International Journal of Innovative Research and Scientific Studies, 8(3) 2025, pages: 105-114

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

An ensemble model for improving the accuracy and security of biometric identification

Lyailya Cherikbayeva[1], Matkerim Bazargul[1], Dauren Darkenbayev[1*], Nurbolat Tasbolatuly[2,3], Zhanetta Kalaubekova[1]

[1]*Al-Farabi Kazakh National University, Almaty, Kazakhstan.*
[2]*Higher School of Information Technology and Engineering, Astana International University, Astana, Kazakhstan.*
[3]*Department of Computer Engineering, Astana IT University, Astana, Republic of Kazakhstan.*

Corresponding author: Dauren Darkenbayev (*Email: dauren.darkenbayev1@gmail.com*)

Abstract

Face recognition is a key area in computer vision and artificial intelligence. With the advent of deep learning, novel methods have been developed that achieve high accuracy in this field. This study aims to enhance facial recognition accuracy by comparing two state-of-the-art algorithms – DeepFace and FaceNet – and proposing an ensemble approach that integrates their strengths. To this end, we conduct a comparative analysis of DeepFace's deep convolutional neural network–based identification and FaceNet's multidimensional vector embedding, then combine their output probabilities into a single ensemble model. Experimental evaluation on publicly available datasets under varying lighting conditions and head poses reveals that the ensemble consistently outperforms each individual algorithm, demonstrating superior accuracy and robustness to external factors. We conclude that this hybrid ensembling strategy significantly improves recognition performance, validating its potential for complex face matching tasks. These findings indicate that real-world applications – such as surveillance, identity verification, and biometric authentication – can benefit from adopting ensemble methods to achieve higher accuracy and resilience.

Keywords: Biometric authentication, Convolutional neural networks, Deep learning, DeepFace, Ensemble method, Face identification, FaceNet, Recognition accuracy.

DOI: 10.53894/ijirss.v8i3.6439
Funding: This study received no specific financial support.
History: Received: 14 March 2025 / **Revised:** 17 April 2025 / **Accepted:** 21 April 2025 / **Published:** 24 April 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Biometric identification is a fundamental component of security systems aimed at protecting assets. This technology is based on the analysis of unique physiological or behavioral characteristics of individuals, such as iris patterns and facial

features. In recent years, face recognition has emerged as one of the most sought-after methods for biometric identification, finding applications in access control, video surveillance, and authentication.

Modern face recognition methods primarily rely on deep convolutional neural networks (CNNs), which deliver high identification accuracy. Among these, DeepFace and FaceNet are recognized as some of the most effective solutions. However, despite their successes, each model exhibits certain limitations [1, 2]. DeepFace may be sensitive to variations in background and subtle facial details, whereas FaceNet demands significant computational resources. This scenario underlines the need for a combined approach that leverages the strengths of both models while mitigating their weaknesses [3].

A notable challenge for face recognition systems is their vulnerability to spoofing attacks. Malicious actors may use printed photographs, videos, or digital manipulations (such as Deepfake) to circumvent biometric security mechanisms [4]. To enhance system reliability, this work proposes the integration of an anti-spoofing module based on MobileNetV2. Instead of merely rejecting suspicious inputs, the module computes a spoofing probability that is fed into the ensemble model, thereby allowing the authenticity degree of the input data to influence the final decision-making process.

This paper proposes an ensemble method that combines the outputs of DeepFace and FaceNet to enhance face recognition accuracy. The study hypothesizes that ensemble learning with these two models will yield a 1–2% improvement in recognition accuracy over each individual model. To validate this hypothesis, a comparative analysis will be conducted, focusing on accuracy, robustness to external factors, and overall model performance.

The primary goal of this study is to determine whether the ensemble of DeepFace and FaceNet can improve the reliability of face recognition under real-world conditions, thus increasing the system's resilience to factors such as lighting variations, changes in viewing angle, and fluctuations in image quality. Key aspects of the research include:

1. Examination of facial recognition features, including the analysis of unique facial characteristics, geometric structure, skin texture, and dynamic expression variations.
2. Evaluation of current biometric identification methods, with an emphasis on deep learning techniques, biometric sensors, and sample matching systems relevant to the field.
3. Investigation of biometric identification's role in enhancing security measures, encompassing applications in physical building security, access control for information systems, and public safety in critical institutions and public spaces.
4. Formulation of a new architecture designed to optimize the efficacy of biometric identification systems in security applications [4].

The overarching aim of this paper is to review existing biometric identification methods and to develop an innovative architecture that enhances their overall efficiency.

