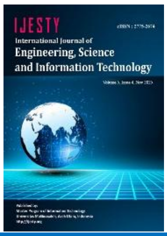
International Journal of Engineering, Science and Information Technology
Volume 5, No. 4 (2025) pp. 660-669
eISSN: 2775-2674
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i4.1765
Research Paper, Short Communication, Review, Technical Paper



Multiperson Automated Attendance System Based on Face Recognition Using YOLO and DeepFace with Active Learning

Didit Suprihanto*, Adi Pandu Wirawan, Kahlil Gibran Saputra, Arif Harjanto, Imam Muhammad Hakim, Happy Nugroho

Department of Electrical Engineering, Faculty of Engineering, Universitas Mulawarman, Samarinda, Indonesia

**Corresponding author Email: didit.suprihanto@ft.unmul.ac.id*

The manuscript was received on 25 April 2025, revised on 27 June 2025, and accepted on 22 October 2025, date of publication 28 December 2025

Abstract

Accurate student attendance tracking is essential in academic environments, yet traditional methods remain inefficient and vulnerable to manipulation. This research presents a classroom attendance system based on facial recognition that integrates YOLO for multiperson face detection and the SFace model from the DeepFace framework for feature extraction and identity matching. A key contribution of this study is the implementation of an Active Learning mechanism that enables the system to update its embedding Database using user-provided corrections, enabling continuous adaptation to real classroom conditions. The system was developed as a Python-based desktop application and evaluated using 38 group images captured with various devices under uncontrolled lighting, diverse head poses, occlusion, and different classroom densities. Performance was assessed using accuracy, False Rejection Rate (FRR), and False Acceptance Rate (FAR) across two scenarios: before and after Active Learning. Experimental results show a substantial improvement after the learning process, with accuracy increasing from 52.0% to 96.6%, while maintaining a low FAR of 0%. These findings demonstrate that Active Learning effectively enhances recognition performance by enriching the embedding Database with real-world facial variations that do not present during initial registration. Overall, the proposed system highlights the importance of integrating Active Learning into face recognition–based attendance applications to improve robustness and adaptability in unconstrained multiperson classroom environments.

Keywords: *Face Recognition, YOLO, DeepFace, Active Learning, Automated Attendance, Deep Learning.*

1. Introduction

Student attendance is a critical component of educational management. Conventional manual methods, such as paper-based lists or roll calls, are not only time-consuming and labour-intensive but also vulnerable to fraud, including proxy attendance [1]. In recent years, biometric technologies have emerged as a more secure and efficient alternative. Among these, face recognition has gained significant traction due to its non-intrusive nature and the uniqueness of facial physiological characteristics, which are difficult to forge [2]. Automated attendance via face recognition can streamline administrative processes, reduce human error, and enhance security in academic environments.

The advancement of deep Learning has revolutionized the field of computer vision, including face recognition. Modern object detection frameworks such as You Only Look Once (YOLO) have become de facto standards for real-time applications due to their remarkable balance between speed and accuracy [3][4]. Concurrently, face recognition libraries like DeepFace provide access to state-of-the-art models, such as SFace, which excels at generating robust, discriminative facial embeddings [5]. SFace employs a sigmoid-constrained hypersphere loss function that mitigates overfitting on noisy data, making it particularly suitable for real-world scenarios where image quality varies [6].

Recent studies have demonstrated the feasibility of multi-face recognition in classroom settings. For instance, Chawla et al. [7] implemented a system using MTCNN for face detection and FaceNet for recognition, achieving 91.9% detection accuracy in real classroom environments. However, their system lacks an adaptive learning mechanism, making it static and potentially vulnerable to performance degradation over time as student appearances change. Our work addresses this limitation by integrating Active Learning, enabling continuous improvement from user feedback. Traditional face recognition systems are typically static, trained on a fixed dataset, and their performance tends to degrade when confronted with unseen variations. This limitation underscores the need for systems that can adapt continuously to changing conditions and improve over time.

To address this, we integrate Active Learning into the attendance system. Active Learning is a machine learning paradigm in which the model selectively queries the user to label data points about which are most uncertain [8]. Recent comprehensive surveys on Deep Active Learning highlight its effectiveness in reducing annotation costs while maintaining model performance in various computer vision tasks, including face recognition [9]. These surveys categorize Active Learning approaches into uncertainty-based, representative-based, and hybrid methods, providing a theoretical foundation for our implementation. In our context, when the system encounters a face it cannot recognize with high confidence, it prompts the user to verify. The correction is then used to update the facial embedding Database, allowing the system to "learn" from real-world interactions and gradually improve its recognition capability. This approach is especially valuable in educational settings where student appearances may change (e.g., hairstyle, accessories) and where lighting and camera conditions are inconsistent.

Despite the growing body of research on face recognition for attendance, few studies have explored integrating Active Learning in multiperson, unconstrained classroom environments. Most existing systems rely on pre-trained models without mechanisms for continuous adaptation, limiting their long-term usability [7]. Moreover, the use of YOLO for face detection combined with SFace for recognition in an Active Learning framework remains underexplored in the context of academic attendance.

Therefore, this research aims to bridge this gap by designing and evaluating an adaptive attendance system that leverages YOLO for face detection, SFace for recognition, and Active Learning for continuous improvement. The specific objectives of this study are:

a. To design and implement a prototype multiperson attendance system that integrates YOLO and DeepFace (SFace) to process group images in real classroom scenarios.
b. To evaluate the system's performance in detecting and recognizing faces under varying environmental conditions.
c. To analyze the effectiveness of the Active Learning mechanism in enhancing recognition accuracy and adaptability over time.

