An IoT-Integrated Deep Learning approach for Real-Time Face Authentication

[1]G.G.Rajput, [2]Durga M Ganjagol and [3]Pavitra Muttur
[1,2,,3]Department of Computer Science Karnataka State Akkamahadevi Women University Vijayapura, Karnataka, India

Abstract

Deep learning has greatly enhanced the performance of numerous recognition systems. and in particular facial recognition. However, achieving robust real-time recognition under diverse facial appearances and integrating Internet of Things (IoT) for real time application is challenging. In this paper we propose Multi-task Cascaded Convolutional Neural Network (MTCNN) deep learning model for robust facial recognition and integrating the proposed method in IoT environment for real time face recognition. MTCNN is used to precisely localize and align facial regions, while FaceNet model is used to generate compact facial embeddings that are subsequently classified using a Support Vector Machine (SVM) classifier. Experiments were carried out on images from LFW facial image dataset. The proposed recognition framework achieved an accuracy of 97.84%, demonstrating reliable facial identity classification on the LFW dataset. The proposed system is integrated with ESP32-CAM and Arduino to enable IoT-based remote monitoring and secure access control. The proposed framework offers an accurate, efficient, and practical solution for smart security, attendance management, and real-time surveillance.

Key words: *Face Recognition, Deep Learning, MTCNN, FaceNet, Support Vector Machine, Internet of Things.*

1 Introduction

The increasing adoption of intelligent digital technologies has accelerated the demand for authentication mechanisms that provide both security and convenience. Among the available biometric techniques, face recognition has emerged as an effective contactless solution by identifying individuals through their facial characteristics without requiring passwords or physical credentials. Face recognition has been widely adopted in applications including smart attendance, access control, surveillance, healthcare, and public safety systems (**8**). Recent surveys have further highlighted its growing role in intelligent vision-based applications (**9**, **16, 17, 21**).

Deep learning approaches based on Convolutional Neural Networks (CNNs) have revolutionized face recognition by replacing traditional handcrafted feature engineering with powerful representation learning. CNNs automatically learn hierarchical features from facial images, enabling robust and accurate identity recognition. Building on this foundation, FaceNet [ ]introduced a discriminative embedding strategy that generates compact facial descriptors, making them highly effective for distinguishing individual identities, while subsequent studies have demonstrated the effectiveness of deep convolutional networks for robust facial representation learning (**2, 3, 4**). In addition, MTCNN has emerged as a popular method for accurate face detection and alignment prior to feature extraction [ ]. Furthermore, modern **margin-based learning methods** have substantially improved recognition accuracy under unconstrained conditions [ ].

The Internet of Things (IoT) technology has extended face recognition to embedded and edge computing platforms. Resource-efficient recognition models have enabled deployment on embedded devices (**10, 11**), while recent studies have demonstrated real-time attendance monitoring and intelligent access control using ESP32-CAM and similar low-cost hardware platforms (**15, 18, 20**).

Although recent face recognition methods have shown promising performance, challenges such as lighting variations, pose changes, occlusion, and hardware limitations still affect recognition accuracy. Hence, developing lightweight and reliable frameworks integrating with IoT applications remains an active area of research (**13, 14, 17, 19, 21**).

Motivated by these challenges, in this paper we propose a deep learning-based smart face recognition framework that combines MTCNN and FaceNet for discriminative feature extraction and a Support Vector Machine (SVM) classifier for recognition.

The proposed method is implemented by integrating the ESP32-CAM with the Arduino platform to enable IoT-enabled real-time monitoring and intelligent access control.

2. 3 Proposed Methodology

Figure 1 illustrates the proposed smart face recognition framework. Facial images are preprocessed through detection, alignment, resizing, and normalization before being processed by MTCNN for face localization. FaceNet extracts discriminative facial embeddings, which are classified using an SVM classifier to recognize registered individuals and estimate the prediction confidence. The system utilizes the LFW dataset for model training and evaluation. An ESP32-CAM is employed for real-time image acquisition in Arduino platform. By combining facial detection, feature learning, identity prediction, and IoT connectivity, the proposed framework supports reliable real-time face recognition and attendance management.

