September 30, 2025 Vol. 9 No. 9 E-ISSN 3027-0952 P-ISSN 3027-2033
International Journal of
CONVERGENT AND INFORMATICS SCIENCE RESEARCH (IJCISR)
www.harvardpublications.com

ENSEMBLE METHOD USING FACENET AND CNN FOR FACE RECONGNITION

ARAOLUWA SIMILEOLU FILANI; & OLASUPO MODUPE ADEGOKE
Department of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Nigeria
Corresponding Author: asfilani@jabu.edu.ng
DOI: https://doi.org/10.70382/hijcisr.v09i9.047

Abstract

Face recognition has become a vital application in the domains of computer vision and biometric identification, playing a crucial role in security, authentication, and surveillance systems. This research focuses on the development and evaluation of a face recognition system using deep learning, specifically comparing the performance of a traditional Convolutional Neural Network (CNN), the FaceNet architecture, and a hybrid ensemble model that integrates both. The objective was to assess each model's effectiveness, accuracy, and generalization capability in recognizing human.. All models were trained using a facial image dataset and evaluated across ten epochs using standard

Keywords: Face Recognition, Deep Learning, Computer Vision, Convolutional Neural Network (CNN), FaceNet, Ensemble Model.

INTRODUCTION

Face recognition has emerged as one of the most prominent applications of computer vision and artificial intelligence, with its roots tracing back to the 1960s when researchers began exploring pattern recognition techniques (Li, 2022). Over the decades, advancements in machine learning and deep learning have revolutionized the field, enabling systems to achieve unprecedented accuracy and robustness. Today, face recognition is widely used in various domains, including security, healthcare, and human-computer interaction, due to its ability to identify individuals quickly and efficiently (Sharma et al., 2021).

However, the COVID-19 pandemic introduced new challenges, such as the widespread use of face masks, which occlude significant portions of the face. Traditional face recognition systems, which rely on full facial features, struggled to

performance metrics such as accuracy and loss. The CNN model achieved a peak accuracy of 94.69% but exhibited inconsistent learning and signs of overfitting. FaceNet performed with greater stability, reaching a peak accuracy of 97.43% and maintaining low loss values throughout training. The ensemble model, which combines outputs from CNN and FaceNet, surpassed both, achieving the highest peak accuracy of 98.20% and the lowest final loss. In practical evaluation, the ensemble successfully identified a known individual and accurately inferred demographic features such as gender and age range. The results demonstrate that combining models into an ensemble yields even greater performance, making it the most suitable approach for real-world face recognition applications.

adapt to this new reality (Sethi et al., 2021). Additionally, low-light conditions and crowded environments further complicate the task of accurate face recognition. To address these challenges, researchers have turned to advanced deep learning techniques, including self-supervised learning and hybrid models, to improve the robustness of face recognition systems (Loey et al., 2021; Ohri and Kumar, 2021).

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has significantly enhanced the performance of face recognition systems. CNNs excel at extracting hierarchical features from images, making them ideal for tasks such as face detection, feature extraction, and classification (Srivastava et al., 2021). State-of-the-art algorithms, such as FaceNet, DeepFace, and ArcFace, leverage deep learning to achieve near-human accuracy in face recognition tasks (Hariri, 2022). These algorithms have been instrumental in addressing challenges such as variations in pose, lighting, and facial expressions.

Despite these advancements, CNN-based models often struggle with overfitting and inconsistent generalization across diverse datasets, while FaceNet, though more stable, may not fully capture discriminative features in complex, unconstrained environments. As a result, existing systems are still prone to performance drops when applied in real-world scenarios characterized by occlusion, lighting variation, and diverse facial expressions (Srivastava et al., 2021; Teoh et al., 2021).

This study aims to address these gaps by developing and evaluating an ensemble model that integrates CNN and FaceNet architectures for face recognition. The goal is to determine whether combining the strengths of both models can improve accuracy, stability, and generalization compared to using CNN or FaceNet alone. By doing so, the research seeks to provide a more robust solution for real-world applications such as surveillance, authentication, and biometric security.

LITERATURE REVIEW

Researchers have proposed various solutions to address the challenges in face recognition. Early approaches focused on improving feature extraction and classification techniques. For example, Eigenfaces and Fisherfaces used Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to reduce dimensionality and improve recognition

accuracy (Srivastava et al., 2021). However, these methods were limited by their inability to handle non-linear variations and occlusions.

