Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

High-Performance Face Spoofing Detection using Feature Fusion of FaceNet and Tuned DenseNet201

Leygian Reyhan Zuama [1], De Rosal Ignatius Moses Setiadi [1,2,*], Ajib Susanto [1], Stefanus Santosa [3], Hong-Seng Gan [4], and Arnold Adimabua Ojugo [5]

[1] Faculty of Computer Science, Univesitas Dian Nuswantoro, Semarang 50131, Indonesia; e-mail : leygianzuama@gmail.com; ajib.susanto@dsn.dinus.ac.id
[2] Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Dian Nuswantoro University, Semarang 50131, Indonesia; e-mail : moses@dsn.dinus.ac.id
[3] Civil Engineering Department, Politeknik Negeri Semarang, Central Java, Indonesia; e-mail : stefanus.santosa@polines.ac.id
[4] School of AI and Advanced Computing, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong -Liverpool University, Suzhou, Jiangsu, P.R. China 215400; e-mail: hongseng.gan@xjtlu.edu.cn
[5] Department of Computer Science, Federal University of Petroleum Resources Effurun, Nigeria; e-mail: ojugo.arnold@fupre.edu.ng
* Corresponding Author : De Rosal Ignatius Moses Setiadi

Abstract: Face spoofing detection is critical for biometric security systems to prevent unauthorized access. This study proposes a deep learning-based approach integrating FaceNet and DenseNet201 to enhance face spoofing detection performance. FaceNet generates identity-based embeddings, ensuring robust facial feature representation, while DenseNet201 extracts complementary texture-based features. These features are fused using the Concatenate function to form a more comprehensive representation for improved classification. The proposed method is evaluated on two widely used face spoofing datasets, NUAA Photograph Imposter and LCC-FASD, achieving 100% accuracy on NUAA and 99% on LCC-FASD. Ablation studies reveal that data augmentation does not always enhance performance, particularly on high-complexity datasets such as LCC-FASD, where augmentation increases the False Rejection Rate (FRR). Conversely, DenseNet201 benefits more from augmentation, while the proposed method performs best without augmentation. Comparative analysis with previous studies further confirms the superiority of the proposed approach in reducing error rates, particularly Half Total Error Rate (HTER), False Acceptance Rate (FAR), and FRR. These findings indicate that combining identity-based embeddings and texture-based feature extraction significantly improves spoofing detection and enhances model robustness across different attack scenarios. This study advances biometric security by introducing an efficient feature fusion strategy that strengthens deep learning-based spoof detection. Future research may explore further optimization strategies and evaluate the approach on more diverse datasets to enhance generalization.

Keywords: Anti-spoofing; Augmentation effect; Biometric security; Data augmentation; FaceNet; Face recognition; Face spoofing detection; Feature fusion.

Received: January, 6th 2025
Revised: February, 6th 2025
Accepted: February, 9sth 2025
Published: February, 12th 2025
Curr. Ver.: February, 12th 2025

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

The application of face recognition technology in the industrial and commercial sectors has increased rapidly[1]. This is supported by increasingly reliable and high-performance face recognition systems[2]–[4], as well as supporting multimodal detection[5], including the ability to detect faces even in the presence of occlusion[6]. However, security challenges such as face spoofing attacks are serious problems in implementing this system[7]. Face spoofing is a type of attack that uses a person's facial data, such as a photo or printed image, to imitate the identity of that individual [8]. Attackers use this technique to infiltrate face recognition

systems, potentially disrupting solutions based on this technology, especially in business and industrial environments[7], [9], [10].

Along with the security challenges in face recognition systems, research in computer vision has focused on developing various anti-face spoofing methods. Various approaches have been applied to detect face spoofing attacks, such as the use of Convolutional Neural Network (CNN) with GoogleNet transfer learning, which achieved an accuracy of 96.74%[11], the Enhanced Face Local Binary Feature (ELBP) method with an accuracy of 95.1%[12], the Radon Transform method with an accuracy of 97.2%[13], the Modified Difference of Gaussian - Local Binary Pattern Variance (DoG-LBPV) method with an accuracy of 99.22%[14], and the Alex-Net method which achieved an accuracy of 91,1%[15]. These approaches have successfully improved detection accuracy but still require testing on more diverse datasets to address various types of attacks.

In addition to method development, research also utilizes datasets specifically designed to support face spoofing classification. The two most popular datasets are the NUAA Photograph Imposter Database[16] and the Large Crowd-Collected Facial Anti-Spoofing Dataset (LCC-FASD)[9]. The NUAA dataset is widely used in anti-face spoofing research because it contains various examples of printed image-based attacks, allowing testing of methods with simple attack scenarios. Meanwhile, LCC-FASD is designed to test models in complex attack scenarios, such as video-based attacks or images with varying lighting. Research using the LCC-FASD dataset reports an accuracy of up to 97% with the application of the CNN method[17].

Although various methods have provided good results in detecting face spoofing attacks, there is still room to improve model performance. This study aims to optimize detection results using feature fusion techniques on the NUAA and LCC-FASD datasets. Feature fusion is a technique that combines features from several models to form a more representative combination of features[18]–[20]. This approach allows the combination of features from various models, such as models A and B, which complement each other with equal weights or according to relevance[19], [20].

Feature fusion techniques can be grouped into two main approaches: feature selection-based and feature extraction-based. Feature selection-based approaches select the most relevant subset of features from the original data. In contrast, feature extraction-based approaches create new features by transforming or combining the original features. Furthermore, the features generated through these approaches are combined using one of two main strategies: serial feature fusion and parallel feature fusion[21]. In serial feature fusion, features are combined sequentially, while in parallel feature fusion, features are combined simultaneously. Other research[22] has developed feature fusion techniques into Deep Feature Fusion Networks (DFFN), which can extract features from low, middle, and high network layers.

The main objective of applying feature fusion techniques in this study is to improve accuracy and reduce loss values by combining features from various models. This approach is expected to overcome the limitations of a single model, as has been shown in research[23]–[27]. This technique has also been applied in various cases, such as multi-label image classification in fundus diseases by combining features from ResNet and Residual Attention Module (RAM)[28]. Another recent innovation is the Binary and Residual Feature Fusion (BARF) technique, which combines ResNet and DenseNet for medical dataset classification[29].

One of the best models designed for face recognition is FaceNet. This model has the advantage of producing highly representative facial feature embeddings and can accurately distinguish identities. Other models, such as DenseNet201, are known for producing rich hierarchical features through dense connection architectures, allowing information from previous layers to be used optimally[30]. By combining features from these two models, it is hoped that a more complete and adequate feature representation will be created for detecting face spoofing attacks.

Flatten is a layer in CNN that changes a feature representation previously in multidimensional form into a one-dimensional vector. Flatten is often implemented after the last convolution or pooling layer to rearrange the spatial feature map to be simpler without changing any features in the extracted features[31]. Flatten's application can help connect spatial information processed in CNN with the classification layer. Flatten makes the classification process more optimal because the model can process all feature information in vector form[32]. Inspired by the literature that has been explained previously, this study contributes to:

1. The use of FaceNet to produce more representative facial identity embeddings allows separation between individuals with a high level of accuracy.
2. Extraction of additional texture features using DenseNet201 with Flatten, which enriches spatial information to improve face spoofing pattern detection.
3. Integrating FaceNet and DenseNet201 features using the Concatenate function creates a richer and more effective combined representation in distinguishing real and fake faces.
4. Comprehensively testing and exploring the proposed method by adding an in-depth analysis of the data augmentation effects on the proposed model and its pre-trained models.

The rest of this paper is organized into several main sections. Section 2 reviews previous research on face spoofing detection methods, including deep learning architectures and feature fusion techniques. Section 3 describes the proposed method, including model configuration, preprocessing, and applied feature fusion techniques. Section 4 presents experimental results and model performance analysis on NUAA and LCC-FASD datasets, while Section 5 compares the proposed method with other pre-trained models and previous studies. Section 6 concludes the main findings and proposes further research directions.

