practical applications, such as online exam proctoring, where devices with limited resources are commonly used [9].

FaceNet and similar machine learning systems have been proposed as solutions to privacy concerns related to identity verification and unauthorized surveillance in online exam supervision. These systems enable secure and automated identity checks, ensuring that only the registered participant can access the exam, while mitigating risks of identity theft or misuse of personal data [7]. These systems improve facial recognition accuracy in online assessments, and anti-cheating technologies that use eye blink detection (e.g., the Eye Aspect Ratio or EAR) have been found effective in combating spoofing attacks [10]. However, these technologies still face challenges with more advanced spoofing techniques, such as deepfake technology.

Previous studies have addressed facial recognition systems in the context of monitoring screen recordings during online exams. For example, Ganidisastra and Bandung reported that FaceNet achieved an accuracy of 98% using an incremental training method. However, these studies primarily focused on improving training efficiency, without addressing potential spoofing threats[11]. In contrast, Akhdan *et al.* incorporated the EAR algorithm for blink detection to mitigate spoofing risks from static images[12]. While effective against such attacks, the performance of this system against more dynamic fraudulent techniques, such as deepfakes, remains uncertain.

These gaps in existing research highlight two key issues. First, although facial recognition systems continue to improve in accuracy, they remain vulnerable to spoofing attacks. Second, EAR-based anti-spoofing methods have demonstrated effectiveness against static image manipulation but have not been fully integrated with facial recognition systems within the context of online examinations. Previous studies have typically evaluated system components in isolation, such as recognition accuracy or spoofing resistance, resulting in a lack of holistic security-oriented solutions.

The novelty of this study lies in the development of an integrated and end-to-end online examination security framework that jointly combines FaceNet-based facial recognition, Dlib-based facial landmark detection, and the Eye Aspect Ratio (EAR) algorithm for real-time blink-based anti-spoofing. Unlike prior works that focus on individual modules independently, this study explicitly integrates identity verification and liveness detection within a single unified pipeline tailored for video-based online exam monitoring. By addressing both recognition accuracy and spoofing resilience simultaneously, the proposed approach provides a more robust and practical solution for securing online examination environments.

2. Literature Review

2.1. Facial Recognition

Facial recognition has been extensively studied as a biometric approach for identity verification, with most modern systems relying on deep learning techniques to extract discriminative facial features. Early approaches focused on handcrafted features, which were sensitive to variations in illumination, pose, and occlusion. More recent studies demonstrate that deep learning–based methods, particularly those employing Convolutional Neural Networks (CNNs), significantly improve robustness by learning hierarchical and invariant facial representations [13], [14].

Despite these advances, prior research largely emphasizes recognition accuracy under controlled or semi-controlled conditions, such as benchmark datasets with limited environmental variability. In practical applications, including online examination monitoring, facial recognition systems must operate under uncontrolled lighting, camera quality variations, and partial occlusions. Several studies report high recognition accuracy; however, they often overlook security vulnerabilities related to presentation attacks, assuming that the detected face corresponds to a live individual [12]. This assumption limits the reliability of facial recognition when deployed as a standalone solution in security-

critical environments. Consequently, while facial recognition technologies have matured in terms of identification performance, their effectiveness in real-world applications depends on complementary mechanisms that address spoofing and liveness verification.

2.2. Face Anti-Spoofing

Face anti-spoofing research focuses on protecting facial recognition systems from presentation attacks, such as printed photos, replayed videos, or synthetic face representations [15]. Existing approaches can generally be categorized into passive and active methods, each with distinct strengths and limitations.

Passive anti-spoofing techniques analyze static facial characteristics, including texture, reflectance, and spatial artifacts, to distinguish genuine faces from spoofed ones [16]. These methods are computationally efficient and do not require user interaction; however, multiple studies report that their performance degrades significantly when confronting sophisticated attacks, such as high-quality video replays or deepfake-generated faces. The visual similarity between real and fake samples in such cases reduces the discriminative power of static features.

In contrast, active anti-spoofing methods leverage dynamic facial behaviors, such as eye blinking, head movement, or facial expressions, to confirm liveness. Blink detection based on the Eye Aspect Ratio (EAR) is one of the most widely adopted active techniques due to its simplicity and real-time feasibility. Compared to passive approaches, EAR-based methods are more effective against static image attacks, as they rely on involuntary physiological responses rather than appearance alone.

Recent studies suggest that hybrid approaches, which combine passive and active methods, can improve robustness against a broader spectrum of attacks, including deepfakes. However, these systems often introduce additional computational complexity and latency, which may not be suitable for real-time online examination environments. Moreover, most existing works evaluate anti-spoofing mechanisms independently of identity recognition, rather than integrating them into a unified authentication pipeline.

In the context of online examinations, where spoofing attempts typically involve static images or prerecorded videos, active liveness detection through blink analysis provides a practical balance between security and efficiency. Therefore, instead of adopting a complex hybrid framework, this study focuses on integrating active liveness detection with facial recognition to address the most relevant threat scenarios in this domain.

2.1. FaceNet

Google FaceNet is a deep learning–based facial recognition model that maps face images into a compact embedding space, where distances directly correspond to facial similarity [17]. Unlike traditional classification-based models, FaceNet formulates face recognition as a metric learning problem using the Triplet Loss function, which enforces intra-class compactness and inter-class separability [18].

Empirical studies demonstrate that FaceNet achieves high recognition accuracy across various datasets, making it a popular choice for identity verification tasks. However, several works report that embedding consistency can be affected by real-world factors such as illumination changes, occlusions, and camera resolution. While data augmentation strategies have been proposed to mitigate these effects, most studies focus on improving recognition robustness rather than addressing spoofing vulnerabilities.

Importantly, FaceNet itself does not include any inherent liveness detection mechanism. As a result, systems relying solely on FaceNet embeddings remain susceptible to presentation attacks if additional safeguards are not implemented. This limitation highlights the need for integrating FaceNet with complementary anti-spoofing techniques when deployed in security-sensitive applications such as online examinations.

2.2. Dlib

Dlib is widely used for facial landmark detection due to its reliability and real-time performance. By employing a pretrained 68-point facial landmark model, Dlib enables precise localization of key facial regions, particularly around the eyes [19], [20]. These landmarks form the foundation for Eye Aspect Ratio (EAR)–based blink detection.

 ISSN 2723-6471

Prior studies demonstrate that EAR provides a robust and computationally lightweight method for detecting eye blinks by analyzing geometric relationships between eye landmarks. Compared to appearance-based eye-state classification, EAR-based methods are less sensitive to lighting variations and image quality, making them suitable for real-time applications.

Nevertheless, Dlib and EAR are typically evaluated as standalone liveness detection components rather than as part of an integrated facial recognition system. Few studies examine how landmark-based blink detection interacts with deep learning–based recognition models in a unified pipeline. This gap suggests an opportunity to combine Dlib's reliable landmark detection with FaceNet's discriminative embeddings to enhance both identity verification and spoofing resistance.

3. Methodology

This study adopts a systematic approach to facial recognition, beginning with data collection and preprocessing stages. The subsequent steps include feature extraction, integration of anti-spoofing systems, and classification of results, as illustrated in figure 1.