Through this work, we demonstrate how adaptive Learning can make face recognition-based attendance systems more robust, practical, and scalable for real-world educational deployment.

2. Literature Review

This chapter presents a comprehensive review of the foundational research and technological advancements relevant to the development of an automated attendance system based on face recognition. The review is structured to systematically examine the key components and methodologies that underpin such a system, beginning with face detection in multiperson environments and then moving to face recognition models designed for unconstrained settings. Subsequently, the discussion shifts to existing automated attendance systems that leverage biometric recognition, highlighting their architectural frameworks and operational challenges. Finally, the review explores the emerging paradigm of Active Learning in computer vision, focusing on its potential to improve data efficiency and model adaptability in face recognition applications. By synthesizing insights from these interconnected domains, this chapter aims to identify research gaps and contextualize the contributions of this study within the broader landscape of intelligent attendance solutions.

2.1. Face Detection in Multiperson Environments

Traditional face detection methods, such as the Haar Cascade Classifier, have been widely used in constrained multiperson scenarios. For instance, [10] applied Haar Cascade combined with VGG16 for prisoner monitoring in jail settings, achieving 87% accuracy. However, Haar Cascades are known to be less robust to occlusion and lighting variations than modern deep learning-based detectors like YOLO. Meanwhile, [11] implemented YOLOv5 for multiperson face detection in crowd surveillance, achieving 94.5% detection accuracy with better real-time performance than Haar Cascade-based approaches. This study illustrates the evolution from traditional methods to more robust deep learning-based approaches in dynamic and complex scenarios.

2.2. Face Recognition Models for Unconstrained Settings

Face recognition in unconstrained settings must handle variations in pose, lighting, and occlusion. While systems perform well on controlled datasets like LFW [1], accuracy drops in real-world scenarios with non-frontal faces or limited samples per person [12]. A key challenge is frontal-to-profile matching, which becomes difficult under extreme pose variations. Research [12] on the ITB Frontal Profile Limited Dataset (IFPLD), which contains only one frontal and one 90° profile image per person, explored solutions for such data-scarce, open-set scenarios. For face verification, their best model combined transfer learning, fine-tuning on a similar profile dataset, and SimCLR contrastive learning, achieving 93% accuracy. For open-set identification, a Prototypical Network trained under an N-way-k-shot setup (with data augmentation) outperformed a Siamese network by up to 17%. A critical finding was that moving from 1-shot to k-shot learning ($k \geq 3$) was essential for performance, as accuracy was near-random without augmentation. This highlights the importance of few-shot Learning and robust augmentation for practical systems that must generalize from very few enrolment images.

2.3. Automated Attendance Systems Based on Face Recognition

Automated attendance systems have become a significant technological innovation in organizational management, with face recognition emerging as a leading method due to its non-intrusive, efficient, and secure nature. These systems primarily serve two core functions: identification (1:N matching to determine who a person is) and verification (1:1 matching to confirm if a person is who they claim to be). A systematic review by Anshari et al. [13] synthesizes the current landscape of face recognition in attendance systems. The review highlights that while traditional methods such as password entry or RFID cards are still used, biometric systems, particularly face recognition, offer significant advantages in preventing proxy attendance and enhancing security. The study maps the common architectures of such systems, which typically involve modules for face detection, face alignment, feature extraction, and finally, recognition (either identification or verification). It notes that Convolutional Neural Networks (CNNs) have become the dominant feature extractors, surpassing earlier methods such as Eigenfaces and Local Binary Patterns (LBP) in accuracy, especially in unconstrained environments.

However, the review also identifies critical challenges for real-world deployment. A primary concern is performance under sub-optimal conditions, including variations in lighting, pose, low-resolution images, and occlusions such as masks or glasses. These factors directly impact the system's reliability. Another significant challenge is scalability and computational cost, as systems must perform rapid and accurate comparisons against a growing database of enrolled individuals. Furthermore, the review underscores the need for robust liveness detection to prevent spoofing attacks using photographs or videos. The study concludes that an ideal attendance system must balance high accuracy, real-time processing speed, and robustness to real-world variations. It points to a growing trend of integrating cloud computing for scalability and employing hybrid models that combine different algorithms to improve overall system resilience. The findings of this review provide a crucial framework for designing a practical attendance system, emphasizing that success depends not only on the core recognition algorithm but also on effectively addressing these deployment-stage challenges.

2.4. Active Learning in Computer Vision and Face Recognition

Active Learning (AL) is a critical paradigm in machine learning that minimizes the costly, time-consuming process of data annotation by iteratively selecting the most informative samples for labelling by an oracle (e.g., a human annotator) [14]. Its core principle is that a model can achieve comparable or superior performance with a significantly smaller, yet intelligently curated, dataset than with a large, randomly labelled set [14].

With the rise of data-hungry deep learning models, Deep Active Learning (DeepAL) has emerged as an essential methodology. A standard DeepAL framework operates cyclically: (1) a model is trained on the current labeled set, (2) an acquisition function scores all samples in a large unlabeled pool based on their potential informativeness, and (3) a batch of the highest-scoring samples is selected, labeled, and added to the training set for the next cycle [14]. This process prioritizes data-centric AI, shifting the focus from solely model architecture design to strategic data iteration.