2. Literature Review

2.1. FaceNet

FaceNet is a model specifically designed for facial recognition. FaceNet is a deep convolutional network-based model that directly learns images of people's faces into Euclidean vector space, which makes this model very efficient[33], [34]. The FaceNet model was developed in 2015 for face detection[35], [36]. The method used in FaceNet is the basic implementation of Euclidean Embedding per image using a deep convolutional network. FaceNet produces its training output into 64, 128, 256, or 512 dimensions using the triple-based loss function technique, which is used to optimize embedding and make the distance between the faces of the same person closer[33], [34].

The application of the FaceNet model is not only for face classification but can also be applied to cases such as verification and face clustering[33]. The Euclidean embedding space approach in this FaceNet model becomes an advantage in the model[33], [35] because using Euclidean creates a face image projected directly into a low-dimensional vector space so that this embedding vector can reflect facial identity with a small Euclidean distance between vectors for faces on the same individual and a large distance for faces of different individuals.

In this Euclidean embedding, the triplet loss function design is applied to optimize the Euclidean embedding, using three types of input in the form of anchors, namely the reference face, positive is the face of the same individual, and also negative, which is the face of a different individual[33], [37]. This triplet loss is designed explicitly for representation in the Euclidean embedding space so that the distance between vectors can accurately represent a relationship between facial identities [6]. The default FaceNet design is presented in Table 1.

Table 1. FaceNet default configuration.

Function	Configuration
Input Image Size	160 × 160 × 3
Network Architecture	Inception-Style Network
Optimizer	SGD (Stochastic Gradient Descent) + AdaGrad + Learning Rate: 0.05
Training Configuration	Loss Function: Triplet Loss

FaceNet uses 160 × 160 × 3 (RGB) facial image input to produce a 128-dimensional embedding vector output. This model is built using the Inception-Style Network architecture, which is known to be efficient in parameter and computational usage and is capable of learning multiple feature scales in a single stage[33], [38], [39]. The loss function used is Triplet Loss, designed to improve inter-class separability and intra-class compactness[33], [40]. The training dataset comprises 200 million images covering eight million facial identities[33]. Optimization is done using a combination of SGD and AdaGrad, using an initial learning rate of

0.05 and a batch size of 1,800 triplets per batch, supported by triplet mining techniques to produce stronger embeddings[33], [41].

The FaceNet model has been used in several research papers, for example, in research[42] and VGGFace2[43]. Training datasets. Based on the training dataset test using the VGGFace2 dataset, it obtained superior performance compared to CASIA-WebFace from the dataset used for comparison. FaceNet was also tested on the low-resolution face image dataset [44]. The development of the FaceNet model using the Labeled Faces in the Wild (LFW)[45] dataset was tested with several low and high-resolution levels. From the use of the FaceNet model in the research conducted by the research, the evaluation results showed that at high resolutions 160×160, 100×100, 80×80 obtained high accuracy results compared to lower resolutions such as 20×20 and 10×10. From the results of the comparison, the resolution of the face image dataset greatly affects the performance evaluation results that will be obtained later.

2.2. DenseNet201

DenseNet201 is a pre-trained model that has been trained using the ImageNet dataset. This model is designed to classify an image[30], [46]. This model uses 201 layers that are interconnected with each other, which makes this model more effective for processing complex image patterns[47], [48]. DenseNet201 has also been implemented in several research studies to solve problems or test the model on existing data in a study [47], [49].

DenseNet201 introduces an efficient network architecture with the concept of dense connectivity[30], where each layer in DenseNet201 is directly connected to all previous layers, which allows the integration of more information and also a better gradient flow during the training process[50], [51]. This direct connection can overcome vanishing gradients and feature redundancy in traditional network architectures[52]. With 201 layers, this pre-trained model can increase efficiency when learning complex features in image data, especially in large datasets such as ImageNet [53]. However, the number of 201 layers does not only come from the number of layers in its density block. Several hidden layers are also counted in the total layers, including batch normalization, transition layers, and pooling layers. The default DenseNet configuration is presented in Table 2.

Table 2. DenseNet201 default configuration.

Function	Configuration
Input Image Size	224 × 224 × 3
Network Architecture	Layer: 201 Layer Growth Rate: 32 Block Configuration: Block 1: 6 Layer. Block 2: 12 Layer. Block 3: 48 Layer. Block 4: 32 Layer Transition Layer: Convolutional 1 × 1 + AveragePooling: 2 × 2
Optimizer	SGD (Stochastic Gradient Descent) + Learning Rate: 0.1
Training Configuration	Loss Function: Categorical cross-entropy. batch_size=32 Weight Decay: 1e-4
Pooling & Output Layer	GlobalAveragePooling + FullyConnectedLayer 1000 ('Softmax')

Based on Table 2, DenseNet201 uses 224 × 224 × 3, which is used as the standard resolution in the ImageNet dataset[48], [53], un to get a balance for image details with efficiency in model computation[53]. The growth rate is 32 to maintain the value of the number of parameters to be efficient by ensuring that each layer of DenseNet201 produces new feature information[54], [55]. The Dense Block configuration (Block 1: 6 layers, Block 2: 12 layers, Block 3: 58 layers, Block 4: 32 layers) is useful for gradually increasing block capacity. While the application of transition layers with convolutional and average pooling is 2 × 2, the

transition layers will be used to reduce the dimensions of the feature[30], [50], reduce memory requirements, and also prevent overfitting[30].

The DenseNet-121 model has been used to detect face spoofing in research [56], demonstrating its effectiveness on public datasets such as CASIA-FASD and laboratory datasets. The densely connected layer structure allows DenseNet-121 to retain and propagate features well, thus capturing image characteristics such as texture and facial distortion due to spoofing. This makes DenseNet-121 superior to manual feature-based methods. Research [57] compared DenseNet-121 and ResNet-18 in detecting face spoofing under various conditions. DenseNet-121 excels because it can retain initial feature information while combining it with information from deeper layers. With datasets such as CASIA-FASD, Replay-Attack, and PR-FSAD, DenseNet-121 has higher precision, recall, and accuracy than ResNet-18. The structure of DenseNet-121 preserves facial shape information, which makes it more effective in recognizing distortions and high-frequency characteristics in fake images.

2.3. Feature Fusion

Feature fusion is a technique used to combine features from multiple sources or models to improve system performance in a particular task[58]. This technique allows the model to capture richer and more complex information than conventional approaches that use only one feature type [58], [59]. In general, feature fusion is categorized into three main approaches: Early Fusion (Data-Level Fusion), Intermediate Fusion (Feature-Level Fusion), and Late Fusion (Decision-Level Fusion).

Early Fusion combines raw data from multiple modalities before the feature extraction stage. This approach requires additional preprocessing, such as normalization or alignment, to ensure alignment between the data before being used in a machine-learning model[60]. On the other hand, Late Fusion combines the final decisions from several different models. These models produce independent predictions using methods like majority voting, probability averaging, or other ensemble strategies to improve overall prediction accuracy[61].

Intermediate Fusion is a method that combines features that have been extracted from multiple models to form a richer feature representation[19]. One of the main techniques in this approach is Concatenation, which combines features from multiple models into a larger feature vector without losing important information. For example, if there are two feature vectors, $h_1 = [5,7,9]$ and $h_2 = [4,8,10]$, then with the concatenation technique, the result is $h_{fusion} = [5,7,9,4,8,10]$, where all features are retained without information reduction[62]. The main advantage of Intermediate Fusion, especially Concatenation, is its ability to capture a more complete representation of various data sources, increase model flexibility, and maintain detailed information from each combined feature. In addition, this approach is simpler and more efficient than other more complex methods but can still provide optimal performance in many scenarios.