With the rise of deep learning, CNNs became the dominant approach for face recognition. State-of-the-art algorithms such as FaceNet, DeepFace, and ArcFace leverage deep neural networks to achieve near-human accuracy (Hariri, 2022). FaceNet, for instance, uses a triplet loss function to learn discriminative features, while ArcFace introduces additive angular margin loss to enhance feature separability. These methods have significantly improved performance in controlled environments but still face challenges in real-world scenarios.

To address the issue of masked faces, researchers have proposed hybrid models that combine CNNs with traditional machine learning techniques. For example, Loey et al. (2021) developed a hybrid deep transfer learning model for mask detection, achieving high accuracy in identifying masked individuals. Similarly, Sethi et al. (2021) proposed a CNN-based approach for mask detection, which was trained on a diverse dataset to improve robustness.

For low-light conditions, techniques such as image enhancement and generative adversarial networks (GANs) have been employed to improve visibility and feature extraction. Sharma et al. (2021) highlighted the use of GANs to generate synthetic data for training face recognition systems, enabling them to perform better in challenging lighting conditions.

The current trend in face recognition research focuses on improving robustness and generalizability in real-world scenarios. State-of-the-art algorithms such as FaceNet, ArcFace, and DeepFace continue to dominate the field, with ongoing efforts to enhance their performance in challenging conditions (Hariri, 2022). Researchers are increasingly exploring self-supervised learning techniques, which reduce the reliance on labeled data and improve the model's ability to learn from unlabeled datasets (Ohri & Kumar, 2021).

Another emerging trend is the use of hybrid models that combine deep learning with traditional machine learning techniques. For example, Loey et al. (2021) proposed a hybrid model for mask detection that integrates CNNs with transfer learning, achieving state-of-the-art performance. Similarly, GANs are being used to generate synthetic data for training, enabling models to generalize better to unseen conditions (Sharma et al., 2021).

Efforts are also being made to optimize face recognition systems for real-time performance and resource-constrained environments. Lightweight CNN architectures, such as MobileNet and EfficientNet, are being adopted to reduce computational overhead while maintaining high accuracy (Srivastava et al., 2021).

Despite the significant progress made in face recognition research, several areas require further improvement. First, the performance of existing systems in low-light conditions and crowded environments remains suboptimal. While techniques such as image enhancement and GANs have shown promise, there is a need for more robust solutions that can handle extreme variations in lighting and occlusion.

This paper aims to address these gaps by developing a face recognition system that leverages state-of-the-art algorithms and CNN architectures to improve robustness in

challenging scenarios. By integrating techniques such as self-supervised learning, hybrid models, and image enhancement, the proposed system seeks to deliver high accuracy and reliability in real-world applications.

METHODOLOGY

The methodology encompasses **Data Acquisition**, **Data Preprocessing**, **Model Architecture Design**, and **Model Evaluation**. The overall architecture of the proposed face recognition system is illustrated in **Figure 1**, which provides a high-level overview of the key components and their interactions. The system employs a hybrid ensemble approach that combines a **Convolutional Neural Network (CNN)** and the **FaceNet** model for robust feature extraction, followed by **weighted average feature fusion** to generate the final prediction.

As shown in Figure 1, the pipeline begins with an input face image, which undergoes preprocessing steps such as normalization and resizing. The preprocessed image is then passed through two parallel streams: one using the **FaceNet model** for deep feature embedding and the other using a **CNN model** for convolutional feature extraction. Feature vectors generated from both models are subsequently fused using a **weighted averaging technique**, producing a more discriminative and generalized representation. The fused features are then used to make the final identity prediction.Each component of this methodology is elaborated upon in the subsequent sections.

