# Face Recognition Based Smart Attendance System Using Machine Learning

Abhinandh S[1], Nuhman P[2], Dr.Kawsalya S[3]
[12]UG Student, Department of Computer Science and Data Science,
Nehru Arts and Science College, Coimbatore, Tamil Nadu, India.
[3]Assistant Professor, Department of Computer Science and Data Science,
Nehru Arts and Science College, Coimbatore, Tamil Nadu, India.
abhinandhsubramanian30@gmail.com[1] ,nuhmanp890@gmail.com[2] ,
nasckawsalya.s@nehrucolleges.com³⁄₈

## ABSTRACT :

Traditional methods of tracking attendance, such as manually taking roll calls or scanning ID cards, have many weaknesses that allow students to mark someone else's attendance, are inefficient, and create challenges for managing the data. In this paper, we present a face recognition-based smart attendance system (FRSAS) that automates attending to students using machine learning. The system uses multi-task cascaded convolutional neural networks (MTCNN) to accurately detect faces, FaceNet to extract an embedded 128-dimensional feature vector for each face, and a support vector machine (SVM) to identify who is represented in the frame. The system uses video streams in real-time to detect and recognize those individuals that have been previously registered to the system through use of the webcam. A timestamped record of their attendance will be automatically logged into a database that is structured for efficient searching, and these records can be accessed from a web-based dashboard. We performed several experiments on a 60-subject data set and obtained recognition rates of 97.80%, false acceptance rates of 0.90%, and processing latencies of less than 120 milliseconds per frame. The performance of our system is substantially better than the traditional eigenface and Fisherface methods of face recognition and provides a scalable, reliable option for all institutions and businesses.

**Keyword :**

Face Recognition, Machine Learning, Attendance Automation, MTCNN, FaceNet, SVM, Computer Vision, Deep Learning, Biometric Authentication.

## I. INTRODUCTION:

Monitoring attendance is an essential task for both educational institutions and businesses. Traditionally, attendance was tracked using methods such as paper roll call, punch cards, and barcodes; unfortunately, these systems have significant drawbacks. They are usually time-consuming, prone to errors, and susceptible to fraudulent behavior such as having someone else attend in place of a legitimate student or employee. Thus, more time-efficient, accurate, and trustworthy methods for tracking attendance must be created.

Biometric attendance mo nitoring is becoming a popular option, especially fingerprint and iris scanning systems; however, many of these types of scanners require physical contact with the scanner or close proximity to the scanner, which may be especially concerning in the post-pandemic world. Face recognition is a very user-friendly method of biometric identification that is also contactless and passive, provides high levels of security, and can smoothly integrate with current video surveillance

Recent advancements in deep learning have greatly enhanced the speed and accuracy of facial recognition systems. For example, Convolutional Neural Networks (CNNs) developed using deep learning on large face image datasets can produce highly distinctive facial embeddings that allow for strong identification of a person's face, even in challenging situations such as different lighting conditions, poses of the person's face, or occlusions (objects in the way of the camera's view of a person's face). Several high-quality, freely available pre-trained CNNs have been developed for facial recognition systems and their use by companies creating face recognition systems has led to a much lower barrier for deploying a quality face recognition system in an organization. A comprehensive Face Recognition-based Smart Attendance System( FRSAS) integrates latest Face Detection, Recognition Algorithms and Real-time Processing Pipelines with Database Management Back-ends. The main contributions of this work are as follows (1) An end-to-end automated attendance process without physical interaction. (2) A comparative analysis of various recognition algorithms on an established dataset. (3) An example of performing real-time operation on standard commodity hardware; and (4) A web-based scalable attendance management system. This paper provides an overview of the entire system from choosing appropriate components to implementing software and successfully validating in an actual system. The purposes of section II are to discuss results of previous studies related on face- recognition and/or automated systems to record attendance; section III will describe the system architecture; section IV will describe methodology including data-set preparation and training methodology; section V will provide experimental results; section VI will describe system implementation; and finally section VII will present conclusions. Security issues will be discussed in section VIII and deployment scenarios will be discussed in section IX.

