ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

An Intelligent Multi Modal Classroom Monitoring System Using Deep Learning for Secure Attendance, Engagement Analysis and Predictive Analytics

Mohammed Omer Farooq[1]; Mohammed Saad[2]; Mohammed Mueez Gous[3]; Mir Mushtaq Ali[4]; Mohazzeba Tanveer Raza[5]

[1]Department of Computer Science and Engineering HKBK College of Engineering Bangalore, India
[2]Department of Computer Science and Engineering HKBK College of Engineering Bangalore, India
[3]Department of Computer Science and Engineering HKBK College of Engineering Bangalore, India
[4]Department of Computer Science and Engineering HKBK College of Engineering Bangalore, India
[5]Department of Computer Science and Engineering HKBK College of Engineering Bangalore, India

Publication Date: 2026/06/12

Abstract: Automated attendance management has emerged as a critical area of research in the domains of computer vision and artificial intelligence, offering efficient, contactless, and proxy-free alternatives to conventional manual methods. This survey presents a comprehensive review of face recognitionbased automated attendance systems, covering classical machine learning approaches such as Haar Cascade, Local Binary Pattern Histograms (LBPH), and Histogram of Oriented Gradients (HOG), as well as advanced deep learning methods including Multi-task Cascaded Convolutional Networks (MTCNN), FaceNet, ResNet-50, YOLOFace, and RetinaFace. The paper systematically examines fifteen recent works, analyzing their architectures, datasets, evaluation metrics (accuracy, precision, recall, F1-score, and inference time), and deployment environments ranging from classroom settings to corporate contexts. Key challenges including variable illumination, occlusion, multi-face detection, and real-time processing are discussed. A comparative analysis reveals that deep learning-based systems, particularly those integrating MTCNN for detection and FaceNet or ResNet- 50 for recognition, consistently outperform classical methods, with accuracies exceeding 98% under normal conditions. Cloudbased solutions leveraging AWS Rekognition demonstrate 100% detection accuracy in controlled scenarios. The survey concludes by identifying open research gaps and future directions including edge deployment, liveness detection, multi-modal biometrics, and privacy-preserving attendance systems.

Keywords: *Face Detection, Face Recognition, Automated Attendance System, Haar Cascade, MTCNN, FaceNet, ResNet-50, LBPH, YOLOFace, RetinaFace, Deep Learning, Convolutional Neural Networks, Classroom Environment.*

How to Cite: Mohammed Omer Farooq; Mohammed Saad; Mohammed Mueez Gous; Mir Mushtaq Ali; Mohazzeba Tanveer Raza (2026) An Intelligent Multi Modal Classroom Monitoring System Using Deep Learning for Secure Attendance, Engagement Analysis and Predictive Analytics. *International Journal of Innovative Science and Research Technology*, 11(5), 4124-4133. https://doi.org/10.38124/ijisrt/26may1752

I. INTRODUCTION

In the rapidly changing digital age, moving from manual to automated attendance systems is no longer only a convenience; it is now a crucial step in enhancing operational accuracy, efficiency, and security across a wide range of businesses [15]. Traditional attendance methods such as paper-based registers, roll calls, punch cards, fingerprint scanning, and Radio Frequency Identification (RFID) systems are not only time-consuming and laborious but are also prone to human error, resulting in vulnerabilities such as proxy attendance and erroneous data [2], [4].

In educational settings, automated attendance tracking has become particularly important. “Taking attendance in classrooms of universities, colleges, schools, any meeting or seminar room is necessary because it is related to academic performance, ensures accountability, encourages discipline, improves safety and security” [2]. Manual systems require teachers to spend valuable class time calling names or circulating sheets, reducing instructional efficiency. Moreover, students can easily mark attendance on behalf of absent peers — a problem commonly referred to as proxy attendance.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

Among the biometric technologies employed in automated attendance systems (AAS), face recognition has emerged as the most widely adopted due to its non-intrusive and contactless nature [5]. Unlike fingerprint or iris scanners that require physical contact, face recognition cameras can identify multiple individuals simultaneously from a distance without any active cooperation from the subject [4]. This property makes facial recognition particularly suited for large classroom environments where queuing at a biometric device is impractical.

Face recognition systems in attendance applications generally follow a four-stage pipeline: (1) face detection, (2) image preprocessing, (3) feature extraction, and (4) face matching and recognition [5]. The performance of each stage depends heavily on the algorithm employed and the environmental conditions — including lighting variability, occlusion, pose variation, and the number of faces in a single frame.

Early face detection methods such as the Viola-Jones Haar Cascade classifier offered real-time performance with modest computational requirements, making them popular for resource-constrained deployments [8], [9]. However, these classical approaches suffer from high false-negative rates under challenging conditions including low illumination, nonfrontal poses, and partial occlusions [1]. The emergence of deep learning has transformed the field, enabling models such as MTCNN, YOLOFace, RetinaFace, and ResNet-50 to achieve near-perfect detection and recognition accuracy even under adverse conditions [1], [15].

This survey is motivated by the need to provide a structured and comprehensive review of the diverse face recognition methodologies applied to automated attendance systems, evaluated across a consistent set of performance metrics. The specific contributions of this paper are:

- A systematic classification and review of fifteen recent works on face recognition-based attendance systems, spanning both classical and deep learning approaches.
- A comparative analysis of algorithms across key metrics:
- accuracy, precision, recall, F1-score, and inference time.
- An examination of real-world deployment challenges including occlusion, multi-face detection, variable illumination, and cloud versus edge processing.
- Identification of research gaps and future directions for robust, scalable, and privacy-conscious attendance systems.