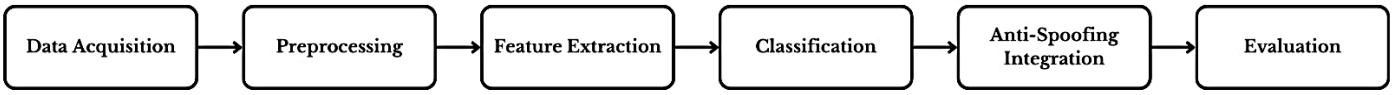

Figure 1. Research workflow

The data processing and face recognition system is designed to enhance identity recognition accuracy by leveraging FaceNet's deep learning-based feature extraction and Dlib's blink detection for anti-spoofing. By combining these components, the system aims to increase the security of online exams by verifying the identity of participants while detecting potential spoofing attempts, as demonstrated by the system's performance in cross-validation and anti-spoofing tests. The system begins with data acquisition to obtain the face dataset and then proceeds to face embedding. In the preprocessing stage, the data is first processed using MTCNN for face detection. MTCNN detects and aligns the faces, after which FaceNet is used to generate embeddings, numerical representations of the facial features. These embeddings serve as feature vectors that enable comparison and identification of faces [18]. The system also includes a mechanism to differentiate real faces from photographs with an anti-spoofing system that detects blinking.

3.1. Data Acquisition

This research utilizes a dataset comprising facial images of students extracted from video recordings. A total of 150 extracted images were obtained, with 10 images per student identity [21]. These images represent various facial orientations, including forward-facing, right turn, left turn, upward gaze, and downward gaze, to capture pose variation within controlled acquisition settings. The video recordings were acquired using a 48-megapixel smartphone camera at 1920 × 1080 pixel resolution and 30 frames per second. The recorded footage was subsequently converted into a sequence of JPEG images, and the resultant image files were organized into directories, as illustrated in figure 2.

Although the dataset size is relatively small compared to large-scale facial datasets commonly used for end-to-end convolutional neural network training (often comprising tens of thousands of images), the dataset is considered appropriate for the proposed experimental design. In this study, the FaceNet model is employed solely as a pre-trained embedding extractor, and its network parameters are not retrained. Only the classification stage, implemented using a Support Vector Machine (SVM), is trained on the collected dataset.

Such a transfer learning framework, where discriminative embeddings are extracted from a model pre-trained on large and diverse facial datasets, has been widely reported to perform effectively even when the downstream task involves limited sample sizes. Moreover, the use of K-fold cross-validation further supports the reliability of performance estimation by reducing the risk of overfitting and providing repeated validation across different data partitions. Therefore, within the context of embedding-based recognition and controlled acquisition settings, the dataset size is considered methodologically justified.

 ISSN 2723-6471

Note: Figure blurred due to privacy restrictions, to view the original dataset please contact the corresponding author.

Figure 2. Dataset

During the extraction process, the image size was adjusted according to the resolution of the recorded video, and a prefix-based filename convention was applied to identify the extracted images. From the extracted data, ten representative images were selected for each student identity based on standardized quality criteria. These criteria included image clarity (sharp focus without blurring), diversity of poses (covering frontal, profile, and slight angular variations), and minimal occlusion (ensuring that facial regions were not obstructed by objects or shadows). To enhance consistency and reproducibility, an automated image quality assessment algorithm was employed to evaluate the images using objective measurements, such as sharpness metrics and pose variation analysis, rather than relying on manual selection. All selected images were required to satisfy these predefined criteria to maintain consistent quality and controlled variability across the dataset.

3.2. Preprocessing

In the preprocessing stage, three steps are performed, namely detection, resizing, and cropping of faces using the Multi-Task Cascaded Convolutional Neural (MTCNN) library. Detection is used to find the position of the face in the given image and mark it with a bounding box. Then, the image is cropped based on the bounding box. After that, the cropped face image is resized according to the dimensions required by the model as shown in figure 3.

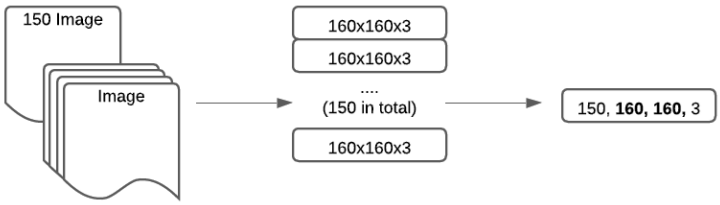

Figure 3. Preprocessing workflow

This array format reflects the data structure used in deep learning, where each image is represented as a 3D array with three color channels. An illustration of the ndarray representation (160,160,3) and the cropped face, showing how the data is prepared before entering the embedding stage using FaceNet.

3.3. Feature Extraction

In this feature extraction stage, the FaceNet model is used to extract the important facial. This process results in an embedding vector with a dimension of 512 which is a numerical representation of the facial features. The main objective of this embedding vector is to minimize the inter-vector distance of the same face, as well as maximize the inter-vector distance of different faces. The feature extraction process starts by processing the face data that has gone through the crop-ping stage with a size of 160x160x3. Next, the face image will pass through several layers in the deep learning-based FaceNet architecture, including Conv2D, Batch Normalization, PReLU, Max-Pooling2D, and Dense (Fully Connected). These layers aim to process the image data to produce a face embedding vector that has 512 dimensions as shown in figure 4.

 ISSN 2723-6471

The resulting embedding vector is a 512-dimensional representation of the facial features, which encapsulates unique information about the face's structure. Each dimension in the vector corresponds to a specific feature of the face, learned during the training process. However, these dimensions are not directly interpretable, as they result from the learned weights of the neural network. The vector is then normalized using L2 normalization, which ensures that the magnitude of the embedding is consistent across all faces.

Handling noise in the vector space is crucial, as external factors such as lighting, facial expression, and occlusion can introduce variability into the embeddings. To mitigate this, the Triplet Loss function is employed. This function minimizes the distance between embeddings of the same identity (positive pairs) while maximizing the distance between embeddings of different identities (negative pairs). This process helps the model focus on the most relevant features of the face and reduces the influence of noise, such as minor changes in lighting or pose, that could otherwise affect the distance between embeddings. As a storage step, the results of this feature extraction are saved in the form of a compression file using the numpy library with the *.npz file format, which makes it easy to save and reuse at a later stage.

