Asistencia: An Intelligent Attendance System

Course Name: Project Based Learning on CSE

Course Code: CSE 321

Submitted by:

Name	ID	Contribution
Jana Elsalhy	120220093	UI/UX & Frontend
Shiref Ashraf	120220099	Models
Ahmed Nagah	120220116	Backend & Database
Tasbih Neamatalla	120220117	Data Exploration & Preprocessing
Kareem Mazrou	120220118	Backend & Database

Submitted to:

Prof. Ahmed Gomaa

May 28, 2025

Table of Contents

List of Figures

1 Introduction

Time management in classrooms is very important, but even now, many institutions still rely on manual attendance methods. These traditional approaches are prone to errors and are not an efficient way to manage class time. So, we took it upon ourselves to address this problem—especially in today's fast-paced academic environment—where minimizing mistakes that affect both students and professors is critical.

Asistencia is our solution to this challenge: an attendance system that uses facial recognition to automatically record student presence. It leverages state-of-the-art facial recognition under real-world surveillance conditions. By integrating YOLOv11-based face detectors with existing CCTV infrastructure, Asistencia automatically identifies students as they enter the classroom and marks their attendance. The records are saved in the system, and in case of any unexpected errors, instructors can manually edit or add students. It also allows instructors to view or update any student's attendance with just a click.

In developing Asistencia, we have:

- Balanced simplicity and accuracy through a modular design that separates the face recognition service from the web frontend.
- Solved the challenges of noisy and low-resolution CCTV footage by selecting models and preprocessing steps that have proven effective on SCface data.
- Addressed dataset bias with fairness in mind, ensuring equitable performance across gender and age groups.
- Explored multiple architectures—including ResNet variants and several YOLO models—to develop a system that delivers both high accuracy and real-time performance.

2 Related Work

Face recognition under surveillance conditions, exemplified by the SCface dataset, remains a challenging task due to factors such as low image resolution, variations in illumination, pose differences, and occlusions. Several recent studies have proposed different strategies to tackle these challenges with varying degrees of success.

Aghdam et al. [1] addressed the problem of resolution mismatch between high-resolution gallery images and low-resolution probe images by proposing to downsample the gallery images to the resolution of the probes rather than applying super-resolution techniques to enhance the probe images. By training deep convolutional neural networks such as ResNet-50 and SENet-50 on large-scale datasets including MS-Celeb-1M and VGGFace2, their approach significantly improved recognition accuracy on the SCface dataset. They reported Rank-1 recognition rates ranging from 78.5% to nearly 100% depending on the camera distance, demonstrating the effectiveness of resolution matching combined with diverse and large training data. However, while this approach is simple and avoids the need for SCface-specific fine-tuning, it potentially sacrifices fine discriminative details due to downsampling. Additionally, the method does not explicitly address other complicating factors such as pose variation or occlusion, which are prevalent in surveillance imagery.

Building upon the need for improved feature extraction in low-resolution conditions, Mishra et al. [8] proposed the multiscale parallel deep CNN (mpdCNN), which captures facial features at multiple spatial scales simultaneously. This design is particularly suited to handle the sparsity and variability of low-resolution surveillance images. Their model demonstrated superior performance compared to traditional single-scale architectures such as the Inception network, which often suffer from overfitting in such contexts. Achieving an accuracy of 88.6% on SCface, the mpdCNN effectively balances feature representation and generalization. Nevertheless, this increased capability comes at the cost of architectural complexity and higher computational requirements. Furthermore, although the multi-scale approach improves robustness, it remains vulnerable to challenges such as extreme pose variations and occlusions.

Complementing these architectural improvements, Tuvskog [9] evaluated the performance of pretrained FaceNet models on SCface and similar datasets, highlighting significant accuracy degradation when models trained on primarily high-quality datasets encountered the low-quality, occluded, and variably posed images common in surveillance data. Her analysis emphasizes the importance of including low-quality or augmented images during training to improve robustness. While this work provides valuable insights into the factors affecting recognition accuracy, it does not propose new model architectures or training strategies. Additionally, the evaluations rely on pretrained models without extensive fine-tuning on the target surveillance datasets, limiting their practical effectiveness.

Both Aghdam et al. [1] and Mishra et al. [8] also noted the limited utility of super-resolution techniques in this domain. Although intuitively appealing for enhancing low-resolution images, super-resolution often introduces artifacts that can confuse recognition models and does not consistently improve accuracy on authentic surveillance images such as those in SCface. These observations suggest that approaches directly addressing resolution mismatch and multi-scale feature extraction are more effective in real-world scenarios.

Despite these advances, face recognition on SCface remains a formidable problem due to the extreme variability in image quality and subject appearance. Continued research efforts are required to develop models that generalize well across diverse conditions without heavy dependence on dataset-specific fine-tuning.

3 Methodology

3.1 System Overview

This system introduces an intelligent, vision-based attendance monitoring solution designed for university classrooms. By leveraging facial recognition powered by the YOLOv11 model and integrating it with existing CCTV camera infrastructure, the system automates the process of identifying students and recording their attendance in real time.

The core functionality is managed through an admin interface that enables login authentication, classroom selection, live camera feed monitoring, and real-time attendance marking. When students are recognized, they are automatically recorded in the attendance sheet. The admin can also manually add unrecognized students and generate statistical reports on student attendance performance.

The system's structure outlines the relationships between key entities such as `Student`, `Admin`, `Class`, and `CCTV Camera`. To provide a clear understanding of the system work-

flow, an activity diagram is included below (Figure 1), illustrating the step-by-step interaction between the admin and the system, from login to logout.