Feature fusion has improved model performance in various tasks, such as small object detection[63]. For example, in a study that overcomes the limitations of the Single Shot MultiBox Detector (SSD) in detecting small objects, feature fusion is applied by combining features from the DenseNet model as a backbone with SSD, resulting in improved detection performance. In addition, in another study, feature fusion was applied to combine local and global features to overcome the limitation of CNN in capturing information from different convolutional layers, which resulted in improved accuracy on CASIA-WebFace and Labeled Faces in the Wild (LFW) datasets[64].

Although there are various more complex feature fusion techniques, Intermediate Fusion with Concatenation remains an attractive choice due to its simplicity and better computational efficiency. A method is proposed in this study to combine FaceNet and DenseNet201 models, which require feature fusion without losing essential information, as an optimal strategy to improve the performance of detection models, especially in the tasks of face recognition and face spoofing detection.

3. Proposed Method

The proposed approach in this study aims to improve the accuracy of face spoofing detection by combining features from two models, namely FaceNet and DenseNet201, by utilizing the serial feature fusion technique. The diagram of the proposed method is shown in Figure 1. The stages carried out in this study are explained as follows:

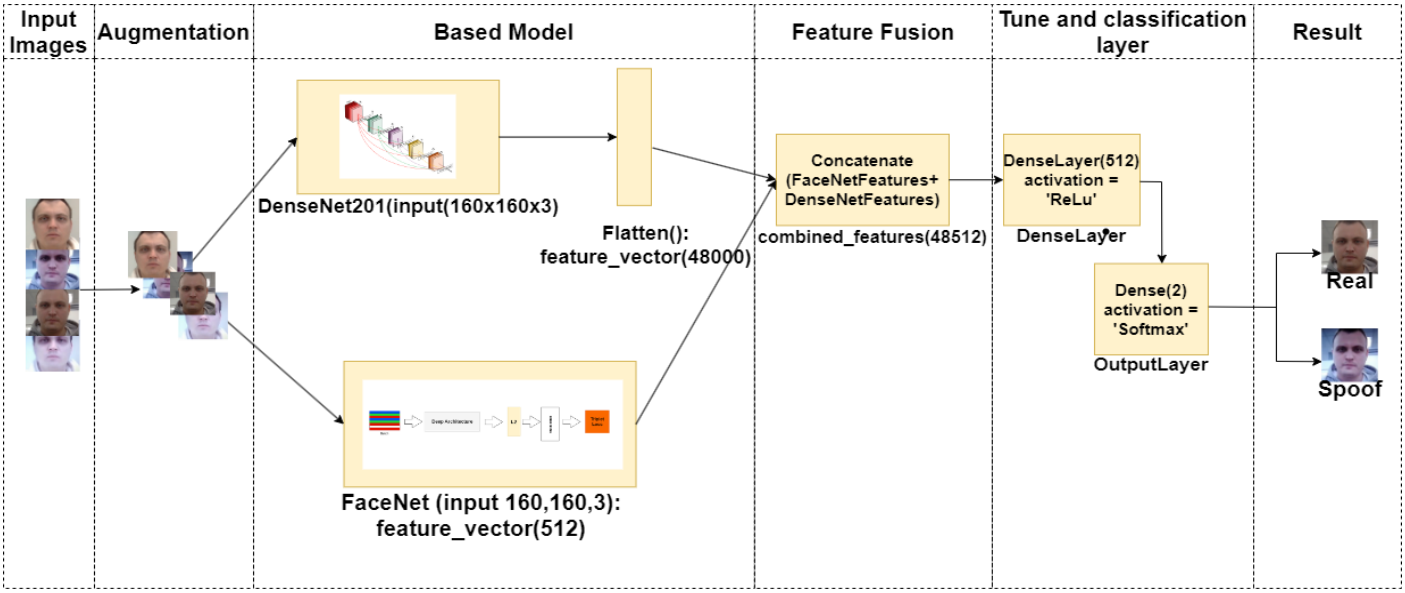


Figure 1. Proposed Method

3.1. Preprocessing

The preprocessing stage is carried out to prepare the dataset for optimal use in the model training process. This process is designed to ensure that the images have sufficient quality and variation so that the model can learn better from the data. The preprocessing steps include:

1. The dataset is divided into three subsets with a proportion of 80:10:10 for training, validation, and testing. This approach aims to ensure the model can be validated and tested separately, reducing the risk of overfitting. The dataset division uses a random method to ensure an even data distribution.
2. In this study, the dataset used is 224 × 224 pixels. To fulfill the fourth research contribution, namely the exploration of the augmentation effect. All images are augmented at their original size to ensure the augmentation is more effective and retains important details. Data augmentation is often considered an important step in improving the generalization and robustness of the model [65]–[68], especially when dealing with various face spoofing scenarios. On the other hand, other studies also explain that data augmentation introduces unwanted artifacts or excessive transformations that reduce model performance[69]–[71]. The augmentation step is carried out to explore whether the augmented data has a positive or negative effect in this case. Data augmentation is only performed on the training dataset using ImageDataGenerator in TensorFlow/Keras, with the parameter settings presented in Table 3.

Table 3. Augmentation techniques and parameters.

Augmentation Type	Parameter Values
Rotation	20 degrees
Horizontal flip	True
Height shift range	0.1
Width shift range	0.1
Brightness range	[0.5, 1.5]

3. After augmentation, the images are resized to 160 × 160 pixels to match the input specifications of FaceNet and DenseNet201.
4. Lastly, image pixel value normalization is performed with a scale (1./255) to reduce data variability.

3.2. Model Configuration

The proposed approach integrates FaceNet and DenseNet201 models with the configuration in Table 4. FaceNet is designed to produce facial identity-based embedding with 512

dimensions, effectively distinguishing individual faces. Meanwhile, DenseNet201 captures hierarchical features such as textures and complex patterns that are relevant in detecting visual anomalies in fake faces. In contrast, by default, DenseNet201 produces a complex spatial feature map in the last convolutional layer. Therefore, DenseNet201 is set with include_top=False, which removes the fully connected layer (top) for classification on the ImageNet dataset so that the model is only used as a feature extractor. By default, DenseNet201 has a 224×224 input. Still, with the include_top=False setting, making DenseNet flexible to 160×160 pixels input produces a feature map in the last convolutional layer resized into one vector with a dimension of 48000 through the Flatten layer.

Table 4. Model Configuration.

Function	Configuration	
Feature Extraction	FaceNet: Input size 160×160×3 Embedding: 512 dimensions	DenseNet201: Input size 160×160×3 +Flatten: 48000 dimensions
Feature Fusion	The Concatenate function produces a combined feature vector with 1152 dimensions.	
Classification Layers	Dense Layer: A dense layer with 512 units and a ReLU activation function. Output Layer: A dense layer with two units and Softmax activation function.	
Training Configuration	Optimizer: Adam with a learning rate of 2×10^{-4} Loss Function: Categorical cross-entropy. Metrics: Accuracy. Epochs: 15 batch_size=32 Validation: 5-fold stratified cross-validation.	

The features from FaceNet and DenseNet201 are combined using the Concatenate function as a feature fusion sequence, producing a combined feature vector with a dimension of 48512. This vector is processed through a dense layer with 512 units and ReLU activation, then forwarded to the output layer with two units and the Softmax activation function for face classification into real and spoof. The Adam optimizer was chosen to optimize model training because of its adaptive ability to handle the dynamics of the loss function during the training process. It can also achieve convergence faster without requiring a lot of manual adjustments to the hyperparameters. Meanwhile, the learning rate of 2×10^{-4} was chosen to maintain a balance between learning speed and parameter update stability. This relatively small value prevents overfitting or divergence during training, especially on complex models such as FaceNet and DenseNet201 combined. This approach leverages the strengths of both models to create more prosperous and more accurate feature representations, thereby improving the performance of face spoofing detection, especially in the face of various variations and challenges of spoofing attacks.