2. Materials and Methods

Face recognition technologies have undergone rapid evolution, transitioning from classical methods to modern deep learning-based approaches. Although ensemble methods are widely applied in computer vision, their utilization for integrating Deep Face and FaceNet has not been fully explored. Early algorithms, such as Eigenfaces and Fisherfaces, rely on linear image analysis techniques-including Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [1]. While these approaches delivered satisfactory results under controlled conditions, they experienced significant declines in accuracy when confronted with variations in pose, lighting, and facial expressions [5].

2.1. Ensemble Methods in Face Recognition

Ensembling is widely used to improve the accuracy and reliability of deep learning models. This technique is applied across various computer vision tasks, including image classification [6, 7], object detection, and segmentation [8, 9]. The main ensembling techniques include prediction averaging, which combines the outputs of multiple models by averaging their probabilities [10-12] majority voting, where the final decision is made based on the majority of the models' predictions [13]; and weighted probability fusion, which takes into account each model's accuracy to reduce the impact of weaker predictions [6]. In face recognition, ensembling has mostly been used to combine different architectures of the same model. However, there are a number of studies exploring the combination of two distinct models.

- Ensemble method combining CNN and SVM – a study where a convolutional neural network (CNN) is used together with a support vector machine (SVM) to enhance face recognition accuracy [14].
- Hybrid face recognition system – a combination of FaceNet and a K-Nearest Neighbors (KNN) classifier to improve classification performance [3, 15].
- Combination of DeepFace and VGG-Face – an approach that merges the embeddings of both models at the feature level [16].
- Dual CNN model with feature-level fusion – an ensemble of ResNet and MobileNet architectures designed to improve robustness in challenging conditions [2, 17].

Despite existing ensemble approaches, the proposed FaceZ method has key distinctions. Unlike the fusion of DeepFace and VGG-Face, which relies on simple embedding merging, FaceZ applies embedding concatenation followed by Meta-Learner training. This allows the model not only to preserve the unique features of each method but also to adapt to varying lighting conditions and head poses. Additionally, unlike hybrid methods using KNN, FaceZ employs full deep neural network training on the combined representations, which enables higher accuracy and robustness.

The aim of this study is to assess the effectiveness and applicability of biometric identification methods, specifically DeepFace, FaceNet, and their ensemble. The research investigates whether ensembling can improve face recognition accuracy and enhance model robustness against external factors. To achieve this aim, the following objectives must be addressed.

- Conduct a study of biometric identification methods: review the main approaches to personal identification, including both classical techniques and modern deep learning-based methods.
- Analyze DeepFace and FaceNet: examine their core working principles, characteristics, and evaluate their accuracy and robustness in biometric identification tasks.
- Assess applicability under various conditions: investigate how suitable DeepFace, FaceNet, and their ensemble are for use in diverse security environments, including varying lighting and camera angles.
- Develop and test an ensemble method combining DeepFace and FaceNet to improve recognition accuracy and minimize the limitations of individual models.
- Compare the three methods: DeepFace, FaceNet, and their ensemble, through a comparative analysis of their accuracy, performance, and robustness.
- Based on the research findings, draw conclusions regarding the applicability and effectiveness of the ensemble approach for real-world biometric identification tasks to enhance system reliability.

For face recognition, we consider three methods: DeepFace, FaceNet, and an ensemble method combining both. The first two methods are based on deep neural networks and provide high recognition accuracy, but they are built on different principles. DeepFace is more robust to variations in lighting and head pose, while FaceNet effectively generates vector representations of faces optimized for accurate comparison.

Below, we examine the architecture, key components, and operational principles of each method in detail. Understanding these differences will support the justification for using an ensemble approach to improve face recognition accuracy.

2.2. DeepFace

The DeepFace method is one of the first successful deep learning models applied to face recognition. Developed by Facebook, this model achieved an accuracy of 97.47% on the Labeled Faces in the Wild (LFW) dataset, which is comparable to human-level performance (97.65%).

A key feature of DeepFace is the use of a deep convolutional neural network (CNN) combined with a 3D face alignment pre-processing step. This significantly improves robustness to changes in pose, lighting, and facial expression. The model takes an RGB image of size 152×152 pixels as input. It applies 3D alignment to convert the face to a frontal view, reducing the effects of head tilts and rotations [15].

Figure 1 shows the architectural scheme of DeepFace. Based on the adjusted input data, integration and alignment are performed through the appropriate interface. This is followed by three locally connected layers and two fully connected layers. The color scheme in the diagram represents the feature maps generated at each processing level.

Convolutional layers are used to extract features, highlighting low-level facial characteristics such as edges, textures, and contours. A key feature of the model is the use of locally connected layers, which enable more precise processing of specific facial areas such as the eyes, nose, and mouth. In the final layers of the neural network, fully connected layers are employed to generalize the extracted features for identity recognition.

