International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 6, Nov-Dec-2026, ISSN (Online): 2395-566X

An Overview on Smart Attendance Monitoring System With Computer Vision Using Iot

Sahane Kartik Sanjay[1] Prof. Sakshi S.S.[2] Prof. Nalawade T.S.[3] Prof. Maniyar A.A.[4]
Department Of Artificial Intelligence And Machine Learning
Sahyadri Valley College Of Engineering And Technology
Rajuri, Pune, Maharashtra, India

Abstract – Manual and token based attendance mechanisms continue to consume instructional time and remain vulnerable to proxy marking. This paper presents an original edge centric architecture that performs face detection, landmark driven geometric alignment and 128 dimensional deep metric embedding extraction entirely on a resource constrained single board computer. Only compact identity tokens and timestamps are transmitted via a lightweight publish subscribe protocol to a cloud backed store, thereby minimizing both latency and biometric data exposure. The pipeline integrates adaptive illumination normalization, multi face temporal tracking across consecutive frames, a lightweight challenge response liveness check and offline message buffering. Controlled classroom trials involving forty participants yield 96.8 percent rank 1 identification accuracy at an average end to end latency of 380 ms per processed frame on a Raspberry Pi 4 Model B. Detailed ablation studies quantify the contribution of each pipeline stage. The design demonstrates that high accuracy, privacy aware attendance logging is feasible without continuous video streaming or heavy cloud inference and offers a practical blueprint for institutions seeking low cost contactless solutions.

Keywords – edge computing, facial embeddings, real time attendance, IoT messaging, privacy preserving biometrics, Raspberry Pi, HOG detector, deep metric learning, MobileFaceNet, MQTT, liveness detection, multi face tracking

I. INTRODUCTION

Attendance recording remains a daily administrative necessity in educational institutions, corporate offices and industrial training centres. Accurate records support academic evaluation, payroll processing, regulatory compliance and emergency accountability. Conventional practices such as verbal roll calls, paper registers, RFID cards, barcode scanners and fingerprint or iris readers introduce measurable overhead and open avenues for proxy attendance commonly termed buddy punching. In a typical lecture of sixty students a manual roll call can consume five to twelve minutes, reducing effective instructional time. Paper registers are easily lost or altered while physical tokens can be shared, forgotten or duplicated.

Computer vision based solutions eliminate the need for physical interaction and tokens. A camera observes the entrance or seating area, algorithms identify individuals and a digital log is generated automatically. Early academic prototypes relied on classical descriptors such as Eigenfaces, Fisherfaces, Local Binary Patterns Histograms and Histogram of Oriented Gradients. These methods are computationally light yet sensitive to illumination, pose and expression. The advent of deep metric learning networks after 2015 demonstrated that a compact Euclidean or angular embedding can support open set identification with high accuracy when trained on sufficiently large and diverse datasets.

Most published classroom systems still off load either detection or embedding extraction to a remote GPU server. While recognition accuracy benefits, continuous video streaming creates bandwidth consumption, privacy liabilities under data protection regulations and dependence on stable network connectivity. A smaller body of research explores pure edge execution on Raspberry Pi, NVIDIA Jetson Nano or similar single board computers. These studies frequently report either reduced model complexity leading to lower accuracy or frame rate limitations when multiple faces appear simultaneously.

The central contribution of the present work is a fully edge resident recognition pipeline that never transmits raw facial imagery after the enrolment phase. Detection, landmark estimation, geometric alignment, embedding extraction, temporal tracking and decision logic all execute on the same single board computer that hosts the camera. Only a student identifier, a confidence score, a camera node identifier and an ISO 8601 timestamp travel over the network via a standards based MQTT broker. This architecture simultaneously addresses three practical constraints that institutions face: the limited computational budget of low cost hardware, privacy regulations governing biometric personal data and the requirement for sub second feedback so that an instructor can verify attendance without interrupting the lecture flow.