This approach minimizes manual effort, enhances accuracy, and facilitates data-driven decision-making for academic administration.

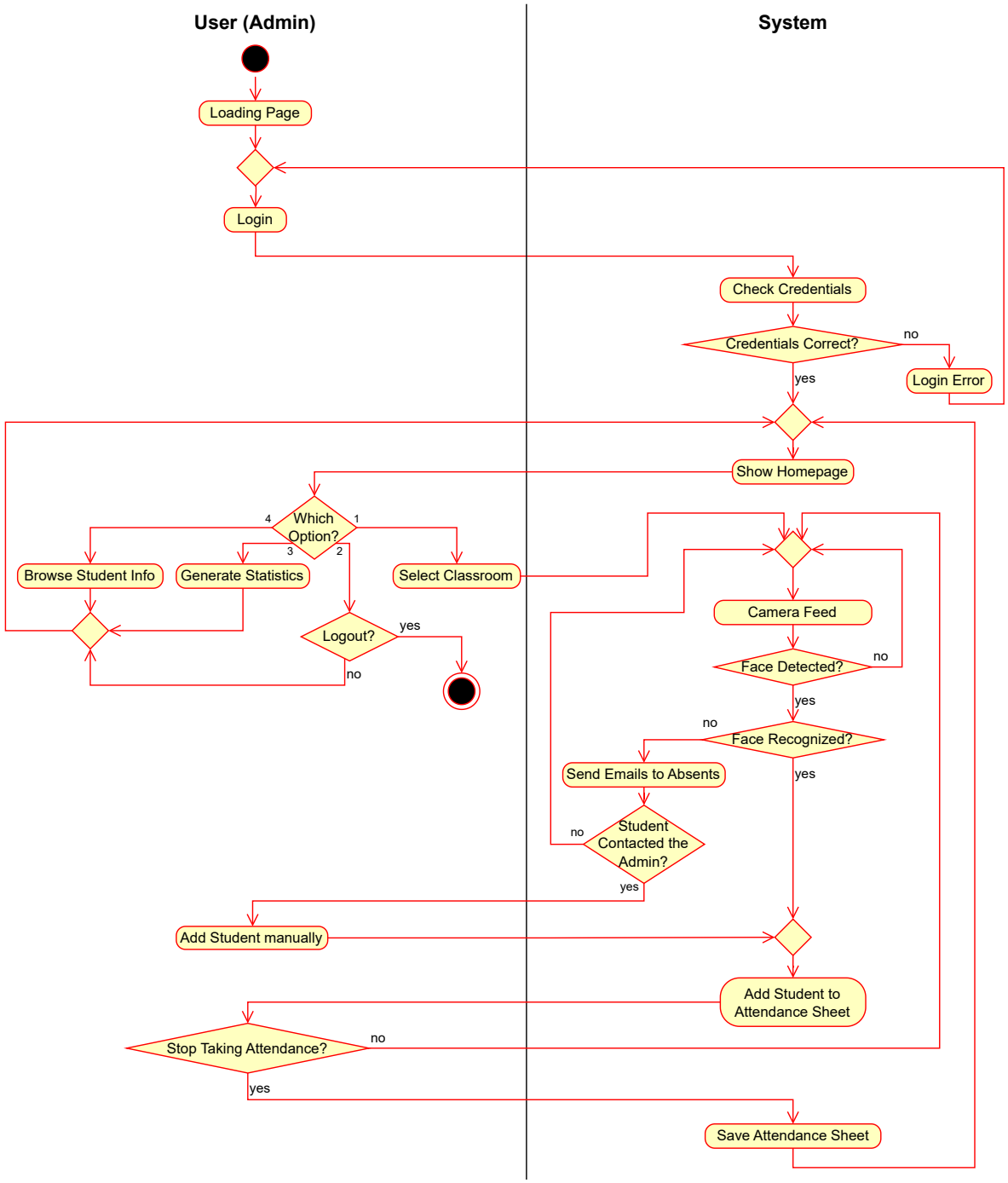
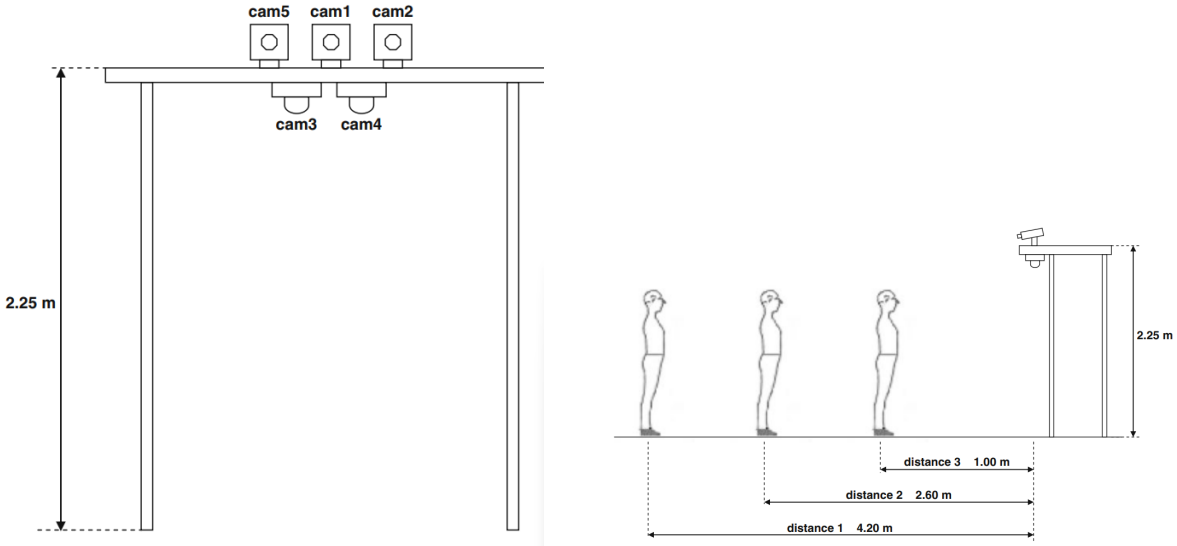


