



Automated Multi-Student Attendance Using Video-Based Facial Recognition and Embedding Matching

Likhitha Alajangi¹, Dr. Kuda Nageswara Rao²

Department of Computer Science and Engineering, Andhra University, Visakhapatnam, India¹

Department of Computer Science and Engineering, Andhra University, Visakhapatnam, India²

Abstract: A major aspect of determining student attendance is having accurate records; unfortunately, traditional attendance systems are inefficient, not accurate, and can be manipulated for "proxy" attendance. A way around these obstacles is to create an intelligent, automated attendance method using deep learning facial recognition technology based on a pretrained model. In addition to traditional still-image approaches, our system begins with recording a short video of each participating student, leading to frame-by-frame extraction for constructing a robust and diverse dataset containing facial image samples of each student. This new pipeline from video to image will improve the representation of students' faces by using the three ways students' faces are captured (pose, expression, and lighting) to provide a more realistic photo of each student. In this approach, we will utilise a pretrained deep learning face recognition algorithm to extract the distinctive features of student faces from our generated dataset. A single classroom image containing multiple students will be used for attendance. The system will identify faces in the group image and extract their unique features by using three face detection methods, namely Haar Cascade, MTCNN, and InsightFace, which are evaluated and compared. For recognition, FaceNet and ArcFace embedding models are investigated using cosine similarity matching to perform both extraction and recognition, which helps us generate face embeddings for each person, compare those features to the database using cosine similarity, and mark attendance for recognized students with no manual intervention. This approach should successfully connect individual-level training data with group-level testing conditions, allowing for accurate facial recognition in the classroom.

Keywords: Automated Attendance System, Facial Recognition, Pretrained Deep Learning Model, Video Frame Extraction.

I. INTRODUCTION

Traditional attendance management systems use methods like roll calls, sign-in sheets and other methods that are not totally reliable and are time-consuming, so with advances in deep learning and computer vision, Facial recognition technology has emerged, which enables automated attendance systems. The benefits of using this type of system include the following: (i) attendance can be taken without the need for contact with the student; (ii) it reduces the manual effort by enabling automated recognition and attendance generation; and (iii) there is minimal disruption to the class when the student is recognised through this process. Historically, traditional computer vision methods employed techniques such as Haar Cascade for face detection and LBPH or PCA-based approaches for facial representation and recognition. Whereas newer approaches have been developed utilising deep learning-based methods (e.g. CNN, FaceNet, InsightFace) for extracting features of faces. These methods use CNNs to generate testing conditions in a real classroom environment. This work aims to address these challenges by implementing a video-based deep learning framework that combines multi-frame enrollment, face detection, facial embedding extraction, and cosine-similarity matching to support automated multi-student attendance.

II. RELATED WORK

Automated attendance systems across educational organisations using facial recognition technology have recently become a reliable way to decrease the time of manually recording attendance, prevent proxy attendance, and allow for better record-keeping. The framework presented in [1] used facial recognition techniques via OpenCV and provided a foundation for automating attendance processes through facial recognition. However, it showed that the performance of the attendance processes was subject to changing lighting conditions, different face orientations and varying image quality. Many researchers have begun using deep learning methods to help increase the accuracy of facial recognition results. The attendance system in [9] used deep transfer learning to extract facial features from images and showed an increased recognition rate in comparison to prior systems. An MTCNN- and FaceNet-based attendance framework was also presented in [15], demonstrating the use of deep face detection and embedding techniques for automated attendance. The attendance system presented in [12] used facial recognition methods for identifying students in school through deep learning algorithms and recording student attendance. Both of these studies demonstrated the benefits of



using learned feature representations, and they primarily relied on static image-based methods of registering facial images. One key to functional classroom attendance systems is the ability to identify multiple students at once. We can see the applicability of face detection and embedding models for recognising multiple students [16]. One means of addressing this issue was outlined in [4], which presented a multi-face recognition system that can detect and identify multiple people in one photo. Their study also offered support for the use of deep learning models to record classroom attendance and identified limitations related to pose differences and occlusion. It was also shown that using strong feature extraction techniques can increase the reliability of facial recognition even when changes occur in lighting conditions, background, and viewing perspective [11] [17]. Various research conducted on the subject has proposed similar attendance systems based on the