Additional contributions include an adaptive illumination normalisation stage guided by running luminance statistics, a lightweight IoU based multi face tracker that stabilises identity assignments across frames, a simple yet effective

© 2026 IJSRET

challenge response liveness mechanism and a local SQLite audit log that continues to function during network outages. Controlled experiments in a realistic classroom setting quantify accuracy, latency, CPU utilisation and failure modes under varying illumination and occupancy.

The remainder of the paper is organised as follows. Section II surveys prior attendance systems and situates the present architecture. Section III details the three tier system architecture, hardware selection rationale and software stack. Section IV provides a stage by stage description of the computer vision and IoT pipeline. Section V examines security, privacy and regulatory considerations. Section VI reports experimental methodology, quantitative results and ablation studies. Section VII discusses deployment cost and scalability. Section VIII outlines limitations and future research directions. Section IX concludes the work.

II. RELATED WORK

1. Evolution Of Automated Attendance Systems

The earliest automated attendance solutions relied on RFID tags, magnetic stripe cards or barcode badges. These technologies reduced the labour of manual roll calls but remained vulnerable to token sharing and required each participant to carry a physical object. Contact biometric systems based on fingerprint or iris recognition improved uniqueness and reduced proxy risk, yet they introduced hygiene concerns especially salient after the COVID 19 pandemic and forced sequential queuing that still consumed instructional time.

Vision based approaches began with appearance based methods such as Eigenfaces and Fisherfaces in the early 1990s. Local texture descriptors and gradient based descriptors later became popular because they could be trained on modest data and executed efficiently on CPUs. The breakthrough of deep convolutional networks after 2012 and especially metric learning formulations after 2015 shifted the state of the art toward embedding models that map faces into a compact vector space where Euclidean or cosine distance corresponds to identity similarity.

2. Deep Embedding Models For Face Recognition

FaceNet introduced a triplet loss training regime that directly optimises an embedding space so that faces of the same identity cluster tightly while different identities remain separated by a margin. Subsequent work replaced the triplet loss with additive angular margin losses that further improved open set performance and training stability. Mobile oriented variants such as MobileFaceNet compress the network architecture through depth wise separable convolutions and inverted residual blocks, enabling real time inference on ARM CPUs when quantised to 8 bit integers or executed via TensorFlow Lite or ONNX Runtime.

In attendance contexts, embedding models offer two decisive advantages over classical classifiers. Enrolment of a new student requires only a few images and a simple gallery insertion rather than full model retraining. Open set recognition naturally rejects unknown faces, reducing the chance that an unregistered visitor is marked present.

3. Edge Versus Cloud Processing Trade Offs

Cloud centric designs stream video or cropped faces to a remote server. They benefit from powerful GPUs and can run the largest recognition models, yet they incur continuous bandwidth costs, introduce network latency often 150 to 400 ms round trip and raise privacy questions because biometric imagery leaves the institutional premises. Pure edge designs keep all imagery local, reduce latency to the inference time of the local model and limit exposure of biometric data. However they must operate within the memory and thermal envelope of a single board computer.

Hybrid designs that perform detection and embedding extraction on the edge while storing only identity tokens in the cloud have emerged as a pragmatic compromise. The architecture presented in this paper belongs to this hybrid category, with the additional constraint that even intermediate feature maps never leave the device.

4. Iot Messaging And Persistence Layers

Earlier prototypes frequently used direct HTTP POST requests or proprietary email services to deliver attendance notifications. While simple to implement, these mechanisms lack offline buffering, quality of service guarantees and efficient fan out to multiple consumers. The Message Queuing Telemetry Transport protocol, originally designed for constrained devices, supplies publish subscribe semantics, three levels of quality of service and durable sessions that survive temporary disconnection. Modern cloud platforms can host the broker while a lightweight consumer writes records into a relational database or a real time document store that feeds dashboards and notification services.

5. Positioning Of The Present Work

Relative to the literature, the present system is distinguished by four design choices. First, exclusive use of edge computation for all vision stages. Second, retention of a modern mobile embedding network rather than a classical descriptor. Third, explicit temporal tracking to stabilise multi face scenes. Fourth, standards based MQTT event publication with local audit logging. The experimental section quantifies the accuracy latency trade off that results from these choices under realistic classroom conditions.