Key acquisition strategies in DeepAL fall into three main categories [14]:

a. Uncertainty-based Methods: These select samples where the current model is most uncertain, often measured by metrics such as prediction entropy and margin, or via Bayesian approaches like Monte Carlo Dropout.
b. Diversity-based Methods: These aim to select a representative subset of the data distribution to prevent redundancy. Core-set selection and adversarial methods, such as VAAL, are prominent examples.
c. Hybrid Methods: These combine uncertainty and diversity to select samples that are both informative for the model and representative of the underlying data manifold, as seen in BatchBALD and BADGE.

Active Learning has been successfully applied across major computer vision tasks. In image classification and object detection, it reduces the need for exhaustive bounding box annotation. For semantic segmentation, where pixel-wise labelling is extremely expensive, region-based AL methods (e.g., superpixel selection) have shown great efficiency. In video recognition and person re-identification, AL helps manage the complexity of temporal data and the challenge of identifying individuals across non-overlapping cameras with minimal labelled tracklets. Despite its promise, integrating AL into real-world face recognition systems presents unique challenges and opportunities. While general DeepAL principles apply, face-specific factors must be considered: extreme pose/lighting variations create hard samples ideal for uncertainty sampling, the need for open-set identification aligns with diversity-based selection to capture intra-class variance, and the high cost of labeling facial identities makes AL's efficiency argument particularly compelling. Recent industrial platforms like YMIR [15] demonstrate the practical viability of a DeepAL-powered, data-centric pipeline for vision tasks, underscoring its potential to revolutionize the development of robust face recognition systems by maximizing model performance per unit of annotation effort.

2.5. Research Gap and Contribution

The literature demonstrates significant advancements in core technologies for automated attendance systems, yet a practical integration gap persists. While face recognition models and active learning strategies have matured independently, their combined application to solve the critical, real-world problem of efficient face database curation for attendance remains underexplored. Specifically, existing approaches often overlook the sequential, open-set nature of attendance environments, where the system must continuously and cost-effectively decide which new, unlabeled face captures are most valuable for expanding and refining the enrolled gallery.

To bridge this gap, this research contributes a novel Active Learning framework specifically designed for face recognition in attendance systems. The proposed work develops and validates a hybrid query strategy that intelligently selects informative samples, such as low-confidence matches or appearances of new individuals, from the operational data stream for human annotation. By embedding this AL mechanism into the attendance pipeline, the system aims to achieve robust recognition performance while substantially reducing the manual effort required for initial enrollment and continuous database updates. This contribution provides a practical, data-efficient methodology for deploying and maintaining accurate attendance systems in dynamic, multiperson settings.

3. Methods

This research is experimental and was conducted at the Electrical Engineering Computational Laboratory, Universitas Mulawarman. The research methodology follows a systematic flow as illustrated in Figure 1.

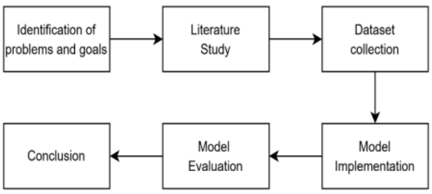


Fig 1. Research Stage Flowchart

3.1. Data Collection

The data used in this study consisted of two distinct sets: an initial enrollment database of formal student portraits and a single group photograph used for system testing and active Learning. This two-stage data configuration was designed to evaluate the system's ability to recognize individuals in unconstrained group settings, using only pre-enrolled formal portraits as references. The initial enrollment data were obtained from the university's official academic portal, consisting of formal portrait photographs of 10 subjects. These images were captured under controlled conditions with frontal poses, uniform lighting, and minimal occlusion, providing high-quality facial representations for generating baseline embeddings.

Fig 2. Example of Formal Student Portraits

An example of the group photograph used for system testing is shown in Figure 3 below. This image depicts a real-world scenario in which the system must detect and recognize multiple faces simultaneously within a single frame, despite variations in pose, facial expression, lighting, and camera distance.

Fig 3. The group photograph is used for enrollment, testing, and active learning phases.

This group photograph illustrates the type of images used in the evaluation, containing both enrolled and unregistered individuals under unconstrained conditions. A total of 38 group photographs were collected for testing, with each image containing between 8 to 20 faces. These images exhibit natural variations in lighting, facial pose, expression, occlusion, and image quality, representing realistic multiperson scenarios commonly encountered in academic and social settings. By using such diverse group photographs for detection, recognition, and Active Learning evaluation, this study assesses the system's robustness and adaptability in challenging, real-world environments.

3.2. System Architecture

To provide a clearer overview of the operational workflow, a process flow diagram of the attendance system is presented in Figure Y. This diagram illustrates the sequential stages, including face detection using YOLO, embedding extraction using SFace, identity matching, and the Active Learning correction mechanism.