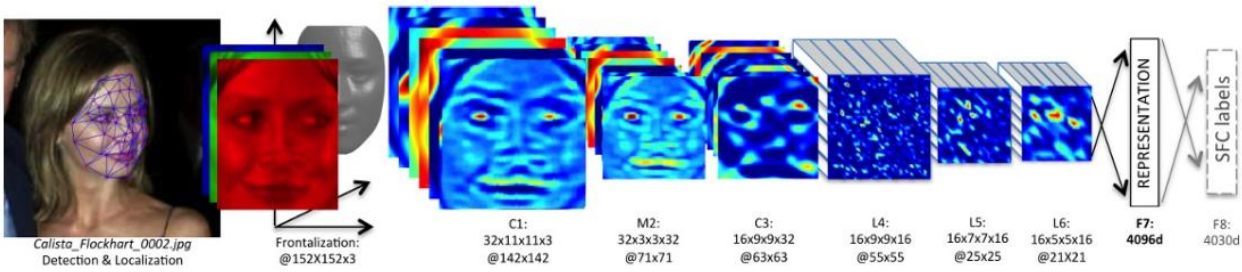

Figure 1.
Architectural scheme of DeepFace.

The model is trained using the softmax function, which provides classification among a large number of faces.

The architecture of DeepFace achieves high accuracy due to 3D alignment and the use of locally connected layers, allowing the model to be robust to changes in lighting and pose. However, the method remains sensitive to background variations and fine facial details, and it requires substantial computational resources when processing large datasets. DeepFace was one of the first models to demonstrate the effectiveness of deep neural networks in facial recognition. Its architectural principles formed the foundation for subsequent advancements such as FaceNet, which uses face embeddings to enable more accurate image matching.

2.3. FaceNet

FaceNet was trained on a dataset containing between 100 and 200 million face images, ranging in size from 96×96 to 224×224 pixels, and representing approximately 8 million distinct identities [2]. FaceNet generates 128-dimensional vector representations (embeddings) for each face by utilizing a triplet loss function. This function trains the model to minimize the distance between embeddings of the same person while maximizing the distance between embeddings of different individuals. This approach enables effective clustering of faces in a multidimensional space, ensuring high recognition accuracy.

Several approaches have been developed based on FaceNet to enhance performance while maintaining identification accuracy. For example, to improve FaceNet's performance, a deep compression algorithm was introduced, using a

convolutional network architecture and a softmax loss function. Figure 2 shows the architectural scheme of FaceNet.This allowed the model to achieve 93.5% accuracy on the 250rf dataset. In a newer version of FaceNet, intensive steps were added to enhance texture structure, resulting in a recognition accuracy of 99.07% on the LFW dataset.

FaceNet transforms its input data directly into a 128-dimensional vector representation using a triplet loss function based on Large Margin Nearest Neighbor (LMNN). The triplet loss function consists of two similar face samples (a positive pair) and one dissimilar sample (a negative pair). It aims to separate the positive pair from the negative one by maximizing the difference in distances between them.

Figure 2.
FaceNet model architecture.

2.4. Triplet Loss

The Triplet Loss is considered suitable for face verification because it teaches the network to increase the distance between images of different people and decrease it between images of the same person. Thus, the goal of triplet loss is to ensure that the distance between images of the same person is at least smaller than the distance between an image of the same person and another person.

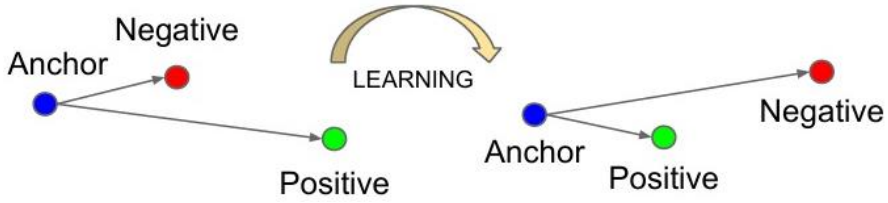

Figure 3.
Triplet Loss: Minimizing Anchor-Positive Distance, Maximizing Anchor-Negative Distance

Triplet loss minimizes the distance between the anchor and the positive, both having the same identity, and maximizes the distance between the anchor and the negative with a different identity.