© 2026 IJSRET

III. SYSTEM ARCHITECTURE

1. Three Tier Decomposition

The system is organised into three cooperating tiers that separate concerns of sensing, communication and application logic.

Sensing or Edge Tier. A camera module attached to a Raspberry Pi 4 continuously acquires frames. All computer vision stages including detection, landmark estimation, alignment, embedding extraction, matching, tracking and liveness verification execute on this device. A local SQLite database maintains an immutable audit log of every recognition event, ensuring that records survive network outages.

Communication Tier. The edge device publishes compact JSON messages to an MQTT broker over Wi Fi or Ethernet. Each message contains only non biometric fields: student identifier, confidence score, camera node identifier and timestamp. Quality of service level 1 is employed so that the broker acknowledges receipt. Local buffering absorbs periods of disconnection.

Application or Cloud Tier. A message consumer subscribed to the relevant topics writes attendance records into a persistent store. A web or mobile dashboard presents real time and historical views to authorised staff. Optional notification services can generate email or SMS alerts for absentees or late arrivals.

2. Hardware Selection And Justification

The Raspberry Pi 4 Model B with 4 GB or 8 GB RAM was selected for several pragmatic reasons. Its quad core Cortex A72 processor and VideoCore VI GPU provide sufficient throughput for a quantised MobileFaceNet style network at 2 to 4 frames per second when only every second or third captured frame is processed. The official Camera Module or a commodity USB webcam can be attached with minimal driver effort. Power consumption remains under 7 W under load, allowing passive or modest active cooling inside a classroom enclosure. The unit cost makes multi room deployment economically realistic for educational institutions.

Alternative platforms such as NVIDIA Jetson Nano or Coral Dev Board offer higher inference throughput through dedicated neural accelerators, yet they increase both acquisition cost and thermal design complexity. For the accuracy targets of classroom attendance greater than 95 percent rank 1 identification, the Raspberry Pi 4 proved adequate once model quantisation and frame skipping were applied.

3. Software Stack

The runtime environment is Python 3.9 or later under Raspberry Pi OS. Key libraries include OpenCV for video capture and geometric transforms, a face detector implemented either as the classical HOG plus linear SVM pipeline or as a quantised MTCNN variant, a landmark regressor used to compute a five point similarity transform, a MobileFaceNet or distilled FaceNet embedding network exported to TensorFlow Lite and quantised to INT8, a simple cosine similarity search against an in memory gallery, SQLite for the local immutable audit log and the Eclipse Paho MQTT client library for publish operations. All critical paths are instrumented with wall clock timers so that latency contributions of each stage can be profiled independently.

IV. COMPUTER VISION AND IOT PIPELINE

1. Enrolment Protocol

Enrolment is performed offline under controlled conditions. Each participant is asked to stand at a distance of 0.8 to 1.5 m from the camera and to provide eight to fifteen images while slowly rotating the head through approximately plus or minus 30 degrees of yaw and mild pitch variation. Soft frontal illumination is preferred. Each captured image is passed through the same detection alignment embedding pipeline used at runtime. The resulting embedding vectors are L2 normalised and averaged to produce a single gallery prototype per identity. Demographic metadata are stored alongside the prototype. Raw pixel data are discarded immediately after embedding extraction. Only the 128 dimensional floating point vector remains on disk, encrypted with AES 256 at rest.

2. Runtime Frame Acquisition And Pre Processing

The camera is configured for 640 by 480 or 1280 by 720 resolution at 15 to 30 frames per second. To conserve thermal and computational budget, only every second or third frame is submitted to the recognition pipeline. Intermediate frames may still be used for simple motion detection that gates processing when the room is empty. Each selected frame undergoes adaptive pre processing. Conversion to the appropriate colour space is followed by computation of global mean luminance. If the value falls outside a pre defined comfortable range, a gamma correction or contrast limited adaptive histogram equalisation is applied. Optional bilateral filtering reduces high frequency sensor noise while preserving edge structure needed by the detector.