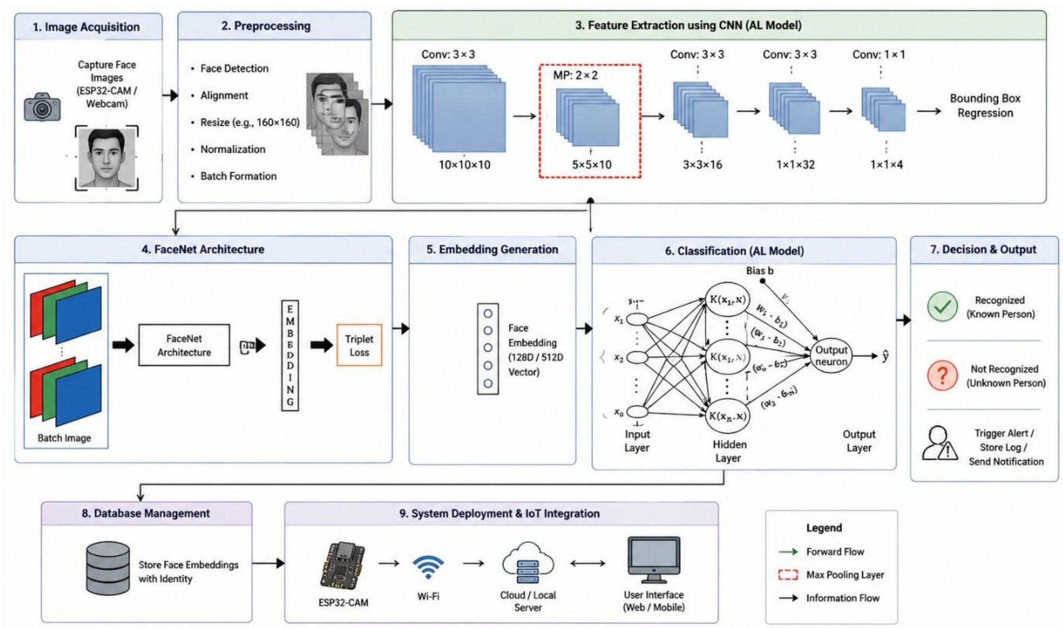


Fig. 1 Architecture of the proposed smart face recognition system

3.1 Image Preprocessing

Image preprocessing is carried out to prepare the facial images before the feature to reduce variations that may affect recognition performance. The image dataset used for training is enhanced through augmentation techniques, including horizontal flipping, rotation, zooming, brightness adjustment, and image translation, thereby improving the robustness and generalization capability of the recognition model. These preprocessing operations increase the variability of the training samples, reduce the risk of overfitting, and provide a well-prepared dataset for the subsequent face detection stage.

Facial localization is carried out using MTCNN to identify the dominant face from the input image for subsequent processing. The adopted facenet-pytorch implementation is configured to retain only the largest detected face, providing a single aligned facial sample for subsequent processing. The detected face is automatically aligned based on the estimated facial landmark locations and extracted with a 20-pixel margin to retain sufficient facial context.The alignment process can be expressed as

$$I_a = M(I),$$

The extracted region is resized to 160×160 pixels, producing an aligned RGB facial image that serves as the input to the FaceNet model for the next stage of feature learning.

Fig 2. Image pre-processing pipeline

3.2 Feature Extraction

The proposed system employs the pretrained InceptionResnetV1 model as a feature encoder. Once the facial region is aligned, it is forwarded through the network to obtain a fixed-size descriptor that summarizes the identity-related information of the input sample. This operation is represented by

$$\mathrm{z} = \mathrm{F}(\mathrm{I_a}), \qquad \mathrm{z} \in \mathbb{R}$$

where I denotes the aligned input image, $\mathrm{F}(\cdot)$ is the pretrained encoder, and z is the resulting descriptor. To maintain a consistent feature scale, the descriptor is converted to a unit vector as

$$\bar{\mathrm{z}} = \frac{\mathrm{z}}{\| \mathrm{z} \|_2}.$$

The normalized descriptor, $\bar{\mathrm{z}}$, is retained for the subsequent identity prediction step.