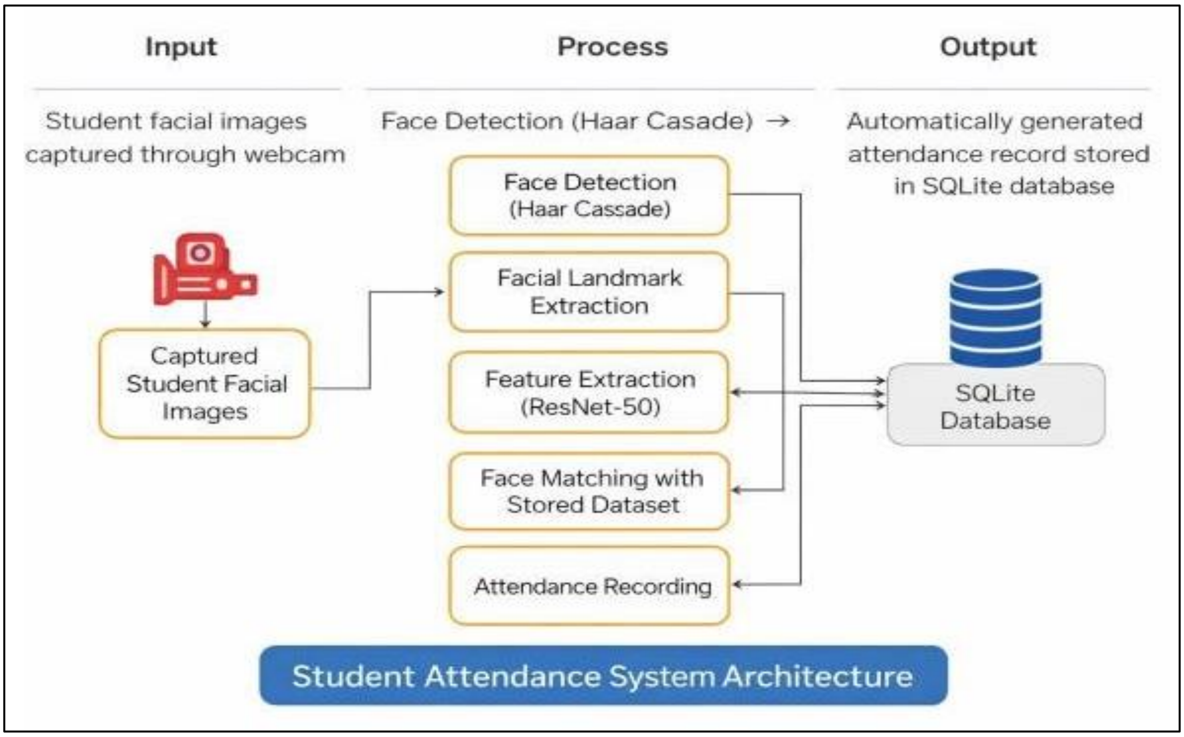

Fig 1 Intelligent Multimodal Classroom Monitoring — System Architecture.

The rest of this paper is organized as follows: Section II provides background on face detection and recognition algorithms. Section III presents the literature review of surveyed works. Section IV provides a comparative analysis and discussion. Section V discusses open challenges and future directions, and Section VI concludes the paper.

II. BACKGROUND: FACE DETECTION AND RECOGNITION ALGORITHMS

A face recognition system for attendance management operates through a well-defined pipeline. Understanding the core algorithms at each stage is essential for appreciating the strengths and limitations of each surveyed system. This section briefly describes the major families of algorithms that appear across the reviewed literature.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

A. Classical Face Detection Methods

➢ *Haar Cascade Classifier:*

The Haar Cascade classifier, developed by Viola and Jones [16], is one of the earliest and most influential face detection algorithms. It uses a series of simple Haar-like features evaluated using integral images to detect faces with real-time performance [1]. The algorithm is lightweight, computationally efficient, and available as a pretrained model in OpenCV. However, it struggles with high false-negative rates under challenging conditions such as nonfrontal poses, low lighting, and occlusions [1], [8].

➢ *Histogram of Oriented Gradients (HOG):*

HOG is a feature descriptor used for object detection in computer vision. It counts occurrences of gradient orientations in localized portions of an image [4]. In face recognition pipelines, HOG is used to detect faces in images before encoding, offering robustness to local geometric and photometric transformations.

➢ *Local Binary Pattern Histogram (LBPH):*

LBPH is a texture-based face recognition algorithm that labels pixels by thresholding the neighborhood of each pixel and considers the result as a binary number. It divides the face image into multiple small regions and builds histograms that are subsequently concatenated into a feature vector [10], [11]. LBPH is computationally efficient and robust to monotonic gray-scale transformations, making it suitable for real-time systems [13].

B. Deep Learning-Based Face Detection

➢ *Multi-Task Cascaded Convolutional Networks (MTCNN):*

Introduced by Zhang et al., MTCNN jointly addresses face detection and facial landmark localization through three cascaded convolutional networks: P-Net (Proposal Network), RNet (Refinement Network), and O-Net (Output Network) [3], [12]. MTCNN excels in detecting faces and facial landmarks with high precision across variations in scale, pose, and facial expressions, making it a robust choice for diverse classroom environments [1], [5].

➢ *YOLOFace:*

YOLOFace is an adaptation of the YOLOv3 object detection architecture specifically designed for face detection tasks [1]. It applies a single neural network across the entire image, dividing it into regions and simultaneously predicting multiple bounding boxes and their associated probabilities. This architecture enables YOLOFace to detect faces more accurately and efficiently, even in complex environments, by leveraging spatial information more effectively [1].