Figure 1: Activity Diagram of the Attendance System Workflow

3.2 Datasets and Processing

3.2.1 Dataset Description

In this study, we utilize the SCface (Surveillance Cameras Face) database, a publicly available dataset designed to evaluate face recognition algorithms in real-world surveillance scenarios [4]. The SCface dataset contains 4,160 images of 130 subjects captured using five commercially available surveillance cameras of varying quality and resolution as shown in Figure 2a. The images were collected in uncontrolled indoor environments with natural lighting and at three distinct distances (1.0 m, 2.6 m, and 4.2 m) to simulate typical surveillance conditionsas shown in Figure 2b.



(a) Camera Positions and Orientations

(b) Different distances of footage capture

Figure 2: SCface Setup

While SCface includes several types of images, such as high-quality frontal mug shots, infrared (IR) images, and various pose images, our study specifically uses only the visible spectrum surveillance images captured by the surveillance cameras shown in Figure 3. These images are most relevant to our system, which is designed for face recognition under typical surveillance camera conditions. We excluded the mug shot images as they represent controlled conditions and do not reflect the low resolution and uncontrolled setting characteristic of real surveillance footage. Similarly, we ignored the IR images for preprocessing and training, as the infrared modality differs significantly from visible light imagery and is not the focus of our system.

3.2.2 Preprocessing

The surveillance images were preprocessed to ensure consistency and mitigate some common real-world data issues. A significant challenge in the SCface dataset is the gender imbalance, where the male subjects vastly outnumber females (114 males vs. 16 females). To address this imbalance and avoid bias during model training, we applied oversampling techniques to augment the female samples, ensuring more balanced gender representation in the training data. Exploratory data analysis (EDA) visualizations, such as gender

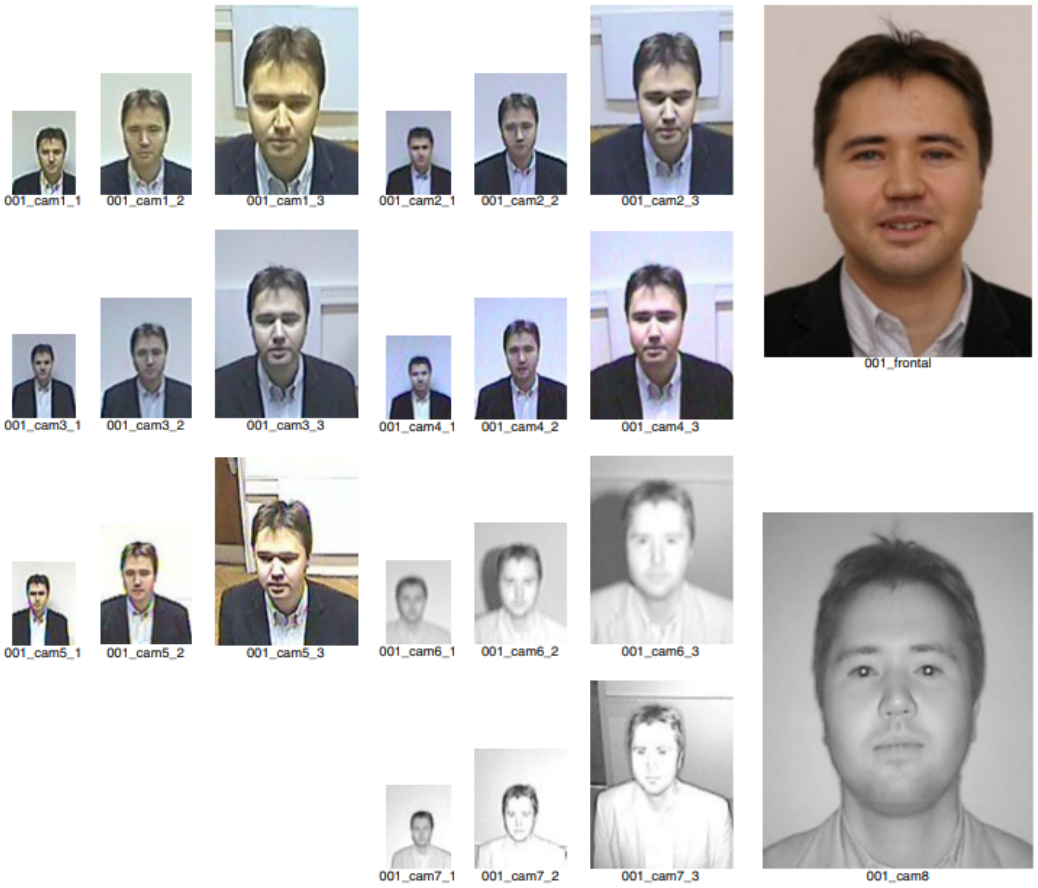


Figure 3: Different Footage from the Same Subject

distribution histograms, confirm the initial imbalance and illustrate the effectiveness of oversampling in balancing the dataset in Figure 4.