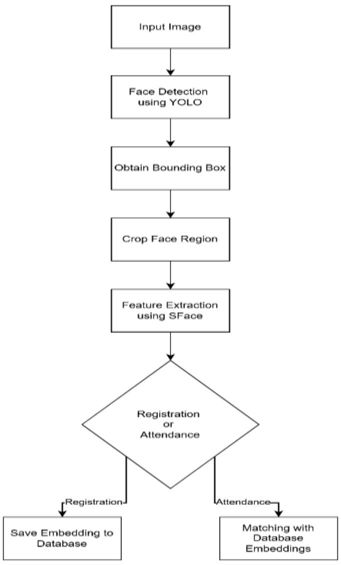


Fig 4. Flowchart System

The flowchart presented delineates the operational pipeline of a dual-purpose face recognition system designed for both user registration and attendance verification. The process commences with an Input Image, which is subsequently analyzed by a YOLO model for Face Detection to locate facial regions. The coordinates from this detection are used to Obtain a Bounding Box, enabling the system to Crop the Face Region precisely. This isolated facial image is then processed using the SFace model to generate a unique, high-dimensional feature vector, known as an embedding. At this critical juncture, the system's workflow diverges based on its operational mode. In the Registration mode, the newly generated embedding is stored directly in the Database for future reference. Conversely, in the Attendance mode, the system performs a matching procedure by comparing the live embedding against all stored templates in the Database. This matching process specifically relies on calculating the Cosine Similarity between the feature vectors to determine their degree of resemblance. The resulting similarity score is finally used for Identity Verification, where it is evaluated against a predefined threshold to either accept or reject the claimed identity, thereby completing the attendance logging or access control procedure.

3.3. Preprocessing and Face Detection (YOLO)

In the preprocessing stage, facial regions in both formal portraits and group photographs were detected using the YOLOv8 face detection algorithm. Upon detection, each face was cropped using its bounding box coordinates. No additional augmentation or color-space conversion was applied. The resized facial images were then directly passed to the SFace model for embedding generation.

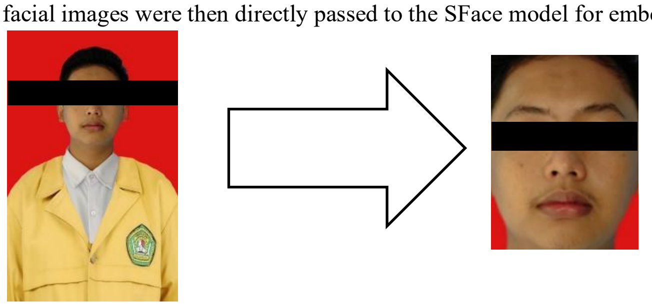


Fig 5. Process Crop Face

For face detection in this study, we employed YOLOv8-Face [16], a specialized variant of YOLOv8 pre-trained on the WIDERFace dataset for face detection. This model was selected for its demonstrated robustness and efficiency in real-time face localization. According to the official repository, YOLOv8-Face achieves state-of-the-art performance on the WIDER Face benchmark, with the YOLOv8x-Face variant reaching Average Precision (AP) scores of 96.33%, 95.16%, and 85.80% on the Easy, Medium, and Hard validation sets respectively. For this research, we selected the YOLOv8n-Face model, the smallest variant in the family, which balances computational efficiency with adequate accuracy, achieving AP scores of 93.79%, 91.82%, and 79.38% on the same benchmarks. Results by line can be seen in Table 1.

Table 1. YOLOv8-Face Model Performance on WIDER Face Benchmark

Model Variant	Easy Set AP	Medium Set AP	Hard Set AP	Parameters	GFLOPs
YOLOv8n-Face	93.79%	91.82%	79.38%	3.0M	8.1
YOLOv8s-Face	95.13%	93.62%	82.90%	11.1M	28.4
YOLOv8m-Face	95.73%	94.47%	84.55%	25.8M	78.7
YOLOv8l-Face	96.26%	95.03%	85.43%	43.6M	164.8
YOLOv8x-Face	96.33%	95.16%	85.80%	68.1M	257.4

Source: Lindevs (2023), YOLOv8-Face GitHub repository

The YOLOv8n-Face model was particularly suitable for our experimental setup, which primarily involves clear, frontal face images under controlled lighting conditions, scenarios that align with the "Easy" category of the WIDERFace benchmark, where the model achieves its highest performance. This model selection ensures real-time processing capability while maintaining detection accuracy appropriate for our research context.

3.4. Feature Extraction with SFace

Feature extraction is the process of transforming facial images into compact numerical representations, known as embedding vectors. In this study, the SFace model generates discriminative 128-dimensional embeddings from detected face regions. As illustrated in Figure 6, this transformation enables efficient comparison and identification of individuals through mathematical similarity measures.

34.1. From Pixels to Embeddings

The SFace model receives cropped face regions from the YOLOv8 detection module and preprocesses them into standardized 112×112-pixel images. Through its ResNet-based architecture enhanced with sigmoid-constrained hypersphere loss, the model extracts deep facial features and encodes them into a fixed-length vector of 128 floating-point values. This 128-dimensional embedding compactly represents each individual's unique facial characteristics while remaining invariant to variations in pose, expression, and lighting within reasonable bounds.

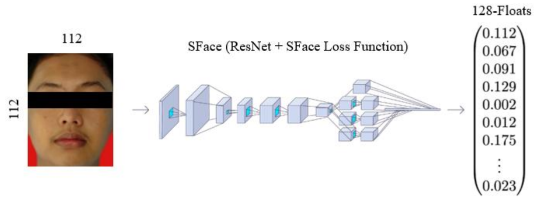


Fig 6. Feature Extraction Process with SFace

3.4.2. Recognition Pipeline

The attendance system implements a comprehensive recognition pipeline that integrates database matching with active learning mechanisms to ensure robust identification. When a face is detected and processed into its 128-dimensional embedding, the system first compares this vector against all enrolled embeddings stored in the Database using cosine similarity [17]. This comparison yields similarity scores that indicate how closely the new face matches each registered identity. Following the initial matching phase, the system incorporates an active learning component where uncertain matches, specifically those with similarity scores falling between predetermined acceptance and rejection thresholds, are flagged for human verification. This creates a continuous feedback loop where ambiguous cases are resolved through human intervention, and the corrected labels are used to retrain and refine the recognition model, thereby progressively enhancing system accuracy and adaptability over time.