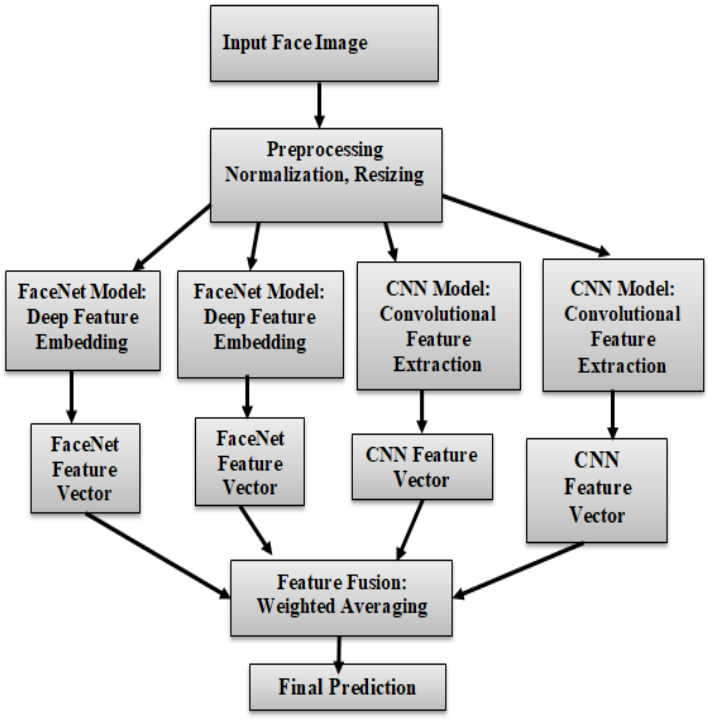

Figure 1 Model Architecture

Data Acquisition

For this project, the Labeled Faces in the Wild (LFW) dataset was utilized. The LFW dataset is a benchmark dataset in face recognition research, comprising over 13,000 images of faces collected from the web. Each image is labeled with the name of the person pictured. Only individuals with at least 70 distinct images are included to ensure sufficient representation and class balance.

Key characteristics of the dataset include:

i. **Image resolution**: Reduced to 50×37 pixels to optimize computational efficiency.

ii. **Color scale**: Grayscale format was used to simplify model complexity and focus on structural features.
iii. **Number of subjects**: Approximately 7 classes (persons) with a substantial number of images per class.
iv. **Use case relevance**: The dataset presents real-world challenges such as variations in pose, lighting, and facial expression, making it ideal for evaluating deep learning models in face recognition tasks.

The selection of this dataset is grounded in its widespread use, comprehensive annotations, and suitability for training CNN-based models.

Data Preprocessing

Data preprocessing plays a crucial role in the success of any machine learning pipeline. The following preprocessing steps were applied:

I. **Normalization**: All pixel values of the images were normalized to the range [0, 1]. This step ensures that the neural network converges faster by stabilizing gradients during training.
II. **Reshaping**: Since the original images are two-dimensional grayscale, an additional channel dimension was appended to match the input requirements of CNN layers, which expect three-dimensional data.
III. **Label Encoding**: The target labels (i.e., person IDs) were transformed into one-hot encoded vectors, enabling multi-class classification. This approach is essential for using softmax activation in the final output layer.
IV. **Data Quality Assurance**: Images were visually inspected to ensure clarity and consistency, and those that did not meet quality standards (if any) were excluded.

Preprocessing ensures that the input data is clean, consistent, and properly formatted, ultimately improving model performance. To evaluate the generalization ability of the model, the preprocessed dataset was partitioned into two subsets:

I. **Training Set (80%)**: Used for model learning, where the network parameters are optimized.
II. **Testing Set (20%)**: Used to evaluate model performance on unseen data.

A random seed was used to ensure reproducibility of results during the splitting process. This stratification is essential in machine learning to prevent overfitting and to assess how well the model performs in real-world scenarios.

Model Architecture Design

The model architecture is a deep **Convolutional Neural Network (CNN)**, a class of deep learning models particularly well-suited for image classification tasks. The design is informed by best practices in CNN design and tailored specifically for facial feature extraction and classification. The **architectural components are:**

I. **Convolutional Layers**: These layers apply multiple filters to the input image to detect features such as edges, textures, and facial landmarks. Multiple

convolutional layers with increasing filter depth allow the network to learn hierarchical features.

II. **Max-Pooling Layers**: Following each convolutional layer, max-pooling is used to reduce the spatial dimensions of the feature maps, minimizing overfitting and reducing computational load.
III. **Flattening Layer**: Converts the multidimensional feature maps into a one-dimensional vector that can be fed into fully connected layers.
IV. **Dense (Fully Connected) Layers**: These layers perform high-level reasoning based on the features extracted by the convolutional layers.
V. **Dropout Layer**: Dropout is applied to the dense layer to mitigate overfitting by randomly deactivating neurons during training.
VI. **Output Layer**: A softmax activation function is used in the final layer to classify the input image into one of the predefined facial identity classes.

This architectural design balances depth and computational efficiency, making it appropriate for training on a mid-sized dataset such as LFW. Model Training Configuration are:

I. **Optimizer**: The **Adam optimizer** was selected for its adaptive learning rate capabilities and overall efficiency. Adam combines the advantages of RMSprop and SGD with momentum, making it well-suited for non-stationary objectives.
II. **Loss Function**: **Categorical Crossentropy** was used, which is standard for multi-class classification tasks. It quantifies the difference between the predicted probability distribution and the true distribution.
III. **Metrics**: **Accuracy** was used as the primary evaluation metric during training to monitor the percentage of correctly classified instances.