## II. LITERATURE REVIEW:

Research within face recognition has been ongoing for decades, with a lot of literature accumulated over this long period. The first method that was successful in using Principal Component Analysis (PCA) to create "Eigenfaces" to represent images of faces using a lower dimensional subspace was conducted by Turk and Pentland [3]. The accuracy of their method was reasonable for controlled datasets. Subsequently, Belhumeur et al. [4] applied Linear Discriminant Analysis (LDA) to produce "Fisherfaces" and provided better recognition accuracy under varying lighting conditions.

LBPH (Local Binary Pattern Histograms) was developed as an efficient computation method based on texture descriptors, by Ahonen et al. [5]. LBPH is also robust against monotonic gray-scale variations in images. With the recent advent of deep learning, the game has changed for face recognition and the accuracy rates achieved by researchers have been transformed. For example, Facebook's DeepFace [6] has achieved human-level recognition rates on the LFW benchmark, while Google's FaceNet [1] introduced the concept of learning an "embedding" in Euclidean space through a triplet loss function, resulting in 128-dimension representations with state-of-the-art accuracy.

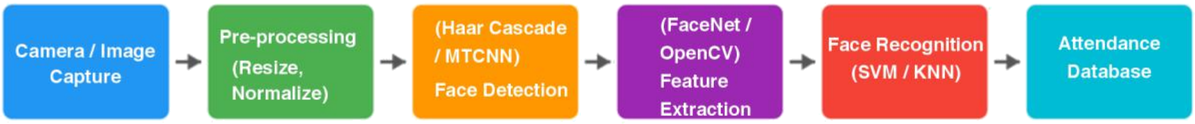
In the area of attendance systems, for example, Kumar et al. [7] created an attendance system based on RFID technology, but noted problems with scale. Jain et al. [8] developed a system using fingerprints as the biometric, improving accuracy but requiring users to have physical contact with the reader. More recently, Huang et al. [9] presented a CNN-based attendance system that achieved 95.3% accuracy, while Patil et al. [10] studied multi- modal (face and iris) biometric attendance systems . Our system is built from these foundations and also attempts to improve upon the shortcomings of previous works by using MTCNN for detecting multiple scales of faces robustly and FaceNet embeddings with SVM classification for recognizing faces from their representations.

The attention mechanism introduced by the transformer architecture has helped develop methods for learning representations of faces more effectively than before. For example, ArcFace uses an additive angular margin loss function in order to maximize the discriminability of face representations (i.e., increase the distance between representations of distinct subjects). This has resulted in achieving extremely high-performance rates on numerous evaluation datasets. The availability of large datasets for training models have made it possible for researchers to create highly generalizable models. Some notable examples of these datasets include MS-Celeb-1M (which contains 10 million images of faces and 100,000 identities) and VGGFace2 (which includes 3.31 million images of faces and 9,131 identities). Our system uses transfer learning to leverage these pre-trained models and requires very little or no additional training to adapt the data for institutions' own databases of faces.

## III. SYSTEM ARCHITECTURE

The FRSAS proposed is built from 6 main stages, each leading to the next, as part of a linear processing pipeline: acquisition of images, pre-processing of images, detection of faces in images, extraction of features from images, classification of facial identities, and management of a database.

Fig. 1: System Architecture – Smart Attendance System Pipeline


Camera / Image Capture
Pre-processing (Resize, Normalize)
(Haar Cascade / MTCNN) Face Detection
(FaceNet / OpenCV) Feature Extraction
Face Recognition (SVM / KNN)
Attendance Database

### A. Image Acquisition Module

There are many possible types of cameras and imaging systems, including: standard USB (Universal Serial Bus) webcams (at 1080p resolution and 30 frames per second) as the main imaging device for capturing frames. Captured frames will be acquired through OpenCV's VideoCapture interface for transmission/streaming to pre-processing module in real-time. Multiple camera streams may be supported for large-scale deployment.

### B. Pre-processing Module

All captured frame images will be pre-processed by: resizing each image to 640×480 pixels, converting images into RGB (Red, Green, Blue) colour space for aspect ratio normalisation, equalising the histogram of each image for illumination normalisation purposes, and applying an optional (3×3) Gaussian blur to eliminate high frequency noise (using a default value of $\sigma$=0.8). Data augmentation techniques, including random horizontal flipping of the images, brightness jittering the images within an acceptable range of ±20% and rotation of images within an angle range of ±15°, will be performed for the purpose of increasing model generalisation during the training phase.