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 14s 14s/step
Face Embedding: [[ 5.19384956e-03  3.88375893e-02 -4.38757204e-02
   7.38230348e-02  2.56002625e-03  1.18242763e-02  4.83232364e-03
  -5.13288518e-03  4.06345576e-02 -1.67379491e-02 -1.36599066e-02
   2.00084504e-03 -3.57626528e-02  3.27086076e-03 -3.48893330e-02
   8.63807928e-03  7.04570813e-03 -9.87051986e-03  3.80739346e-02
   4.52246517e-02  7.13659450e-02  2.11375505e-02 -1.91964786e-02
  -1.03330985e-02  9.38101560e-02  6.61508515e-02 -5.95844984e-02
  -5.19852038e-04 -1.52706867e-02  3.53425853e-02  6.39391690e-02
  -1.55130420e-02  5.45448549e-02  4.29209620e-02 -5.72598726e-02
  -1.21937525e-02 -9.90338326e-02 -1.10246819e-02 -2.51479503e-02
   1.74074993e-02 -6.51243702e-02  2.87800771e-03 -2.11830959e-02
  -9.13853645e-02 -1.18339770e-02 -4.47591245e-02  1.98462699e-02
  -1.02526061e-01 -3.43789831e-02 -3.61553505e-02 -1.07920252e-01
   1.79148931e-02 -1.81389209e-02  6.07692404e-03 -7.55900368e-02
  -2.19287761e-02 -4.76215892e-02  2.96855457e-02  1.60755217e-02
```

Figure 4. 512-Dimensional Face Feature

Figure 4 illustrates the 512-dimensional face embedding vector generated by FaceNet, which numerically represents key facial features. Each point in the 512-dimensional space corresponds to a feature of the face, and the relative distances between different vectors indicate the degree of similarity between faces.

3.4. Classification

The face recognition classification in this study employs a linear Support Vector Machine (SVM) [22], [23]. Given a training dataset.

$$\{(x_i, y_i)\}_{i=1}^{N} \tag{1}$$

$x_i \in \mathbb{R}^{512}$ represents the 512-dimensional embedding vector extracted by FaceNet and $y_i \in \{-1, +1\}$ denotes the class label, the objective of SVM is to determine an optimal separating hyperplane defined as:

$$f(x) = w^T x + b \tag{2}$$

$w \in \mathbb{R}^{512}$ is the weight vector and $b \in \mathbb{R}$ is the bias term. For the linear SVM, the optimization problem is formulated as:

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i \tag{3}$$

Subject to the constraints:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0 \tag{4}$$

ξ_i are slack variables allowing soft-margin classification, $C > 0$ is the regularization parameter controlling the trade-off between maximizing the margin and minimizing classification errors.

 ISSN 2723-6471

A linear kernel is used in this study, defined as:

$$K(x_i, x_j) = x_i^T, x_j \tag{5}$$

This kernel is suitable for high-dimensional feature representations such as the 512-dimensional embeddings generated by FaceNet. For prediction, a new sample xxx is classified according to:

$$\hat{y} = sign\ (w^T x + b) \tag{6}$$

After solving the optimization problem in (3)–(4), the resulting hyperplane defined in (2) is used for classification. A linear kernel is adopted due to its effectiveness in handling high-dimensional embedding vectors generated by FaceNet. In this study, the regularization parameter C is set to its default value. Future work may explore hyperparameter tuning using grid search combined with cross-validation to further optimize classification performance and reduce overfitting.

The face dataset was split into 80% for training and 20% for testing to ensure that the model can learn optimally while being evaluated on data it has never seen before. This training/testing split helps assess the model's generalization performance. The model's classification results will be evaluated using the Confusion Matrix, which provides key metrics such as accuracy, precision, recall, and F1 score, offering insights into how well the model distinguishes between different classes [24], [25].

3.5. Face Anti-Spoofing

This study utilizes both OpenCV and Dlib for face anti-spoofing, with each library playing a distinct role. OpenCV is responsible for face detection and image preprocessing, such as converting video frames to grayscale and resizing images to appropriate dimensions. The Haar Cascade Classifier in OpenCV, specifically the haarcascade_frontalface_default.xml, performs the initial face detection step to locate faces in the input video frames [26].

Dlib, on the other hand, handles facial landmark detection and Eye Aspect Ratio (EAR) calculation, which are crucial for blink detection and liveness verification. By tracking the movement of key landmarks around the eyes, Dlib ensures the system can differentiate between a real person and a static image. Dlib uses the frontal face detector and the shape_predictor() function to extract facial feature points such as the eyes, nose, and mouth. These points are essential for identifying the positions of facial parts, particularly the eyes, which are central to blink detection, as shown in figure 5 [27]. Therefore, OpenCV handles the initial face detection and image preprocessing, while Dlib focuses on precise facial landmark localization and blink detection, ensuring the authenticity of the face during the anti-spoofing process.

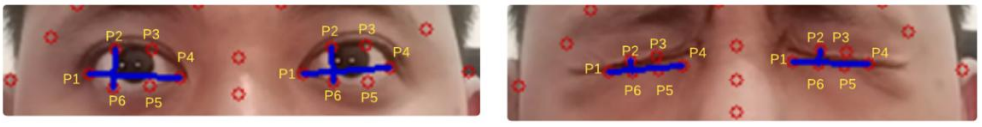

Figure 5. Eyes Landmark

Once the facial features have been extracted successfully, the next stage is Eye Detection. Landmarks are extracted for the left and the right eyes. Detection of a possible blink is done using the Eye Aspect Ratio (EAR) [28]. EAR is calculated based on 6 major points that surround the eye. The Eye Aspect Ratio (EAR) is calculated by measuring the distances between specific facial landmarks around the eyes, specifically the vertical and horizontal distances [29]. A blink is detected when the EAR value drops below a threshold for a sustained period. In this study, the system evaluates the EAR over a time window of 2 consecutive frames. If the EAR remains below the threshold for this duration, the system concludes that a blink has occurred. This short time window ensures that transient fluctuations in EAR are not mistakenly identified as blinks.

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \times \|p_1 - p_4\|} \tag{7}$$

Lastly, in the Classification step, the system examines how the EAR value changes over time to determine if there has been a blink. If the EAR value continues to be below the threshold for more than the specified time, the system will detect a blink, and then will update the status of the face. The algorithm in this classification is designed in such a way that if the number of winks in a time frame is above a certain threshold and the detection time exceeds the time threshold, the system will assume that spoofing is in progress and that the detected face is, in fact, a face that is not

 ISSN 2723-6471