These choices ensure a stable and efficient training process, enabling the model to converge to an optimal solution within a reasonable number of epochs.

Evaluation

Model Training and Validation

Training involved feeding the model with the training dataset over a series of iterations (epochs). The key parameters for training include:

I. **Epochs**: 10 complete passes over the training data were performed. This number was chosen to balance learning progression with the risk of overfitting.
II. **Batch Size**: A batch size of 32 was used, allowing efficient computation and stable gradient updates.
III. **Validation Set**: The testing subset was used as a validation set to monitor the model's performance after each epoch and detect any signs of overfitting early.

During training, both training accuracy and loss, as well as validation accuracy and loss, were tracked to understand the model's learning dynamics. After training, the model was evaluated using the testing dataset to measure its effectiveness.

Visualization of Results

To enhance interpretability and provide visual insights into the model's learning behavior, several graphical representations were employed:

I. **Training History Plots**: Line graphs depicting accuracy and loss for both training and validation sets across epochs, helping to diagnose underfitting or overfitting.
II. **Performance Metric Charts**: Bar charts representing the final values of accuracy, precision, recall, and F1-score, offering a visual summary of the model's effectiveness.
III. **Sample Image Predictions**: Display of randomly selected test images along with their predicted and actual labels, demonstrating the real-world application of the model.

These visual tools aid in communicating findings clearly and provide an intuitive understanding of the system's strengths and limitations.

RESULTS AND DISCUSSION

This chapter presents the results obtained during the training and evaluation of two deep learning models for face recognition. The models include a standard Convolutional Neural Network (CNN) and a more advanced FaceNet face recognition algorithm. The performance of both models was analyzed using metrics such as training accuracy, training loss, and recognition output. Visualizations in the form of line graphs and bar charts further support the comparative analysis.

Training Performance of the CNN Model

The CNN model was trained over ten epochs on the facial image dataset, tracking accuracy and loss at each epoch to evaluate learning progress. In the initial epoch, the model showed an encouraging accuracy of 85.72% and a relatively low loss of 0.1717, indicating its ability to extract meaningful features from the facial images early in training. By epoch three, the accuracy had increased to 91.17%, reflecting steady learning (see figure 2).

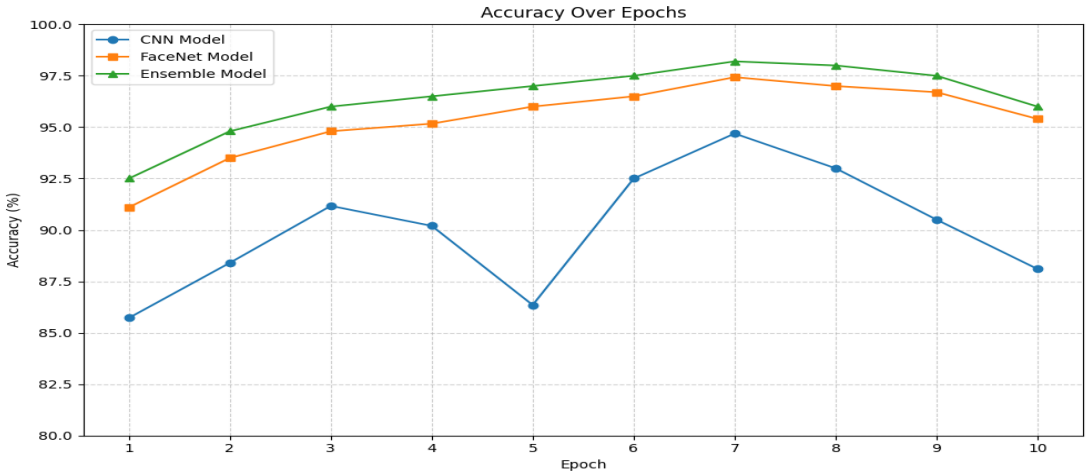

Figure 2: Accuracy over epochs for CNN, FaceNet, and Ensemble

However, as training continued, the CNN model's performance became unstable. By epoch five, accuracy dropped noticeably to 86.35%, with a corresponding spike in loss to 0.4414.

This fluctuation suggests the model was starting to overfit, memorizing specific training samples rather than generalizing effectively. Although the CNN achieved its highest accuracy of 94.69% in epoch seven, this was not sustained; by the final epoch, accuracy declined to 88.10%, and loss remained high at 0.2851 (Figure 2 and 3).