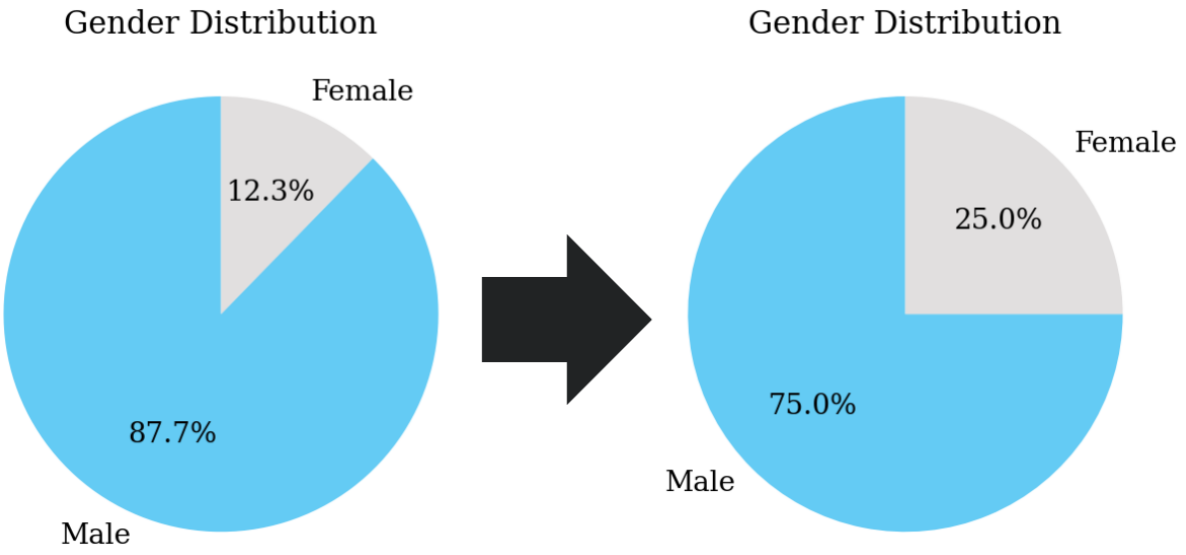


Figure 4: Gender Distribution before and after oversampling

Additionally, we analyzed the age distribution of subjects in SCface, which predominantly covers young adults, a demographic consistent with university populations. This makes SCface particularly well-suited for applications targeted at academic or campus environments. The dataset also includes metadata on facial hair presence, providing further demographic variety and realism.