Adam optimizer is chosen to optimize the model training because of its adaptive ability to handle the dynamics of the loss function during the training process. It can also achieve faster convergence without requiring much manual adjustment of hyperparameters. Meanwhile, the learning rate of 2×10^{-4} is chosen to maintain a balance between learning rate and parameter update stability. This relatively small value prevents overfitting or divergence during training, especially on complex models such as FaceNet and DenseNet201 combined. This approach leverages the strengths of both models to create more prosperous and more accurate feature representations, thereby improving the performance of face spoofing detection, especially in the face of various variations and challenges of spoofing attacks.

3.3. Evaluation Metrics

Model evaluation in face spoofing classification uses various metrics to assess its effectiveness and resilience to spoofing attacks. Proper evaluation allows the identification of system weaknesses and the improvement of model performance. The metrics used include accuracy, precision, recall, and F1-score. Accuracy measures the percentage of correct predictions from the entire test data. Precision assesses the extent to which the positive predictions

generated are spoofing attacks. Recall shows the model's ability to recognize all spoofing attacks.

Meanwhile, the F1-score balances the two as the harmonic mean of precision and recall, especially in imbalanced class distributions. In addition to conventional metrics, face spoofing detection is also evaluated using the False Acceptance Rate (FAR), False Rejection Rate (FRR), and Half Total Error Rate (HTER). FAR mismeasures the percentage of spoofing attacks classified as real faces, which risks causing the system to accept fake identities. FAR is calculated using Equation (1). FRR measures the percentage of real faces wrongly categorized as spoofing, which can cause user discomfort, calculated by Equation (2). HTER is the average of FAR and FRR, which reflects the overall balance of the system, which is calculated by Equations (3) to (6).

$$FAR = \frac{FP}{FP + TN} \tag{1}$$

$$FRR = \frac{FN}{FN + TP} \tag{2}$$

$$HTER = \frac{FAR + FRR}{2} \tag{3}$$

Where FP (False Positive) is the number of spoofing cases classified as real faces; TN (True Negative) is the number of real faces correctly classified; FN (False Negative) is the number of real faces incorrectly classified as spoofing, and TP (True Positive) is the number of real faces correctly classified.

A low HTER value indicates a more reliable model regarding face spoofing detection because it has a good balance in accepting real faces and rejecting spoofing attacks. Therefore, in this study, HTER is used as the main metric to evaluate the effectiveness of the proposed method[72]–[74]. In contrast, accuracy, precision, recall, and F1-score are used to provide additional analysis regarding the classification performance[75], [76].

4. Results and Discussion

This research uses two datasets, the NUAA Photograph Imposter Database[16]. The NUAA dataset has 12,614 images, with 5105 as client images and 7509 imposter images. With each photo session having different lighting and a photo session on the results of the face spoofing attack, such as the angle of the tilted print face, this dataset was taken using a webcam with a resolution of 640×480 pixels. The second dataset is LCC-FASD[9], with 18,827 high-quality images. This dataset is taken from a collection of short videos edited by making the face in the image in the center by cropping the image. The image size in the LCC-FASD dataset varies from 150 pixels to 1350 pixels. Figures 2 and 3 show sample images from the NUAA and LCC-FASD datasets, respectively.

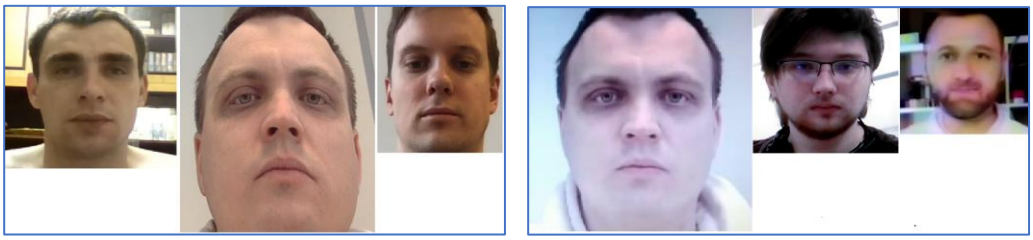
After all images are resized to 224 × 224 × 3, the dataset is split into three parts with a composition of 80% training, 10% validation, and 10% testing. This study added a data augmentation process to obtain a more detailed analysis and determine the augmentation's effect. It should be emphasized that augmentation was not performed on the testing data. Sample results of the augmentation process are shown in Figure 4.

(a) (b)

Figure 2. Sample of NUAA dataset (**a**) Client Face/Real Image; (**b**) Attack Photo/Imposter/ Spoof Image.

The sample datasets NUAA and LCC-FASD visually differ due to differences in attack methods. For NUAA, the method used is face spoofing by engineering the face using a

printed face from paper or printed face. In LCC-FASD, an attack method is more difficult to distinguish, namely by screen capturing from the device that displays the face. This makes the LCC-FASD dataset challenging to distinguish with human vision and complex because spoofing in LCC-FASD is relatively "smoother" and focuses on the face object. So, it can be concluded that the LCC-FASD dataset is more complex and difficult to distinguish than the NUAA dataset.



(a) (b)

Figure 3. Sample of LCC-FASD dataset (**a**) Real Image; (**b**) Spoof Image.

(a)

(b)

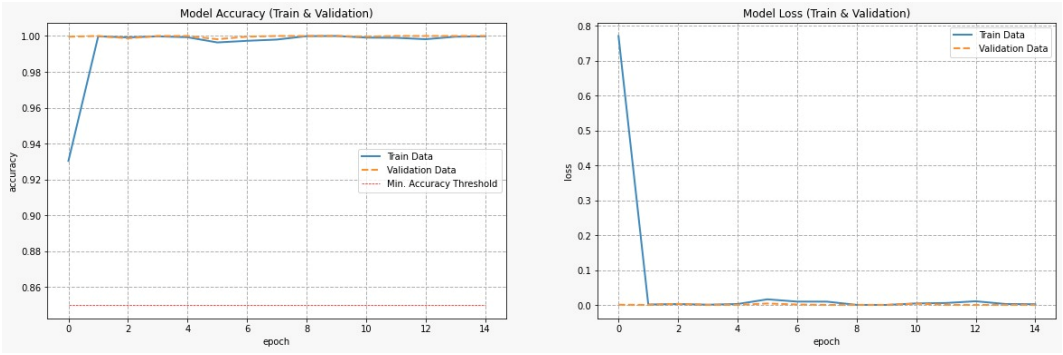
(c)

(d)

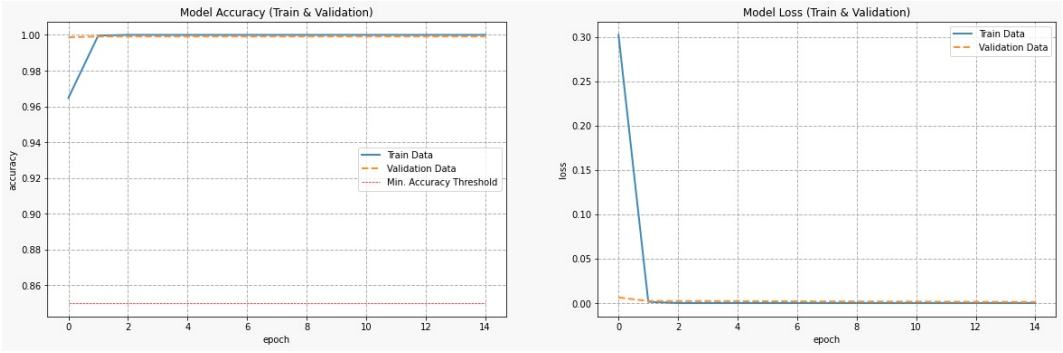
Figure 4. Sample of data augmentation results (**a**) Real Image of NUAA dataset; (**b**) Spoof Image of NUAA dataset; (**c**) Real image of LCC-FASD dataset; (**d**) Spoof image of LCC-FASD dataset.