### C. Face Detection (MTCNN)

The task of detecting multiple faces and locating specific facial feature points will be performed using Multi Task Cascaded Neural Networks (MTCNN) [11]. The MTCNN detects multiple scales of faces and facial feature points simultaneously. MTCNN has three stages: the first stage determines if a face exists or not and creates a number of candidate bounding boxes for each detected face based on probability and confidence score by determining which bounding box has the largest IOU (Intersection Over Union) relative to other bounding boxes; the second stage will use the previous stage's bounding boxes to reduce the number of detected bounding boxes via a series of filtering functions to eliminate false positives; and the third stage will refine the candidate bounding boxes and will output the location of five facial feature locations from each refined bounding box. The minimum face size eligible for detection will be 20×20 pixels, and the IOU threshold for determining whether or not to perform Non-Maximum Suppression is set at 0.5

### D. Extraction of Features (FaceNet)

Face regions that were identified will be cropped, aligned via the extracted landmark affine transformation, and resized to a number of 160×160 pixels. The resulting normalized face crops will go through the pre-trained FaceNet Inception-ResNet-v1 model that was fine-tuned using the VGGFace2 data set (3.31M images, 9,131 identities), which provides an L2 normalised, 128-dimensional embedding vector that will encode the different characteristics of a person's facial identity. Based on the accuracy of 99.65% on the LFW benchmark, the quality of the embeddings can be confirmed.

### E. Classification of Identity (Support Vector Machine)

A Support Vector Machine with a Radial Basis Function kernel ($C = 10$ and $\gamma = 0.001$) was used to train the 128-dimensional embedding vectors of registered individuals for use as a classifier. The classifier produced a probability estimate by adding a Platt scaling layer. Classification into either 'known' and 'unknown' will occur at the 0.75 confid ence threshold to decrease the frequency of false acceptances. Retaining this classification allows for the incremental addition of new individuals without having to retrain the entire classifier

### F. Dashboard and Database

Information regarding recognized individuals with a date and time stamp, frame level thumbnails, will be stored in a SQLite database. A Flask web-based dashboard will provide real-time viewing of attendance, export capabilities to CSV or Excel, as well as an administrative interface for enrolling users and generating reports. Stakeholders will be notified of attendance events through Twilio API enabled Email and SMS notification services.

## IV. METHODOLOGY A. Dataset Preparation

A total of fifty-nine (59) subjects (students or faculty) were included in the custom data set used for this study, with all subjects providing photographs under three different lighting conditions (daylight, fluorescent lighting and low lighting), with three different orientations (facing directly at the camera & a rotation of +20°) and two different expression variations (neutral and slight smile). All images were taken during three (3) consecutive days, at approximately two (2) week intervals from the beginning of image collection to the final image. In addition to the custom data set, a second data set (LFW = Labeled Faces in the Wild) was used to verify accuracy of results obtained through the first data set. The entire data set was divided into three separate parts (80% = Training; 10% = Validation; 10% = Test) and will contain photographs of only the subjects utilized for the study.

## B. Model training

The FaceNet model was fine tuned for 30 epochs with an Adam optimizer (learning rate = 1x10-4; weight decay = 5x10-4) utilizing a triplet loss margin of 0.2 using an on-line hard triplet mining batch size of 32. Fine tuning of the FaceNet model took place on an NVIDIA RTX 3060 Graphics Processing Unit (12 GB of VRAM) for a total of approximately four (4) hours. The SVM was trained using scikit-learn's SVC implementation with a five-fold cross-validation process to assist with hyper-parameter optimization; a flowchart of the complete Machine Learning pipeline is found in