➢ *RetinaFace:*

RetinaFace introduces pixel-level face localization tasks into the detection process, significantly improving detection performance [1]. Utilizing a deep residual network backbone, RetinaFace achieves state-of-the-art accuracy on benchmark datasets such as WIDER FACE and FDDB. Its high accuracy, however, comes at the cost of longer inference times [1].

C. Deep Learning-Based Face Recognition

➢ *FaceNet:*

FaceNet is a deep convolutional neural network that generates 128-dimensional face embeddings using a triplet loss function [5]. Two photographs of the same person produce similar embeddings, while those of different persons produce very different ones. FaceNet's embeddings enable fast, accurate one-shot recognition and scale well to large datasets.

➢ *ResNet-50:*

ResNet-50 is a 50-layer residual network comprising one initial convolutional layer, sixteen residual blocks, one global average pooling layer, and one fully connected layer [15]. Its residual connections mitigate the vanishing gradient problem, allowing very deep networks to be trained effectively. When applied to face recognition, ResNet50 extracts 128 unique face discriminating features that enable robust matching even under occlusion conditions [15].

➢ *AWS Rekognition:*

AWS Rekognition is a cloud-based deep learning face detection and recognition service that uses Convolutional Neural Networks (CNNs) to analyze pixel patterns and identify key landmarks such as the eyes, nose, and mouth [2]. It enables large-scale multi-face detection and matching against stored face databases without requiring local computational infrastructure.

III. LITERATURE REVIEW

This section reviews fifteen recent works on face recognition-based automated attendance systems. The works are organized thematically: classical algorithm-based systems, deep learning-based systems, multi-face and classroomspecific systems, and cloud-based systems.

A. Classical Algorithm-Based Attendance Systems

➢ *Haar Cascade and LBPH with OpenCV (Yadav et al., 2022) [8]:*

Yadav et al. proposed an attendance management system based on facial recognition using the Haar Cascade algorithm for face detection and the Local Binary Pattern Histogram (LBPH) algorithm for face recognition. The system captures live video from a webcam, applies Haar Cascade to detect individual faces, and uses LBPH to match detected faces against a pre-registered dataset. "The Haar Cascade algorithm is applied to the video to identify the unique features and characteristics of each student, such as eyes, nose, ears and lips" [8]. Upon recognition, attendance is automatically recorded into an Excel sheet with the current date as the header. The system achieved a student face recognition rate of 77% with a false-positive rate of 28%, demonstrating the limitations of Haar Cascade under challenging conditions such as non-frontal poses and varying lighting.

➢ *Attendance Management using Facial Recognition with Haar-Cascade and LBPH (Rajput et al., 2022) [9]:*

Rajput et al. developed an automated attendance system using OpenCV with Haar Cascade for face detection and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

LBPH for face matching. The system comprises four phases: database creation, face detection, face matching, and attendance marking. Real-time video from a webcam is processed frame by frame; detected faces are matched against stored histogram values, and attendance is marked and stored in an Excel file.

The authors note that "the Haar Cascade algorithm works with many positive and negative images and provides the surface properties that are most important to the identification process, namely ROI (Region of Interest)" [9]. The system demonstrated practical utility but acknowledged susceptibility to illumination variation and pose changes.

➢ *Attendance System Using HOG and SVM Classifier (Krishnan and Manikuttan, 2022) [4]:*

Krishnan and Manikuttan proposed a face recognition-based attendance marking system using Histogram of Oriented Gradients (HOG) for face detection, a 68-point facial landmark model for face alignment, and 128-dimensional face encodings compared via a Support Vector Machine (SVM) classifier. The system converts images to black and white, computes gradient directions for each pixel, and generates HOG representations that remain consistent across varying illumination. "Face encoding is a method of representing the face with 128 computer-generated measures. Two photographs of the same people would have similar encoding, whereas two photographs of two distinct persons would have completely different encoding" [4]. The final output is stored as a CSV file containing recognized names and timestamps. The system demonstrated higher recognition rates when the difference between training and test images was small.

➢ *Multi-Facial Attendance System with Haar Cascade and LBPH (Rathi et al., 2023) [11]:*

Rathi et al. proposed a multifacial automated attendance system using OpenCV's Haar Cascade for face detection and LBPH for face recognition.

The system is capable of detecting and recognizing up to 15 faces simultaneously in a single frame, addressing the multi-student scenario of classroom settings. The authors noted that LBPH "is a straightforward yet very effective texture operator that labels pixels of a picture by thresholding the area around each pixel" [11]. The system features separate modules for faculty management, student management, subject allocation, and attendance marking. The study concluded that combining LBPH and Haar Cascade creates a cost-effective face recognition platform capable of handling the demands of classroom attendance.

➢ *Smart Attendance Using Face Recognition with FRAMS (Kumar et al., 2025) [10]:*

Kumar et al. developed the Face Recognition-Based Attendance Management System (FRAMS) using Haar Cascade classifiers for face detection and LBPH for face recognition, implemented in Python with OpenCV and Tkinter for the GUI. The system captures student images via webcam, builds a training dataset, and marks attendance by comparing live faces against the stored database. "Haar Cascade classifiers are incorporated for face detection while the facial recognition aspect is conducted through an efficient algorithm referred to as Local Binary Patterns Histograms (LBPH) that is resistant to changes in lighting, facial emotions and facial poses" [10]. The system was tested under various conditions and achieved high recognition accuracy in controlled environments, though challenges remained in busy multi-face images and low-light conditions.