real. Being able to detect eye blinks is crucial for spoofing detection systems because eye blinks cannot be easily imitated, and therefore, become a behavioral response that is invaluable in distinguishing real organisms from an attempt to spoof using a photograph's static image [30].

3.6. System Testing

This evaluation compares the performance of FaceNet and DeepFace using confusion matrix–based metrics, including accuracy, precision, recall, and F1-score. Both models are evaluated under identical testing conditions to ensure a fair comparison.

FaceNet is selected due to its suitability for real-time applications, as its multi-dimensional embeddings remain robust against variations in viewpoint, facial expression, and illumination. DeepFace is included as a comparative baseline because it is a widely used open-source model that demonstrates reliable performance under suboptimal facial conditions. In this study, identity recognition performance is not evaluated in isolation but is integrated with an active anti-spoofing mechanism, as described in Algorithm 1, to ensure that only live and verified identities are accepted.

Algorithm 1. Integrated Face Recognition and Liveness Verification Framework

Input:
Video sequence $\boldsymbol{V} = \{\boldsymbol{F_t}\}_{\boldsymbol{t=1}}^{\boldsymbol{T}}$
Trained SVM classifier $\boldsymbol{f}(.)$
Recognition threshold $\boldsymbol{\tau_r}$
Blink threshold $\boldsymbol{\tau_b}$
Output:
Final decision $\boldsymbol{D} \boldsymbol{\epsilon} \{\boldsymbol{Accept}, \boldsymbol{Reject}\}$

1: Frame Extraction
$\boldsymbol{V} = \{\boldsymbol{F_t}\}_{\boldsymbol{t=1}}^{\boldsymbol{T}}$

2: Face Detection
For each frame $\boldsymbol{F_t}$:
$B_t = D(\boldsymbol{F_t})$

If $B_t \neq \emptyset$, crop face region:
$x_t = C(\boldsymbol{F_t}, B_t)$

3: Feature Embedding
Extract 512-dimensional embedding:
$z_t = \boldsymbol{\mathcal{F}}(x_t), \quad z_t \in \mathbb{R}^{512}$

4: Identity Classification
Compute decision function:
$\hat{y}_t = f(z_t) = (w^T z_t + b)$

Let classification confidence be p_t.
Recognition indicator:
$I_t = \begin{cases} 1, & p_t \geq \boldsymbol{\tau_r} \\ 0, & Otherwise \end{cases}$

5: Eye Landmark Detection
Detect eye landmarks:
$(p_1, p_2, p_3, p_4, p_5, p_6)_t = \mathcal{L}(x_t)$

6: Eye Aspect Ratio (EAR)
$EAR_t = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \times \|p_1 - p_4\|}$

7: Blink Detection
For k consecutive frames:
$L = \begin{cases} 1, & if\ EAR_t < \boldsymbol{\tau_b}\ for\ k\ frames \\ 0, & Otherwise \end{cases}$

8: Final Decision Rule
$D = \begin{cases} Accept, & I_t = 1 \wedge L = 1 \\ Reject, & Otherwise \end{cases}$

9: Return *D*

4. Results and Discussion

4.1. Dataset Split and Validation Technique

The dataset was partitioned using an 80:20 hold-out strategy, resulting in 120 images for training and 30 images for independent testing. Although the dataset comprises only 150 samples, it is important to emphasize that the convolutional neural network backbone (FaceNet) was not trained from scratch. Instead, it was utilized as a fixed pretrained embedding extractor. Consequently, the learning process in this study operates on 512-dimensional discriminative embeddings rather than raw image pixels.

ISSN 2723-6471

In transfer learning settings, the number of required training samples is substantially reduced because high-level facial representations have already been learned from large-scale datasets during the pretraining phase. Therefore, the training stage in this work involves only the optimization of a linear Support Vector Machine (SVM) classifier in the embedding space, which requires significantly fewer samples compared to end-to-end CNN training.

To ensure robust performance estimation, K-fold cross-validation was applied during the model development phase. This approach mitigates the risk of overfitting to a single data partition by repeatedly evaluating the model across multiple training–validation splits. It should be noted that cross-validation was used solely for model selection and internal validation. The final performance metrics were computed using the independent 20% hold-out test set, which remained completely unseen during training and model selection.

Within this embedding-based recognition framework and controlled acquisition environment, the dataset size is considered methodologically adequate for evaluating classifier performance. Nevertheless, future studies may expand the dataset to assess scalability and robustness under more diverse real-world conditions. During K-fold cross-validation, the dataset is randomly partitioned into k equal-sized folds. In each iteration, one-fold is used as the validation set while the remaining k − 1 folds are used for training. This process is repeated k times so that each fold serves as validation exactly once. The validation accuracies from all iterations are then averaged to obtain a reliable estimate of predictive performance.

In the first scenario, the dataset contained 150 total images, and each fold contained 30 images. Therefore, the dataset was divided into 5 folds, and the training and validation process was conducted 5 times. The performance of the training and validation of each fold of the k cross-folds has been illustrated in table 1.

Table 1. 5-fold Cross Validation Results

Fold	Accuracy FaceNet	Accuracy Deepface
1	100%	80%
2	90%	83.33%
3	100%	96.67%
4	96.67%	80%
5	100%	70%
Average	97.33%	82%

The classification model Testing FaceNet has a cross-validation k of 10. The dataset has been split into 10 subsets containing 150 data images each with 15 images in each. This process has been repeated 10 times for training and validation. The results for the cross validation of k 10 are shown in table 2.

Table 2. 10-fold Cross Validation Results

Fold	Accuracy FaceNet	Accuracy Deepface
1	100%	86.67%
2	100%	73.33%
3	100%	93.33%
4	93.33%	73.33%
5	100%	93.33%
6	100%	100%
7	100%	86.67%
8	93.33%	80%
9	100%	80%
10	100%	73.33%
Average	98.67%	84%

 ISSN 2723-6471

Although FaceNet achieves consistently high accuracy across multiple folds, this result should be interpreted with caution due to the relatively small dataset size. To reduce overfitting, K-fold cross-validation was applied and FaceNet was used as a pretrained feature extractor without retraining network weights, which limits model memorization. During each fold, training and validation data were kept mutually exclusive to avoid data leakage. However, since all samples originate from the same acquisition environment, the model's generalization to unseen conditions may be limited. Therefore, the reported results primarily reflect performance under controlled experimental settings, and future work will focus on evaluation using larger and more diverse datasets to further assess generalization capability.