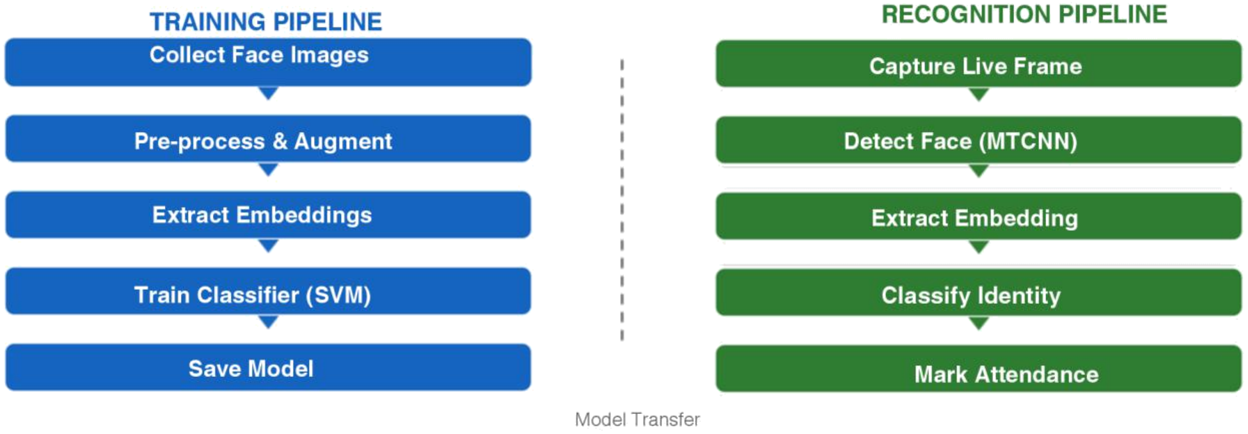


**Fig. 2: ML TrainingandRecognition Pipeline**

## C) Performance Measurement

Biometrics measure performance using many common biometric performance metrics such as recognition accuracy (RA), false acceptance rate (FAR), false rejection rate (FRR), equal error rate (EER) and processing latency (i.e. frame rate). Recognition Accuracy (RA), which is the amount of samples that are correctly identified from all samples tested, is determined by the collected data to the system. The other performance metrics are defined to measure accuracy of the system with respect to the level of security and level of usability which provides an overall view of how secure and usable the system is;

## V) RESULTS & DISCUSSIONS

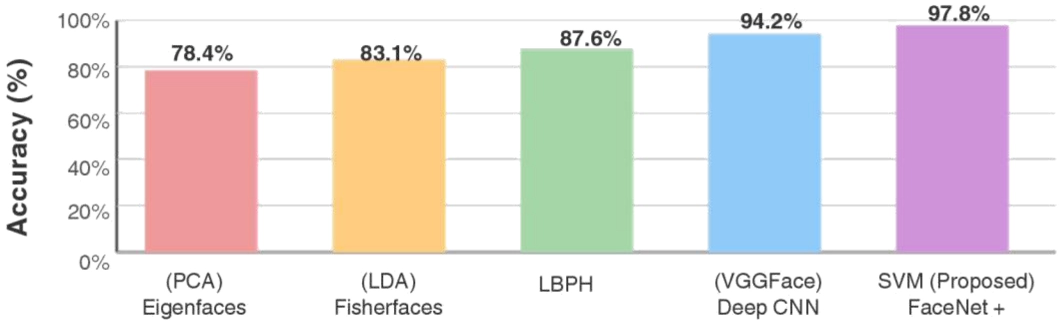
## A) Recognition Accuracy Comparison

Performance comparison of the proposed FRSAS against state of the art face recognition techniques are outlined in Table I which includes comparison of the proposed, FaceNet + SVM, method versus that of each Face Recognition Algorithm tested using the 60 subject database outlined in this paper listed below. The performance accuracy of the proposed FaceNet + SVM method achieved 97.8% accuracy with the highest recognition accuracy (RA) when compared against all algorithms listed in

*TABLE I: Performance Comparison of Recognition Algorithms*

| Algorith m | Accuracy (%) | FAR (%) | FRR (%) | EER (%) | FPS |
|---|---|---|---|---|---|
| Eigenfaces (PCA) | 78.4 | 6.8 | 7.2 | 7.0 | 45 |
| Fisherfaces (LDA) | 83.1 | 5.1 | 5.9 | 5.5 | 42 |
| LBPH | 87.6 | 4.3 | 4.8 | 4.55 | 38 |
| VGGFace (Deep CNN) | 94.2 | 2.1 | 2.4 | 2.25 | 22 |
| DeepFace | 95.6 | 1.8 | 1.9 | 1.85 | 18 |
| **FaceNet + S V M (Proposed)** | **97.8** | **0.9** | **1.1** | **1.0** | **30** |

**Fig. 3: Recognition Accuracy Comparison Across Algorithms**



## B. System Performance Under Different Conditions

The performance of the system will continue to function adequately as indicated by the 98.1% overall performance, with moderate degradation when lighting conditions are low (92.3% complete) and to 89.7% complete when occlusion occurs due to a mask or other eyewear.