➢ *Attendance Monitoring System using LBPH and Haar Cascade (Reddy et al., 2025) [13]:*

Reddy et al. implemented a facial recognition-based attendance monitoring system using Haar Cascade for face detection and LBPH for recognition, built with Python, OpenCV, Tkinter, and Pandas. The system captures approximately 100 images per student during registration, trains an LBPH-based classifier, and marks attendance in real-time via webcam. "The training module preprocesses images and trains an LBPH-based classifier using OpenCV. The recognition module utilizes a Haar Cascade-based face detection approach to identify individuals in real-time" [13]. Attendance is logged to CSV files with timestamps. The study acknowledged that inadequate illumination impacted feature extraction quality and partial occlusions reduced accuracy, recommending future integration of deep learning models for improved robustness.

B. Deep Learning-Based Attendance Systems

➢ *MTCNN-Based Intelligent Education Management System (Li and Zhang, 2024) [3]:*

Li and Zhang designed an intelligent education management system for universities based on an improved MTCNN face recognition algorithm. The improvement involved modifying the loss function to cross-entropy, introducing parameter fine-tuning at the O-Net stage, abandoning the facial alignment module, and adding a discriminant formula to reduce false detections. The system architecture uses a B/S network structure with cluster deployment and load balancing, comprising four main modules: classroom attendance, status monitoring, report analysis, and background management. "When there was no occlusion, recognition accuracy was 99.4%. However, when occlusion was presented at 10%, 20%, and 30%, the accuracy dropped to 92.3%, 84.25%, and 73.4%, respectively" [3]. The improved MTCNN system had a maximum GPU utilization rate of 75% at 1000 concurrent users, which was 11% lower than the traditional MTCNN system, demonstrating improved scalability.

➢ *MTCNN and FaceNet for Automated Attendance (Baker et al., 2025) [5]:*

Baker et al. proposed an automated attendance and security monitoring system comparing three experimental configurations: (1) Haar Cascade with LBPH, (2) Haar Cascade with LDA-PCA, and (3) MTCNN with FaceNet. Experiment 3, using MTCNN and FaceNet, achieved the best performance with a face detection accuracy of 86.67% and a recognition accuracy of 86.91%. "MTCNN demonstrated consistent processing times across varying numbers of faces: 143.28 milliseconds for one person, 139 milliseconds for two people, and 139.94 milliseconds for three people" [5]. The

FaceNet model generated 128-dimensional embeddings and was evaluated using ROC curves, achieving an area of 0.92. The study concluded that MTCNN and FaceNet, though computationally heavier, provided superior accuracy and scalability over classical methods.

- *MTCNN-Based Deep Learning Attendance System (Angulakshmi and Susithra, 2024) [12]:*

Angulakshmi and Susithra proposed an automated attendance system using MTCNN for both face detection and recognition. The MTCNN pipeline operates through three stages: P-Net for candidate window generation, R-Net for candidate refinement, and ONet for precise facial landmark localization. Face embeddings are generated using the Insight Face model, producing 128dimensional feature vectors that are classified using a SoftMax classifier. The proposed MTCNN achieved a classification accuracy of 98.2%, sensitivity of 99%, specificity of 97.5%, and precision of 97.5%, outperforming Haar Cascade with CNN (84% accuracy). Table **??** summarizes this comparison.

- *AI-Driven Attendance Tracking with Haar Cascade and ResNet (Lalitha et al., 2025) [15]:*

Lalitha et al. presented a novel attendance management system combining Haar Cascade for face detection and ResNet-50 for face recognition. The dlib frontal face detector uses Haar Cascade to locate facial regions, followed by extraction of 68 facial landmarks which serve as input to the ResNet-50 model. ResNet-50 extracts 128 unique face discriminating attributes that are stored in a CSV file and used for matching. “The system achieves excellent accuracy in a variety of difficult situations, such as dim lighting, occlusions, and a wide range of facial emotions, by combining the deep learning architecture of ResNet with the reliable facial identification capabilities of Haar Cascade” [15]. Under normal conditions, the system achieved 100% accuracy, precision, and recall. Under low lighting and occlusion conditions (wearing glasses, headgear, or facial hair), accuracy was 91.6%, precision 93.8%, and recall 95.8%. The system also supports video input for offline attendance marking and stores records in an SQLite database with a Flask web interface.

- *Face Recognition Attendance System Using Deep Learning and MTCNN (Ghodekar et al., 2023) [6]:*

Ghodekar et al. proposed a smart student attendance and activeness monitoring system using the OpenFace recognition model for attendance and the Haar Cascade model for monitoring student engagement. The system employs graph-clustering with the OpenFace recognition model for error correction and dataset refinement, selecting one center image per student and eliminating duplicates. For activeness monitoring, the Haar Cascade model analyzes facial expressions and eye movements (gaze tracking), classifying students as “active” or “nonactive” based on adjustable thresholds. “The introduction of the activeness metric is a significant advancement. It helps instructors identify students who are actively participating in the class and those who might need additional support” [6]. The system tracks entry and exit times and classifies students as active if detected for more than 30 minutes per session.

C. Contactless and Hardware-Integrated Systems

- *Automated Attendance System Using Contactless Facial Recognition (Venkatakrishnan et al., 2024) [14]:*

Venkatakrishnan et al. developed a contactless attendance system integrating ESP32-CAM hardware with Python-based facial recognition using OpenCV and the face_ recognition library. The ESP32-CAM module captures real-time facial images which are transmitted for processing. The recognition algorithm encodes detected faces into unique 128-dimensional vectors and compares them against pre-stored encodings. Attendance data is securely managed in a MongoDB Atlas cloud database. The system demonstrated successful multiperson attendance tracking and “contactless approach stream- lined the attendance tracking process, eliminating the need for manual check-ins and minimizing physical contact” [14]. The integration of cloud storage ensures real-time access for administrators and students through a web-based interface.