The embedding is represented as $(x) \in R^d$. It maps the image x into a d -dimensional Euclidean space. Additionally, we constrain this embedding by ensuring that it lives on a d -dimensional hypersphere, i.e., $||f(x)_2|| = 1$. Here, we want to ensure that the image x_i^a (anchor) of a specific person is closer to all other images x_i^p (positive) of the same person than to any image x_i^n (negative) of any other person. This is clearly shown in Figure 3.

$$||f(x_i^a) - f(x_i^p)||_2^2 + a < ||f(x_i^a) - f(x_i^n)||_2^2$$
$$\forall \left(f(x_i^a) f(x_i^p) f(x_i^n)\right) \in T \quad (1)$$

Where a a is the distance difference introduced between positive and negative pairs. T is the set of all possible triplets in the training set, with a cardinality of N.

$$L = \sum_{i=1}^{N} [||f(x_i^a) - f(x_i^p)||_2^2 - ||f(x_i^a) f(x_i^n)||_2^2 + \alpha] \quad (2)$$

Generating all possible triplets will lead to a set of triplets that are easy to satisfy (i.e., fulfilling the constraint in the equation). These triplets will not contribute to the learning process and will slow down convergence, as they will still be passed through the network. It is crucial to select hard triplets that are active and can help improve the model.

2.5. Anti-Spoofing Methods for Face Recognition Attacks

Anti-spoofing in face recognition systems refers to a methodology aimed at defending against identity substitution attacks, designed to detect attempts to use counterfeit images, videos, or digital manipulations. Modern anti-spoofing approaches are based on various principles, including the analysis of facial texture characteristics, depth estimation, infrared scanning, and micro-movement detection.

One of the most commonly used anti-spoofing methods is the analysis of skin texture, which helps identify distinctive features that differentiate a live face from printed images or screen projections. One such method is the use of Local Binary Patterns (LBP), which analyze the microtextures of the face by decomposing the image into local fragments and then encoding their structure into binary features. This method effectively detects anomalies arising in artificial images but may exhibit low robustness to changes in lighting and image quality. Additionally, Gabor filters can be used, which analyze the frequency-space characteristics of the image, detecting subtle changes in skin texture. The advantage of this approach lies in its ability to extract hidden patterns that are difficult to replicate in printed photos.

Modern anti-spoofing systems can combine multiple methods simultaneously, leveraging the advantages of each. One of the most effective solutions is the use of the deep neural network model MobileNetV2, which is trained on a large number of authentic face images and counterfeits, highlighting the most significant features that distinguish a real face from a forged one. Unlike traditional methods based on manual features, MobileNetV2 automatically extracts and analyzes complex spatial dependencies, enhancing the model's robustness to varying lighting conditions and image quality.

The integration of anti-spoofing into the face recognition system can be done in various ways. One approach involves a preliminary check of the image before it is passed to the main recognition model. In this case, the system blocks any images that are classified as fakes, reducing the likelihood of a successful attack. However, this method may result in false rejections of real faces if the anti-spoofing model lacks sufficient accuracy. A more effective solution is to pass the probability of spoofing to the ensemble face recognition system. In this case, the authenticity probability becomes one of the factors considered in the final classification, allowing the system to more flexibly adapt to various conditions.

The application of anti-spoofing is a crucial part of modern biometric identification systems, significantly enhancing their security. The optimization of anti-spoofing methods includes the use of pre-trained neural networks, increasing the resolution of input data, and the combined application of different approaches, such as infrared analysis and motion detection. Further research may focus on integrating multispectral methods, using hybrid neural network architectures, and developing algorithms that are resistant to new types of attacks, including high-quality digital manipulations and deepfakes. After discussing these components, we will describe the proposed ensemble model, FaceZ.

3. Results

In face recognition tasks where high accuracy is required, the ensemble of FaceNet and DeepFace methods can significantly improve the system's effectiveness. Both approaches are based on deep neural networks that provide high accuracy in biometric identification tasks.

For training and testing the ensemble model, a combined dataset was used, including images from LFW (Labelled Faces in the Wild) and a self-collected sample (Figure 4). The self-collected part of the dataset consisted of 5,970 images of different faces. All images were preprocessed, normalized, and augmented (reflection, slight rotations, and brightness adjustments) to enhance the model's robustness.

Figure 4.
Images of people used in the project.

Ensembling allows combining their advantages.

- Improved accuracy by combining differences in architectures.
- Robustness to errors of one of the models.
- Improved results on heterogeneous data.

3.1. Description of the Ensembling Architecture

The main idea of ensembling is to combine the results of several models to improve overall prediction accuracy.

The system receives pre-normalized and scaled RGB images of size 224×224 pixels as input. These images are simultaneously passed through two independent neural network models, each of which analyzes the image and generates predictions. Each of the two architectures receives the input image x and computes the probabilities of belonging to different classes.

$$P_{FaceNet1} = FaceNet_1(x), P_{FaceNet2} = FaceNet_2(x), P_{DeepFace1} = Deepface_1(x).$$

Here, P is the probability matrix, where each element $P_{i,j}$represents the probability of the input image $x_i$belonging to class j.