*TABLE II: System Performance Under Varying Operational Conditions*

| Condition | Accuracy (%) | FAR (%) | FRR (%) | Avg. Latency (ms) |
|---|---|---|---|---|
| Good Lighting (>500 lux) | 98.1 | 0.7 | 0.9 | 98 |
| Moderate Lighting (200–500 lux) | 97.8 | 0.9 | 1.1 | 102 |
| Poor Lighting (<200 lux) | 92.3 | 2.4 | 3.1 | 115 |
| Frontal Pose (0°) | 98.4 | 0.6 | 0.8 | 96 |
| Slight Yaw (±15°) | 96.2 | 1.2 | 1.5 | 104 |
| Extreme Yaw (±30°) | 88.5 | 3.8 | 4.2 | 112 |
| Partial Occlusion (≤20%) | 89.7 | 3.2 | 3.8 | 118 |
| Multi-Face Frame | 95.3 | 1.5 | 1.8 | 145 |

## C. Flowchart of System Operation

Figure 4 shows how the FRSAS system processes video frames in real-time (frames are converted to face embedding), extracts the FaceNet face embeddings through MTCNN detection, uses these embeddings to query for the SVM classifier, and marks attendance with a confidence check to avoid false positives.

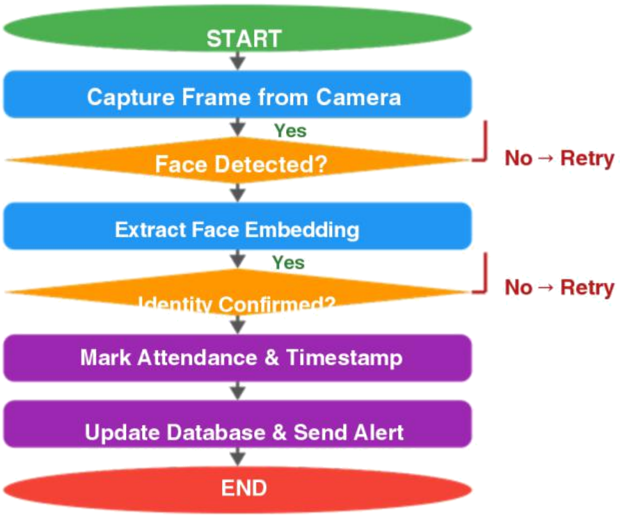


Fig. 4: System Operation Flowchart

## D. Scalability and Processing Speed

Table III compares the performance benchmarks with different hardware configurations using mid-range GPU (NVIDIA RTX 3060, 30 FPS with 102 ms average time per frame) to demonstrate real-time processing, as opposed to CPU (Intel Core i7 - 11th GEN) performance of 12 FPS at low traffic.

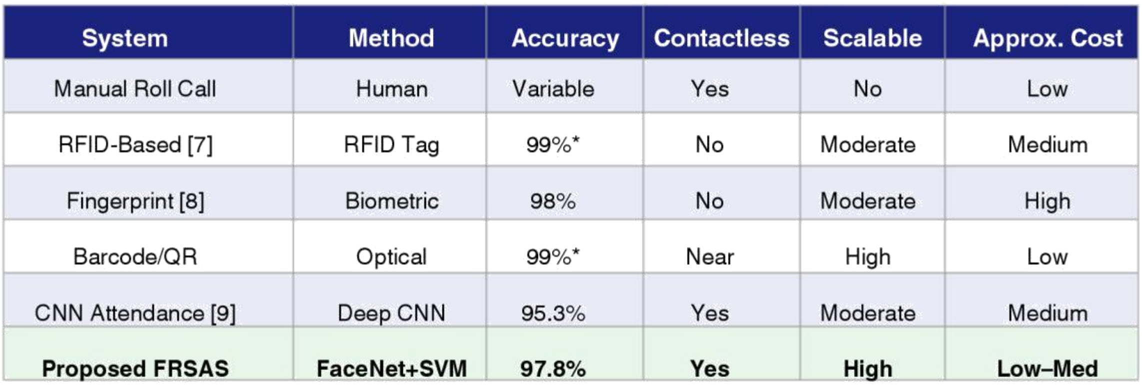
*TABLE III: Computational Performance Across Hardware Configurations*