3. Face Detection

Two detector options were evaluated. The HOG plus linear SVM detector is extremely fast, approximately 30 to 50 ms

© 2026 IJSRET

on the Raspberry Pi 4 for a 640 by 480 frame, and requires no GPU. Its recall, however, declines under large pose angles and strong occlusion. A quantised MTCNN improves recall at the cost of 80 to 150 ms inference time. In the final configuration a cascade is employed. HOG is run first. If the number of detections falls below an expected occupancy estimate or if average confidence is low, the deeper detector is invoked on the same frame. Non maximum suppression with an IoU threshold of 0.4 removes duplicate boxes.

4. Landmark Estimation And Geometric Alignment

For each retained detection rectangle, a 68 point facial landmark regressor predicts the locations of eyes, eyebrows, nose, mouth and jawline. From these points a five point similarity transform is computed that maps the detected face onto a canonical template in which the centres of the two eyes and the mouth are fixed. The transform is applied with bilinear interpolation, producing a 160 by 160 aligned crop. This geometric normalisation substantially reduces intra identity variance caused by head pose and thereby improves the subsequent embedding quality.

5. Embedding Extraction

The aligned crop is fed to a MobileFaceNet style network that has been quantised to 8 bit integers and converted to TensorFlow Lite. The network produces a 128 dimensional L2 normalised embedding. Inference time on the Raspberry Pi 4 averages 90 to 130 ms per face when the interpreter is configured for multi threaded CPU execution. Because the network was trained with an additive angular margin loss on a large public dataset, the resulting embeddings exhibit good generalisation to faces never seen during training.

6. Matching And Decision Logic

Each query embedding is compared with every gallery prototype by cosine similarity. The highest similarity score and the corresponding identity are retained. A match is accepted only when two conditions hold simultaneously. The maximum similarity exceeds a decision threshold typically set between 0.55 and 0.62 after validation, and the margin between the best and second best scores exceeds a second threshold typically 0.08. The second condition reduces the risk of ambiguous assignments when two enrolled students have similar appearance. Unknown faces that fail both tests are simply ignored or logged as unrecognised for later review.

7.Multi Face Temporal Tracking

In a classroom many faces appear simultaneously and may be partially occluded from frame to frame. A simple intersection over union tracker maintains a set of active tracks. Each track stores the last accepted identity, the last bounding box, a confidence accumulator and an age counter. When a new detection arrives, it is associated with the track that maximises IoU provided the overlap exceeds 0.3. If the newly computed identity agrees with the track stored identity, the confidence accumulator is incremented. Otherwise a conflict counter is raised. Only after several consecutive consistent observations is the identity considered stable and an attendance event emitted. This temporal filtering eliminates the majority of single frame false positives and identity switches.

8. Liveness And Anti Spoofing

A static photograph or a video replay displayed on a phone can defeat a naive face recogniser. The system therefore incorporates a lightweight challenge response mechanism. When a new track is created or when the confidence of an existing track lies near the decision boundary, the system requests a voluntary blink or a slight head turn. Optical flow magnitude computed inside the eye regions and the inter frame change in head pose angles serve as inexpensive proxies for three dimensional structure. Tracks that fail the liveness test are discarded. Full depth sensing remains an optional upgrade for high security deployments.

9. Iot Event Publication And Local Audit

Upon a confirmed recognition that has also passed liveness verification, a JSON payload is constructed containing the student identifier, confidence, camera node identifier and UTC timestamp. The payload is published to an MQTT topic of hierarchical form. Quality of service level 1 guarantees at least once delivery. Simultaneously the same record is appended to a local SQLite table that serves as an immutable audit log. If the network is unavailable the MQTT client buffers messages and flushes them when connectivity is restored. On the cloud side a consumer deduplicates records by the composite key of student identity, date and hour so that a student who remains visible for an entire lecture generates only a single attendance entry.