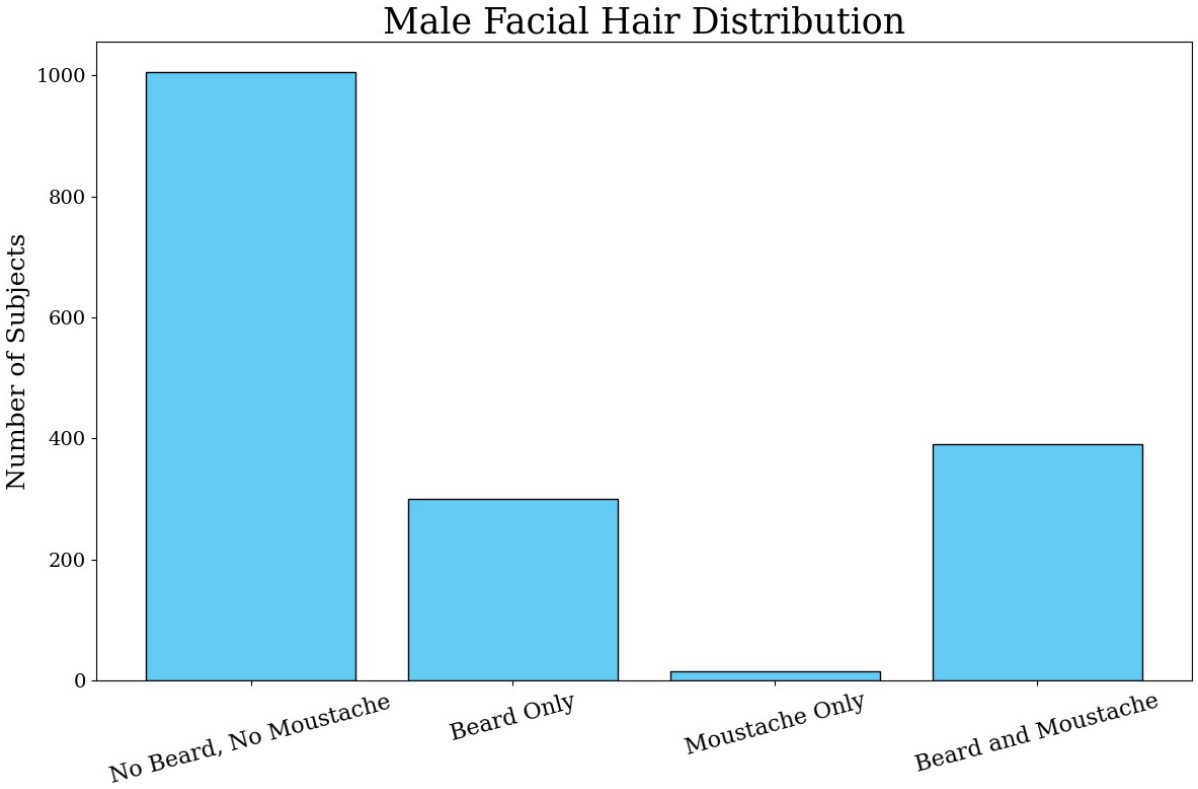


Figure 5: Males Facial Hair

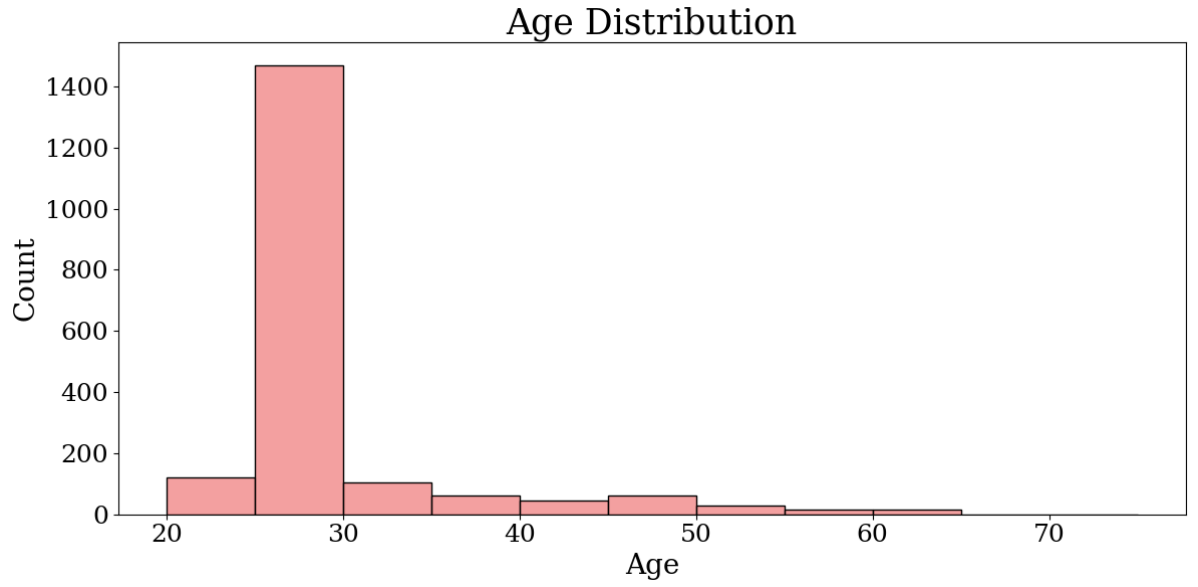


Figure 6: Age Distribution in SCface

All images were resized to a consistent shape of 416×416 pixels to fit the input requirements of our model. The dataset was divided into training, validation, and test sets with proportions of 70%, 15%, and 15% respectively, to allow for robust model training and unbiased evaluation.

The SCface dataset offers several advantages that make it well-suited for evaluating face recognition systems designed for real-world surveillance applications. Its images are captured under realistic, uncontrolled indoor conditions using commercially available surveillance cameras of varying quality and at different distances. This setup realistically simulates the variability in image resolution and quality encountered in practical surveillance scenarios. Furthermore, the dataset's demographic information, such as age distribution and facial hair presence, provides a diverse representation aligned with real populations, particularly those found in academic environments, making SCface especially relevant for university-focused applications. By focusing on visible spectrum surveillance images and excluding mug shots and IR images, the dataset directly reflects the challenging conditions that a surveillance system would face.

However, the dataset also presents some limitations that require careful consideration during preprocessing and model development. One major issue is the significant gender imbalance, with male subjects far outnumbering females, which can introduce bias and affect the model's fairness. While oversampling female images can mitigate this to an extent, the relatively small number of female subjects remains a constraint. Additionally, the dataset consists exclusively of Caucasian subjects, limiting ethnic diversity and potentially affecting the generalization of models trained on it. Another challenge stems from the varying image sizes caused by different camera distances; resizing these images to a consistent input size may introduce distortion or loss of important facial details. Lastly, by excluding mug shots and infrared images, the dataset focuses solely on visible light surveillance, which could restrict the model's applicability in multimodal or low-light environments.

3.3 Face Recognition Models

We experimented with several state-of-the-art deep learning models to evaluate their effectiveness for face recognition on surveillance images. The models range from classical convolutional networks to modern object detection architectures, each selected to capture different strengths relevant to our task.

To explore more specialized architectures, we incorporated several models from the YOLO family, well-regarded for their efficiency and accuracy in object detection tasks. The first, **YOLOv8n (Nano)** [10], is a lightweight model designed for fast inference with a compact backbone and spatial pyramid pooling. It effectively captures features at multiple scales, which is vital for detecting faces of varying sizes in surveillance footage. The model was trained with stochastic gradient descent (SGD) and a composite loss combining classification and localization (CIoU).

Building upon this, **YOLOv8s (Small)** [10] introduces a deeper backbone with additional convolutional blocks, complemented by enhanced neck and detection head layers. This model benefits from data augmentation techniques like mosaic and mixup, which improve generalization by exposing the network to more varied training examples. YOLOv8s was trained using SGD with a cosine learning rate scheduler and a combined loss function balancing objectness, classification, and bounding box regression.

We also evaluated **ResNet50** [5], a deeper residual network with 50 layers that em-

ploys bottleneck blocks to learn more complex representations efficiently. Similar to ResNet18, it was trained using the Adam optimizer with a learning rate of 0.001 and weighted cross-entropy loss.