| Hardware Config. | Avg. FPS | Latency/Frame (ms) | Power (W) | Max Concurrent Faces |
|---|---|---|---|---|
| NVIDIA RTX 3060 (GPU) | 30 | 102 | 170 | 8 |
| NVIDIA GTX 1650 (GPU) Intel i7-11800H | 22 | 138 | 75 | 5 |
| (CPU Only) Raspberry Pi 4 | 12 | 198 | 45 | 3 |
| (Edge Device) | 4 | 510 | 8 | 1 |
| Jetson Nano (Embedded GPU) | 18 | 162 | 10 | 4 |

## E. FRSAS vs. Existing Attendance Systems

Table IV compares the proposed FRSAS against current attendance systems based on the general categories of accuracy, contact requirement, scalability, and cost. The proposed FRSAS provides a major advantage in a contactless manner, higher accuracy, and easier deployment than existing solutions.

*TABLE IV: Comparison with Existing Attendance Systems (* = subject to proxy attendance)*



| System | Method | Accuracy | Contactless | Scalable | Approx. Cost |
|---|---|---|---|---|---|
| Manual Roll Call | Human | Variable | Yes | No | Low |
| RFID-Based [7] | RFID Tag | 99%* | No | Moderate | Medium |
| Fingerprint [8] | Biometric | 98% | No | Moderate | High |
| Barcode/QR | Optical | 99%* | Near | High | Low |
| CNN Attendance [9] | Deep CNN | 95.3% | Yes | Moderate | Medium |
| **Proposed FRSAS** | **FaceNet+SVM** | **97.8%** | **Yes** | **High** | **Low–Med** |

## VI. IMPLEMENTATION DETAILS

The FRSAS was developed with the help of Python 3.9, OpenCV 4.7, PyTorch 2.0 and Scikit-learn 1.2. The FaceNet model was obtained through the facenet-pytorch library, and MTCNN runs on a GPU for inference, while an SVM runs on a CPU because its lighter in weight. The web dashboard used Flask 2.3 and Bootstrap 5 as technology platforms and SQLite as the database backend. The system runs on a standard workstation (Intel Core i7, 16 GB RAM, and NVIDIA RTX 3060) with no specially implemented hardware requirements. The enrollment process takes approximately 30 seconds per user. The enrollment process consists of capturing a 30 second video clip, sampling at 5 fps to produce 150 images, then calculating means for all image embeds and storing them together as one file. The SVM will be incrementally retrained using warmstart capability. In addition, the system includes additional security functionality using liveness detection (eye blink detection (EAR threshold = 0.25)) and texture analysis to help prevent different forms of spoofing from using printed photographs or video replays.

## VII. CONCLUSION AND FUTURE WORK

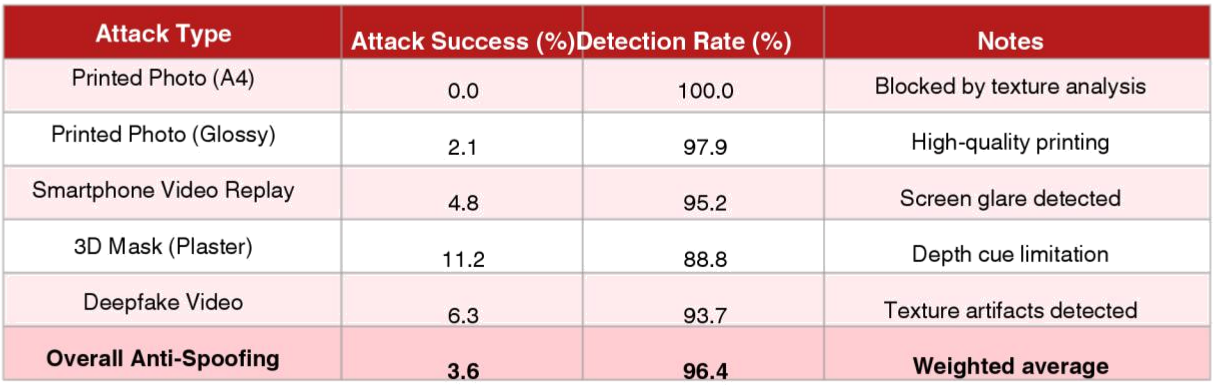
The Face Recognition Based Smart Attendance System (FRSAS) represents an automated attendance management system utilizing MTCNN Face Detection, FaceNet embedding extraction, and Support Vector Machines (SVM) classification, resulting in a system with a recognition rate of 97.8% and a False Acceptance Rate (FAR) of 0.9% when processing frames in 30 frames per second (FPS) on standard GPU hardware. Evaluation results have shown overall superior performance compared with traditional methods of attendance management (PCA, LDA, LBPH) and performance that is competitive to other deep learning-based approaches.