In Figure 5, it can be seen that the models with and without augmentation show almost identical accuracy results. The training and validation accuracies reach their maximum values in the first few epochs and remain stable until the end of training. This indicates that the NUAA dataset is relatively easier to classify, and augmentation does not significantly change the learning process. Instead, it has a slight negative impact where the accuracy in the first

epoch is around 0.93, while without augmentation, it starts from 0.97. The loss graph also shows the same pattern, namely a rapidly decreasing trend for training and validation data, indicating that the model does not experience overfitting.

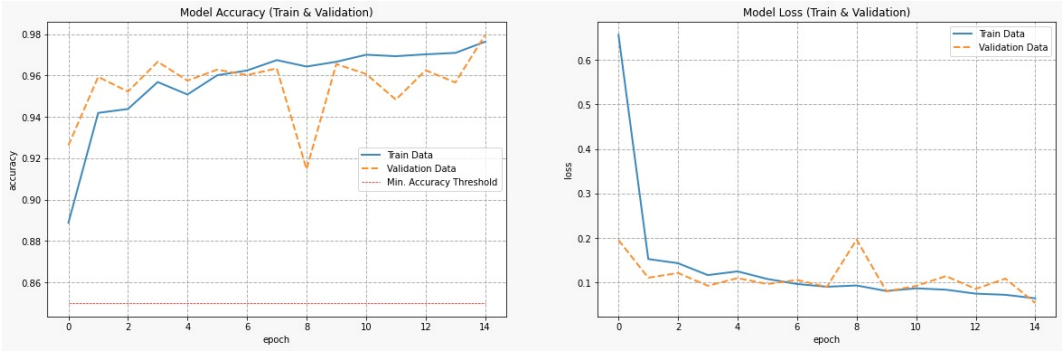


(a)

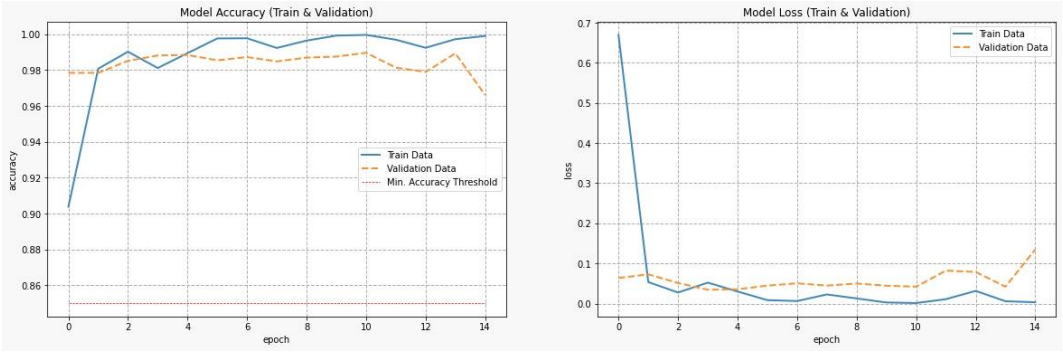


(b)

Figure 5. Training and validation results of NUAA dataset (**a**) with augment; (**b**) without augment.



(a)



(b)

Figure 6. Training and validation results of LCC-FASD dataset (**a**) with augment; (**b**) without augment.

In contrast, Figure 6, which depicts the training and validation results on the LCC-FASD dataset, shows a relatively more complex performance. The model without augmentation shows higher variation in validation accuracy, with relatively large fluctuations in some of the final epochs, although overall, it is better. In contrast, the model with augmentation has a more stable trend, although there is instability at some points. The loss graph also shows that although augmentation helps in learning stability, there are some epochs where the validation loss spikes, which may indicate challenges in model convergence on more complex datasets. Although overall, the results are still better without augmentation. The results of the training and validation evaluation of the proposed method are presented in Figure 5 for the NUAA dataset and Figure 6 for the LCC-FASD dataset, while the testing evaluation is presented in Table 5.

Table 5. Testing result of the proposed method.

Metrics	Without Augmentation		With Augmentation	
	NUAA	LCC-FASD	NUAA	LCC-FASD
Accuracy	1.00	0.99	1.00	0.98
F1-Score	1.00	0.98	1.00	0.95
Precision	1.00	0.98	1.00	0.94
Recall	1.00	0.98	1.00	0.96
HTER	0.000	0.0183	0.000	0.0616
FAR	0.000	0.0053	0.000	0.0072
FRR	0.000	0.0313	0.000	0.1159

Table 5 shows that data augmentation impacts the two datasets used in this study differently. On the NUAA dataset, the model has shown perfect performance both with and without augmentation, with an accuracy of 1.00 and F1-score, precision, and recall values reaching 1.00. In addition, the HTER, FAR, and FRR values remain zero. In contrast to NUAA, the results on the LCC-FASD dataset show that augmentation does not provide any performance improvement. After augmentation, the accuracy decreased slightly from 0.99 to 0.98, while the F1-score, precision, and recall also decreased from 0.98 to 0.95, 0.98 to 0.94, and 0.98 to 0.96, respectively. A more striking improvement is seen in the error metrics, where HTER increases from 0.0183 to 0.0616, FAR increases from 0.0053 to 0.0072, and FRR experiences a significant jump from 0.0313 to 0.1159. This increase in FRR indicates that augmentation causes the model to be more stringent in classifying faces, leading to more errors in recognizing real faces.

This finding confirms that augmentation does not always improve model performance, especially on high-complexity datasets such as LCC-FASD. Augmentation can suboptimally alter the data distribution, making the model overly sensitive to certain features and increasing the likelihood of misclassification. For datasets with screen-based attacks that are more difficult to distinguish from real faces, the augmentation approach needs to be adjusted to avoid disrupting the key characteristics of the training data.

5. Ablation Study and Comparison

An ablation study was conducted to evaluate the impact of data augmentation on model performance and to compare the proposed method with the base models, namely DenseNet201 + Flatten and FaceNet. This evaluation aims to understand how data augmentation contributes to improving model generalization on the NUAA and LCC-FASD datasets. The results in Table 6 show that on the NUAA dataset, the proposed method achieves 100% accuracy, both before and after augmentation, with HTER, FAR, and FRR remaining at 0.000. This indicates that augmentation does not have a significant impact because the model is already optimal without additional synthetic data. In contrast, FaceNet experienced a decrease in accuracy from 0.99 to 0.96 after augmentation, followed by an increase in HTER and FRR, indicating that augmentation reduced its performance. DenseNet also experienced a decrease in performance after augmentation, but the decrease was still relatively minimal.

In Table 7, which compares the performance on LCC-FASD, augmentation has a more complex impact. The proposed method without augmentation shows the best accuracy (0.99),

higher than FaceNet (0.93) and DenseNet (0.96). However, after augmentation, the accuracy of the proposed method decreases to 0.98, while DenseNet increases to 0.97, indicating that augmentation is more beneficial for DenseNet than other models. The HTER and FRR of the proposed method increase sharply after augmentation, indicating the model becomes tighter in classification but more prone to false rejection. FaceNet experiences a decrease in recall and F1-score after augmentation, indicating the model's difficulty in correctly recognizing spoofing attacks. Overall, these results indicate that augmentation does not always improve model performance uniformly.

Table 6. Comparison with base models for NUAA dataset.

Metrics	Without Augmentation			With Augmentation		
	FaceNet	DenseNet	Proposed	FaceNet	DenseNet	Proposed
Accuracy	0.99	1.00	1.00	0.96	0.9992	1.00
F1-Score	0.99	1.00	1.00	0.96	0.9992	1.00
Precision	0.99	1.00	1.00	0.96	0.9992	1.00
Recall	0.99	1.00	1.00	0.97	0.9992	1.00
HTER	0.0099	0.000	0.000	0.0405	0.00065	0.000
FAR	0.0080	0.000	0.000	0.0124	0.000	0.000
FRR	0.0117	0.000	0.000	0.0686	0.0013	0.000

Table 7. Comparison with base models for LCC-FASD dataset.