III. PROPOSED METHODOLOGY

In this system, processes are defined to achieve the goal of the system, which consists of face detection, extraction of facial embeddings, and recognition based on cosine similarity calculation. The whole system works in two main parts: the student registration stage and the classroom attendance stage, which are connected through a common student database

A. System Architecture

As illustrated in Fig. 1, the architecture consists of two independent stages with a common facial embedding database. The registration stage involves an enrollment video of each student being processed frame by frame. The faces are detected using one of the evaluated methods (Haar Cascade, MTCNN, or InsightFace) and are converted into embeddings using either FaceNet or ArcFace, so we construct a student-specific facial template. During attendance generation, we process the image and compare the resulting embeddings with already stored data using cosine similarity, which helps to identify the students and unknown people and generate a report. The six detector–embedding combinations are evaluated under a common experimental protocol, while the final application uses the selected configuration for operational attendance generation.

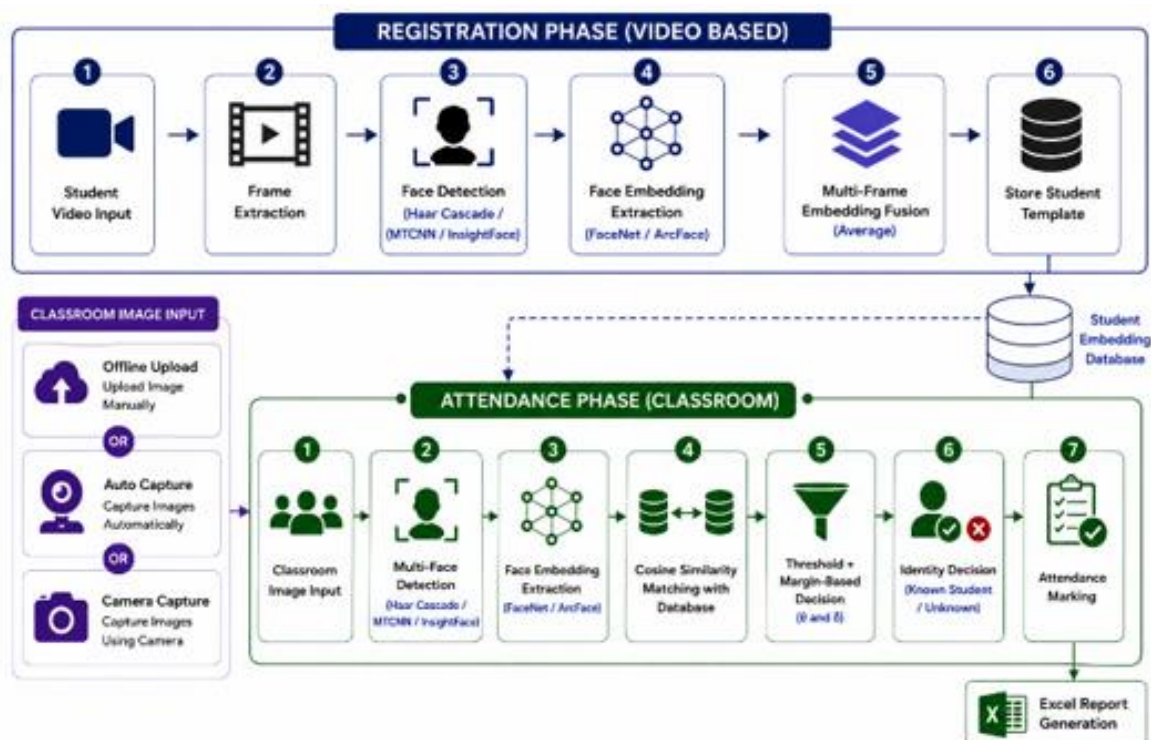


Fig. 1. Overall System Architecture of the Proposed Video-Based Attendance Framework