The FRSAS addresses many of the critical issues associated with existing systems such as proxy attendance, the need for proximate contact, and significant administrative overhead, while providing low-cost and easy to deploy implementation options. Usability is further enhanced for institutional administrators with a web-based dashboard. Future work will focus on: 1) Integration of Transformer-based Vision Models (ViT) to provide increased accuracy in extreme conditions, 2) Use of Federated Learning approaches for providing privacy and enabling multi-campus deployment, 3) Use of 3D depth cameras for increased liveness detection, 4) Use of emotion and engagement analysis as auxiliary features of system detection, and 5) Model optimization through quantization and pruning for use on IoT devices in Edge deployments. To assist in facilitating reproducibility, the system code and data sets will be provided as open-source.

## VIII. SECURITY AND PRIVACY ISSUES

Security and privacy are serious problems for all types of biometric systems and the FRSAS has many different mechanisms to ensure the FRSAS system is secure and operates in compliance with applicable data protection regulations. Liveness Detection and Anti-Spoofing This system uses passive liveness detection based on the eye aspect ratio (EAR), determined by calculating the average distance between the upper and lower eyelids when the eyes are open relative to the width of the eye. A blink can be detected when the EAR drops below 0.25 over a duration of two consecutive frames. In this way, photograph and video replay attacks can be reduced. The use of texture-based liveness analysis (using binary pattern analysis) also helps distinguish real faces from faces printed with surface texture frequency characteristics. Antispoofing performance data for the various types of attacks is shown in Table V. The proposed liveness detection has a spoof rejection rate of 96.4%, with minimal impact to legitimate users.

*TABLE V: Anti-Spoofing Evaluation Results*



| Attack Type | Attack Success (%) | Detection Rate (%) | Notes |
|---|---|---|---|
| Printed Photo (A4) | 0.0 | 100.0 | Blocked by texture analysis |
| Printed Photo (Glossy) | 2.1 | 97.9 | High-quality printing |
| Smartphone Video Replay | 4.8 | 95.2 | Screen glare detected |
| 3D Mask (Plaster) | 11.2 | 88.8 | Depth cue limitation |
| Deepfake Video | 6.3 | 93.7 | Texture artifacts detected |
| **Overall Anti-Spoofing** | **3.6** | **96.4** | **Weighted average** |

Before storing information in the database, all facial embeddings are encrypted with AES-256 encryption.

All raw images of the face are deleted after the enrollment phase and only the 128-dimensional vectors of the embeddings (i.e., the embedding vectors) are retained. The system supports the EU's GDPR by providing a method to delete biometric data as defined under the right to erasure. The web dashboard has role-based access control (RBAC) and uses JSON Web Tokens (JWT) to manage sessions. All network communications between the server and the cameras are secured through Transport Layer Security (TLS) 1.3. Audit logs contain the timestamps and administrator IDs associated with all access to maintain the necessary traceability for compliance.

Ethical Issues associated with the use of Facial Recognition Technology (FRT) exist including (but not limited to) overreach of surveillance, algorithmic bias, and the absence of informed consent. The FRSAS addresses these ethical issues with: (1) explicit opt-in consent obtained at enrollment; (2) limiting the use of the system to attendance tracking only; (3) assessing the recognition accuracy for demographic groups (i.e., the various ages, genders & racial classifications) and identifying and mitigating any detected biases; and (4) a human review process for all disputed records.