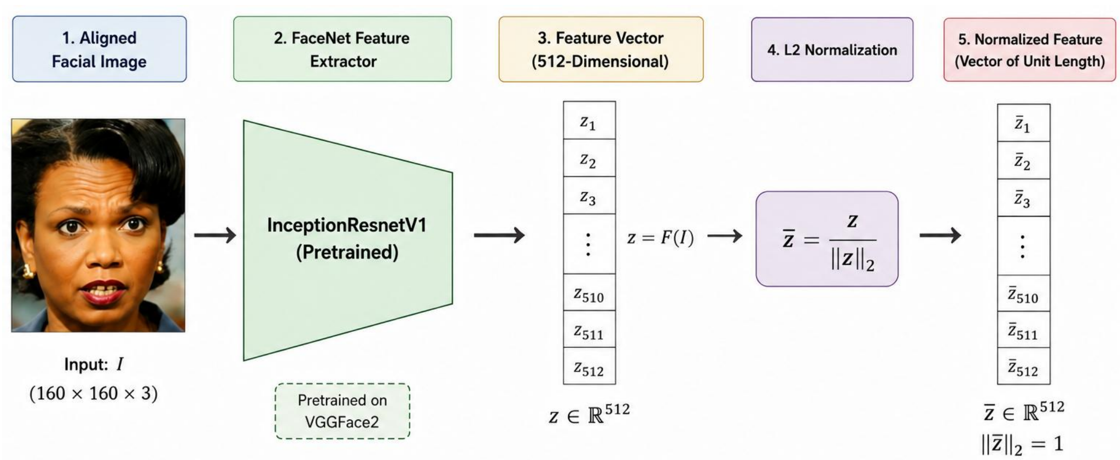


Fig. 3 FaceNet-based feature extraction pipeline

3.3 Recognition Using SVM classifier

In the recognition stage, the facial feature vectors generated in previous stage are utilized to identify the corresponding individual through a supervised learning approach. A linear Support Vector Machine (SVM) is trained using the normalized feature vectors, enabling it to learn the optimal separation boundaries among the enrolled identities. By projecting the features into a higher-dimensional space, the SVM effectively distinguishes between individuals, ensuring robust and reliable recognition performance (Fig. 4).

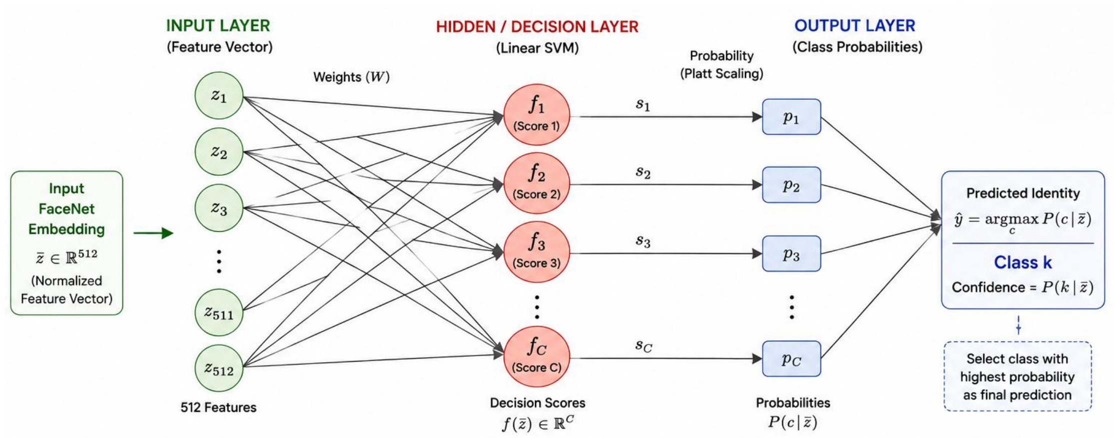


Fig. 4 Classification using SVM classifier

For an input feature vector $\bar{\boldsymbol{z}}$, the predicted identity is obtained as

$$\hat{y} = f(\bar{z}),$$

where $f(\cdot)$ denotes the classification model obtained during learning and $\hat{y}$ represents the identity predicted by the model. A confidence score is generated for each candidate identity before the final class is determined. The final recognition result is obtained by selecting the class with the largest confidence value,

$$\hat{y} = \arg\max_{c \in \mathcal{C}} P(c|\bar{z}),$$

where $\mathcal{C}$ denotes the set of enrolled identities and $P(c|\bar{z})$ indicates the confidence associated with class c. The selected identity and its confidence score are then forwarded to the subsequent verification stage for the final recognition process.