B. Dataset Creation

A customizable dataset shown in Table I has been created by recording 20 students on video, which are later extracted into approximately 500 images, which are used exclusively for identification. In addition, we considered 60 validation set images and 60 known test images, with three of each registered student. An additional set of unknown subjects to evaluate unknown subject rejection.



C. Video-based Student Registration

The following are the continuing components of the registration process:

- Video for each enrolled student obtained.
- Frame-by-frame extraction via OpenCV.
- Detection of faces on each extracted frame.
- Facial embedding obtained for detected face regions.
- Multi-frame facial embeddings are combined to create an overall template.
- Fused facial embedding stored in the facial database.

The multi-frame embedding fusion technique is used in the enrollment phase of this proposed framework. Thus, multiple frames from the registration video are used to extract their respective embeddings. Let $E_1, E_2, E_3, \dots, E_n$ be the set of facial embeddings obtained from n frames of video. $E_{\text{fused}} = \frac{1}{n} \sum_{i=1}^n E_i$ (1)

D. Face Detection

Correct face detection is very important for facial recognition since it has a direct influence on the quality of extracted face features. We have evaluated three approaches. a) Haar cascade classifier: It is a traditional machine learning-based method of face detection that is based on the appearance of Haar-like features, and uses multiple classifiers creating face detection and recognition [1], [3], but the defect of the method is its low efficiency in case of changes in head position, light conditions, and partial occlusion of faces. b) MTCNN: It works on face localisation and landmark detection using three complementary stages of networks. The distinctive feature of MTCNN is its resistance to head orientation and light variability [5]. c) InsightFace: works to achieve successful face detection and landmark localisation on difficult backgrounds like different head poses, light changes, and partial occlusion [14].

E. Face Embeddings

After the detection of a face, the detected face is represented in terms of a face embedding, which is numerical in nature. The face embeddings identify the specific features of a face, making it easier to compare with other faces. This section takes a look at two facial embedding models in order to evaluate their performance. a) Face Net: It is known for its ability to generate discriminative face embeddings because of the way it has been developed in accordance with the triplet loss function, which minimizes the distance between a pair of images that belong to the same identity while maximizing the distance between pairs of face images belonging to different identities. b) Arc Face: This is known for creating discriminative face embeddings using the Additive Angular Margin loss. The advantage of this is that it will increase the angular distance between two different faces while at the same time reducing the distance between two faces belonging to the same individual. These are evaluated independently with each of the three face detection methods.

F. Face Embedding Database

The combined embedding data obtained during the registration process is saved in an organised face database that serves as the registration repository for the attendance processing activity. Each entry contains Student Id, Student Name, their embedding vector and others. The data is saved in structured JSON files to enable prompt access during the identification process.

G. Student Identification Via Cosine Similarity

Student identification is accomplished through matching their facial image with the stored image using a cosine similarity metric. Cosine similarity between the two vectors A and B is represented as follows: $\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$ (2) The match with the highest similarity will be chosen as the likely one. To avoid confusion between real faces and unknown faces, thresholds are introduced for every detector-embedded combination separately as an independent task according to some uniqueness tests. The threshold will be chosen from a variety of values between 0 and 1 based on the characteristics of the operating point of FAR-FRR and fixed before the evaluation on the test set. If the maximum similarity score is not high enough, the face will be marked as Unknown.

H. Attendance And Report Generation

The process of creating attendance records can be summarised in the following steps :

- Take one or more pictures that can be either captured manually or automatically triggered by the system.
- Determine all the user faces in the picture through face detection.
- Generate embeddings for the faces in the picture.