- *Efficient Face Recognition Attendance System with Spectra Sense Algorithm (Deborah et al., 2023) [7]:*

Deborah et al. proposed an automated attendance system using a novel Spectra Sense algorithm alongside Haar Cascade for face detection, Dlib for face alignment, and CNN for feature extraction. The Spectra Sense algorithm integrates camera setup, face detection, alignment, feature extraction, face recognition, time-stamping, and cloud integration into a unified pipeline. The system supports continuous learning by updating the face recognition model with new faces over time. “The spectra sense-based attendance mechanism has been reported to achieve over 98.3% accuracy” [7], compared to 97.41% for Viola-Jones, 92% for Neural Network-based detection, and 98% for SVM-based detection. The proposed algorithm also achieved a precision of 0.37961 and recall of 0.29311.

D. Multi-Face Detection in Classroom Environments

- *Comparative Analysis of Multi-Face Detection Methods (Ananda et al., 2024) [1]:*

Ananda et al. conducted a comprehensive comparative analysis of four advanced face detection algorithms — Haar Cascade, MTCNN, YOLOFace, and RetinaFace — evaluated across four real classroom scenarios with varying student counts (32–40 students) and illuminance levels (40–152 Lux). Experiments were performed on Google Colab with an NVIDIA T4 GPU. “YOLOFace demonstrated exemplary performance with perfect precision and recall in most scenarios” [1], achieving F1-score of 1.0 in three of four scenarios with inference times of 2.1–3.3 seconds. RetinaFace also achieved perfect scores in most cases but had the longest inference times (up to 12.73 seconds). MTCNN exhibited high precision (up to 1.00) and recall (up to 0.97) but had higher inference times and occasionally missed some faces. Haar Cascade was the fastest (down to 0.54 seconds) but displayed the lowest precision (0.89) and recall (0.25). This study is particularly valuable as it tests algorithms specifically in real classroom conditions rather than standard benchmarks.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

- *Cloud-Based Multi-Face Attendance with AWS Rekognition (Kamruzzaman et al., 2025) [2]:*

Kamruzzaman et al. presented a cloud-based attendance system using AWS Rekognition for face detection and matching, Amazon S3 for image storage, AWS Lambda for serverless processing, and DynamoDB for face print storage. The system captures a single classroom image, detects and crops individual faces, uploads them to AWS Rekognition for comparison against the registered face database, and generates attendance reports sent via email. The system was tested on three practical images containing 49, 30, and 14 people respectively, and "achieved 100% accuracy" [2] in all three cases, correctly identifying all present, absent, and unknown individuals. This demonstrates the powerful scalability of cloud-based approaches for largegroup attendance without local computational overhead.

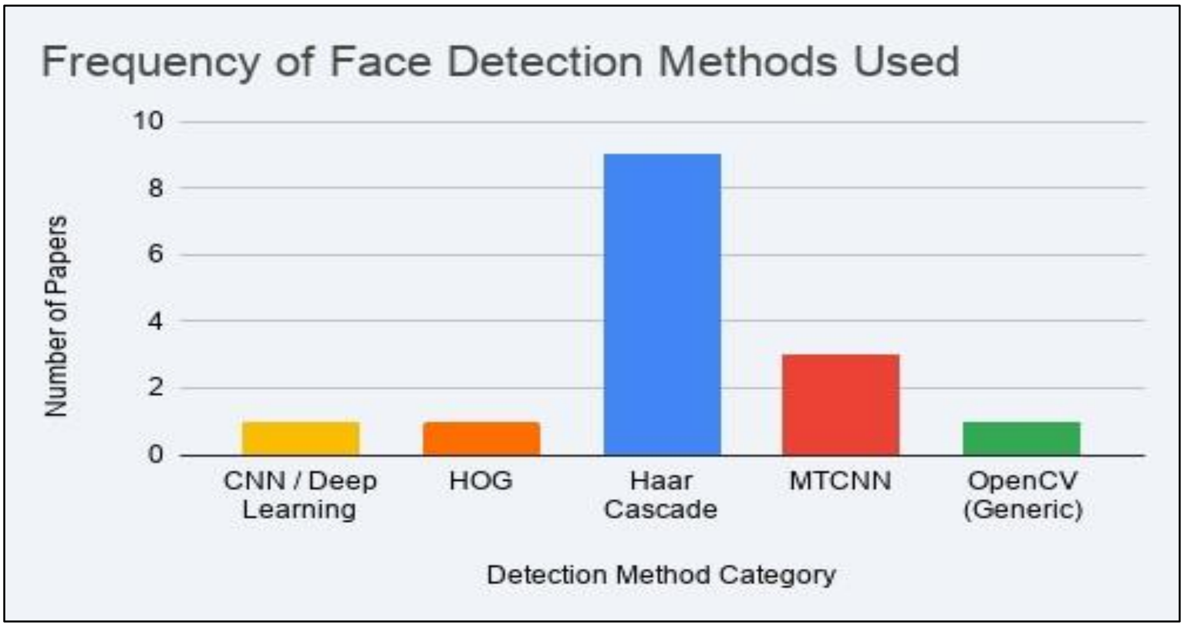

Fig 2 Frequency of Face Detection Methods Used.