The predictions of the FaceNet and DeepFace models are combined at the probability level.

$$\begin{gathered} P_{FaceNet} = P_{FaceNet1} + P_{FaceNet2} \\ P_{DeepFace} = P_{DeepFace1} + P_{DeepFace2} \\ P_{ensemble} = P_{FaceNet} + P_{DeepFace} \end{gathered} \quad (3)$$

However, simple summing of probabilities can blur the boundaries between classes, so Softmax with temperature scaling is applied:

$$P_{final} = Softmax\left(\frac{P_{ensemble}}{T}\right) \quad (4)$$

where T is the temperature parameter that reduces the uncertainty of predictions. Instead of simple argmax, as in classical ensembling, a Meta-Learner (MLP) is used, which is trained on the predictions from FaceNet and DeepFace.

A new training set is created from the predicted classes.

$$X_{meta} = [\arg max\,(P_{FaceNet}), \arg max\,(P_{DeepFace})] \quad (5)$$

The Meta-Learner is a multi-layer perceptron (MLP).

$$y_{pred} = \arg max\,(Softmax(W_2 * ReLU(W_1 * X + b_1) + b_2)) \quad (6)$$

where $W_1, W_2$are the weights of the network, and X represents the input features (the predictions from FaceNet and DeepFace). The Meta-Learner is trained using the formula.

$$Meta = MLP.fit(X_{meta}, y_{train})$$
$$y_{final} = Meta.predict(X_{meta}) \quad (7)$$

The advantage of this approach is the adaptive determination of the importance of each prediction. The interaction between the models significantly increases the recognition accuracy (see Figure 5), as each model complements the results of the other.

The proposed ensemble method demonstrates high efficiency in face recognition by combining the predictions of FaceNet and DeepFace. Each model independently analyzes the image, extracting key features. Their predictions are then passed to the Meta-Learner, which processes the results of the base models and selects the final class with the highest probability. In this way, the functional advantages of FaceNet and DeepFace are combined, enhancing the accuracy and stability of the system. The ensemble model combines the results of the separately trained neural networks, FaceNet and DeepFace, by concatenating their embeddings.

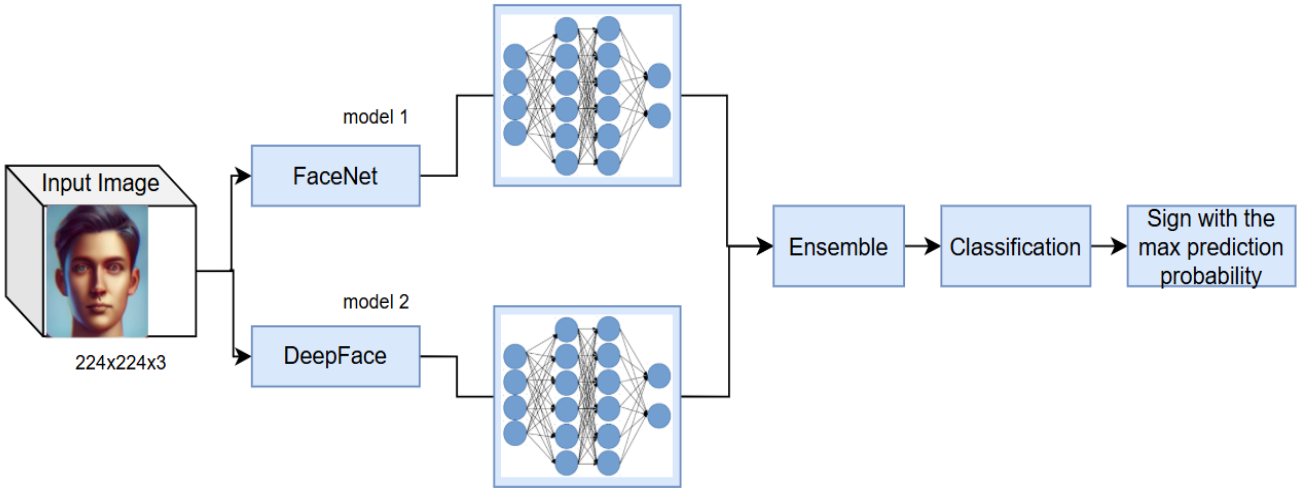

Figure 5.
Architecture of the ensemble model.

The final prediction is made using soft voting, and the class with the highest probability is selected as the final result. The proposed ensemble method is described in Algorithm 1.

Algorithm 1 Pseudo code for the proposed method

```
Class FaceZEnsembleModule(Module)
procedure init(modelA, modelB)
init()
    classifier ← linear transformation(128 * 2, num_classes)
end procedure

procedure forward(x):
    x1 ← modelA(x)
    x2 ← modelB(x)
    x ← concatenate((x1, x2), dim=1)
    final_out ← softmax(x)
returnfinal_out
endprocedure
endclass

ensemble ← FaceZEnsembleModule(FaceNet, DeepFace)

for param in ensemble model parameters():
   param.requires_grad ← False
```

```
    endfor

    for param in ensemble model classifier parameters():
        param.requires_grad ← True
    endfor
    ensemble_training_results=training(ensemble_model, 300epoch)
```

The parameters of the neural networks in the ensemble method can vary depending on the specific implementation and model requirements. Table 2 presents the parameters of the proposed ensemble method based on FaceNet and DeepFace. The Adam optimizer has been chosen because it effectively adjusts the weights and biases of the neural network during training, aiding in the minimization of the loss function and accelerating the model's convergence.