Metrics	Without Augmentation			With Augmentation		
	FaceNet	DenseNet	Proposed	FaceNet	DenseNet	Proposed
Accuracy	0.93	0.96	0.99	0.92	0.97	0.98
F1-Score	0.75	0.91	0.98	0.73	0.93	0.95
Precision	0.88	0.88	0.98	0.85	0.90	0.94
Recall	0.98	0.96	0.98	0.68	0.97	0.96
HTER	0.1199	0.0360	0.0183	0.1524	0.0347	0.0616
FAR	0.1753	0.0359	0.0053	0.0677	0.0410	0.0072
FRR	0.1199	0.0361	0.0313	0.2371	0.0284	0.1159

Table 8. Comparison with related works.

Dataset	Method	Accuracy (%)	HTER
NUAA Photograph Imposter	Ref [11]	96.74%	0.0312
	Ref [12]	95.1%	0.055
	Ref [13]	97.2%	-
	Ref [14]	99.22%	0.0039
	Ref [15]	91.1%	-
	Ours	100%	0.0000
LCC-FASD	Ref [17]	97%	-
	Ours	99%	0.0183

Comparisons are also made with several previous studies, as presented in Table 8. On the NUAA dataset, the proposed method achieves 100% accuracy with HTER 0.0000, which is better than the methods in Ref[11]–[15], which have 91.1% to 99.22% accuracy. On the LCC-FASD dataset, the proposed method achieves 98% accuracy, which is higher than the study in Ref [17] (97%), indicating that the feature fusion technique applied in this study provides significant advantages in handling more complex datasets. These results confirm that the proposed approach is superior in accuracy performance and more reliable under various dataset conditions, making it a more robust solution for face spoofing detection.

6. Conclusions

This study proposes a face spoofing detection approach based on feature fusion between FaceNet and DenseNet201. Experimental results show that this method achieves perfect accuracy (100%) on the NUAA dataset and 98% on LCC-FASD, outperforming previous methods. The ablation study confirms that data augmentation does not consistently improve model performance. On the NUAA dataset, augmentation does not make a significant difference because the model has achieved optimal performance without additional synthetic data. In contrast, on the LCC-FASD dataset, augmentation causes a slight decrease in accuracy and an increase in HTER, FAR, and FRR values, indicating that augmentation makes the model tighter in classification and increases false rejection errors.

Comparative analysis with baseline models shows that DenseNet201 benefits from data augmentation, while the proposed model performs best without augmentation. FaceNet, which relies more on identity embedding, also experiences a decrease in performance after augmentation, especially on more complex datasets such as LCC-FASD. These findings confirm that augmentation does not always improve model performance uniformly, especially in the case of datasets with higher difficulty levels.

Compared with previous studies, the proposed method successfully outperforms previous approaches in accuracy and lower error rate. Integrating FaceNet and DenseNet201 with a flattening and concatenation-based feature fusion strategy improves the model's robustness in detecting face spoofing attacks. The main contribution of this study includes the Fusion of identity features from FaceNet and texture features from DenseNet201 to improve the detection accuracy. In addition, the ablation study provides additional insights into the impact of augmentation on the model with different dataset complexity levels. These results indicate that the proposed approach offers a more robust and reliable solution for biometric-based security systems. For further research, optimization of the loss function, testing in a broader dataset, and exploration of attention-based feature fusion can be important steps in improving the model's generalization to more complex face spoofing scenarios.

Author Contributions: Conceptualization: LRZ and DRIMS; Methodology: LRZ and DRIMS; Software: LRZ; Validation: all; Formal analysis: all.; Investigation: all; Resources: LRZ; Data curation: LRZ.; Writing—original draft preparation: LRZ; Writing—review and editing: All; Visualization: LRZ and DRIMS; Supervision: DRIMS; Project administration: DRIMS.; Funding acquisition: all. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] M. Prasad, S. Jain, P. Bhanodia, and A. Priya, "Influence of Standalone and Ensemble Classifiers in Face Spoofing Detection using LBP and CNN Models," *Eur. J. Electr. Eng. Comput. Sci.*, vol. 8, no. 2, pp. 17–30, Mar. 2024, doi: 10.24018/ejece.2024.8.2.604.

[2] L. Li, P. L. Correia, and A. Hadid, "Face recognition under spoofing attacks: countermeasures and research directions," *IET Biometrics*, vol. 7, no. 1, pp. 3–14, Jan. 2018, doi: 10.1049/iet-bmt.2017.0089.

[3] I. U. W. Mulyono, D. R. I. M. Setiadi, A. Susanto, E. H. Rachmawanto, A. Fahmi, and Muljono, "Performance Analysis of Face Recognition using Eigenface Approach," in *Proceedings - 2019 International Seminar on Application for Technology of Information and Communication: Industry 4.0: Retrospect, Prospect, and Challenges, iSemantic 2019*, Sep. 2019, pp. 12–16. doi: 10.1109/ISEMANTIC.2019.8884225.

[4] V. Maheswari, C. A. Sari, D. R. I. M. Setiadi, and E. H. Rachmawanto, "Study Analysis of Human Face Recognition using Principal Component Analysis," in *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Sep. 2020, pp. 55–60. doi: 10.1109/iSemantic50169.2020.9234250.

[5] A. Pathirana *et al.*, "A Reinforcement Learning-Based Approach for Promoting Mental Health Using Multimodal Emotion Recognition," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 124–142, Sep. 2024, doi: 10.62411/faith.2024-22.

[6] K. Muhamada *et al.*, "Exploring Machine Learning and Deep Learning Techniques for Occluded Face Recognition : A Comprehensive Survey and Comparative Analysis," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 160–173, Sep. 2024, doi: 10.62411/faith.2024-30.

[7] A. Hadid, "Face Biometrics Under Spoofing Attacks: Vulnerabilities, Countermeasures, Open Issues, and Research Directions," in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2014, no. Cmv, pp. 113–118. doi: 10.1109/CVPRW.2014.22.

[8] J. Komulainen, A. Hadid, M. Pietikainen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," in *2013 International Conference on Biometrics (ICB)*, Jun. 2013, pp. 1–7. doi: 10.1109/ICB.2013.6612968.

[9] D. Timoshenko, K. Simonchik, V. Shutov, P. Zhelezneva, and V. Grishkin, "Large Crowdcollected Facial Anti-Spoofing Dataset," in *2019 Computer Science and Information Technologies (CSIT)*, Sep. 2019, pp. 123–126. doi: 10.1109/CSITechnol.2019.8895208.

[10] J. Komulainen, A. Hadid, and M. Pietikäinen, "Face Spoofing Detection Using Dynamic Texture," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7728 LNCS, no. PART 1, 2013, pp. 146–157. doi: 10.1007/978-3-642-37410-4_13.

[11] R. Ruschel, L. R. Schardosim, and J. Scharcanski, "On the Data Conditioning for Facial Spoofing Attacks Detection using Deep Learning," in *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, May 2019, vol. 2019-May, no. September, pp. 1–6. doi: 10.1109/I2MTC.2019.8827094.

[12] X. Liu, R. Lu, and W. Liu, "Face liveness detection based on enhanced local binary patterns," in *2017 Chinese Automation Congress (CAC)*, Oct. 2017, vol. 2017-Janua, pp. 6301–6305. doi: 10.1109/CAC.2017.8243913.

[13] R. D. Albu, "Face anti-spoofing based on Radon transform," *2015 13th Int. Conf. Eng. Mod. Electr. Syst. EMES 2015*, no. April, pp. 1–4, 2015, doi: 10.1109/EMES.2015.7158435.

[14] M. R. Hasan, S. M. Hasan Mahmud, and X. Y. Li, "Face Anti-Spoofing Using Texture-Based Techniques and Filtering Methods," *J. Phys. Conf. Ser.*, vol. 1229, no. 1, p. 012044, May 2019, doi: 10.1088/1742-6596/1229/1/012044.