Finally, **YOLOv11s (Small)** [6], the latest in the YOLO series, features enhanced convolutional blocks and a task-aligned anchor-free detection head. These improvements allow the model to achieve better spatial accuracy, which is crucial when detecting and recognizing faces in cluttered or challenging scenes. It was trained with SGD and momentum, using the CIoU loss to optimize classification and localization simultaneously.

Each model brings unique advantages to the task, from the simplicity and reliability of ResNet architectures to the speed and multi-scale detection capabilities of YOLO models. Through comprehensive evaluation, we assess how these differing architectures perform on surveillance face recognition, guiding future improvements.

4 Software Design

The *Asistancia* system is designed to automate student attendance tracking using facial recognition technology integrated into a web-based management platform. The architecture follows a client-server model, combining computer vision and modern web technologies to offer a scalable, secure, and modular solution.

4.1 System Architecture

The system consists of three core components:

4.1.1 Frontend (Client-Side)

The frontend delivers a web-based interface for administrators and teachers.

- **Technologies:** HTML, CSS, JavaScript.
- **Main Features:**
  - `login.html`: User authentication via MySQL credentials.
  - `classes.html`: Class creation and student enrollment.
  - `student.html`: Student record display and search.
  - `live.html`: Real-time attendance view and manual input.
- **Design:** Static files are served with responsive CSS. Assets such as logos enhance UX.

4.1.2 Backend (Server-Side)

Implemented using Node.js with Express, the backend manages sessions, APIs, and database interactions.

- **Technologies:** Node.js, Express.js, MySQL, `express-session`, `dotenv`.
- **API Endpoints:**

  - `/api/login`, `/api/session`, `/api/logout`
  - `/api/attendance`, `/api/ai-attendance`
  - `/api/students`, `/api/classes`
- **Security:** Environment variables for sensitive data; role-based access control.

4.1.3 Face Recognition Service

This service processes classroom images and identifies student faces using YOLOv11.

- **Technologies:** Flask, OpenCV, YOLOv11 (Ultralytics), NumPy.
- **Model:** Pre-trained YOLOv11 with a 0.5 confidence threshold.
- **Endpoints:**
  - `/health`: Model and service check.
  - `/detect`: Processes base64 images and returns recognized student IDs.
- **Face Recognition:** Uses DeepFace embeddings stored in JSON format.

4.2 Database Schema

The MySQL database supports user roles, student data, class management, and attendance logging.

- `users(username, password, role, name)`
- `classes(id, name, description, schedule, teacher_id)`
- `class_students(class_id, student_id)`
- `attendance(id, student_id, class_id, timestamp)`

Foreign keys and indexes maintain referential integrity and optimize query performance.

4.3 System Workflow

1. **Authentication:** Users log in via `login.html`; sessions are role-managed.
2. **Management:** Teachers create classes and enroll students using `classes.html`.
3. **Attendance:**
   - Manual: Entered through the UI.
   - Automated: Recognized faces are sent via `/api/ai-attendance`.
4. **Reporting:** Data is retrieved and shown on `live.html` and `student.html`.

4.4 Scalability and Security

- **Scalability:** Independent scaling of backend and face recognition service is supported. MySQL replication or sharding enables large-scale deployment.
- **Security:** Sensitive configurations are stored in environment variables. Middleware enforces access control. HTTPS is recommended for deployment.

4.5 System Implementation

4.5.1 Frontend

Static HTML pages interact with backend APIs. Key pages include:

- `login.html`: Authenticates users and redirects based on role.
- `classes.html`: Enables class creation and student enrollment.
- `student.html`: Shows searchable student records.
- `live.html`: Displays real-time attendance with manual override.

4.5.2 Backend

Implemented in `server.js`:

- Uses `express`, `mysql2`, `body-parser`, and session middleware.
- API endpoints handle authentication, attendance, and student/class management.
- Utility functions (e.g., `toMySQLDatetime`) and error handling ensure robustness.

4.5.3 Face Recognition Service

Implemented in Python using Flask:

- Loads YOLOv11 model via Ultralytics.
- Uses OpenCV and NumPy for image preprocessing.
- Endpoints return student IDs and bounding box data.
- Embeddings are stored in JSON with one face per student enforced.

4.5.4 Database

MySQL schema supports constraints and indexing:

- Tables include `users`, `students`, `classes`, `class_students`, `attendance`.
- Constraints and indexing improve integrity and speed.
- Duplicate checks prevent redundant attendance entries.

5 Results

We calculated the accuracy of the proposed models and compared between them based on accuracy as a first filteration. ResNet50 (98.83%) and YOLOv11 (99.5%) outperformed ResNet18 (95.43%) and YOLOv8 (92.75%). Then, we compared between ResNet50 and YOLOv11 based on frames per seconds. ResNet50 acheived detected 125 frames per second while YOLOv11 outperformed it by detecting 127 frames per second. Figures 7 and 8 shows the results of both YOLOv11 and ResNet50 during the training process.