- Find the student face with the greatest similarity using the cosine similarity method.
- Identify the student with the most similar face in the comparative process. After attendance has been recorded, attendance records are exported to an Excel file with the student details.

IV. EXPERIMENTAL SETUP

All experiments were conducted on Google Colab using a GPU (NVIDIA Tesla T4)-backed runtime.

A. Software Libraries/Dependencies

The implementation of the attendance system used only open-source libraries written in Python. Dependencies were installed at the start of the Google Colab session. Each library, its function in the pipeline, and version are provided in Table II.

B. Dataset Splitting For Testing And Training

The data was collected with video footage from 20 students enrolled in the classroom registration process, producing 500 image captures to be used for the dataset. To provide an independent evaluation, we captured 60 validation images, 60 known-subject test images, and 7 unknown-subject test images.

C. Recognition Threshold Selection

The threshold for recognition was identified from the validation dataset exclusively. There were no testing datasets were used in threshold identification. For each configuration of detector and embedding, cosine similarity thresholds were computed from 0.0 to 1.0 in increments of 0.01. For each of the thresholds, the value of the false acceptance rate (FAR) and the false rejection rate (FRR) was obtained. The chosen threshold values were kept constant before testing on both known and unknown test sets in Table 3.

D. Evaluation Protocol And Metrics

The evaluation of the system included the use of nine supplementary metrics. All classification metrics (Accuracy, Precision, Recall, F1-Score) were computed with the use of macro averaging across all classes of enrolled students, with Zero division = 0 for handling Unknown predictions. FARFRR curves were produced by sweeping the threshold of cosine similarity from 0.0 to 1.0 in 101 equally spaced points and computing the False Accept Rate (FAR) and the False Reject Rate (FRR) at each point.

V. RESULTS AND DISCUSSION

The results of evaluating the proposed framework on a dataset of 20 students' enrollment videos, which were later extracted into 500 images and an independent set of 60 images for each validation and known testing.

A. Face Detection performance

Table III shows how well the three face detectors performed over the complete dataset of 627 images. Insight Face has shown a high detection rate by detecting 100% of the faces. Whereas MTCNN and Haar Cascade have 98.5% and 91.3%, respectively. However, this improvement also leads to an increase in computational cost.

B. Recognition performance

The recognition results are analysed by combining the detector and embedding methods jointly. MTCNN with FaceNet has shown stronger results than the rest of the set by obtaining 100% accuracy, 94.83% threshold accuracy, and an F1 score of 95.83% although InsightFace with FaceNet also has 100% accuracy; they have lower threshold accuracy and F1 score. These show that face localisation affects downstream embedding-based recognition, which is shown in detail in Table V

C. FAR, FRR and EER analysis

The selected settings were not fixed for all combinations but were independently determined through validation. MTCNN paired with ArcFace had the lowest validation EER of 3.03%, while InsightFace with FaceNet came in at 3.25% and MTCNN with FaceNet at 3.64%. The Haar-based option had much higher EER over 19%, indicating that deep-learning classifiers achieve better distinction between the genuine and the imposter cases in this case, as given in Table VI

D. ROC analysis

ROC analysis further shows that MTCNN and InsightFace show the highest results compared to others. The Haarbased configurations show lower AUC values, which indicates greater overlap between genuine and imposter similarity scores shown in Fig 2