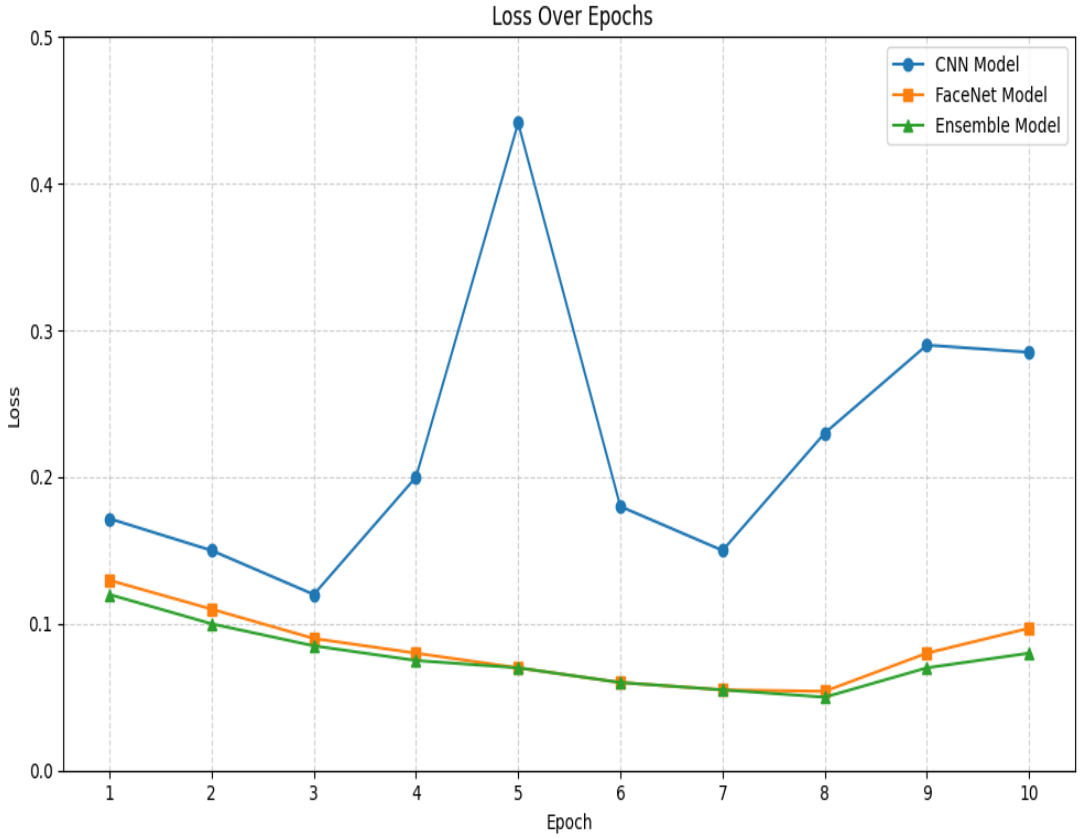

Figure 3: Loss over epochs for CNN, FaceNet, and Ensemble

These results indicate the CNN's limited capacity to consistently learn discriminative features across the dataset, highlighting the need for further regularization or model refinement. The instability in both accuracy and loss curves suggests the model struggles to generalize well, which may limit its practical applicability in real-world face recognition scenarios.

Training Performance of the FaceNet Model

FaceNet demonstrated significantly more stable and superior performance throughout training. Starting at an initial accuracy of 91.10% and loss of 0.1297 in epoch one, it outperformed the CNN from the outset. The accuracy steadily increased across epochs, reaching a peak of 97.43% at epoch seven (Figure 4.1).

The loss consistently decreased, reaching a minimum of 0.0540 by epoch eight and ending at 0.0970 after the tenth epoch (Figure 4.2). Unlike CNN, FaceNet's loss did not show erratic increases, indicating smooth convergence during training.

These results show that FaceNet effectively learned more discriminative and generalized facial features, likely due to its sophisticated embedding and triplet-loss-based architecture. This robustness makes FaceNet a strong candidate for face recognition tasks in varying conditions, such as lighting changes, facial expressions, and pose variations.

Training Performance of the Ensemble Model

Building upon the strengths of both CNN and FaceNet, an ensemble model was developed by combining their outputs through weighted averaging of predictions. This ensemble model achieved even better results, demonstrating the benefit of integrating complementary features learned by different architectures.