The final stage of recognition depends on the similarity between two normalized facial feature vectors. Let $\bar{\boldsymbol{z}}_q$ denote the normalized feature representation extracted from the input face, while $\bar{\boldsymbol{z}}_r$ corresponds to the stored facial representation used for comparison.Their similarity is obtained from

$$S(\bar{\boldsymbol{z}}_q, \bar{\boldsymbol{z}}_r) = \frac{\bar{\boldsymbol{z}}_q^T \bar{\boldsymbol{z}}_r}{\| \bar{\boldsymbol{z}}_q \|_2 \| \bar{\boldsymbol{z}}_r \|_2},$$

where $S(\cdot)$ indicates the similarity value derived from the normalized query and reference feature vectors.

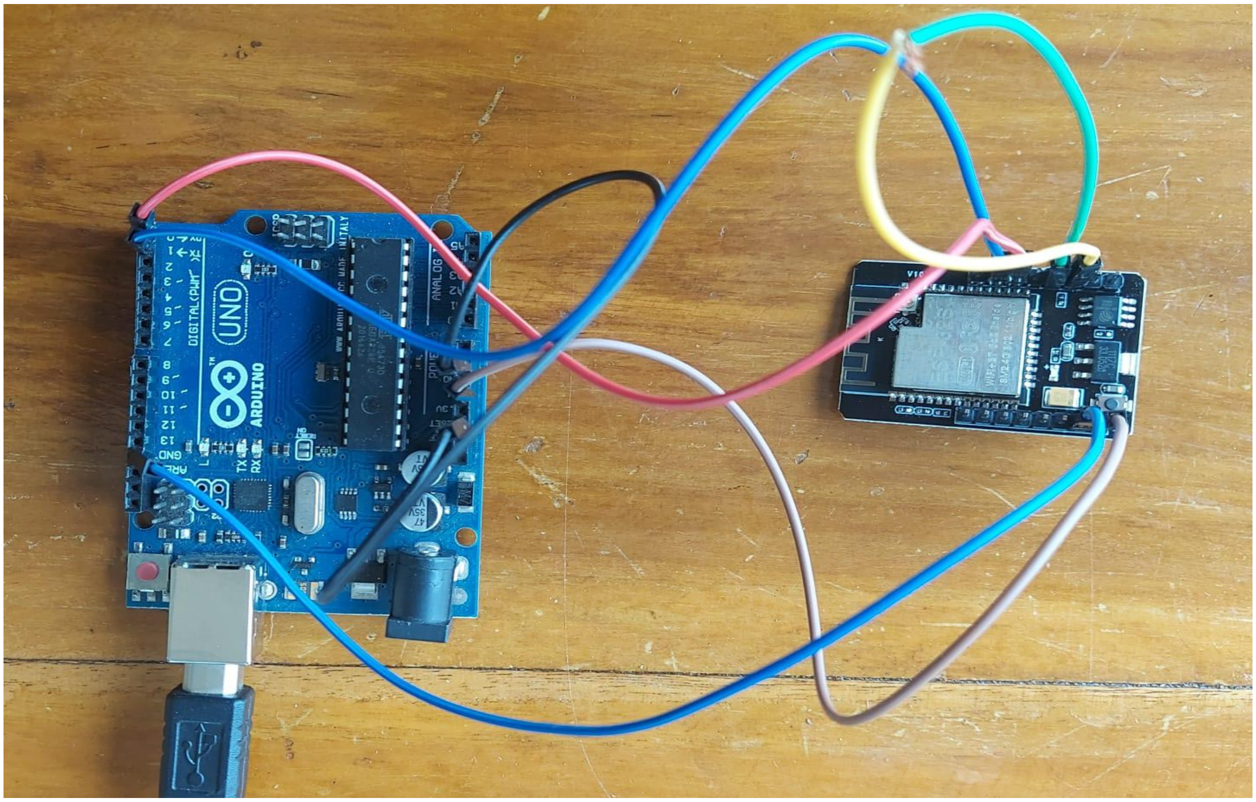
The final recognition decision is obtained by comparing the similarity score with a predefined threshold value τ,

$$\hat{y} = \begin{cases} \text{Known}, & S \geq \tau, \\ \text{Unknown}, & S < \tau. \end{cases}$$

where $\tau = 0.50$ denotes the matching threshold and $\hat{y}$ represents the predicted identity. Faces satisfying the threshold are accepted; otherwise, they are classified as unknown.