4.2. Confusion Matrix

The confusion matrix shown in figure 6 is presented in normalized form, where each value represents the proportion of predictions for a given class, rather than absolute counts. A value of 1.0 indicates that all samples of a class were correctly classified, meaning the model accurately identified every instance of that class. A value of 0.5 indicates partial correctness, where half of the samples from the class were correctly classified, and the remaining samples were misclassified into other classes. This normalized approach provides a clearer understanding of the model's performance, particularly in cases where the model may have difficulty distinguishing between certain classes.

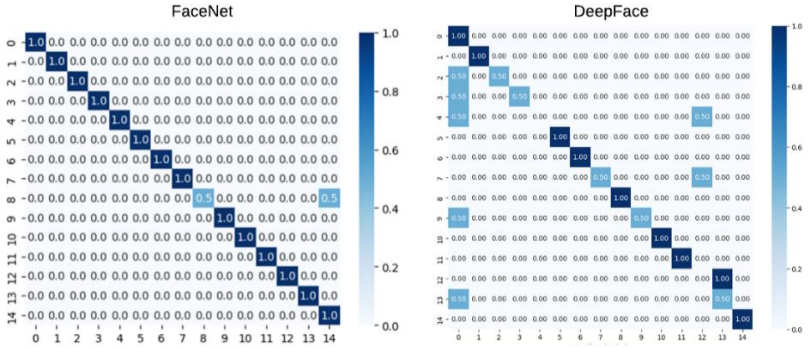

Figure 6. Confusion Matrix

In every confusion matrix, the main diagonal contains the number of correct predictions for each class. For FaceNet, the confusion matrix indicates 14 fully correct classifications and one partially correct classification, resulting in only a single misclassification case. This demonstrates that FaceNet is highly consistent in distinguishing between different student identities. In contrast, the DeepFace confusion matrix shows 8 correct classifications (1.0), and 5 are partial classifications (0.5), indicating that DeepFace struggled with some identities but was still able to partially classify them correctly.

Overall, FaceNet achieves substantially higher performance than DeepFace, with an accuracy of 96.67% compared to 70%. This indicates that FaceNet performs better than DeepFace because DeepFace might be having trouble with facial variations like expression and lighting conditions. Therefore, DeepFace is most likely worse because of high sensitivity in those areas. Thus, FaceNet is superior in face recognition, especially for applications that need high accuracy and stability.

Although the FaceNet confusion matrix shows predictions that are highly concentrated along the diagonal, this result should be interpreted cautiously given the limited dataset size. The high normalized values indicate strong performance under the current experimental setup, but they may also reflect limited data diversity rather than full generalization capability. To prevent data leakage, training and testing samples were strictly separated, and FaceNet was employed as a pretrained embedding extractor without retraining model weights. Nevertheless, since the data originate from a controlled acquisition environment, the results primarily demonstrate robustness within this setting. Future work will include evaluation on larger and more diverse datasets to further assess generalization performance.

4.3. Classification Reports

The model's predictive accuracy regarding the prediction of identities is assessed by using a classification report which contains precision, recall, and F1-Score as presented in table 3.

 ISSN 2723-6471

Table 3. Classification Report FaceNet and DeepFace

	FaceNet					DeepFace			
	Precision	**Recall**	**F1-Score**	**Support**		**Precision**	**Recall**	**F1-Score**	**Support**
Abi	1.00	1.00	1.00	2	**Abi**	0.29	1.00	0.44	2
Alvin	1.00	1.00	1.00	2	**Alvin**	1.00	1.00	1.00	2
Apta	1.00	1.00	1.00	2	**Apta**	1.00	0.50	0.67	2
Fafa	1.00	1.00	1.00	2	**Fafa**	1.00	0.50	0.67	2
Fahmi	1.00	1.00	1.00	2	**Fahmi**	0.00	0.00	0.00	2
Fania	1.00	1.00	1.00	2	**Fania**	1.00	1.00	1.00	2
Felix	1.00	1.00	1.00	2	**Felix**	1.00	1.00	1.00	2
Hani	1.00	1.00	1.00	2	**Hani**	1.00	0.50	0.67	2
Ken	1.00	0.50	0.67	2	**Ken**	1.00	1.00	1.00	2
Muftah	1.00	1.00	1.00	2	**Muftah**	1.00	0.50	0.67	2
Rafli	1.00	1.00	1.00	2	**Rafli**	1.00	1.00	1.00	2
Rambe	1.00	1.00	1.00	2	**Rambe**	1.00	1.00	1.00	2
Rizky	1.00	1.00	1.00	2	**Rizky**	0.00	0.00	0.00	2
Ryan	1.00	1.00	1.00	2	**Ryan**	0.33	0.50	0.40	2
Vico	0.67	1.00	0.80	2	**Vico**	1.00	1.00	1.00	2
Accuracy			0.97	30	**Accuracy**			0.70	30
Macro Avg	0.98	0.97	0.96	30	**Macro Avg**	0.77	0.70	0.70	30
Weighted Avg	0.98	0.97	0.96	30	**Weighted Avg**	0.77	0.70	0.70	30

Overall, FaceNet achieves near-perfect precision, recall, and F1-scores for most identity classes. While this indicates strong discriminative capability of the pretrained FaceNet embeddings, such uniformly high performance should be interpreted with caution given the limited dataset size and the small number of samples per identity. The misclassification observed for label 8 (Ken), which resulted in reduced recall, indicates that the model is not entirely immune to identity confusion and suggests uneven performance across classes with similar facial characteristics.

In contrast, DeepFace exhibits more varied precision and recall values across identities, reflecting a more conservative and realistic performance profile. The observed trade-off between precision and recall indicates that DeepFace is less prone to memorization but more sensitive to facial variations and class imbalance.

It is important to note that each identity in this study is represented by a limited number of samples, which constrains the reliability of per-class metrics and may contribute to optimistic performance estimates. To reduce the risk of data leakage, training and testing sets were strictly separated, and both FaceNet and DeepFace were used solely as pretrained embedding extractors without retraining their internal parameters. Nevertheless, the reported results primarily reflect performance under controlled conditions, and further evaluation on larger and more diverse datasets is required to fully assess generalization capability.

The t-SNE visualization as illustrated in figure 7 reveals that FaceNet does a better job of clustering the facelogs with clear separations between student identities, indicating that it is able to reliably discriminate faces with minimal overlap. In contrast, DeepFace shows more dispersed and less organized clusters, with noticeable overlap between student identities, which could be due to class imbalance and overfitting in the training phase. This suggests that FaceNet is more effective in clustering and identifying facelogs, whereas DeepFace struggles with class imbalance and may not perform as well in real-world scenarios where data distribution is more varied [31].

ISSN 2723-6471

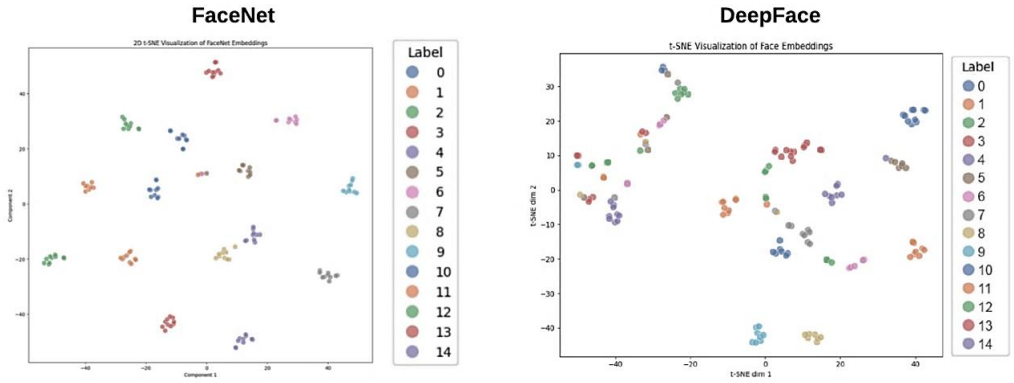

Figure 7. T-SNE visualization FaceNet and DeepFace embedder