Table 1.
Characteristics of the ensemble model.

Parameters	Ensemble
Number of layers	72 layers of FaceNet + 10 layers of DeepFace
Number of neuron sinfully connected layers	512
Activation function	ReLu
Optimizer	Adam
Activation function	ReLu
Batch size	32
Loss function	Cross Entropy
Number of epochs	300

Figure 6 shows the values of Precision, Recall, and F1-score for three models: FaceNet, DeepFace, and the proposed ensemble method FaceZ. It can be seen that the ensemble model demonstrates the highest scores, outperforming the base models across all metrics. This confirms that combining models enhances the accuracy and reliability of face recognition.

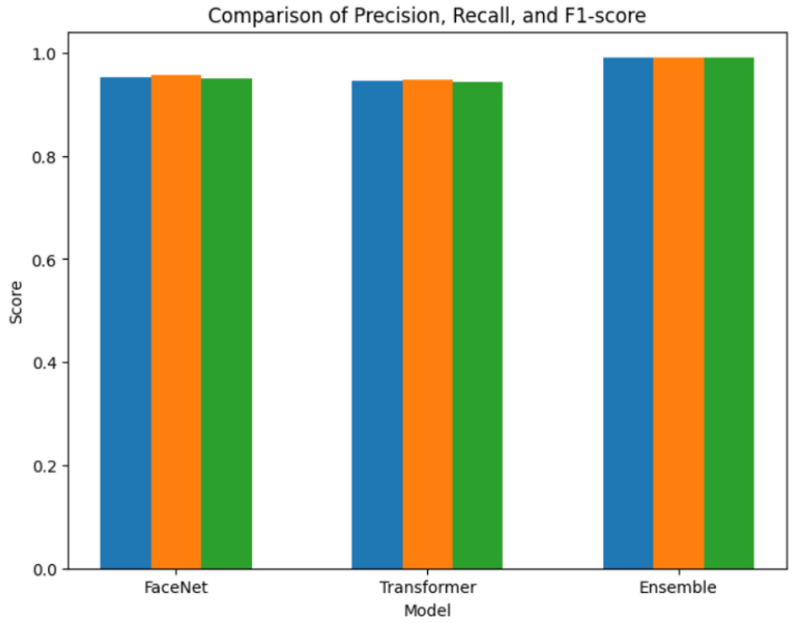

Figure 6.
Comparison of Precision, Recall, and F1-score metrics for FaceNet, DeepFace, and the ensemble model.

Table 2.
Characteristics of the ensemble model.

Model	Accuracy	Precision	Recall	F1
FaceNet	99%	0.95	0.93	0.95
DeepFace	98%	0.92	0.9	0.91
Ensemble	99.6%	0.98	0.97	0.97

If the accuracy of FaceNet reached 97% in the first iteration, the model demonstrated stability in subsequent training stages, maintaining high accuracy and F1-score values. However, it may still produce errors under varying lighting conditions and face angles.

Experimental results showed that the ensemble of FaceNet and DeepFace improves recognition accuracy. The Meta-Learner enabled the combination of their predictions, while the integration of anti-spoofing techniques reduced the number of errors caused by attacks.

The proposed ensemble method (FaceZ) outperformed both individual models across all metrics, achieving 99.6% accuracy, 98% precision, and 97% recall. This proves that the combination of models compensates for their weaknesses, delivering maximum accuracy and reliability in face recognition. The ensemble also demonstrated a high F1-score of 98%, confirming a balanced trade-off between precision and recall. Thus, the ensemble not only improved accuracy but also enhanced the security of the system.

3.2. Integration of Anti-Spoofing into the Ensemble Method

Anti-spoofing plays a key role in biometric identification systems by preventing face spoofing attacks. However, using it solely as a filtering mechanism may reduce the overall effectiveness of the system if genuine faces are mistakenly blocked. This study proposes integrating the anti-spoofing probability into the face recognition ensemble model, allowing the system to more flexibly account for the likelihood of spoofing.

The integration of anti-spoofing into the face recognition process is based on modifying the input vector of the ensemble method. Previously, the ensemble considered only the predictions of FaceNet and DeepFace. Now, an additional parameter is introduced—the face authenticity probability, calculated using MobileNetV2.