3.4 IoT-Based Real-Time Face Recognition

The hardware design for an IoT-enabled real-time face recognition system integrates multiple modules to ensure seamless data acquisition, processing, and communication. A camera module captures facial images, which are processed by a microcontroller or embedded board (Arduino and ESP32 CAM module). Extracted facial features are then classified using a Support Vector Machine (SVM) algorithm implemented in the system's software layer. The circuit includes essential components such as GPIO interfaces, and communication modules to enable connectivity with cloud platforms for storage and monitoring. The final circuit design shown in Fig 5 shows proper sensor interfacing and IoT-based data transmission using UART serial communication. This hardware integration provides reliable communication between the image acquisition device and the control unit, supporting continuous face recognition and IoT-enabled authentication.



3. Results: The proposed face recognition model is validated using the LFW deep-funneled dataset [ ]. Its diverse collection of facial images captured under real-world conditions enables comprehensive evaluation of the proposed method. The dataset specifications and partitioning strategy adopted are summarized in Table 1.

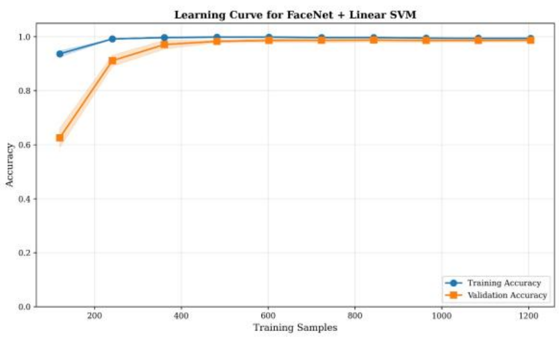
Table 1. LFW Dataset Specifications

Attribute	Description
Dataset	LFW (Deep-Funneled)
Images	13,233
Identities	5,749
Resolution	250 × 250 pixels

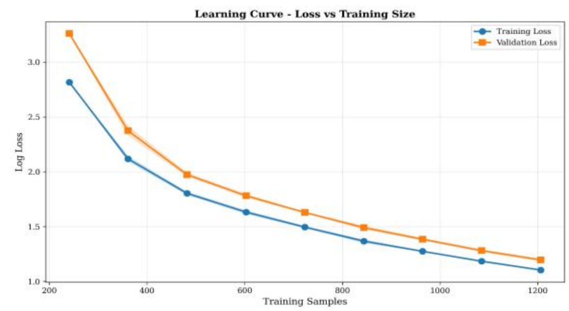
Only those identities having 20 or more facial images were included in the experiment to provide sufficient samples for learning. The dataset was divided into training, validating, and testing groups in an 80:10:10 ratio using stratified sampling. After training, the model was tested on desktop, Flask-based web, and ESP32-CAM with Arduino Uno implementations to examine its performance in different environments. The complete workflow was developed in Python with the help of PyTorch, facenet-pytorch, and Scikit-learn libraries, and the execution was carried out on a GPU machine.

Model Training Performance

The learning performance of the proposed FaceNet–SVM framework was assessed using the training and validation accuracy together with the corresponding loss values. Figure 8 illustrates the accuracy and loss curves obtained during model training.



(a)Training and Validation Accuracy



(b)Training and Validation Loss

Fig 6: Training performance of the proposed FaceNet-SVM model.

As illustrated in Fig. 8(a), the training and validation accuracies reached 99.34% and 96.30%, respectively, indicating good generalization with minimal overfitting. Figure 8(b) shows a consistent reduction in training and validation losses to 0.9368 and 1.1002, demonstrating stable convergence and effective learning of facial features.

Table 2 summarizes the final performance metrics.