The t-SNE visualizations provide a qualitative overview of the embedding distribution produced by each model. The visualization suggests that FaceNet embeddings tend to form more compact clusters associated with student identity labels, with less apparent overlap between different identities. In contrast, DeepFace embeddings appear more dispersed, with several identities exhibiting overlapping regions in the reduced feature space. These visual patterns suggest that FaceNet embeddings are more consistently grouped by identity compared to DeepFace. However, it is important to note that t-SNE is primarily a visualization tool, and these observations are intended as qualitative insights rather than quantitative measurements of clustering quality.

4.4. Testing Using Different Video Frame Rates

In this study, the performance of the face recognition system was evaluated at two different video frame rates, 15 fps and 30 fps. The primary performance indicator considered was recognition confidence, which reflects the classifier's certainty in identifying a participant. As shown in table 4, the system achieved correct recognition at both frame rates, with confidence values remaining above the defined acceptance threshold.

It should be noted that this evaluation focused on recognition reliability rather than computational latency or processing speed. Although higher frame rates such as 30 fps provide more frequent visual information, this study did not explicitly measure latency, processing time per frame, or real-time computational constraints. Consequently, no quantitative comparison of system responsiveness between 15 fps and 30 fps is reported.

Nevertheless, the results demonstrate that the proposed system remains functionally stable at both frame rates, indicating its practical applicability in typical online examination scenarios. Future work will extend this evaluation by incorporating latency measurements and real-time performance analysis, including frame processing time and system responsiveness, to better assess deployment feasibility under different hardware and bandwidth conditions.

Table 4. Testing Using Different Video Frame Rates

No	Name	15 FPS		30 FPS	
		Recognized	Confidence	Recognized	Confidence
1	Abi	Yes	45.44	Yes	45.89
2	Alvin	Yes	33.09	Yes	35.26
3	Apta	Yes	40.39	Yes	41.77
4	Fafa	Yes	42.36	Yes	39.68
5	Fahmi	Yes	36.21	Yes	36.67
6	Fania	Yes	37.61	Yes	39.43
7	Felix	Yes	28.93	Yes	40.17
8	Hani	Yes	40.83	Yes	39.65
9	Ken	Yes	47.04	Yes	38.90
10	Muftah	Yes	44.91	Yes	45.09
11	Rafli	Yes	29.38	Yes	33.65

12	Rambe	Yes	31.34	Yes	31.78
13	Rizky	Yes	52.13	Yes	52.30
14	Ryan	Yes	35.86	Yes	35.81
15	Vico	Yes	41.99	Yes	42.06

4.5. Testing with Unrecognized Faces

This test is conducted to evaluate the model's performance in dealing with unrecognized faces, i.e. faces that are not included in the training data. The model still carries out the detection process on the face video, but the final result is not expected to show a high confidence level, as an indicator that the model is not sure about the identity match. The confidence values presented in table 5 are derived from the SVM's probability output, which shows that all faces that are not recognized by the system produce a fairly low range of confidence values, ranging from 12.04% to 19.29%. These values consistently show a low confidence level, indicating that the system is able to distinguish between recognized and untrained faces.

Table 5. FaceNet Testing with an Unrecognized Face

No	Name	Recognized	Confidence
1	Identity 1	No	13.39
2	Identity 2	No	13.92
3	Identity 3	No	19.29
4	Identity 4	No	19.22
5	Identity 5	No	14.55
6	Identity 6	No	12.04
7	Identity 7	No	15.50
8	Identity 8	No	16.89
9	Identity 9	No	18.49
10	Identity 10	No	17.09
11	Identity 11	No	13.62
12	Identity 12	No	15.80
13	Identity 13	No	18.78
14	Identity 14	No	14.42
15	Identity 15	No	16.26

4.6. Face Anti-Spoofing Test

The face anti-spoofing test was conducted using a video containing footage of each student's face while blinking consciously in front of the camera. The model utilizes the EAR (Eye Aspect Ratio) measurement algorithm to detect changes in eye shape when opening and closing. The following results can be seen in Table 6. Based on Table 6, the test results show that all tested identities were successfully detected to blink correctly. The range of EAR values recorded varies for each individual, with the lowest value being 0.240, and the highest being 1.197. This range indicates that there is a significant change in eye shape during the blinking process that the model is able to recognize.

The face anti-spoofing test is carried out using static images containing the faces of each student. This method aims to make the system able to distinguish which faces are real and which faces are fake faces through an eye blink detection approach. The following results can be seen in table 6. Based on table 6, a static image or fake face test was conducted to detect spoofing attempts by utilizing the EAR parameter. In this case, although the EAR value was detected below the threshold, there was no significant change in the EAR value over a period of 2 frames in a period of 10 seconds. This indicates the absence of eye blinking movement which is an indicator of liveness. Therefore, the system concludes that the face is a fake or spoofing face. Thus, the pro-gram can distinguish between a real face that moves and blinks and a static image that shows no signs of life despite having an EAR value close to the threshold.

Table 6. Face Anti-Spoofing Test on Real Face and Fake Face (Static Image)

No	Name	Real Face		Fake Face (static image)	
		Blink Detected	Confidence	Blink Detected	Confidence
1	Abi	Yes	0.482 – 1.053	No	0.416
2	Alvin	Yes	0.260 – 0.634	No	0.638
3	Apta	Yes	0.440 – 0.838	No	0.681
4	Fafa	Yes	0.420 – 0.948	No	0.870
5	Fahmi	Yes	0.390 – 0.879	No	0.685
6	Fania	Yes	0.257 – 0.925	No	0.830
7	Felix	Yes	0.448 – 0.934	No	0.761
8	Hani	Yes	0.303 – 1.197	No	0.775
9	Ken	Yes	0.312 – 0.812	No	0.759
10	Muftah	Yes	0.468 – 1.039	No	0.724
11	Rafli	Yes	0.312 – 1.055	No	0.520
12	Rambe	Yes	0.271 – 0.909	No	0.707
13	Rizky	Yes	0.359 – 0.994	No	0.855
14	Ryan	Yes	0.453 – 0.742	No	0.675
15	Vico	Yes	0.240 – 0.797	No	0.696

4.7. Threshold

The determination of the threshold is done to provide a reference limit in the decision-making process by the system, both in recognizing facial identity and in detecting eye blinks as an indicator of liveness. This threshold value becomes the basis for distinguishing between conditions that are considered valid and those that are not, so that the system can work objectively and consistently on various data tested.