Each model provides its own prediction of the probability that a face belongs to a certain class.

$$\text{FaceNet: } P_{FaceNet} = f_{FaceNet}(X)$$
$$\text{DeepFace: } P_{DeepFace} = f_{DeepFace}(X) \quad (8)$$
$$\text{Anti-spoofing (MobileNetV2): } P_{spoof} = f_{MobilNetV2}(X)$$

Where P_{spoof} is the probability of spoofing (the higher the value, the more likely the face is fake).

Thus, the final prediction of the model is determined by the following formula.

$$X_{meta} = \left[P_{FaceNet}, P_{DeepFace}, 1 - P_{spoof}\right] \quad (9)$$

Here, $1-P_{spoof}$ is used to ensure that a higher probability of face authenticity increases the overall score in the ensemble. The final prediction is defined as.

$$P_{final} = g(W, X_{meta}) \quad (10)$$

Where W represents the weights of the ensemble classifier, and $g(W, X_{meta})$ is the function executed by the Meta-Learner.

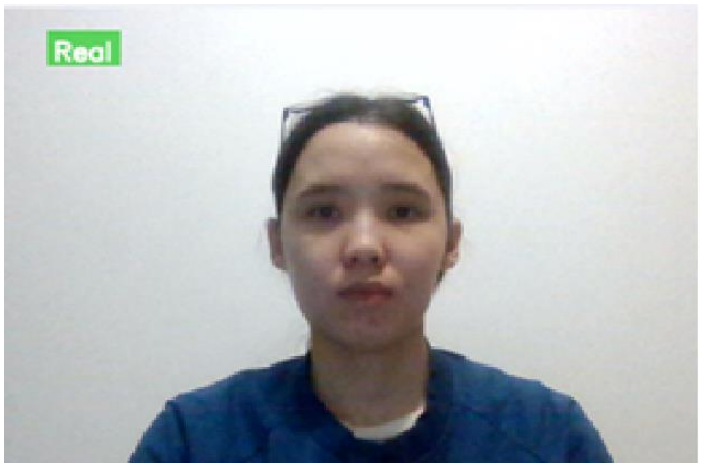

Figure 7.
Anti-spoofing result: a real face detected.

Figure 7 shows the result of the anti-spoofing system, demonstrating the successful verification of face authenticity. The system processed an image captured in real time and classified it as "Real," which is confirmed by the green marker in the top-left corner. This indicates that the algorithm correctly identified a live face, distinguishing it from potential attacks such as the use of photos or videos.

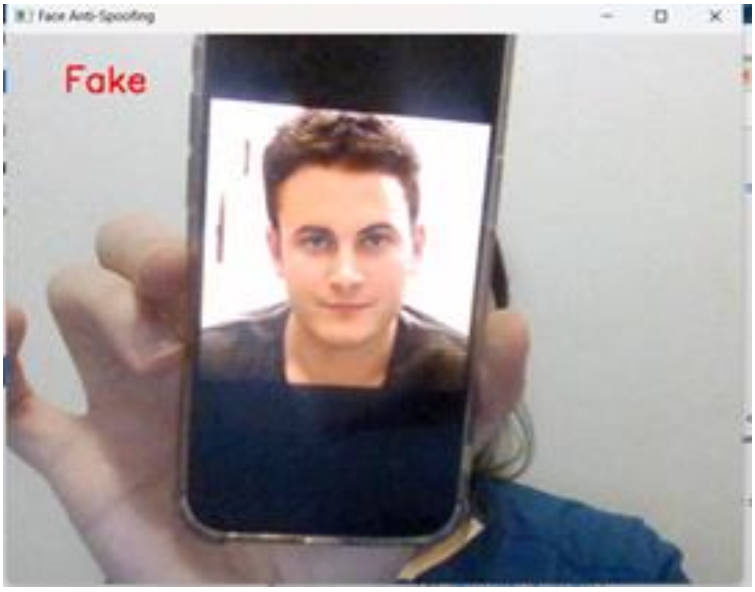

Figure 8.
Anti-spoofing result: spoof detected.

Figure 8 illustrates the result of the anti-spoofing module integrated into the proposed ensemble method, FaceZ. A person is showing a photo of a face on a smartphone screen to the camera, simulating an identity spoofing attempt. However, the system correctly classifies this as a fake face, marking it with a red "Fake" label in the top-left corner of the image.

The integration of anti-spoofing into the ensemble model has significantly impacted the overall accuracy of the face recognition system. The primary goal of anti-spoofing is to detect fake images (such as printed photos, smartphone screens, or Deepfakes) and prevent their false authentication.

A comparative analysis of the ensemble model with and without anti-spoofing revealed that including anti-spoofing improves the reliability of the system, although it may slightly reduce overall accuracy due to false negatives under challenging conditions. Optimizing anti-spoofing parameters, such as the classification probability threshold, as well as adapting the model to various lighting and capture conditions, are important areas for future research.