Table 2 : Training and validation accuracy of Proposed FaceNet–SVM Model

Metric	Value
Training Accuracy	99.34%
Validation Accuracy	96.30%
Training Loss	0.9368
Validation Loss	1.1002

Recognition Performance

Experimental validation was carried out on the Labeled Faces in the Wild (LFW) dataset to assess the recognition capability of the proposed FaceNet–SVM model. The quantitative evaluation results obtained from the test dataset are summarized in Table 3.

Table 3; Recognition Performance on the LFW Test Dataset

Metric	Value
Test Accuracy	97.84%
Precision	98.01%
Recall	97.84%
F1-score	97.79%
ROC–AUC	99.76%
Test Loss	1.0258

The experimental results demonstrate that the proposed FaceNet-SVM model provides reliable face recognition performance on the LFW test dataset. The proposed model attained a test accuracy of 97.84%, while the weighted precision, recall, and F1-

score were recorded as 98.01%, 97.84%, and 97.79%, respectively. In addition, the model attained a macro ROC–AUC of 99.76%, indicating excellent discrimination among different facial identities. The obtained test loss of 1.0258 further reflects consistent prediction capability and good generalization to previously unseen face images.

Classification Report and Confusion Matrix Analysis

The quantitative evaluation is summarized in Table 4, and the corresponding prediction outcomes are illustrated in Fig. 6. The obtained results indicate consistently high classification performance across all evaluation measures. Only a limited number of samples were assigned to incorrect classes, suggesting that the developed recognition system provides reliable and stable identification.

Table 4: Summary of Classification Performance

Metric	Value
Overall Accuracy	97.84%
Weighted Precision	98.01%
Weighted Recall	97.84%
Weighted F1-score	97.79%
Macro ROC–AUC	99.76%

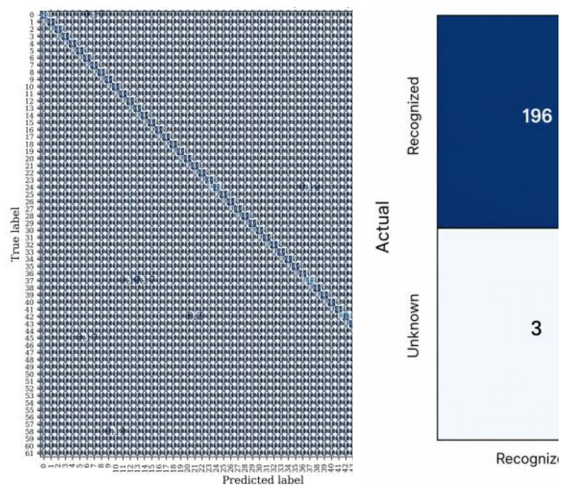


Fig 7: Confusion matrix for known and unknown face recognition.

Receiver operating characteristic (ROC) Curve Analysis

The graphical evaluation presented in Fig. 10 supports the effectiveness of the developed recognition system. An AUC score of 99.76% demonstrates the strong discriminative ability of the developed classifier, enabling accurate recognition of facial identities across the evaluation dataset.

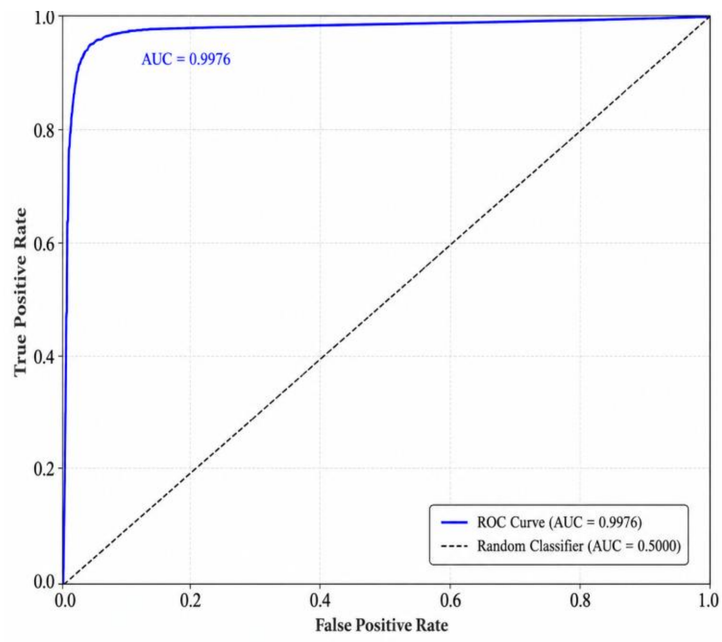


Fig 8: Receiver operating characteristic (ROC) curve of the proposed face recognition system.