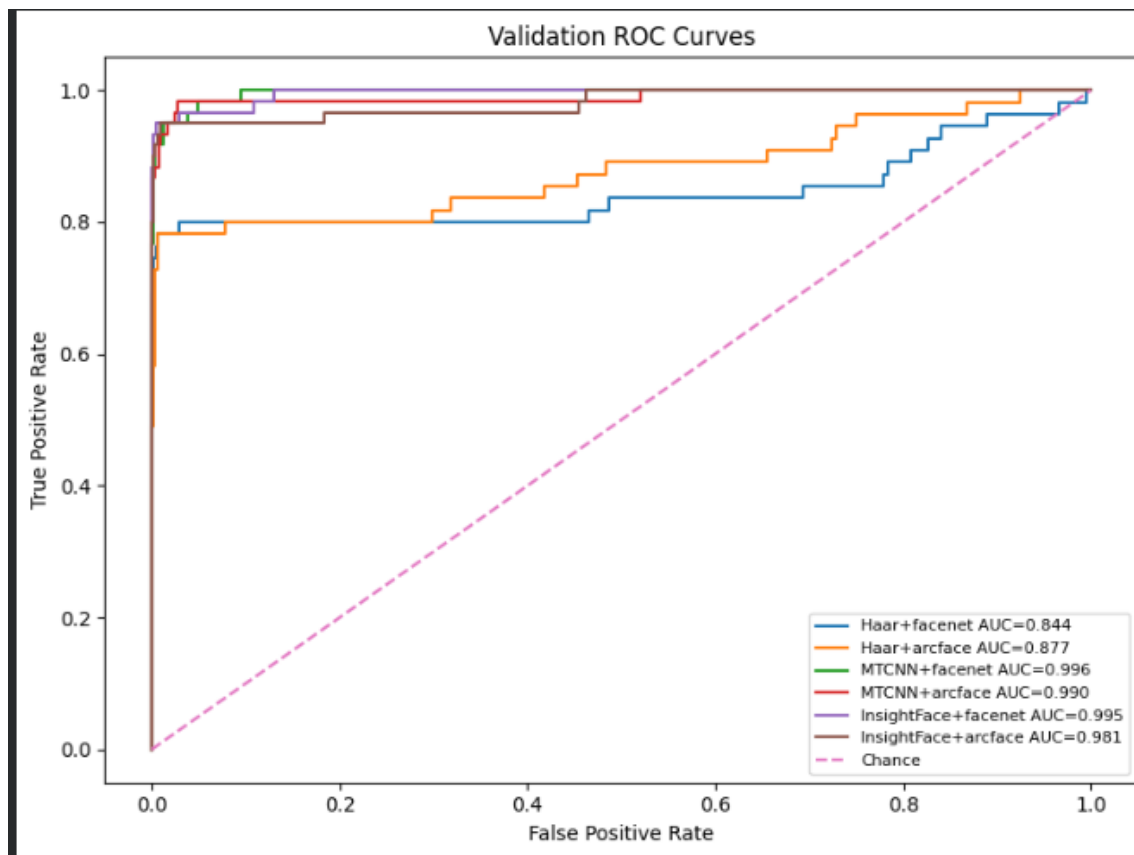


Fig. 2. ROC curves for six configurations

E. Unknown-Subject Rejection

Open-set evaluation was performed using seven images of unregistered students. MTCNN with ArcFace achieved the highest result in identifying the unknown students and rejecting them, these show important distinction between closed-set recognition and open-set rejection, as given in Table VII

F. Multi-Frame Fusion and similarity analysis

a) Multi-frame fusion performance : The benefit of utilising several enrollment frames is reflected in the case of multiframe fusion ablation. Thus, for MTCNN and FaceNet, the Rank-1 accuracy increased from 93.10% when using a single enrollment frame to 98.28% when using three frames, with the percentage achieving the maximum (100%) value possible when utilising ten enrollment frames. Further increasing the number of enrollment frames results in no improvement in the accuracy value. b) Similarity score analysis : This gives that potential confusing identities. So analysing similar pairs gives us more information about recognition errors shown in Table VIII

VI. SYSTEM IMPLEMENTATION

The attendance system that has been proposed is a web application that combines all processes of student registration, facial template generation, and attendance logging in one platform. The application is built as an online authentication mechanism to allow access only to registered users. Once the user logs in to the system, the student gets registered by entering input data and uploading the enrollment video. The video is processed, and several frames are produced, providing information about facial features of the student. To generate attendance, classroom pictures or frames are fed as input to the recognition pipeline. The faces are detected with the help of the detection algorithm and converted into embeddings. The algorithm contains several steps for processing the embedding input. The application also allows seeing an attendance report, which includes records of each student and an attendance report for each class.