The integration of anti-spoofing into the ensemble not only enhanced the security of the system but also reduced the number of false positives that could occur with strict image filtering. As a result, higher accuracy was achieved in real-time conditions, since the model no longer simply blocks faces but instead analyzes their authenticity probability during the recognition process.

Summarizing the obtained results and the comparison of methods, the main conclusions of the study are formulated below.

4. Conclusion

This paper examined two leading deep learning algorithms for face recognition - DeepFace and FaceNet-as well as the proposed ensemble method with integrated anti-spoofing. All three approaches possess unique characteristics and advantages, allowing for effective face identification under various conditions.

The analysis of individual models showed that FaceNet outperforms DeepFace in terms of accuracy (99% vs. 95%), while DeepFace demonstrates better robustness to changes in lighting and head orientation. FaceNet's ability to represent faces as vectors in a multidimensional space ensures high accuracy and stability in identification tasks. Meanwhile, DeepFace, thanks to its architecture, proves effective in applications requiring adaptability to diverse image types and challenging capture conditions.

The developed ensemble method, combining the strengths of both models along with the MobileNetV2 anti-spoofing module, demonstrated the best results. Integrating spoof probability into the classification process improved system security and reduced errors caused by spoofing attacks. The final accuracy of the proposed model reached 99.6%, with 98% precision and 97% recall, confirming the effectiveness of the approach.

However, despite the high accuracy, the proposed ensemble method requires increased computational resources, which may affect processing speed in real-time scenarios. Future work will focus on optimizing the model to enhance speed and adapt to varying lighting conditions and image quality.

Thus, the proposed FaceZ method can be applied in real-world security systems where high accuracy is critical. The ensemble of FaceNet and DeepFace, combined with MobileNetV2-based anti-spoofing, makes it suitable for biometric access control systems, online identification, and digital security. However, additional optimization of processing speed will be necessary for commercial deployment.

References

[1] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1-127, 2009. https://doi.org/10.1561/2200000006

[2] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701-1708.

[3] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, "Fast high dimensional vector multiplication face recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1960-1967.

[4] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006.

[5] T. Berg and P. N. Belhumeur, "Tom-vs-Pete classifiers and identity-preserving alignment for face verification," in *Proc. British Machine Vision Conf. (BMVC)*, 2012, vol. 2, p. 7.

[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991. https://doi.org/10.1162/jocn.1991.3.1.71

[7] L. Cherikbayeva, A. Yerimbetova, and E. Daiyrbayeva, "Research of cluster analysis methods for group solutions of the pattern recognition problem," in *2021 6th International Conference on Computer Science and Engineering (UBMK)*, 2021: IEEE, pp. 1-4.

[8] Y. Amirgaliyev, A. Ataniyazova, Z. Buribayev, M. Zhassuzak, B. Urmashev, and L. Cherikbayeva, "Application of neural networks ensemble method for the Kazakh sign language recognition," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3275-3287, 2024. https://doi.org/10.11591/eei.v13i5.7803

[9] A. Yeleussinov, Y. Amirgaliyev, and L. Cherikbayeva, "Improving OCR accuracy for Kazakh handwriting recognition using gan models," *Applied Sciences*, vol. 13, no. 9, p. 5677, 2023. https://doi.org/10.3390/app13095677

[10] M. Telmanov, M. Suchkov, Z. Abdiakhmetova, and A. Kartbayev, "Strategic processor task allocation through game-theoretic modeling in distributed computing environments," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 2, pp. 1371-1380, 2025.

[11] K. Alimhan, M. N. Kalimoldayev, A. A. Adamov, O. Mamyrbayev, N. Tasbolatuly, and A. Smolarz, "Further results on output tracking for a class of uncertain high-order nonlinear time-delay systems," *Przeglad Elektrotechniczny*, vol. 95, no. 5, pp. 88-91, 2019. https://doi.org/10.15199/48.2019.05.22

[12] G. Bakhadirova, N. Tasbolatuly, A. Tanirbergenova, A. Dautova, A. Akanova, and Y. Ulikhina, "Computer simulation of control of high-order nonlinear systems using feedback," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1096-1109, 2024. https://doi.org/10.47738/jads.v5i3.275

[13] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215-244, 2021. https://doi.org/10.1016/j.neucom.2020.10.081

[14] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012: IEEE, pp. 2518-2525.

[15] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815-823.

[16] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5325-5334.

[17] Y. Liu, Z. Wen, and H. Xu, "Dual CNN-based face recognition model using feature-level fusion," *Appl. Sci.*, vol. 13, no. 9, pp. 5677-5689, 2023.