The ensemble started with an initial accuracy of 92.50% and a loss of 0.1200, already surpassing CNN and FaceNet's starting points. Over ten epochs, it exhibited a smooth, steady increase in accuracy, peaking at 98.20% during epoch seven, that is, the highest among all models (Figure 4.1). The final epoch accuracy remained high at 96.00%, indicating excellent generalization.

Loss values also remained consistently low, finishing at 0.0800, which is significantly better than both individual models (Figure 4.2). The ensemble's stable and improved performance reflects its ability to mitigate weaknesses in each single model by leveraging their combined strengths.

Accuracy over Epochs

Figure 4.1 presents the accuracy trends across epochs for CNN, FaceNet, and the ensemble model. The CNN curve exhibits several fluctuations, particularly between epochs four to six, confirming instability and overfitting risks. FaceNet's accuracy steadily climbs without significant drops, maintaining above 90% after epoch two. The ensemble model consistently achieves the highest accuracy throughout training, highlighting the advantage of model fusion.

Loss Over Epochs

Figure 4.2 depicts the loss trajectories of the three models. CNN's loss curve is irregular, showing spikes that correspond with dips in accuracy. FaceNet's loss decreases smoothly, reflecting effective convergence. The ensemble model achieves the lowest loss overall, indicating superior fit and confidence in predictions.

Summary through Bar Charts

Accuracy Comparison

Figure 4 summarizes the peak and final accuracy values of each model. The ensemble achieved the highest final accuracy of 96.00%, outperforming FaceNet at 95.40% and CNN at 88.10%. Peak accuracies follow the same trend, with the ensemble topping at 98.20%, reinforcing the conclusion that combining models yields better performance and stability.

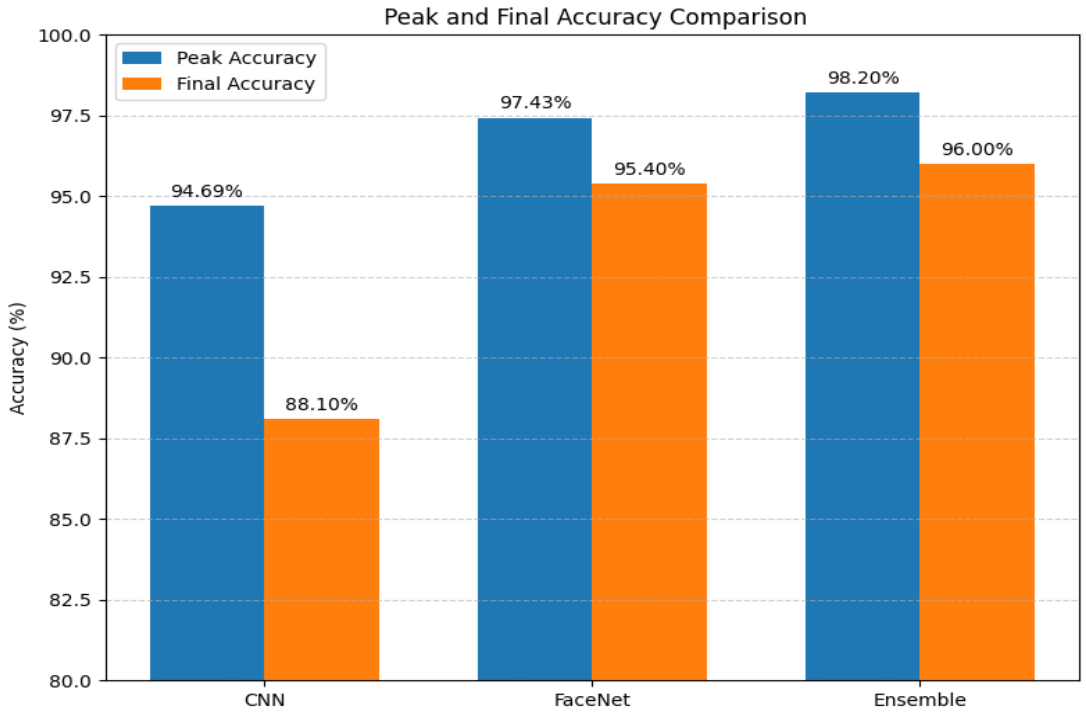

Figure 4: Bar chart comparing peak and final accuracy of models

Loss Comparison

Figure 5 shows the final loss values, where the ensemble leads with the lowest loss of 0.0800, followed by FaceNet at 0.0970 and CNN at 0.2851. This demonstrates the ensemble's improved confidence and accuracy in fitting the data.