[15] S. Kumar *et al.*, "Face Spoofing, Age, Gender and Facial Expression Recognition Using Advance Neural Network Architecture-Based Biometric System," *Sensors*, vol. 22, no. 14, p. 5160, Jul. 2022, doi: 10.3390/s22145160.

[16] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6316 LNCS, no. PART 6, 2010, pp. 504–517. doi: 10.1007/978-3-642-15567-3_37.

[17] P. Chaudhari, N. Kulkarni, and P. More, "AntiSpoofing for Facial Recognition Through Separable Convolution," in *2021 IEEE Pune Section International Conference (PuneCon)*, Dec. 2021, pp. 1–4. doi: 10.1109/PuneCon52575.2021.9686544.

[18] D. Sudha and M. Ramakrishna, "Comparative Study of Features Fusion Techniques," in *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*, Mar. 2017, pp. 235–239. doi: 10.1109/ICRAECT.2017.39.

[19] H. Kibriya, R. Amin, A. H. Alshehri, M. Masood, S. S. Alshamrani, and A. Alshehri, "A Novel and Effective Brain Tumor Classification Model Using Deep Feature Fusion and Famous Machine Learning Classifiers," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–15, Mar. 2022, doi: 10.1155/2022/7897669.

[20] B. Fernando, E. Fromont, D. Muselet, and M. Sebban, "Discriminative feature fusion for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 3434–3441. doi: 10.1109/CVPR.2012.6248084.

[21] J. Yang, J. Y. Yang, D. Zhang, and J. F. Lu, "Feature fusion: Parallel strategy vs. serial strategy," *Pattern Recognit.*, vol. 36, no. 6, pp. 1369–1381, 2003, doi: 10.1016/S0031-3203(02)00262-5.

[22] X. Shang, M. Song, and C. Yu, "Hyperspectral Image Classification With Background," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Jul. 2019, pp. 2714–2717. doi: 10.1109/IGARSS.2019.8898520.

[23] U. Mangai, S. Samanta, S. Das, and P. Chowdhury, "A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification," *IETE Tech. Rev.*, vol. 27, no. 4, p. 293, 2010, doi: 10.4103/0256-4602.64604.

[24] T. Janani and A. Ramanan, "Feature Fusion for Efficient Object Classification Using Deep and Shallow Learning," *Int. J. Mach. Learn. Comput.*, vol. 7, no. 5, pp. 123–127, Oct. 2017, doi: 10.18178/ijmlc.2017.7.5.633.

[25] J. Singh, D. Thakur, T. Gera, B. Shah, T. Abuhmed, and F. Ali, "Classification and Analysis of Android Malware Images Using Feature Fusion Technique," *IEEE Access*, vol. 9, pp. 90102–90117, 2021, doi: 10.1109/ACCESS.2021.3090998.

[26] M. A. Aslam, M. N. Salik, F. Chughtai, N. Ali, S. H. Dar, and T. Khalil, "Image Classification Based on Mid-Level Feature Fusion," in *2019 15th International Conference on Emerging Technologies (ICET)*, Dec. 2019, pp. 1–6. doi: 10.1109/ICET48972.2019.8994721.

[27] S. Zhong *et al.*, "Behavior Prediction for Unmanned Driving Based on Dual Fusions of Feature and Decision," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3687–3696, Jun. 2021, doi: 10.1109/TITS.2020.3037926.

[28] Z. Li, M. Xu, X. Yang, and Y. Han, "Multi-Label Fundus Image Classification Using Attention Mechanisms and Feature Fusion," *Micromachines*, vol. 13, no. 6, p. 947, Jun. 2022, doi: 10.3390/mi13060947.

[29] M. Abdar *et al.*, "BARF: A new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification," *Inf. Sci. (Ny).*, vol. 577, pp. 353–378, 2021, doi: 10.1016/j.ins.2021.07.024.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[31] M. S. Sunarjo, H. Gan, and D. R. I. M. Setiadi, "High-Performance Convolutional Neural Network Model to Identify COVID-19 in Medical Images," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 19–30, Aug. 2023, doi: 10.33633/jcta.v1i1.8936.

[32] A. Susanto, I. U. W. Mulyono, C. A. Sari, E. H. Rachmawanto, D. R. I. M. Setiadi, and M. K. Sarker, "Improved Javanese script recognition using custom model of convolution neural network," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 6, p. 6629, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6629-6636.

[33] 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)*, Jun. 2015, vol. 07-12-June, pp. 815–823. doi: 10.1109/CVPR.2015.7298682.

[34] I. William, D. R. I. M. Setiadi, E. H. Rachmawanto, H. A. Santoso, and C. A. Sari, "Face Recognition using FaceNet (Survey, Performance Test, and Comparison)," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, Oct. 2019, doi: 10.1109/ICIC47613.2019.8985786.

[35] E. Jose, M. Greeshma, T. P. Mithun Haridas, and M. H. Supriya, "Face Recognition based Surveillance System Using FaceNet and MTCNN on Jetson TX2," *2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019*, pp. 608–613, 2019, doi: 10.1109/ICACCS.2019.8728466.

[36] H.-C. Li, Z.-Y. Deng, and H.-H. Chiang, "Lightweight and Resource-Constrained Learning Network for Face Recognition with Performance Optimization," *Sensors*, vol. 20, no. 21, p. 6114, Oct. 2020, doi: 10.3390/s20216114.

[37] W.-C. Cheng, H.-C. Hsiao, Y.-Q. Hong, and D.-Y. Wang, "Masked face recognition based on facenet and genetic algorithm," *Int. J. Appl. Sci. Eng.*, vol. 20, no. 3, pp. 1–10, 2023, doi: 10.6703/IJASE.202309_20(3).005.

[38] S. Peng, H. Huang, W. Chen, L. Zhang, and W. Fang, "More trainable inception-ResNet for face recognition," *Neurocomputing*, vol. 411, pp. 9–19, Oct. 2020, doi: 10.1016/j.neucom.2020.05.022.

[39] Y. Jiang, M. Ye, P. Wang, D. Huang, and X. Lu, "MRF-IUNet: A Multiresolution Fusion Brain Tumor Segmentation Network Based on Improved Inception U-Net," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–8, Aug. 2022, doi: 10.1155/2022/6305748.

[40] Z. Ming, J. Chazalon, M. M. Luqman, M. Visani, and J.-C. Burie, "Simple Triplet Loss Based on Intra/Inter-Class Metric Learning for Face Verification," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2017, vol. 2018-Janua, pp. 1656–1664. doi: 10.1109/ICCVW.2017.194.

[41] M. Tang, Z. Huang, Y. Yuan, C. Wang, and Y. Peng, "A Bounded Scheduling Method for Adaptive Gradient Methods," *Appl. Sci.*, vol. 9, no. 17, p. 3569, Sep. 2019, doi: 10.3390/app9173569.

[42] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 8697–8710. doi: 10.1109/CVPR.2018.00907.

[43] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A Dataset for Recognising Faces across Pose and Age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, May 2018, pp. 67–74. doi: 10.1109/FG.2018.00020.

[44] M. R. Golla and P. Sharma, "Performance Evaluation of Facenet on Low Resolution Face Images," in *Communications in Computer and Information Science*, vol. 839, Springer Singapore, 2019, pp. 317–325. doi: 10.1007/978-981-13-2372-0_28.

[45] H. L. Erik Learned-Miller, Gary B. Huang, Aruni RoyChowdhury, Hua, and Gang, *Advances in Face Detection and Facial Image Analysis*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-25958-1.

[46] Z. Zhou, M. Liu, W. Deng, Y. Wang, and Z. Zhu, "Clothing Image Classification with DenseNet201 Network and Optimized Regularized Random Vector Functional Link," *J. Nat. Fibers*, vol. 20, no. 1, Apr. 2023, doi: 10.1080/15440478.2023.2190188.