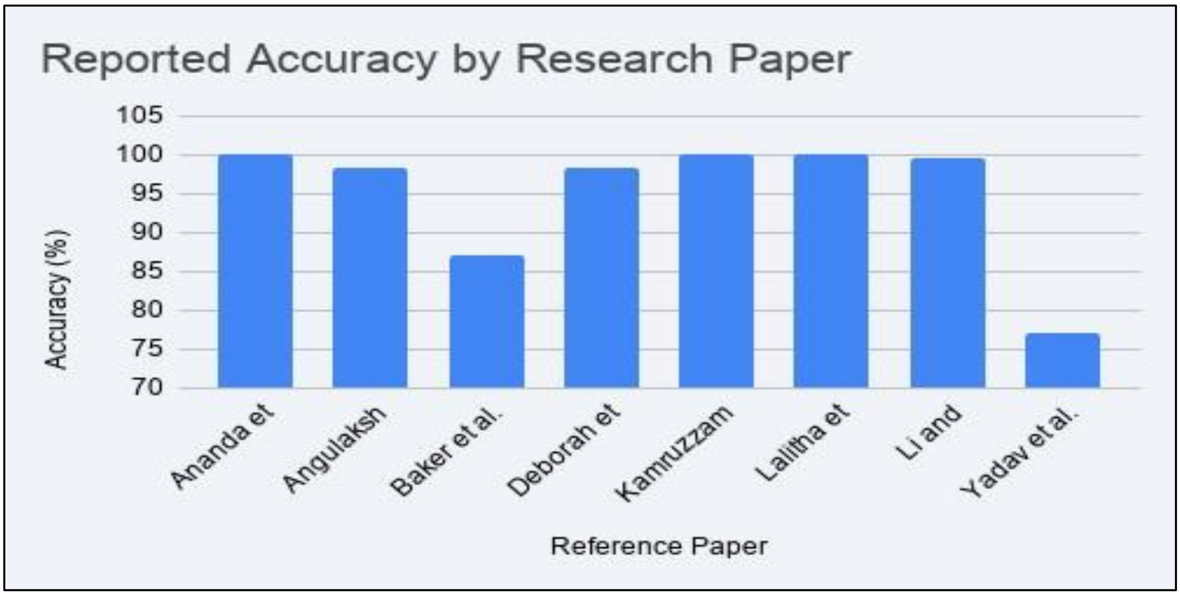

Fig 3 Reported Accuracy by Paper.

IV. COMPARATIVE ANALYSIS AND DISCUSSION

- *Algorithm Performance Comparison*

Table I provides a consolidated comparison of the surveyed works across key dimensions: detection algorithm, recognition algorithm, accuracy, and primary deployment environment.

- *Face Detection Algorithm Comparison*

The comparative study by Ananda et al. [1] provides the most rigorous evaluation of face detection algorithms in real classroom environments, testing four algorithms across four scenarios with student counts of 32–40 and illuminance of 40–152 Lux. Table II summarizes the precision, recall, F1-score, and inference time ranges across all scenarios. YOLOFace achieves the best balance between detection accuracy and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

speed, with perfect precision and recall in most scenarios and inference times of 2.15–3.33 seconds. Haar Cascade is fastest but achieves only 0.25–0.34 recall, meaning it misses the majority of faces in a classroom. RetinaFace achieves near-perfect accuracy but is prohibitively slow at up to 12.73 seconds per frame for real-time use.

Table 1 Comparative Summary of Face Recognition-Based Attendance Systems

Reference	Detection Method	Recognition Method	Accuracy (%)	Real-time?	Storage/Output
Ananda et al. [1]	Haar, MTCNN, YOLOFace, RetinaFace	Same (detection only)	YOLOFace: 100% (F1)	Yes	N/A (detection study)
Kamruzzaman et al. [2]	AWS Rekognition (CNN)	AWS Rekognition	100%	Yes	DynamoDB, Email
Li and Zhang [3]	Improved MTCNN	MTCNN + Cross-entropy	99.4% (normal) 73.4% (30% occlusion)	Yes	MySQL, Web/App
Lalitha et al. [15]	Haar Cascade + Dlib	ResNet-50	100% (normal) 91.6% (occlusion)	Yes	SQLite, Flask Web
Deborah et al. [7]	Haar Cascade + Dlib	CNN (Spectra Sense)	98.3%	Yes	Cloud + Database
Angulakshmi & Susithra [12]	MTCNN	MTCNN + SoftMax (Insight Face)	98.2%	Yes	Excel/Database
Baker et al. [5]	MTCNN	FaceNet (128-D embeddings)	86.91%	Yes	XLSX
Ghodekar et al. [6]	Haar Cascade	OpenFace	Not specified	Yes	Database
Venkatakrishnan et al. [14]	OpenCV face recognition	128-D face encodings	Not specified	Yes	MongoDB Atlas, CSV
Krishnan & Manikuttan [4]	HOG	SVM (128-D encodings)	Not specified	Yes	CSV
Rajput et al. [9]	Haar Cascade	LBPH	Not specified	Yes	Excel
Yadav et al. [8]	Haar Cascade	LBPH	77%	Yes	Excel
Kumar et al. [10]	Haar Cascade	LBPH	High (controlled)	Yes	CSV
Rathi et al. [11]	Haar Cascade	LBPH	Not specified	Yes	Database
Reddy et al. [13]	Haar Cascade	LBPH	Not specified	Yes	CSV

Table 2 Face Detection Performance in Classroom Scenarios [1]

Method	Precision	Recall	F1	Inf. Time (s)
Haar Cascade	0.89–0.93	0.25–0.34	0.39–0.50	0.54–0.80
MTCNN	0.97–1.00	0.84–0.97	0.92–0.99	4.27–10.08
YOLOFace	1.00	0.97–1.00	0.98–1.00	2.15–3.33
RetinaFace	0.97–1.00	1.00	0.99–1.00	9.15–12.73

➢ *MTCNN vs. Haar Cascade with CNN*

Angulakshmi and Susithra [12] directly compare MTCNN with Haar Cascade-CNN classification on the same dataset. The results are reproduced in Table 3.

➢ *Comparison Across Occlusion Conditions*

A critical real-world challenge is the recognition of faces under occlusion — students wearing glasses, masks, or head-gear. Table 4 compares systems that explicitly report performance under occlusion conditions.