3.4.3. Cosine Similarity

Cosine similarity is a method for measuring the similarity between two objects expressed as vectors [17]. In the context of face recognition, each face is represented as an embedding vector extracted from a CNN model such as ResNet-50 enhanced with SFace. These vectors are then compared using cosine similarity to determine whether two faces belong to the same identity. The basic formula is:

$$cos\ cos\ (A, B) = \frac{\sum_k (A_k \times B_k)}{\sqrt{\sum_k A_k^2} \times \sqrt{\sum_k B_k^2}} \quad (1)$$

The cosine similarity method measures similarity based on the angle between two vectors, not their magnitudes, making it suitable for normalized data such as face embeddings in hyper-spherical space [18]. The output value ranges from −1 to 1, with 1 indicating perfect similarity, 0 indicating no similarity, and −1 indicating opposite direction. In a face verification system, a threshold is used to determine whether a match occurs. For example, if the cosine similarity value between two embeddings exceeds 0.7, the two faces are considered to belong to the same person. This approach is not only computationally efficient but also aligns with loss functions such as SFace, which optimize angular distances on a hyperspherical manifold. The use of cosine similarity has been shown to improve accuracy across various face recognition benchmarks, including LFW, MegaFace, and IJB-C, due to its ability to handle variations in pose, lighting, and expression while maintaining consistency in similarity measurements.

3.4.4. Deep Face Framework

The Deep Face Framework [19] is an open-source platform designed to streamline the integration of various deep learning-based face recognition models. This framework supports multiple state-of-the-art models including FaceNet, FaceNet512 [20], ArcFace [21], VGG-Face, OpenFace, SFace [6], and DeepID [22], enabling the deployment of face recognition technology across diverse applications without the need to develop algorithms from scratch. Trained on well-established datasets such as Labelled Faces in the Wild (LFW) and YouTube Faces (YTF), DeepFace demonstrates robust performance under various testing conditions.

This framework was selected for the present study due to its flexibility and ability to support multiple face recognition models through a straightforward interface, allowing rapid integration into the developed system, particularly in data-constrained scenarios such as having only one image per individual. The DeepFace Framework also supports face detection with efficient models like YOLOv8, ensuring reliable face localization before feature extraction by models such as SFace. This combination ensures the system remains effective even with limited data, a primary challenge in academic face recognition research.

3.4.5. SFace Model Specifications

SFace (Sigmoid-constrained Hypersphere Loss) is a novel loss function designed to enhance the robustness of deep face recognition models by adopting a moderated optimization strategy for intra-class and inter-class distances [6]. Unlike traditional loss functions that enforce strict minimization of intra-class distance and maximization of inter-class distance, SFace introduces adaptive constraints controlled by sigmoid-based gradient rescaling functions. These functions enable the model to effectively optimize clean samples while reducing overfitting to noisy or low-quality training data.

The loss operates on a hypersphere manifold, where deep features and classifier weights are normalized and optimized using angular distance. SFace consists of two components: an intra-class loss term that pulls features toward their corresponding class center, and an inter-class loss term that pushes features away from other class centers. The key innovation lies in using sigmoid functions to rescale the gradients for both terms, enabling precise control over the optimization speed. This results in a balanced training dynamic in which clean samples converge faster, while potentially mislabeled or ambiguous samples are optimized more cautiously.

SFace is applied on top of a well-established CNN architecture, specifically ResNet-50, which has been widely used in image classification research including in the fashion domain [23]. The ResNet-50 architecture is chosen for its ability to address the vanishing gradient problem via residual connections, enabling the training of deeper, more effective networks. In implementation, SFace does not modify the basic structure of ResNet-50; instead, it replaces the loss function used in the final embedding layer. Thus, SFace leverages the feature extraction capability of ResNet-50 while applying a more moderate and noise-tolerant optimization mechanism through its sigmoid-constrained formulation

3.5. Active Learning Mechanism

The system implements a human-in-the-loop active learning mechanism, following a pool-based paradigm in which the model selects the most uncertain samples for human verification to optimize annotation efficiency [24]. This iterative workflow (Figure 7) enables continuous model refinement in real-world attendance settings.

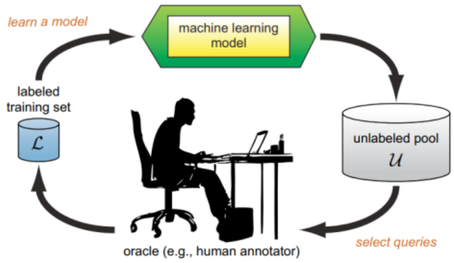


Fig 7. Human-in-the-Loop Active Learning Workflow

In operation, the trained face recognition model processes unlabeled face captures from daily attendance sessions. Uncertainty sampling is applied to identify low-confidence predictions such as challenging cases with non-frontal poses or poor lighting which are then queued for human review. After correction, these samples are added to the training dataset and the model is fine-tuned. This cycle repeats, allowing the system to progressively adapt to new identities and appearance variations with minimal manual labeling effort.