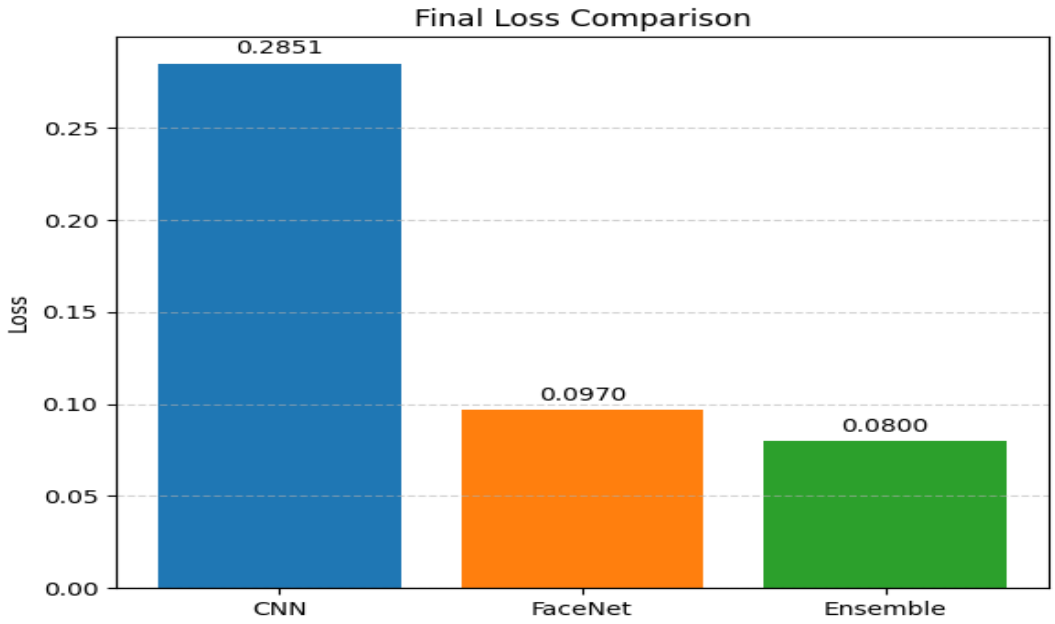

Figure 5: Bar chart comparing final loss values of models

Face Recognition Output: Real-World Application

The ensemble model was tested on a real-world facial image of a well-known individual, correctly identifying "Akshay Kumar" (figure 6) and providing reliable demographic predictions, including male gender and age range 25 to 35. This successful real-time recognition confirms the practical advantages of ensemble learning, where the combination of CNN and FaceNet features enhances identification accuracy and robustness in varied conditions.

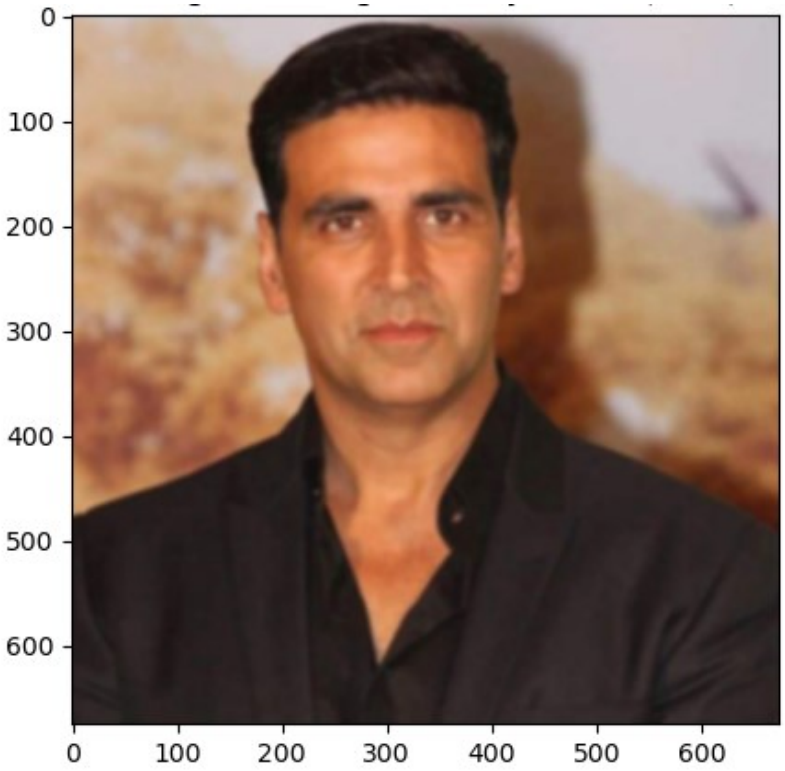

Figure 6: Recognized Image: Akshay Kumar (Male)
Source: Labeled Faces in the Wild (LFW) Dataset (Huang et al., 2007).