Table 3 MTCNN vs. Haar Cascade + CNN Performance [12]

Metric	MTCNN	Haar Cascade + CNN
Accuracy	98.2%	84.0%
Sensitivity	99.0%	83.7%
Specificity	97.5%	99.0%
Precision	97.5%	95.0%

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

Table 4 Performance Under Occlusion Conditions

Reference	Model	Normal Acc.	Occlusion Acc.
Li & Zhang [3]	Improved MTCNN	99.4%	73.4% (30%occ.)
Lalitha et al. [15]	ResNet-50 + Haar	100%	91.6%

The ResNet-50 + Haar Cascade system of Lalitha et al. [15] significantly outperforms the improved MTCNN system under heavy occlusion, demonstrating that deep residual archi-tectures trained end-to-end on 68-landmark features provide better robustness to partial face visibility.

➢ *Storage and Scalability*

The surveyed systems employ a range of storage solu-tions. Classical systems predominantly use CSV files or Excel spreadsheets, which are simple but lack scalability and query capabilities. More advanced systems employ re-lational databases (SQLite, MySQL) or cloud-based NoSQL databases (DynamoDB, MongoDB Atlas). The cloud-based AWS Rekognition system of Kamruzzaman et al. [2] repre-sents the extreme end of scalability, capable of processing classroom-scale images with 100% accuracy and automatic email reporting, though it requires reliable internet connectiv-ity. The SQLite + Flask solution of Lalitha et al. [15] offers a practical offline alternative with an intuitive web interface for date-specific attendance retrieval.

➢ *Multi-Face Detection Capability*

One of the most critical requirements for classroom atten-dance is the simultaneous detection and recognition of multiple faces. Table 5 summarizes multi-face capabilities across the surveyed systems.

Table 5 Multi-Face Detection Capability

Reference	Max Faces	Notes
Ananda et al. [1]	32–40	Tested in real class-rooms
Kamruzzaman et al. [2]	49	100% accuracy
Lalitha et al. [15]	Multiple	Shown in Fig. 8 of paper
Rathi et al. [11]	15	Per frame
Li & Zhang [3]	Large scale	1000 concurrent users
Yadav et al. [8]	Limited	Accuracy drops with count

V. CHALLENGES AND FUTURE DIRECTIONS

➢ *Illumination Variability*

Low and variable lighting remains one of the most persistent challenges for all face recognition systems. Classical methods like Haar Cascade and LBPH perform poorly under dim lighting [8], [11]. Even the improved MTCNN system of Li and Zhang [3] shows accuracy degradation under occlusion conditions. The ResNet-50 system of Lalitha et al. [15] demonstrates notable resilience, achieving 91.6% accuracy under combined low-lighting and occlusion conditions, but this still falls short of the 100% achieved under normal conditions. Future systems should integrate adaptive preprocessing techniques such as histogram equalization and retinex-based illumination normalization to improve robustness.

➢ *Occlusion and Pose Variation*

Face occlusion caused by glasses, masks, or headgear represents a significant challenge. The COVID-19 pandemic intensified this concern, as mask-wearing became widespread. Li and Zhang [3] show that recognition accuracy drops to 73.4% when faces are 30% occluded. Future work should focus on training models specifically on occluded face datasets and employing partial face recognition techniques that can identify individuals from visible facial subregions.

➢ *Privacy and Ethical Concerns*

Facial recognition systems inherently collect and process biometric data, raising significant privacy and ethical concerns. Baker et al. [5] acknowledge that "the deployment of facial recognition necessitates stringent data protection measures to address privacy concerns and comply with GDPR." Future systems must incorporate data anonymization, on-device processing to minimize data transmission, explicit consent mechanisms, and compliance with local biometric data protection regulations.

➢ *Real-Time Performance and Edge Deployment*

While deep learning models offer superior accuracy, their computational demands can limit real-time performance on edge devices. The inference time comparison by Ananda et al. [1] shows that RetinaFace requires up to 12.73 seconds per frame — unacceptable for real-time classroom attendance. Future research should explore model compression techniques including pruning, quantization, and knowledge distillation to enable deployment on low-power edge devices such as Raspberry Pi or ESP32-based systems while maintaining acceptable accuracy.

➢ *Liveness Detection and Anti-Spoofing*

None of the surveyed systems explicitly address the risk of spoofing — an attacker presenting a photograph or video of a registered user to falsely mark attendance. Reddy et al. [13] suggest that liveness detection should be implemented to prevent such spoofing. Future systems should integrate liveness detection mechanisms such as eye-blink detection, 3D depth analysis, or challenge-response mechanisms to ensure that only physically present individuals are marked as attending.

➢ *Dataset Bias and Generalization*

Most surveyed systems train and test on small, homogeneous datasets that may not reflect the diversity of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

real-world student populations in terms of race, age, gender, and facial expression variability. Li and Zhang [3] acknowledge that "the performance of the system in facial recognition of different races, ages, and expressions has not been explored, and it is still difficult to completely avoid the problem of dataset bias." Future systems should be trained on large, diverse, and representative datasets to ensure equitable performance across all demographic groups.

➢ *Continuous Learning and Adaptation*

As student populations change each semester, attendance systems must support efficient enrollment of new students and re-training of recognition models. Deborah et al. [7] propose incorporating continuous learning, "updating the face recognition model with new faces and features over time to improve its accuracy and robustness." Implementing incremental learning strategies that update models without full retraining would significantly improve the practical usability of these systems.

➢ *Integration with Educational Technologies*

Future research should focus on seamlessly integrating attendance systems with broader educational management platforms including Learning Management Systems (LMS), student information systems, and engagement monitoring tools. Li and Zhang [3] demonstrate this vision through their intelligent education management system that integrates attendance, status monitoring, and report analysis. Ananda et al. [1] note that future research should focus on "developing face recognition models capable of identifying all students in a classroom and seamlessly integrating with other educational technologies, such as emotion recognition, engagement monitoring, and automated attendance systems."