VII. CONCLUSION AND FUTURE WORK

This work provides a full-scale attendance system. It covers the entire process, including student enrollment, attendance, and Excel report generation. This work was evaluated using 20 registered students, which were extracted into 500 images, 60 independently captured test and validation set images, and 7 unknown-subject test images. Six detector-embedding configurations were evaluated. The results show that MTCNN with FaceNet achieved higher accuracy compared to other detector and embedding models. They achieved the lowest validation EER of 3.03%.



Overall, the proposed system gives a practical workflow from enrollment to attendance generation. However, the results were interpreted for the current custom dataset.

A. Contributions

- Full development and implementation of an end-to-end pipeline, which allows for the processing of video-based attendance without intervention.
- The use of a multi-frame embedding fusion technique to combine all frames obtained from the enrollment video of a particular student and thereby create a better embodiment of the student.
- Comparison of three types of face detectors as well as two types of embedding models based on dataset, which includes controlled variables for the camera and for various angles (0°, 45°, and 90°).
- Experimental analysis demonstrating better improvement for MTCNN-FaceNet configuration.
- Practical testing of the system we developed by working with real-time photos, thus proving the functionality of the system in practice.

B. Limitations

The limited experimental dataset gathered comprised 20 students whose pictures were captured to create a total of 500 images in a controlled environment. Practically, after using the dataset for the experiments, it would become quite clear that such an approach may not be sufficient when assessing the efficiency of the suggested approach, as many other elements affect recognition efficiency in larger institutions. In addition, it is important to note that certain circumstances should also hold true, such as facial occlusions because of a mask, scarf, or hands held up to the face, which are not trained.

C. Future Work

It is suggested that, to improve the existing attendance management systems, the new system of continuous video attendance recording should be applied so as to develop a detailed attendance database and increase the dataset amount. Also, the use of the new student identification techniques involving facial expression recognition may increase the effectiveness of student identification in the classroom dynamics. The additional objectives of the student identification research include other up-to-date technologies, such as encrypted facial recognition and biometric systems. They pose a series of ethical and security problems related to personal data protection.

TABLES

TABLE I
DATASET ACQUISITION AND EVALUATION DETAILS

Parameter	Dataset Acquisition Details
Number of Registered Students	20 students
Enrollment Data	20–30 seconds Video recordings of individual students
Acquisition Devices	Vivo V50 smartphone camera (50 MP front autofocus camera, f/2.0 aperture, 92° field of view; supports 4K, 1080p, and 720p video recording) and HP laptop integrated webcam
Facial Pose Variation	0°–45°, 45°–90°, and 90°–180° pose ranges
Lighting Conditions	Natural indoor classroom illumination
Validation Images	60 images, independently captured; 3 images per student
Known Test Images	60 images, independently captured; 3 images per student
Unknown Test Images	7 images from unregistered subjects

TABLE II
LIBRARIES AND FRAMEWORKS USED IN THE PROPOSED FRAMEWORK

Libraries	Versions	Purpose
Python	3.10	Primary programming language
OpenCV	4.x	Image and video processing
Haar Cascade	OpenCV implementation	Face detection
MTCNN	Python implementation	Face detection and facial landmark localization
InsightFace	Python package	Face detection, landmark localization, and face-processing pipeline



FaceNet	Pre-trained model	Facial embedding extraction
ArcFace	Pre-trained model	Facial embedding extraction
NumPy	1.x	Numerical and vector operations
Pandas	2.x	Data organization and result analysis
scikit-learn	1.x	Evaluation metrics and analysis
Matplotlib	3.x	Performance visualization
ONNX Runtime	GPU-compatible configuration	Model inference
OpenPyXL	3.x	Attendance report generation