[47] K. Adam, I. I. Mohamed, and Y. Ibrahim, "A Selective Mitigation Technique of Soft Errors for DNN Models Used in Healthcare Applications: DenseNet201 Case Study," *IEEE Access*, vol. 9, pp. 65803–65823, 2021, doi: 10.1109/ACCESS.2021.3076716.

[48] S. Kumar, V. P. Singh, S. Pal, and P. Jaiswal, "Energy-efficient model 'DenseNet201 based on deep convolutional neural network' using cloud platform for detection of COVID-19 infected patients," *Epidemiol. Method.*, vol. 12, no. 1, p. 20210047, Jan. 2023, doi: 10.1515/em-2021-0047.

[49] D. Nurfadhillah, G. Santoso, Fatimah, G. Murti Wibowo, Darmini, and Nuryatno, "Effectiveness of Automatic Detection of Osteoarthritis using Convolutional Neural Network (CNN) Method with DenseNet201 on Digital Images of Knee Joint Radiography," *E3S Web Conf.*, vol. 448, p. 02052, Nov. 2023, doi: 10.1051/e3sconf/202344802052.

[50] G. Huang, Z. Liu, G. Pleiss, L. van der Maaten, and K. Q. Weinberger, "Convolutional Networks with Dense Connectivity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8704–8716, Dec. 2022, doi: 10.1109/TPAMI.2019.2918284.

[51] X. Feng, H. Yao, and S. Zhang, "An efficient way to refine DenseNet," *Signal, Image Video Process.*, vol. 13, no. 5, pp. 959–965, Jul. 2019, doi: 10.1007/s11760-019-01433-4.

[52] A. Hess, "Exploring Feature Reuse in DenseNet Architectures," *arXiv*. 2018. [Online]. Available: http://arxiv.org/abs/1806.01935

[53] I. Kreso, J. Krapac, and S. Segvic, "Efficient Ladder-Style DenseNets for Semantic Segmentation of Large Images," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 4951–4961, Aug. 2021, doi: 10.1109/TITS.2020.2984894.

[54] S. A. Samad and J. Gitanjali, "Augmenting DenseNet: Leveraging Multi-Scale Skip Connections for Effective Early-Layer Information Incorporation," *IEEE Access*, vol. 12, no. July, pp. 141344–141360, 2024, doi: 10.1109/ACCESS.2024.3460830.

[55] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2921–2929. doi: 10.1109/CVPR.2016.319.

[56] S.-G. Y. Su-Gyeong Yu, S.-E. K. Su-Gyeong Yu, K. H. S. So-Eui Kim, and E. C. L. Kun Ha Suh, "Effect of Facial Shape Information Reflected on Learned Features in Face Spoofing Detection," 網際網路技術學刊, vol. 23, no. 3, pp. 517–525, May 2022, doi: 10.53106/160792642022052303010.

[57] S.-G. Yu, S.-E. Kim, K. H. Suh, and E. C. Lee, "Face Spoofing Detection Using DenseNet," in *International Conference on Intelligent Human Computer Interaction*, 2021, pp. 229–238. doi: 10.1007/978-3-030-68452-5_24.

[58] D. Wang and X. Chen, "Research on Feature Fusion Method Based on Graph Convolutional Networks," *Appl. Sci.*, vol. 14, no. 13, p. 5612, Jun. 2024, doi: 10.3390/app14135612.

[59] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional Feature Fusion," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2021, pp. 3559–3568. doi: 10.1109/WACV48630.2021.00360.

[60] A. Ahmed and S. Mohamed, "Implementation of early and late fusion methods for content-based image retrieval," *Int. J. Adv. Appl. Sci.*, vol. 8, no. 7, pp. 97–105, Jul. 2021, doi: 10.21833/ijaas.2021.07.012.

[61] Y. Wei, D. Wu, and J. Terpenny, "Decision-Level Data Fusion in Quality Control and Predictive Maintenance," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 1, pp. 184–194, Jan. 2021, doi: 10.1109/TASE.2020.2964998.

[62] K. Hu, L. Li, X. Tao, J. D. Velásquez, and P. Delaney, "Information fusion in crime event analysis: A decade survey on data, features and models," *Inf. Fusion*, vol. 100, no. May, p. 101904, Dec. 2023, doi: 10.1016/j.inffus.2023.101904.

[63] S. Zhai, D. Shang, S. Wang, and S. Dong, "DF-SSD: An Improved SSD Object Detection Algorithm Based on DenseNet and Feature Fusion," *IEEE Access*, vol. 8, pp. 24344–24357, 2020, doi: 10.1109/ACCESS.2020.2971026.

[64] Y. Zhou, Y. Liu, G. Han, and Z. Zhang, "Face Recognition Based on Global and Local Feature Fusion," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec. 2019, pp. 2771–2775. doi: 10.1109/SSCI44817.2019.9003045.

[65] K. P. Zaw and A. Mon, "Enhanced Multi-Class Skin Lesion Classification of Dermoscopic Images Using an Ensemble of Deep Learning Models," *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 256–267, Nov. 2024, doi: 10.62411/jcta.11530.

[66] H. T. Adityawan, O. Farroq, S. Santosa, H. M. M. Islam, M. K. Sarker, and D. R. I. M. Setiadi, "Butterflies Recognition using Enhanced Transfer Learning and Data Augmentation," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 115–128, Nov. 2023, doi: 10.33633/jcta.v1i2.9443.

[67] M. A. Hambali and P. A. Agwu, "Adversarial Convolutional Neural Network for Predicting Blood Clot Ischemic Stroke," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 51–64, Jun. 2024, doi: 10.62411/jcta.10516.

[68] F. M. Firnando, D. R. I. M. Setiadi, A. R. Muslikh, and S. W. Irianandа, "Analyzing InceptionV3 and InceptionResNetV2 with Data Augmentation for Rice Leaf Disease Classification," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 1–11, May 2024, doi: 10.62411/faith.2024-4.

[69] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, Dec. 2022, doi: 10.1016/j.array.2022.100258.

[70] R. Cai, C. Soh, Z. Yu, H. Li, W. Yang, and A. C. Kot, "Towards Data-Centric Face Anti-spoofing: Improving Cross-Domain Generalization via Physics-Based Data Synthesis," *Int. J. Comput. Vis.*, Oct. 2024, doi: 10.1007/s11263-024-02240-2.

[71] P. Kirichenko *et al.*, "Understanding the Detrimental Class-level Effects of Data Augmentation," in *37th Conference on Neural Information Processing Systems (NeurIPS 2023)*, 2023, vol. 36. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/38c05a5410a6ab7eeeb26c9dbebbc41b-Paper-Conference.pdf

[72] A. Bakshi, S. Gupta, A. Gupta, S. Tanwar, and K. Hsiao, "3T-FASDM: Linear discriminant analysis-based three-tier face anti-spoofing detection model using support vector machine," *Int. J. Commun. Syst.*, vol. 33, no. 12, Aug. 2020, doi: 10.1002/dac.4441.

[73] L. Feng *et al.*, "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 451–460, Jul. 2016, doi: 10.1016/j.jvcir.2016.03.019.

[74] H. Kim, J. Lee, Y. Jeong, H. Jang, and Y. Yoo, "Advancing Cross-Domain Generalizability in Face Anti-Spoofing: Insights, Design, and Metrics," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2024, pp. 970–979. doi: 10.1109/CVPRW63382.2024.00102.

[75] F. Mazhar, N. Aslam, A. Naeem, H. Ahmad, M. Fuzail, and M. Imran, "Enhanced Diagnosis of Skin Cancer from Dermoscopic Images Using Alignment Optimized Convolutional Neural Networks and Grey Wolf Optimization," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 368–382, Jan. 2025, doi: 10.62411/jcta.11954.

[76] M. Ha, "Top-Heavy CapsNets Based on Spatiotemporal Non-Local for Action Recognition," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 39–50, May 2024, doi: 10.62411/jcta.10551.