VI. CONCLUSION

This survey has systematically reviewed fifteen recent works on face recognition-based automated attendance systems, covering a broad spectrum of approaches from classical Haar Cascade and LBPH methods to advanced deep learning architectures including MTCNN, FaceNet, ResNet-50, YOLOFace, and RetinaFace.

The key findings of this survey can be summarized as follows. First, classical methods such as Haar Cascade and LBPH remain widely used due to their low computational requirements and ease of implementation, but consistently underperform deep learning approaches, particularly in challenging conditions involving occlusion, low lighting, and non-frontal poses. Second, deep learning-based systems, especially those combining MTCNN for face detection with FaceNet or ResNet-50 for recognition, achieve substantially higher accuracy, with the ResNet-50 + Haar Cascade system achieving 100% accuracy under normal conditions and 91.6% under occlusion. Third, cloud-based approaches such as AWS Rekognition demonstrate remarkable scalability, achieving 100% detection and recognition accuracy for up to 49 people in a single image. Fourth, multi-face detection in real classroom environments remains an area where YOLOFace and RetinaFace excel, though the trade-off between inference speed and accuracy must be carefully managed for real-time applications.

The survey identifies several critical open challenges including illumination variability, occlusion robustness, privacy compliance, edge deployment, liveness detection, and dataset bias. Addressing these challenges will require interdisciplinary efforts combining advances in computer vision, privacy-preserving machine learning, embedded systems, and educational technology.

As artificial intelligence continues to advance, face recognition-based attendance systems are poised to become standard infrastructure in educational institutions and workplaces. The insights provided in this survey aim to guide researchers and practitioners in designing systems that are not only accurate and efficient but also robust, scalable, and ethically responsible.

REFERENCES

[1]. G. F. Ananda, H. A. Nugroho, and I. Ardiyanto, "Comparative Analysis of Multi-Face Detection Methods in Classroom Environments: Haar Cascade, MTCNN, YOLOFace, and RetinaFace," in *2024 7th International Conference on Vocational Education and Electrical Engineering (ICVEE)*, IEEE, 2024, pp. 268–273.

[2]. M. Kamruzzaman, M. E. A. Sheikh, M. R. Hossain, and M. M. B. Mojumdar, "Automated Attendance System for the Entire Classroom with Multi-Face Facial Recognition via AWS Rekognition," in *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, 2025.

[3]. L. Li and Q. Zhang, "Intelligent Education Management System Design for Universities Based on MTCNN Face Recognition Algorithm," *Scalable Computing: Practice and Experience*, vol. 25, no. 4, pp. 2463– 2475, 2024.

[4]. S. P. Krishnan and A. Manikuttan, "Attendance Management System Using Facial Recognition," in *2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*, IEEE, 2022.

[5]. O. Baker, K. Subaramaniam, D. Lopez, S. Palaniappan, and W. Li, "A Deep Learning-Based Approach Using MTCNN and FaceNet for Automated Attendance and Security Monitoring Systems," in *2025 IEEE 7th Symposium on Computers & Informatics (ISCI)*, IEEE, 2025, pp. 104–109.

[6]. A. Ghodekar, A. Phapale, T. Gunjal, D. Thorat, A. Landage, and R. Nikam, "A Face Recognition-Based Smart Student Attendance and Activeness Monitoring System," in *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, IEEE, 2023.

[7]. N. D. R, S. M, A. R. S, P. M. Amirthavarshini, A. Vinora, and K. Swathi, "An Efficient Face Recognition Based Attendance System," in *2023 3rd International Conference on Advancement in Electronics &*

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may1752

Communication Engineering (AECE), IEEE, 2023, pp. 155–160.
[8]. A. Yadav, A. Sharma, and S. S. Yadav, "Attendance Management System Based on Face Recognition Using Haar-Cascade," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, 2022, pp. 1972–1976.
[9]. I. Rajput, S. Srivastava, O. Dahiya, N. Nazir, A. Sarwar, S. Aggarwal, N Kaur, and B. Kaur, "Attendance Management System using Facial Recognition," in *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, IEEE, 2022, pp. 797–801.
[10]. N. Kumar, Sheetal, V. K. N, and Alli, "Smart Attendance by Using Face Recognition," in *2025 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, IEEE, 2025.
[11]. S. Rathi, S. Kapadnis, G. Sonar, G. Sonar, and S. R. Vispute, "MultiFacial Automated Attendance System using Haar Cascade, LBPH, and OpenCV-Based Face Detection and Recognition," in *2023 7th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, IEEE, 2023.
[12]. M. Angulakshmi and V. Susithra, "Face Recognition Based Attendance Management System Using Deep Learning Method," in *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, IEEE, 2024.
[13]. G. V. V. Reddy, Y. Manvitha, G. Ravi Teja, G. Maresh, and A. De, "Development of Attendance Monitoring System using Facial Recognition," in *2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, IEEE, 2025.
[14]. G. R. Venkatakrishnan, R. Rengaraj, R. Jeya, S. K. R, S. Manikandan, N Divya Dharshini, and M. A. Abishek, "Design and Implementation of Automated Attendance System Using Contactless Facial Recognition," in *2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, IEEE, 2024.
[15]. L. S, K. P. Karthik, T. S. Gupta, A. M. V, and A. J. Sreekar, "AIDriven Attendance Tracking with Haar Cascade and ResNet," *Procedia Computer Science*, vol. 258, pp. 3544–3553, 2025.
[16]. P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.