3.6. System Evaluation

The face recognition system is evaluated using standard performance metrics to ensure reliability and robustness in real-world applications. The evaluation follows established protocols used in biometric systems research, incorporating measures such as accuracy, false acceptance rate (FAR), and false rejection rate (FRR) [25]. Accuracy measures the overall correctness of the system and is calculated as the ratio of successful identifications to the total number of trials:

$$\text{Accuracy } (\lambda) = \frac{number\ of\ successful\ trials}{number\ of\ successful\ trials} \times 100\% \tag{2}$$

False Acceptance Rate (FAR) is the frequency with which the system incorrectly accepts an unauthorized user. It is computed as:

$$\text{FAR} = \frac{number\ of\ successful\ trials}{total\ trials} \times 100\% \tag{3}$$

False Rejection Rate (FRR) measures how often the system incorrectly rejects an authorized user. The formula is:

$$\text{FRR} = \frac{number\ of\ false\ trials}{total\ trials} \times 100\% \tag{4}$$

In the experimental setup, tests were conducted with varying distances between the camera and the user's face to evaluate the effect of capture distance on system performance. The camera was placed at multiple positions to simulate real-world usage scenarios, and images were captured under diverse conditions including changes in lighting and natural facial expressions. The performance of the face recognition system was assessed using standard evaluation metrics: accuracy, False Acceptance Rate (FAR), and False Rejection Rate (FRR). This setup is designed to validate the system's robustness for secure access control applications under variable environmental conditions.

4. Result and Discussion

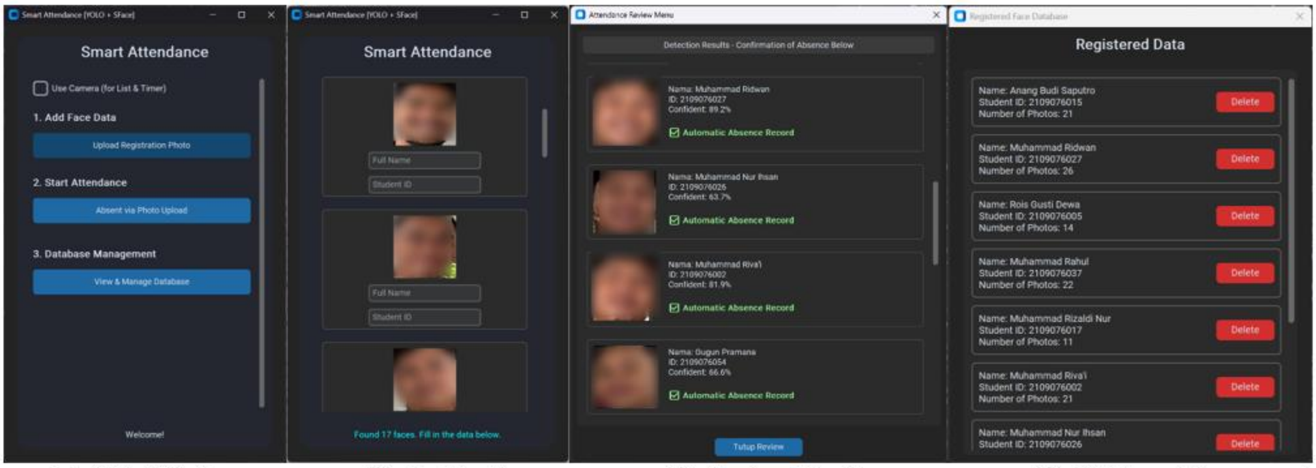
The performance of the implemented face recognition system is quantified using standard biometric evaluation metrics. The results for accuracy, False Acceptance Rate (FAR), and False Rejection Rate (FRR), which are critical for determining the system's reliability and security, are detailed in the following sections.

4.1. System Implementation

The system prototype was successfully implemented as a desktop application with a graphical user interface (GUI) built using CustomTkinter. The GUI consists of a main menu, a registration menu, an attendance menu, and a database management menu. The system can process group images, detect faces with YOLO, and recognize them with adequate accuracy.

4.1.1. Graphical User Interface

The graphical user interface (GUI) of the developed system was implemented using CustomTkinter. The prototype provides several main menus, including registration, attendance processing, and database management. The figure below illustrates the main interface of the system prototype.



(a) **Main Windows** (b) **Registration** (c) **Review Attendance** (d) **Database Management**

Fig 8. User Interface of the Face Recognition-Based Attendance System. (a) Main Window as the central control panel; (b) Face Registration Interface; (c) Attendance Review and Confirmation Interface; (d) Database Management Interface.

a. Main Window
Serves as the central control dashboard, providing access to all core system features through a clear vertical navigation panel on the left side. The navigation buttons include: "Face Registration" for enrolling new students, "Take Attendance" for processing classroom photos, "Manage Database" for data administration, and "Activate Camera" as an alternative option for real-time registration or attendance capture using a webcam. This interface features a clean, intuitive layout to guide users through the main workflow.
b. Face Registration Interface
When a user uploads an individual or group photo for registration, this interface automatically detects all faces using YOLO and displays them in an organized grid layout. Each detected face is presented in a separate box containing: (1) a *preview* of the cropped face, (2) two input fields for Full Name and Student ID, and (3) a "Save" button to store the facial embedding into the Database. This design ensures a structured registration workflow and minimizes input errors before data storage.
c. Attendance Review and Confirmation Interface
After processing a classroom photo, this interface displays the face recognition results comprehensively. For each detected face, the system displays a face preview, the suggested identity, and the corresponding confidence score (as a percentage). The *active learning* feature is implemented here: when the confidence score falls below 79%, the system flags it and provides a dropdown for the user to manually correct it. The corrections provided are then used to update the embedding Database automatically. A "Confirm All" button is available to finalize the process and export attendance data.
d. Database Management Interface
Provides administrative functions for managing stored data. This interface displays a tabular list of all registered students, with columns including Name, Student ID, and the number of embedding samples stored for each profile. A "Delete" button is provided in each row to remove entries that are no longer needed. Additionally, an option to export the entire Database to CSV format is available for backup or further analysis. This interface supports ongoing maintenance and refinement of the stored facial data. Overall, the system's user interface is designed with a consistent colour scheme, clear typography, and linear navigation, enhancing the user experience and ensuring efficient operation in real academic environments.