From the test results, the confidence value used to determine the threshold is obtained from the prediction results of the classification model, namely the SVM. This model provides confidence for each recognized face class. This confidence value is a measure of how confident the model is that the face belongs to a particular class. From the test results in table 4 (Testing Using Different Video Frame Rates) and table 5 (Testing with Unrecognized Faces), the minimum confidence value for recognized faces was 28.93%, and the maximum confidence value for unrecognized faces was 19.29%.

The result of this face threshold calculation is 24.11%. With this value, the system will only confirm a person's identity if the confidence of the prediction result is greater than 24.11%, thus minimizing errors in recognizing foreign or fake faces as illustrated in figure 8.

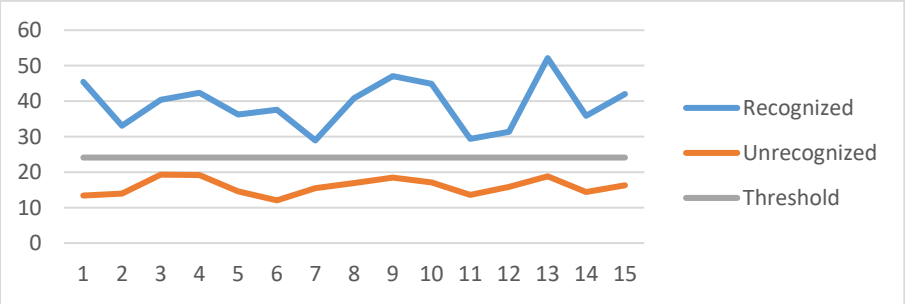

Figure 8. Threshold determination in face recognition

In face anti-spoofing, the EAR (Eye Aspect Ratio) value is used to measure the change in eye shape during blinking. EAR is calculated based on the distance between the landmark points around the eyes detected using the face detection

method and the landmarks from Dlib. From the data in Table 6 (Face Anti-Spoofing Test), the minimum EAR value when the eyes were open was 0.634, and the maximum EAR value when the eyes were closed was 0.482. The result of this EAR threshold calculation is 0.558. This value is used to determine when a blink is considered valid, which is when the EAR value drops below 0.558 and then rises again within a certain time, indicating the presence of real blinking activity from the subject. With this threshold, the system can distinguish between real humans and static or spoofed images as illustrated in figure 9.

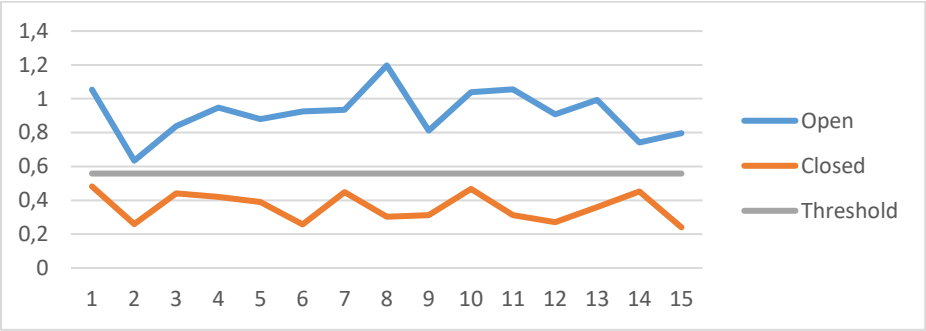

Figure 9. Threshold determination on face anti-spoofing

4.8. Discussion

The experimental results demonstrate that the embedding-based approach using FaceNet combined with a linear SVM classifier achieves superior performance compared to DeepFace under the tested conditions. The high accuracy obtained by FaceNet indicates that its learned feature representations are highly discriminative, even when applied to a relatively small dataset. This supports the effectiveness of transfer learning, where embeddings learned from large-scale datasets can generalize well to controlled experimental settings with limited training samples.

The performance gap between FaceNet and DeepFace suggests differences in embedding robustness. FaceNet appears to generate more stable feature representations under variations in pose and minor illumination changes. In contrast, DeepFace shows a higher number of partial or ambiguous classifications, which may indicate sensitivity to intra-class variation. These findings highlight the importance of selecting an embedding model that is not only accurate but also consistent across different facial conditions.

The integration of liveness detection through Eye Aspect Ratio (EAR) further strengthens the reliability of the proposed system. By combining identity verification and blink-based liveness confirmation, the framework reduces the likelihood of spoofing attacks using static images. This dual-verification mechanism enhances system robustness, particularly for real-time authentication scenarios.

However, several limitations should be acknowledged. First, the dataset size remains relatively small compared to large-scale face recognition benchmarks. Although transfer learning mitigates this limitation, performance in more diverse and unconstrained environments may differ. Second, the liveness detection mechanism relies solely on blink detection, which may be vulnerable to sophisticated spoofing techniques such as high-quality video replays. Future work may incorporate additional anti-spoofing strategies, such as texture-based analysis or depth estimation, to further improve security.

Overall, the results indicate that FaceNet combined with a linear SVM classifier provides a reliable and computationally efficient solution for controlled face recognition applications. The addition of lightweight liveness verification makes the framework suitable for practical authentication systems requiring both accuracy and security.

5. Conclusion

This study demonstrates that FaceNet outperforms DeepFace in terms of face recognition accuracy and anti-spoofing performance. In 5-fold cross-validation, FaceNet achieved an average accuracy of 97.33%, and 98.67% in 10-fold cross-validation, while DeepFace showed a maximum accuracy of 84%. For anti-spoofing evaluation, a blink-based liveness detection approach using the Eye Aspect Ratio (EAR) was employed. Based on the conducted experiments

involving 15 participants and static image spoofing attempts, the system successfully distinguished live faces from spoofed images. No misclassification was observed within the tested dataset, indicating that all live faces were correctly recognized as genuine and all static image attacks were correctly identified as spoofed, even when the EAR value dropped below the threshold of 0.558.

The system also demonstrated robust performance at varying frame rates (15 fps and 30 fps), maintaining accurate recognition with a confidence threshold of 24.11%. This further validated its ability to handle different video conditions effectively. In conclusion, FaceNet, in combination with Dlib's blink detection, offers a reliable and effective solution for securing online exams. The system's high accuracy in face recognition and anti-spoofing detection ensures stronger identity verification and helps prevent impersonation fraud, making online assessments more secure and trustworthy.

6. Declarations

6.1. Author Contributions

Conceptualization: E.I.H.U., I.G.S.M.D. and R.R.F.; Methodology: E.I.H.U., I.G.S.M.D., A.J. and R.R.F.; Software: R.R.F., W.M.P. and A.R.F.S ; Validation: E.I.H.U., I.G.S.M.D. and R.R.F.; Formal Analysis: E.I.H.U., I.G.S.M.D., and R.R.F.; Investigation: A.J. and R.R.F.; Resources: W.M.P. and R.R.F.; Data Curation: I.G.S.M.D.; W.M.P. and A.R.F.S. Writing Original Draft Preparation: E.I.H.U., W.M.P., A.R.F.S. and R.R.F.; Writing Review and Editing: E.I.H.U., A.J., W.M.P. and A.R.F.S; Visualization: E.I.H.U., I.G.S.M.D., W.M.P. and A.R.F.S.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Li, "The Application and Challenges of Different Face Recognition Technologies in the Three Major Fields of Security, Social Media, and Medical Care," *ACE*, vol. 95, no. 1, pp. 174–181, Oct. 2024, doi: 10.54254/2755-2721/95/2024CH0051.

[2] L. Alzubaidi, "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions," *J Big Data*, vol. 8, no. 1, pp. 53-67, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[3] J. Zhang and N. Hu, "Accuracy and Robustness Evaluation of Deep Learning Algorithms in Facial Recognition Systems," *Systems and Soft Computing,* vol. 7, no. 1, pp. 20-52, 2025, doi: https://doi.org/10.1016/j.sasc.2025.200252.