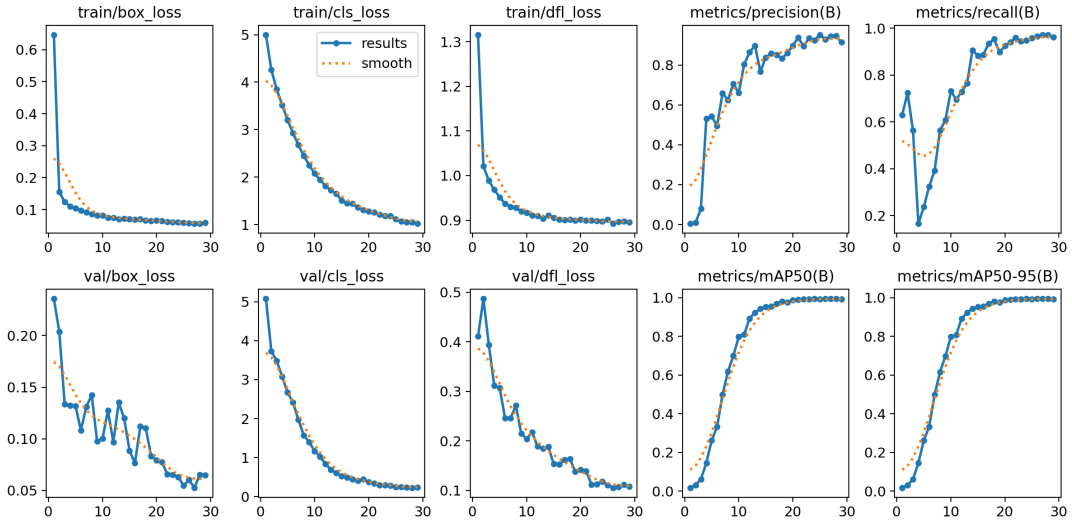


Figure 7: Results of YOLOV11

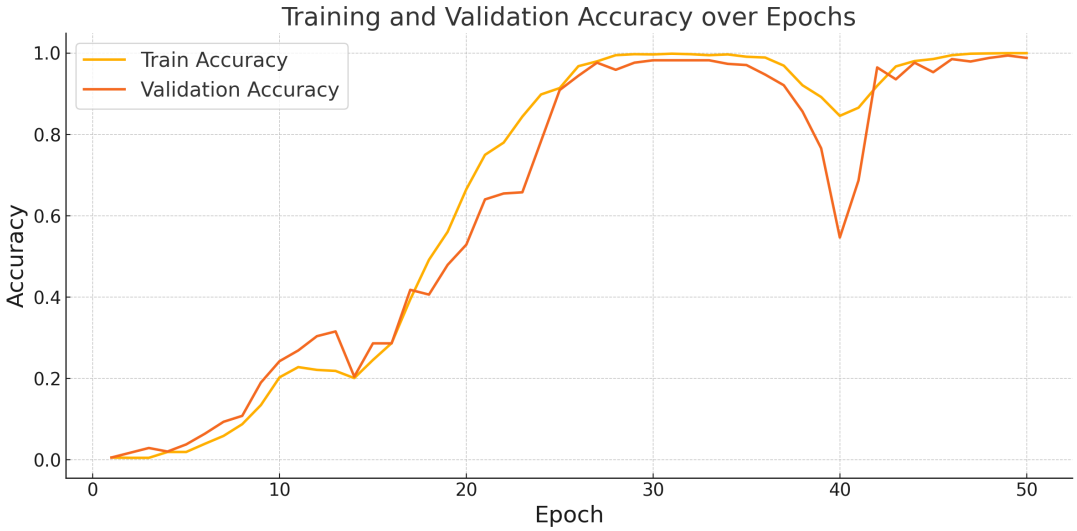


Figure 8: Results of ResNet50

Based on these results, we applied YOLOv11 to our system, however, we compared our models with previous related works': Aghdam et al. [1], Mishra et al. [8], and Tuvskog [9] who worked on the same dataset applying different models: SENet-50, LResNet50E-IR, mpdCNN, 512-VGG, 512-CASIA. Our model - YOLOV11 - clearly outperformed all previous models as shown in Table 1 and Figure 9

Paper / Model	Accuracy (%)	Notes
Aghdam et al. [1]	SENet-50: 97.23 LResNet50E-IR: 98.15	Highest accuracy reported; IR data pretraining.
Mishra et al. [8]	mpdCNN: 88.6	Multiscale CNN architecture.
Tuvskog [9]	512-VGG: 99.49 512-CASIA: 97.61	Pretrained on VGGFace2 and CASIA-WebFace.
YOLOv11	99.50	127 FPS; face alignment and lighting normalization.
ResNet50	98.83	125 FPS.

Table 1: Accuracy comparison across face recognition models.