4.1.2. Detection and Recognition Performance

The YOLO model demonstrated fast, accurate face detection in group images. The SFace model, via DeepFace, successfully generated discriminative embeddings that distinguish different identities. Initial test results (before Active Learning) showed good accuracy, but there were still identification errors, especially for faces with challenging poses or lighting.

4.1.3. Effectiveness of Active Learning

A comparative evaluation of the Before and After Active Learning scenarios demonstrates that the active learning mechanism provides measurable improvement, even in highly unconstrained group-image environments. From 38 collected group images, 8 (approximately 21%) were selected as the test set, representing three classroom density conditions: high-density (18–20 faces), medium-density (12–17 faces), and low-density (8–11 faces). This distribution ensures comprehensive evaluation across real-world scenarios, from crowded to low-occlusion settings. The group images were captured using different devices, leading to inconsistencies in resolution, lighting, sharpness, colour, and subject appearance frequency. These factors, combined with poor initial image quality (e.g., motion blur, low light, and small faces), contributed to lower initial accuracy and a high FRR, preventing reliable automatic recognition.
After user intervention during the active learning phase, the system updated its embedding Database with new samples from real classroom photos, capturing facial variations absent from the original registration portraits. Despite imbalanced subject representation (some individuals appeared in many photos, others only once), the system showed measurable improvement. This highlights the role of Active Learning in enriching and embedding diversity, even with varying devices and subject frequencies. The quantitative results of this improvement are shown in Table 2.

Table 2. System Performance Comparison Before and After Active Learning

Evaluation Metric	Before Active Learning	After Active Learning	Improvement

Accuracy (%)	52	96,6	+44,6%
FRR (%)	48	3,4	-44,6
FAR (%)	0	0	0%

The results clearly demonstrate that the Active Learning mechanism significantly improves system performance, even in highly unconstrained group image environments. Accuracy substantially increased after the system incorporated new embeddings extracted from real classroom photographs, which captured variations in pose, scale, illumination, camera quality, and facial clarity conditions absent from the original single-image registration dataset. Unlike the initial setup, which relied on a single portrait image per subject, the updated embedding Database now contains richer, more diverse representations of each registered individual. This enhancement greatly improves the system's discriminative ability, as evidenced by the sharp reduction in FRR from 48% to 3.4%, while FAR remained stable at 0%. This confirms that Active Learning strengthened recognition performance without increasing misclassification risk.

Certain limitations persist, however, particularly in cases of extremely small facial regions, heavy occlusion, non-frontal poses, and low-resolution inputs. These challenges align with observations in [8], which note that Active Learning is most effective when difficult samples fall outside the initial training distribution. As the system accumulates corrected samples, it becomes increasingly robust and better adapted to real classroom environments. Overall, the experiment verifies that Active Learning is highly valuable for improving face recognition in unconstrained multiperson attendance scenarios, even with diverse photo qualities due to varying cameras, lighting, and layouts. Future improvements may include collecting higher-quality group images, increasing the number of samples per subject, or employing a more powerful recognition backbone to enhance robustness and accuracy further.

5. Conclusion

This research successfully developed and implemented a classroom attendance system that integrates YOLO for face detection and the SFace model from DeepFace for recognition. The system demonstrates the capability to process multiperson group images and perform semi-automatic attendance marking under diverse and unconstrained classroom conditions. The combination of YOLOv8n-face and SFace provided reliable detection and recognition performance, although accuracy was influenced by environmental factors such as lighting, camera quality, face orientation, and occlusion. Another challenge arose from an imbalanced embedding database, caused by unequal subject appearance frequencies in the training photos. Most significantly, integrating the Active Learning feature led to a substantial improvement in system performance. By incorporating user-corrected embeddings, the system achieved a marked increase in accuracy and a significant reduction in the False Rejection Rate (FRR), enabling it to adapt to real-world variations and grow increasingly robust over time. For future work, potential improvements include employing more advanced detection and recognition models, collecting more balanced datasets, integrating anti-spoofing mechanisms, and extending the system into a mobile or web-based platform for broader deployment.

Acknowledgement

The authors would like to express their sincere gratitude to the supervising lecturers for their continuous guidance, constructive feedback, and valuable insights throughout the completion of this research. Appreciation is also extended to the Electrical Engineering Computational Laboratory, Universitas Mulawarman, for providing the facilities and technical support necessary for the implementation and evaluation of this study.