TABLE III

SELECTED THRESHOLDS FOR DIFFERENT DETECTOR-EMBEDDING COMBINATIONS

Detector	Embedding	Selected Threshold
Haar	FaceNet	0.47
Haar	ArcFace	0.15
MTCNN	FaceNet	0.63
MTCNN	ArcFace	0.30
InsightFace	FaceNet	0.63
InsightFace	ArcFace	0.25

TABLE IV

COMPARISON OF FACE DETECTION PERFORMANCE

Detector	Detected	Total	Detection Rate	Avg. Time/Image
Haar Cascade	573	627	0.914	315.6
MTCNN	618	627	0.986	713.5
InsightFace	627	627	1.000	1158.2

TABLE V

COMPARISON OF FACE RECOGNITION PERFORMANCE

Detector	Embedding	Rank-1	Thresh. Acc.	Precision	Recall	F1
MTCNN	FaceNet	1.000	0.9483	1.000	0.9417	0.9583
InsightFace	FaceNet	1.000	0.9167	1.000	0.9167	0.9400
InsightFace	ArcFace	0.9667	0.9167	0.9875	0.9167	0.9429
MTCNN	ArcFace	0.9138	0.8966	0.9750	0.8833	0.9100
Haar	ArcFace	0.7455	0.7455	0.9300	0.7333	0.7892
Haar	FaceNet	0.7455	0.7455	0.9175	0.7333	0.7870

TABLE VI

COMPARISON OF FAR, FRR AND EER

Detector	Embedding	FAR	FRR	EER
Haar	FaceNet	19.71%	20.00%	19.86%
Haar	ArcFace	21.53%	20.00%	20.77%
MTCNN	FaceNet	3.95%	3.33%	3.64%
MTCNN	ArcFace	2.72%	3.33%	3.03%
InsightFace	FaceNet	3.16%	3.33%	3.25%
InsightFace	ArcFace	4.82%	5.00%	4.91%

TABLE VII

UNKNOWN FACE REJECTION PERFORMANCE

Detector	Embedding	Unknown FAR	Correct Rejection
MTCNN	ArcFace	28.57%	71.43%
InsightFace	FaceNet	42.86%	57.14%
MTCNN	FaceNet	57.14%	42.86%
InsightFace	ArcFace	57.14%	42.86%
Haar	FaceNet	71.43%	28.57%
Haar	ArcFace	100.00%	0.00%

TABLE VIII

SIMILARITY SCORE

Detector	Embedding	Genuine	Impostor	Gap
Haar	FaceNet	0.3032	0.7150	0.4118
Haar	ArcFace	0.4216	0.9168	0.4952



MTCNN	FaceNet	0.1565	0.6321	0.4756
MTCNN	ArcFace	0.3952	0.8862	0.4910
InsightFace	FaceNet	0.1517	0.6343	0.4826
InsightFace	ArcFace	0.3935	0.8882	0.4947