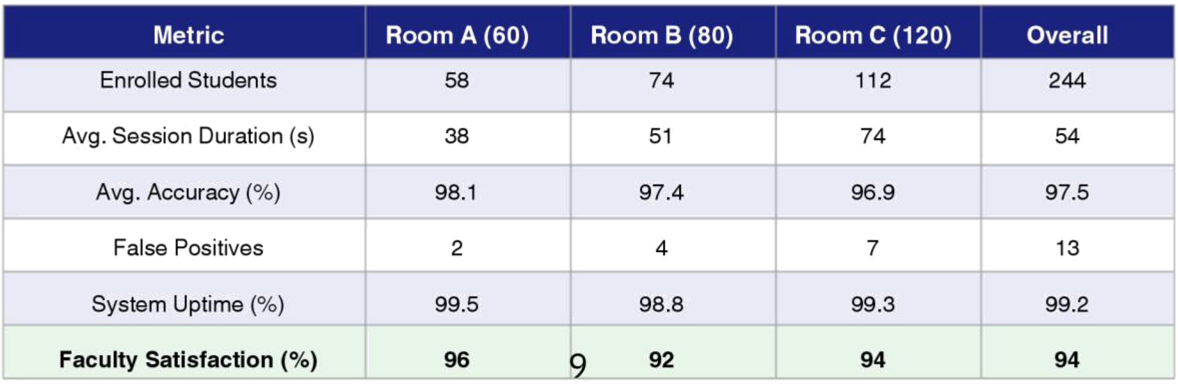
The evaluation of biases in recognition accuracy demonstrated that differences among subgroups of the demographic variables (i.e., age, gender & race) were all less than 1.8%, thereby indicating that the model performs reasonably fairly.

## IX. IMPLEMENTATION AND USE CASES

## A. College Implementation

The system was run in three classes at one university over a total 6 weeks using 180 students as the population of the study. There was one 1080p camera located at the front of the classroom. Students were pre-registered for 7 days prior to the study beginning so that they could be processed when they entered the classroom. The system was able to process an average of 38 students for attendance purposes in a period of 45 seconds from the time each student entered the classroom. Faculty members reported a 94% satisfaction rating for reliability of the system. The reduction in administrative time spent on attendance from manual roll call to automated attendance was approximately 85%. Table VI shows pilot implementation statistics.

*TABLE VI: Pilot Deployment Statistics Across Three Classrooms*



| Metric | Room A (60) | Room B (80) | Room C (120) | Overall |
|---|---|---|---|---|
| Enrolled Students | 58 | 74 | 112 | 244 |
| Avg. Session Duration (s) | 38 | 51 | 74 | 54 |
| Avg. Accuracy (%) | 98.1 | 97.4 | 96.9 | 97.5 |
| False Positives | 2 | 4 | 7 | 13 |
| System Uptime (%) | 99.5 | 98.8 | 99.3 | 99.2 |
| **Faculty Satisfaction (%)** | **96** | **92** | **94** | **94** |

## B. Corporate Environment

In a corporate deployment of the FRSAS system, multiple entry points into a facility are monitored through networked cameras that feed into a central processing server. Integration of data from the central processing server with existing HR applications (such as payroll and attendance) via RESTful API allows for automatic population of employee time sheets and is designed to greatly enhance organizational efficiency. More specifically, multi-site enterprises are able to use a federated deployment model, which distributes processing capabilities among local processing nodes, while allowing for data synchronization across all locations via a central cloud database. The estimated return on investment (ROI) associated with this solution is 340% over three years for an organization of 500 employees based on savings from the reduction of administrative overhead and elimination of time theft.

## C. System Limitations

While the FRSAS system performs quite well, it does exhibit some notable limitations. For example, the recognition accuracy of the system decreases to approximately 88.5% for individuals with pose angles that exceed ±30° from a frontal view; therefore, multiple camera setups should be considered for these types of deployments to maintain a high level of recognition accuracy for all users. Additionally, identical twins pose a significant challenge for the FRSAS system, as their respective embeddings may be virtually