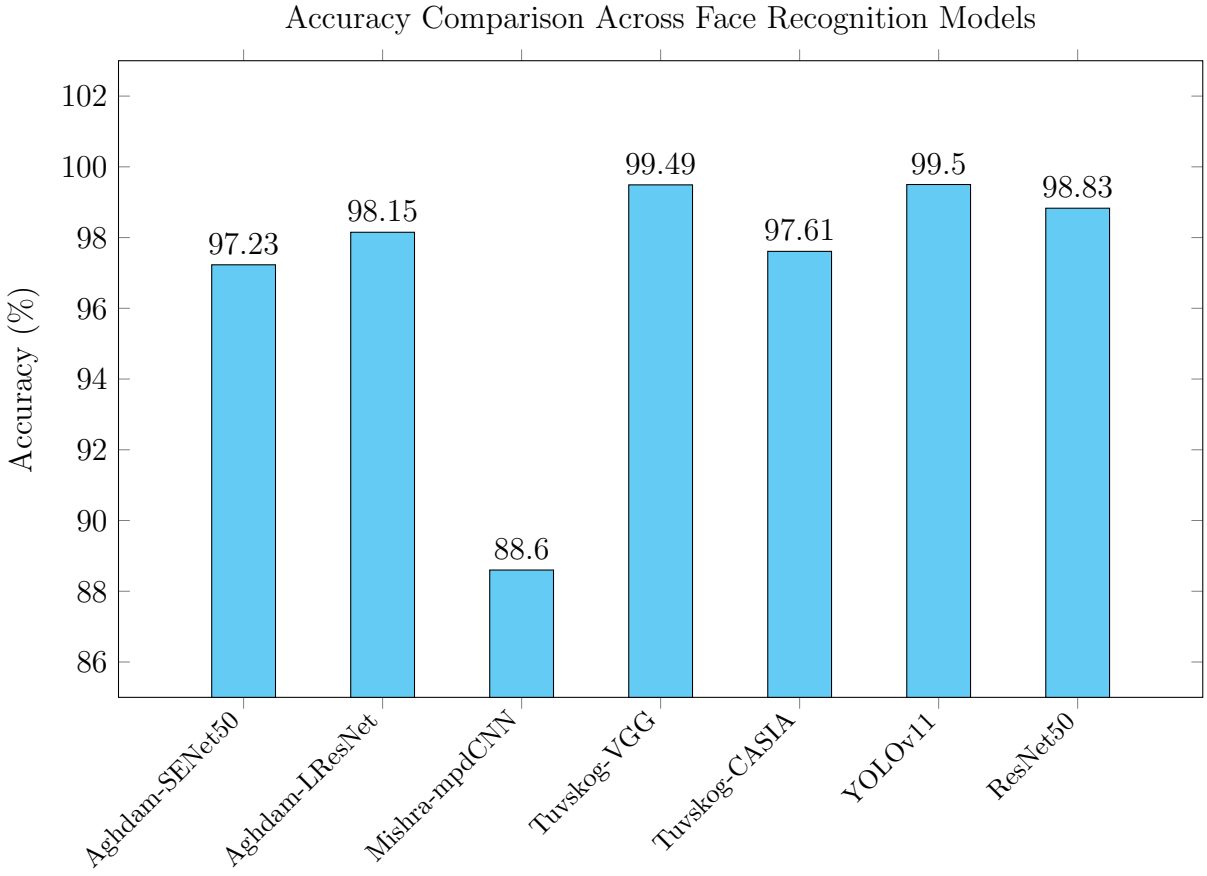


Figure 9: Accuracy comparison across face recognition models.

The model achieved an accuracy of **99.5%**, outperforming several state-of-the-art models referenced in literature, including SENet-50, LResNet50E-IR, and FaceNet-based architectures.

In addition to its high recognition accuracy, our model reached an average of **127 frames per second (FPS)**. This makes it highly suitable for live classroom environments where low latency is critical for practical deployment.

Preprocessing steps included face alignment and lighting normalization to improve robustness under varying classroom conditions. As shown in Table 1, our approach outperformed comparable models in terms of both accuracy and runtime efficiency, validating its effectiveness for attendance automation in academic settings.

6 Conclusion

Asistencia performs better than state-of-the-art facial recognition algorithms coupled with existing CCTV infrastructure as it provides a contactless, automated solution. The system addresses real-world challenges commonly found in surveillance-based environments, such as low-resolution imagery, variable lighting, and diverse facial orientations.

Through the use of the SCface dataset and robust preprocessing techniques—including face alignment and gender balancing—Asistencia was trained to operate reliably under unconstrained classroom conditions. Our evaluation of multiple models showed that YOLOv11 achieved superior performance, with an accuracy of 99.5% and real-time operation at 127 frames per second, making it highly suitable for live deployment.

Furthermore, Asistencia accommodates market needs for scalable and equitable facial recognition systems in educational settings. It demonstrates that with perfect training data and architectural design, real-world constraints can be effectively implemented. In the long term, we intend to scale Asistencia to multi-camera systems, study cross-domain generalization, and enhance privacy defenses for deployment.

References

[1] Omid Abdollahi Aghdam et al. "Exploring Factors for Improving Low Resolution Face Recognition". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2019, pp. 2363–2370. DOI: 10.1109/CVPRW.2019.00290.

[2] Nabeel Ali et al. "Automated attendance management systems: systematic literature review". In: *International Journal of Technology Enhanced Learning* 14 (Jan. 2022), p. 37. DOI: 10.1504/IJTEL.2022.120559.

[3] Zhiyi Cheng, Xiatian Zhu, and Shaogang Gong. *Surveillance Face Recognition Challenge*. Apr. 2018. DOI: 10.48550/arXiv.1804.09691.

[4] Mislav Grgic, Kresimir Delac, and Sonja Grgic. "SCface - Surveillance cameras face database". In: *Multimedia Tools Appl.* 51 (Feb. 2011), pp. 863–879. DOI: 10.1007/s11042-009-0417-2.

[5] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 2016, pp. 770–778.

[6] Ruksar Khanam and Mohammad Hussain. "YOLOv11: An Overview of the Key Architectural Enhancements". In: *arXiv preprint arXiv:2410.17725* (2024). URL: https://arxiv.org/abs/2410.17725.

[7] Pei Li et al. "Face recognition in low quality images: A survey". In: *arXiv preprint arXiv:1805.11519* (2018).

[8] Nayaneesh Kumar Mishra, Mainak Dutta, and Satish Kumar Singh. "Multiscale parallel deep CNN (mpdCNN) architecture for the real low-resolution face recognition for surveillance". In: *Image and Vision Computing* 115 (2021), p. 104290.

[9] Johanna Tuvskog. *Evaluation of Face Recognition Accuracy in Surveillance Video*. 2020.

[10] Ultralytics. *Ultralytics YOLOv8*. https://docs.ultralytics.com/models/yolov8/. Accessed: 2025-05-27. 2023.