REFERENCES

- [1] K. Vyshnavi, S. Shaik, M. S. Kumar, M. S. V. Praveen, and S. Potti, "Facial recognition based attendance system using OpenCV," J. Comput. Sci. Eng., ISSN: 2583-9055, May 2024.
- [2] J. Patel, S. Gandhi, V. Katheriya, P. Pataliya, and A. Majumdar, "Enhancing classroom attendance systems with face recognition through CCTV using deep learning," Procedia Comput. Sci., vol. 258, pp. 3031–3041, 2025.
- [3] A. Lal and N. Krishnamoorthy, "Face recognition system for real-time attendance monitoring using OpenCV," Indian J. Nat. Sci., vol. 16, no. 89, Apr. 2025.
- [4] A. P. Chawla and A. Suman, "Automated multi-face recognition-based attendance system using deep learning models," in Proc. Research Trends in AI, IoT Infrastructure and Industry (AIIII), Anna Univ., Chennai, India, Mar. 2024.
- [5] A. S. Lateef and M. Y. Kamil, "Effortless student attendance: A smart human-computer interactive system using real-time facial recognition," An-Najah Univ. J. Res. A (Nat. Sci.), vol. 39, no. 2, pp. 213–224, Feb. 2025.
- [6] H. Kale, K. Aswar, K. Yadav, and Y. Mali, "Attendance marking using face detection," Int. J. Adv. Res. Sci. Commun. Technol., vol. 4, no. 3, Oct. 2024.
- [7] P. B. Jha, A. Basnet, B. Pokhrel, G. K. Thakur, and S. Chhetri, "An automated attendance system using facial detection and recognition technology," Apex J. Bus. Manage., vol. 1, no. 1, pp. 104–115, 2023.
- [8] A. Z. Lari, F. Khan, A. Ahmad, A. A. Raza, and M. Suaib, "An efficient attendance management system for college environments using machine learning facial recognition technology," Int. J. Innov. Res. Comput. Sci. Technol., vol. 13, no. 3, pp. 8–12, 2025.
- [9] K. Alhanaee, M. Alhammadi, N. Almenhali, and M. Shatnawi, "Face recognition smart attendance system using deep transfer learning," Procedia Comput. Sci., vol. 192, pp. 4093–4102, 2021.
- [10] N. A. Ismail, C. C. Wen, H. Samma, and M. S. H. Salam, "Web-based university classroom attendance system based on deep learning face recognition," KSII Trans. Internet Inf. Syst., vol. 16, no. 2, pp. 503–523, Feb. 2022.
- [11] Y. Ding, X. Ma, and B. Yang, "Research on image feature extraction and environment inference based on invariant learning," Appl. Sci., vol. 14, no. 23, Art. no. 10770, 2024, doi: 10.3390/app142310770.
- [12] O. Oshin, J. Amenaghawon, F. Moninuola, and O. Idowu-Bismark, "Class attendance system using facial recognition," Inf. Stud., vol. 30, no. 6, pp. 1589–1595, 2025, doi: 10.18280/isi.300617.
- [13] C. H. Boe, K.-W. Ng, S.-C. Haw, P. Naveen, and E. A. Anaam, "An automated face detection and recognition for class attendance," JOIV: Int. J. Inform. Vis., vol. 8, no. 3, pp. 1146–1153, Sep. 2024, doi: 10.62527/joiv.8.3.2967.
- [14] T. Mohaimen, N. K. Ibrahim, N. Yusof, Z. Othman, and M. H. A. Ibrahim, "Improving student attendance using a smart biometric system with facial recognition using InsightFace and cosine similarity algorithm," Int. J. Res. Innov. Social Sci. (IJRISS), 2025.
- [15] P. J. Prathap, Y. H. N. Yashaswini, V. S. Varsha, M. I. Shariff, Honnaraju, and A. K. B. Ambika, "MTCNN based automatic attendance system using face recognition," Int. Res. J. Eng. Technol. (IRJET), vol. 10, no. 4, pp. 1827–1831, Apr. 2023.
- [16] V. J. W. Moses, P. Santhiya, G. J. Leelipushpam, and S. S. Kirubakaran, "Multi face recognition based attendance marking system using MTCNN with FaceNet," in Proc. 3rd Int. Conf. Self Sustainable Artificial Intelligence Systems (ICSSAS), 2025, pp. 135–140, doi: 10.1109/ICSSAS66150.2025.11080815.
- [17] K. Vignesh, B. Baranitharan, G. K. Sardar, K. Raj, A. Kumar, and U. Kumar, "Smart face attendance system using FaceNet," in Proc. 2nd Int. Conf. Visual Analytics and Data Visualization (ICVADV), 2026, pp. 1548–1553, doi: 10.1109/ICVADV67766.2026.11470539.