The obtained curve demonstrates a high discriminative capability with an AUC of 99.76%, indicating effective separation between facial identities.

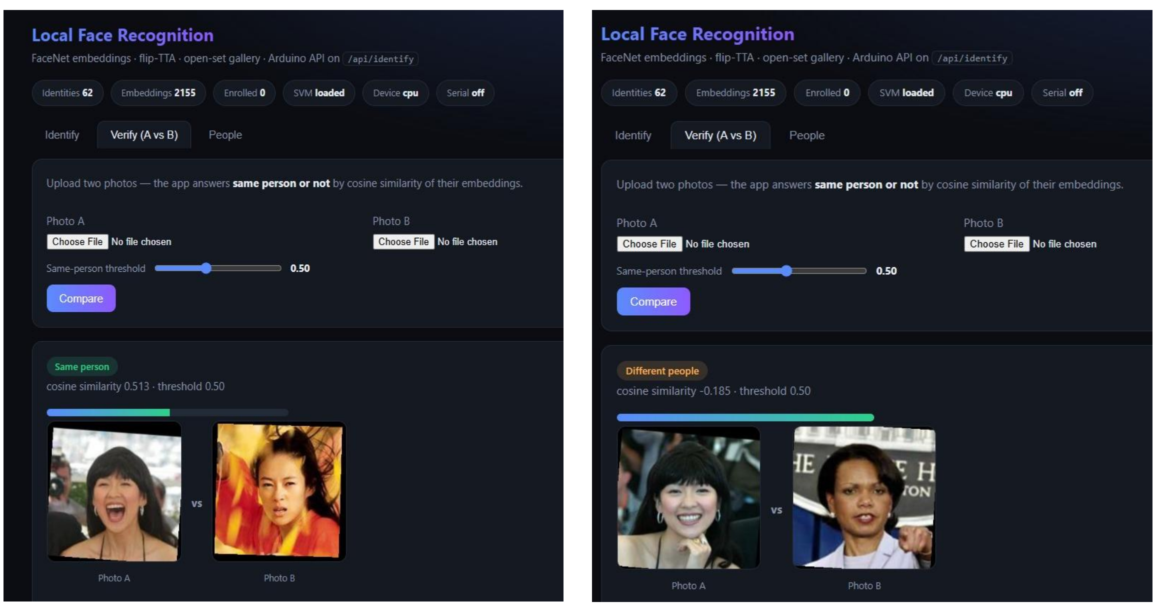
The proposed IoT-enabled face recognition system was validated using an ESP32-CAM module for real-time image acquisition. The captured images were transmitted to the recognition framework, where facial regions were automatically detected and identified using the trained FaceNet-SVM model. As illustrated in Fig. 11, the system correctly recognized multiple registered individuals by displaying their predicted names and corresponding confidence scores above the detected faces. These results demonstrate that the developed framework can perform reliable real-time face recognition, making it suitable for IoT-based smart surveillance and access control applications.

Face Verification Results

The face verification capability of the proposed recognition system was evaluated using pairs of facial images. Figure 14 presents representative verification outcomes for both matching and non-matching image pairs. The system successfully confirmed images belonging to the same individual while correctly rejecting images of different individuals based on cosine similarity between the extracted facial embeddings. These results demonstrate the effectiveness of the proposed approach for accurate and reliable face verification.

Fig 9: Real-time face recognition results for multiple registered individuals.

The proposed system detects facial regions, predicts the corresponding identities, and displays the associated recognition confidence for each detected face (Fig 15).



Same individual *Different individuals*

Fig 10: Face verification results obtained using the proposed recognition system. (a) Successful verification of the same individual. (b) Correct rejection of different individuals.