[4] A. Sakhipov, I. Omirzak, and A. Fedenko, "Beyond Face Recognition: A Multi-Layered Approach to Academic Integrity in Online Exams," *EJEL*, vol. 23, no. 1, pp. 81–95, Feb. 2025, doi: 10.34190/ejel.23.1.3896.

[5] S. Essahraui, "Deep Learning Models for Detecting Cheating in Online Exams," *Computers, Materials and Continua*, vol. 85, no. 2, pp. 3151–3183, 2025, doi: https://doi.org/10.32604/cmc.2025.067359.

[6] T. Lancaster and C. Cotarlan, "Contract Cheating by STEM Students Through a File Sharing Website: A Covid-19 Pandemic Perspective," *Int J Educ Integr,* vol. 17, no. 1, pp. 1-3, Dec. 2021, doi: 10.1007/s40979-021-00070-0.

[7] A. Pramadi, M. Pali, F. Hanurawan, and A. Atmoko, "Academic Cheating in School: A Process of Dissonance Between Knowledge and Conduct," *Mediterranean Journal of Social Sciences*, vol. 8, no. 6, pp. 155–162, Nov. 2017, doi: 10.1515/mjss-2017-0052.

[8] N. Salsabila, A. Siswanto, and L. Bayuaji, "Design of a Smart Home Door Security System with Face Detection and Smart Bell using ESP32-CAM," in *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI),* vol. 2025, no. 1, pp. 124–129, 2025. doi: 10.1109/ICMCSI64620.2025.10883160.

[9] M. S. Assiri and M. M. Selim, "A Swin Transformer-Driven Framework for Gesture Recognition to Assist Hearing Impaired People by Integrating Deep Learning with Secretary Bird Optimization Algorithm," *Ain Shams Engineering Journal*, vol. 16, no. 6, pp. 1-13, May 2025, doi: 10.1016/j.asej.2025.103383.

[10] S. R. Akhdan, R. Supriyanti, and A. S. Nugroho, "Face Recognition with Anti Spoofing Eye Blink Detection," *AIP Conference Proceedings*, vol. 2482, no. 1, pp. 1-16, Feb. 2023, doi: 10.1063/5.0113512.

[11] A. H. S. Ganidisastra and Y. Bandung, "An Incremental Training on Deep Learning Face Recognition for M-Learning Online Exam Proctoring," in *2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob),* vol. 2021, no. 1, pp. 213–219, 2021. doi: 10.1109/APWiMob51111.2021.9435232.

[12] R. V. Virgil Petrescu, "Face Recognition as a Biometric Application," *Journal of Mechatronics and Robotics*, vol. 3, no. 1, pp. 237–257, Jan. 2019, doi: 10.3844/jmrsp.2019.237.257.

[13] M. U. Haq, M. A. J. Sethi, S. Ahmad, N. Ahmad, M. S. Anwar, and A. Kutlimuratov, "A Comprehensive Review of Face Detection/Recognition Algorithms and Competitive Datasets to Optimize Machine Vision," *Computers, Materials and Continua*, vol. 84, no. 1, pp. 1–24, 2025, doi: https://doi.org/10.32604/cmc.2025.063341.

[14] J. Chi, C. Kim On, H. Zhang, and S. S. Chai, "A Review of Deep Convolutional Neural Networks in Mobile Face Recognition," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 23, pp. 4–19, Dec. 2023, doi: 10.3991/ijim.v17i23.40867.

[15] V. H and T. G, "Antispoofing in Face Biometrics: A Comprehensive Study on Software-Based Techniques," *Comput Sci Inf Technol,* vol. 4, no. 1, pp. 1–13, Mar. 2023, doi: 10.11591/csit.v4i1.pp1-13.

[16] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep Learning for Face Anti-Spoofing: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 45, no. 5, pp. 5609–5631, 2023, doi: 10.1109/TPAMI.2022.3215850.

[17] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* vol. 2015, no. 1, pp. 815–823, 2015. doi: 10.1109/CVPR.2015.7298682.

[18] S. Qi, X. Zuo, W. Feng, and I. G. Naveen, "Face Recognition Model Based On MTCNN And Facenet," in *2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC),* vol. 2022, no. 1, pp. 1–5, 2022. doi: 10.1109/ICMNWC56175.2022.10031806.

[19] C. Q. Lai and S. S. Teoh, "An Efficient Method of HOG Feature Extraction Using Selective Histogram Bin and PCA Feature Reduction," *Adv. Electr. Comp. Eng.*, vol. 16, no. 4, pp. 101–108, 2016, doi: 10.4316/AECE.2016.04016.

[20] X. Li, J. Luo, C. Duan, Y. Zhi, and P. Yin, "Real-Time Detection of Fatigue Driving Based on Face Recognition," *J. Phys.: Conf. Ser.,* vol. 1802, no. 2, pp. 1-14, Mar. 2021, doi: 10.1088/1742-6596/1802/2/022044.

[21] J. K. Essel, J. A. Mensah, E. Ocran, and L. Asiedu, "On the Search for Efficient Face Recognition Algorithm Subject to Multiple Environmental Constraints," *Heliyon,* vol. 10, no. 7, pp. 1-18, Apr. 2024, doi: 10.1016/j.heliyon.2024.e28568.

[22] I. G. S. Mas Diyasa, D. A. Prasetya, H. A. Cahyani Kuswardhani, and C. Halim, "Detection of Abnormal Human Sperm Morphology Using Support Vector Machine (SVM) Classification," *Information Technology International Journal,* vol. 2, no. 2, pp. 57–63, Nov. 2024, doi: 10.33005/itij.v2i2.36.

[23] M. Afifudin, A. Junaidi, A. N. Sihananto, and I. Fithriyah, "GWO-SVM: An Approach to Improving Svm Performance Using Grey Wolf Optimizer in Intellectual Disability Classification," *JITET,* vol. 12, no. 3S1, pp. 1-12, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5359.

[24] I. P. Pratama and N. K. Ningrum, "Face Recognition Using MTCNN Face Detection, ResNetV1 Feature Embeddings, and SVM Classification," *Journal of Applied Informatics and Computing,* vol. 9, no. 5, pp. 1-12, 2025, doi: https://doi.org/10.30871/jaic.v9i5.11016.

[25] A. B. Prastyo and A. Setyanto, "Analysis of FaceNet and VGG16 for Blind Face Recognition with MTCNN and HaarCascade Detection Methods," *G-Tech,* vol. 9, no. 2, pp. 570–577, Apr. 2025, doi: 10.70609/gtech.v9i2.6573.

[26] I. G. S. M. Diyasa, A. H. Putra, M. R. M. Ariefwan, P. A. Atnanda, F. Trianggraeni, and I. Y. Purbasari, "Feature Extraction for Face Recognition Using Haar Cascade Classifier," in *Nusantara Science and Technology Proceedings, Galaxy Science*, vol. 2022, no. may, pp. 1-12, May 2022. doi: 10.11594/nstp.2022.2432.

[27] C. Meijerink, "Facial Landmark Detection Under Challenging Conditions," *Thesis, University of Twente, Enschede,* 2021. [Online]. Available: https://purl.utwente.nl/essays/86867

[28] D. Borza, A. Darabant, and R. Danescu, "Real-Time Detection and Measurement of Eye Features from Color Images," *Sensors,* vol. 16, no. 7, pp. 1105-1119, Jul. 2016, doi: 10.3390/s16071105.

[29] H. Qi, C. Wu, Y. Shi, X. Qi, K. Duan, and X. Wang, "A Real-Time Face Detection Method Based on Blink Detection," IEEE Access, vol. 11, no. 1, pp. 28180–28189, 2023, doi: 10.1109/ACCESS.2023.3257986.

[30] T. Jung, S. Kim, and K. Kim, "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern," *IEEE Access*, vol. 8, no. 1, pp. 83144–83154, 2020, doi: 10.1109/ACCESS.2020.2988660.

[31] S. M. H. Abidi, S. A. Hassan, S. M. Raza, and M. J. Beliatis, "Advances in Face Recognition: A Comprehensive Review of Feature Extraction and Dataset Evaluation," *Electronics,* vol. 15, no. 2, pp. 338-349, Jan. 2026, doi: 10.3390/electronics15020338.