References

[1] M. M. Al-dabbagh, "A Review of Students Attendance Management Systems," vol. 10, no. 3, pp. 136–145, 2025.
[2] N. M. Hassan, M. A. Moussa, and M. H. M. Mahmoud, "CNN and Adaboost fusion model for multiface recognition based automated verification system of students attendance," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 35, no. 1, pp. 133–139, 2024, doi: 10.11591/ijeecs.v35.i1.pp133-139.
[3] P. Jiang *et al.*, "ScienceDirect A Review of Yolo Algorithm Developments A Review of Yolo Algorithm Developments," 2022, doi: 10.1016/j.procs.2022.01.135.
[4] G. Jocher, J. Qiu, and A. Chaurasia, "Ultralytics YOLO," 2023. [Online]. Available: https://ultralytics.com
[5] S. Serengil and A. Ozpinar, "A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules," *J. Inf. Technol.*, vol. 17, no. 2, pp. 95–107, 2024, doi: 10.17671/gazibtd.1399077.
[6] Y. Zhong, W. Deng, J. Hu, D. Zhao, X. Li, and D. Wen, "SFace: Sigmoid-Constrained Hypersphere Loss for Robust Face Recognition," *IEEE Trans. Image Process.*, vol. 30, no. 10,4,5,8, pp. 2587–2598, 2021, doi: 10.1109/TIP.2020.3048632.
[7] A. P. Chawla, S. E. Juliet, A. Suman, A. Khan, and G. Manikandan, "Deep Learning based Multi-face Recognition System for Automatic Attendance Registering in Classrooms," in *Proceedings of the 6th International Conference on Deep Learning, Artificial Intelligence and Robotics (ICDLAIR 2024)*, Atlantis Press, 2025, pp. 203–214. doi: 10.2991/978-94-6463-740-3_18.
[8] Y. Gal, R. Islam, and Z. Ghahramani, "Deep Bayesian active learning with image data," *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 3, pp. 1923–1932, 2017.
[9] D. Li, Z. Wang, Y. Chen, R. Jiang, W. Ding, and M. Okumura, "A Survey on Deep Active Learning: Recent Advances and New Frontiers," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 36, no. 4, pp. 5879–5899, 2025, doi: 10.1109/TNNLS.2024.3396463.
[10] G. S. M. Diyasa, A. Fauzi, M. Idhom, and A. Setiawan, "Multi-face Recognition for the Detection of Prisoners in Jail using a Modified Cascade Classifier and CNN," *J. Phys. Conf. Ser.*, vol. 1844, no. 1, p. 12005, Mar. 2021, doi: 10.1088/1742-6596/1844/1/012005.
[11] E. C. A. Mondesir, *Group Face Recognition : Identifying and Tracking Multiple Individuals*, no. Icciet 2024. Atlantis Press International BV. doi: 10.2991/978-94-6463-471-6.
[12] M. Djamaluddin, R. Munir, N. P. Utama, and A. I. Kistijantoro, "Open-Set Profile-to-Frontal Face Recognition on a Very Limited

Dataset," *IEEE Access*, vol. 11, pp. 65787–65797, 2023, doi: 10.1109/ACCESS.2023.3289923.

[13] A. Anshari, S. A. Hirtranusi, D. I. Sensuse, Kautsarina, and R. R. Suryono, "Face Recognition for Identification and Verification in Attendance System: A Systematic Review," in *2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2021, pp. 316–323. doi: 10.1109/COMNETSAT53002.2021.9530817.

[14] R. Takezoe, X. U. Liu, S. Mao, M. T. Chen, S. Zhang, and X. Wang, "Deep Active Learning for Computer Vision : Past and Future," pp. 1–18.

[15] P. X. Huang, W. Hu, and L. L. X. Wang, "YMIR : A Rapid Data-centric Development Platform for Vision Applications," no. NeurIPS, pp. 1–8, 2021.

[16] Lindevs, "Pre-Trained Yolov8-Face models," 2024. [Online]. Available: https://github.com/lindevs/yolov8-face

[17] A. Sanjaya, A. B. Setiawan, U. Mahdiyah, I. N. Farida, and A. R. Prasetyo, "PENGUKURAN KEMIRIPAN MAKNA MENGGUNAKAN COSINE SIMILARITY MEASUREMENT OF MEANING SIMILARITY USING COSINE SIMILARITY AND," vol. 10, no. 4, pp. 747–752, 2023, doi: 10.25126/jtiik.2023106864.

[18] Supiyanto and Sriyono, "Metode Cosine Similarity Untuk Mendeteksi Kemiripan Pada Dokumen Teks," *SAINS J. MIPA*, vol. 1, pp. 1–7, 2023.

[19] V. Tomar, N. Kumar, and A. R. Srivastava, "Single sample face recognition using deep learning: a survey," *Artif. Intell. Rev.*, vol. 56, no. 1, pp. 1063–1111, 2023, doi: 10.1007/s10462-023-10551-y.

[20] D. Sanberg, "Face Recognition using Tensorflow." [Online]. Available: https://github.com/davidsandberg/facenet

[21] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10_Part_1, pp. 5962–5979, Oct. 2022, doi: 10.1109/TPAMI.2021.3087709.

[22] S. Serengil and A. Ozpinar, "LightFace: A Hybrid Deep Face Recognition Framework," 2020. doi: 10.1109/ASYU50717.2020.9259802.

[23] A. Puspitasari, D. Sava, and D. Roliawati, "Indonesian Journal on Data Science Penerapan ResNet-50 CNN untuk Optimalisasi Klasifikasi pada Data Fashion," vol. 3, no. 1, pp. 1–12, 2025.

[24] A. Laurent, "Active Learning and Human Feedback for Large Language Models," 2025. Accessed: Dec. 27, 2025. [Online]. Available: https://intuitionlabs.ai/articles/active-learning-hitl-llms

[25] A. Jalaluddin and S. Mulyono, "MENGGUNAKAN METODE HAAR CASCADE CLASSIFIER DAN LOCAL BINARY PATTERN HISTOGRAM PADA AKSES," vol. 1, no. 1, 2023.