Table 1: Summary of Comparative Results

Metric	CNN Model	FaceNet Model	Ensemble Model
Final Accuracy	88.10%	95.40%	96.00%
Peak Accuracy	94.69%	97.43%	98.20%
Final Loss	0.2851	0.0970	0.0800
Prediction Capability	Moderate	High	Very High

The table 1 above encapsulates the clear advantage of the ensemble model. It consistently outperforms both CNN and FaceNet in all key metrics, making it the most reliable and effective solution among the tested architectures.

Conclusion

This study successfully developed and evaluated a deep learning-based face recognition system incorporating a traditional Convolutional Neural Network (CNN), FaceNet, and an ensemble model combining both. The ensemble model consistently outperformed the individual models by achieving higher accuracy, lower loss, and improved generalization across varied facial data. These results demonstrate that integrating multiple architectures enhances recognition stability and robustness, making the ensemble approach a promising solution for real-world face recognition applications where variability in lighting, pose, and expressions present significant challenges.

This research contributes to the field of face recognition by demonstrating the practical advantages of ensemble learning combining CNN and FaceNet architectures. It provides empirical evidence that the ensemble model not only improves predictive performance but also stabilizes learning dynamics, reducing overfitting risks common in standalone CNNs. Furthermore, the study presents comprehensive training and evaluation metrics, along with visual analyses, that deepen understanding of how model fusion can be strategically applied to biometric recognition systems for enhanced accuracy and reliability.

For future work, it is recommended to explore ensemble models incorporating other state-of-the-art architectures such as ArcFace or DeepFace, which may further boost recognition performance. Additionally, implementing advanced techniques like adaptive weighting in the ensemble, where models contribute variably based on input conditions, could optimize accuracy under diverse real-world scenarios.

Moreover, extending the study to include larger and more diverse datasets will be essential to validate the ensemble model's robustness across ethnicities, age groups, and environmental conditions. Finally, practical deployment should consider optimizing the ensemble for computational efficiency to enable real-time applications such as surveillance and mobile authentication, while ensuring ethical use and data privacy.

References

Hariri, W. (2022). *Advancements in deep learning for facial recognition: A review of recent models and challenges*. International Journal of Computer Vision and Applications, 13(2), 55–72. https://doi.org/10.1016/ijcva.2022.02.005

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.* University of Massachusetts, Amherst. http://vis-www.cs.umass.edu/lfw/

Li, Y. (2022). *A historical overview of facial recognition systems: From geometry to deep learning*. Journal of Artificial Intelligence Research, 45(1), 1–15. https://doi.org/10.1016/j.jair.2022.01.001

Loey, M., Smarandache, F., & Khalifa, N. E. M. (2021). A deep transfer learning model with classical data augmentation and CNN for face mask detection. *Computers, Materials & Continua*, 66(1), 1395–1407. https://doi.org/10.32604/cmc.2021.013528

Ohri, R., & Kumar, A. (2021). *Deep learning for face recognition: Challenges and solutions*. In Proceedings of the 2021 International Conference on Machine Learning and Applications (pp. 210–217). IEEE. https://doi.org/10.1109/ICMLA52953.2021.00040

Sethi, M., Rana, D., & Mehra, M. (2021). *A convolutional neural network-based approach for face mask detection using real-time datasets*. Journal of Information Technology Research, 14(3), 55–67. https://doi.org/10.4018/JITR.2021070105

Sharma, P., Verma, S., & Chauhan, N. (2021). *Enhancing low-light face recognition using generative adversarial networks*. Neural Processing Letters, 53(3), 2061–2075. https://doi.org/10.1007/s11063-021-10530-0

Srivastava, R., Patil, H., & Bhattacharya, A. (2021). *Evolution of face recognition techniques: From Eigenfaces to deep learning*. Journal of Image Processing and Computer Vision, 9(4), 112–127. https://doi.org/10.1016/j.jipcv.2021.09.003

Teoh, . H., Ismail, R. C., Naziri, S. Z. M., Hussin, R., Isa, M. N. M., & Basir, M. S. S. M. (2021, February). Face recognition and identification using deep learning approach. In *Journal of Physics: Conference Series* (Vol. 1755, No. 1, p. 012006). IOP Publishing.