Comparison with the Baseline FaceNet Model

A comparative evaluation between the proposed FaceNet–SVM framework and the baseline FaceNet model is presented in Table 5. The developed framework delivered improved recognition performance on the LFW dataset. The results demonstrate that the proposed FaceNet–SVM framework consistently outperformed the reference FaceNet model across all reported evaluation metrics. This improvement suggests that incorporating an SVM classifier with the extracted facial descriptors contributes to more accurate and reliable identity classification.

Table 5: Comparison of the proposed FaceNet-SVM framework with the baseline FaceNet model reported by Zhalgas et al. (ref21?) on the LFW dataset.

Metric	baseline FaceNet model	Proposed FaceNet–SVM
ROC–AUC	0.9761	0.9976
Accuracy (%)	94.07	97.84
Precision (%)	94.10	98.01
Recall (%)	94.03	97.84
F1-score (%)	94.06	97.79

The experimental evaluation demonstrates the effectiveness of the proposed FaceNet–SVM framework for facial identity recognition. The quantitative results presented in Table 4 and Table 5 indicate that the developed approach achieved consistently high recognition performance and outperformed the baseline FaceNet(ref21?) model across the reported evaluation metrics. The confusion matrices shown in Fig. 9 confirm that most facial samples were assigned to their correct identities with only a few misclassifications. Likewise, the ROC curve in Fig. 10 reflects the strong discriminative capability of the developed model. The face verification examples in Fig. 12 further demonstrate reliable matching for both genuine and non-matching image pairs. In addition, the real-time implementation illustrated in Fig. 11 verifies that the proposed framework can accurately recognize multiple registered individuals using the ESP32-CAM-based IoT platform.

4. Conclusions: A practical face recognition framework integrating MTCNN, FaceNet, and a linear support vector machine was developed and evaluated in this study. The proposed approach effectively performed face detection, feature extraction, and identity classification using the LFW dataset. The evaluation outcomes confirm the effectiveness of the proposed recognition framework, which achieved an overall accuracy of 97.84%, together with precision, recall, and F1-score values of 98.01%, 97.84%, and 97.79%, respectively. An ROC-AUC of 99.76% further reflects the strong discriminative capability of the developed model. The developed model was successfully implemented on Flask and ESP32-CAM platforms for real-time face recognition. The obtained results demonstrate its suitability for practical biometric applications.

Acknowledgement: The research work is partly supported by VGST KFIST L2 funding by GoK sanctioned to the Department of Computer Science, KSAWU Vijayapura.

References:

1. Arsenovic, Marko, et al. "FaceTime—Deep learning based face recognition attendance system." *2017 IEEE 15th International symposium on intelligent systems and informatics (SISY)*. IEEE, 2017.

2. Bhatti, Kaneez Laila, et al. "Smart Attendance Management System Using Face Recognition." *EAI Endorsed Trans. Creative Technol.* 5.17 (2018): e4.

3. Daescu, Ovidiu, Hongyao Huang, and Maxwell Weinzierl. "Deep learning based face recognition system with smart glasses." *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 2019.

4. Hussain, Shaik Asif, and Ahlam Salim Abdallah Al Balushi. "A real time face emotion classification and recognition using deep learning model." *Journal of physics: Conference series*. Vol. 1432. No. 1. IOP Publishing, 2020.

5. Teoh, K. H., et al. "Face recognition and identification using deep learning approach." *Journal of Physics: Conference Series*. Vol. 1755. No. 1. IOP Publishing, 2021.

6. Alhanaee, Khawla, et al. "Face recognition smart attendance system using deep transfer learning." *Procedia Computer Science* 192 (2021): 4093-4102.

7. Dang, Thai-Viet. "Smart home management system with face recognition based on ArcFace model in deep convolutional neural network." *Journal of Robotics and Control (JRC)* 3.6 (2022): 754-761.

8. Dang, Thai-Viet. "Smart attendance system based on improved facial recognition." *Journal of Robotics and Control (JRC)* 4.1 (2023): 46-53.

9. Koodalsamy, Banumalar, Manikandan Bairavan Veerayan, and Vanaja Narayanasamy. "Face recognition using deep learning." *E3S Web of Conferences*. Vol. 387. EDP Sciences, 2023.

10. Medjdoubi, Abdelkader, Meriem Meddeber, and Khadidja Yahyaoui. "Smart city surveillance: Edge technology face recognition robot deep learning based." (2024): 25-36.