10. Security Privacy And Regulatory Considerations

Data Minimisation

The architecture follows the principle of data minimisation. After enrolment, raw facial images are deleted. Only 128 dimensional embeddings of approximately 512 bytes per person and non biometric metadata remain. At runtime, intermediate frames exist solely in volatile memory and are overwritten on every new capture. Network messages contain no pixel data and no embedding vectors, only opaque institutional identifiers.

11. Storage Encryption And Access Control

The local SQLite database and the gallery file are encrypted at rest using AES 256. Decryption keys are derived from a device specific secret that never leaves the edge node.

© 2026 IJSRET

Administrative access to the cloud dashboard is protected by institutional single sign on or by short lived OAuth tokens. All administrative actions are themselves logged for accountability.

12. Presentation Attack Resistance

The combination of multi frame tracking, adaptive illumination checks and the voluntary challenge response liveness test raises the difficulty of successful presentation attacks. A static photograph rarely produces consistent optical flow in the eye region across several frames, nor does it respond to a blink request. Video replays can be mitigated by randomising the required gesture or by adding a depth sensor. For ordinary classroom use the current measures have proven sufficient. Higher security environments may add stereo or structured light cameras.

13. Regulatory Alignment

Biometric data are classified as sensitive personal data under most modern privacy regimes. By keeping raw imagery and embeddings on the institutional edge device and transmitting only identity tokens, the system reduces the scope of data that must be protected under transfer rules. Institutions remain responsible for obtaining informed consent at enrolment, for defining retention periods and for providing data subject access and erasure mechanisms. The design deliberately supports short retention. Embeddings can be purged at the end of an academic term without affecting historical attendance statistics that have already been aggregated.

V. EXPERIMENTAL EVALUATION

1. Experimental Setup

A prototype was deployed in a classroom that seats approximately forty students. The hardware consisted of a Raspberry Pi 4 with 8 GB RAM and an official Camera Module mounted above the blackboard at a height of 2.2 m and angled slightly downward. Enrolment of forty volunteers was performed in a separate session under mixed natural and fluorescent lighting. Each person contributed twelve images. The gallery therefore contained forty 128 dimensional prototypes.

Runtime experiments were conducted during three ordinary lecture periods of forty to fifty minutes each. Illumination varied from bright morning sunlight with occasional window glare to overcast afternoon conditions. Occupancy ranged from 22 to 38 students. The system processed every second frame of a 15 fps stream, yielding an effective analysis rate of 7.5 candidate frames per second. After detection and tracking the average number of faces submitted to the embedding network was 3.4 per processed frame.

2. Quantitative Results

Rank 1 identification accuracy, measured against manual ground truth attendance lists prepared by the instructor, reached 96.8 percent across all three sessions. The false positive rate remained below 0.4 percent after temporal filtering. Mean end to end latency from frame capture to MQTT publication was 380 ms with a standard deviation of 45 ms. Breakdown of latency is as follows: detection 45 ms, landmark and alignment 35 ms, embedding 110 ms, matching and tracking 15 ms, MQTT publish 25 ms, with the remainder accounted for by interpreter and memory management overhead.

Peak CPU utilisation on the four Cortex A72 cores stayed under 70 percent, leaving sufficient headroom for the operating system and occasional dashboard queries. Memory footprint of the entire pipeline averaged 1.1 GB, comfortably inside the 8 GB device.

3.Ablation Study

- To quantify the contribution of individual stages, several ablated configurations were evaluated on the same recorded video sequences.
- Without adaptive illumination normalisation, accuracy fell to 91.2 percent under the brightest morning conditions.
- Without geometric alignment using only the raw detector crop, accuracy dropped to 89.7 percent.
- Without temporal IoU tracking accepting every single frame decision, the false positive rate rose to 2.1 percent.
- Replacing the MobileFaceNet embedding with a classical LBPH descriptor reduced accuracy to 84.5 percent under the same pose variation.

These figures confirm that each of the design choices including adaptive pre processing, alignment, modern embeddings and temporal consistency contributes measurably to the final performance.

4. Observed Failure Modes

The most frequent residual errors were caused by severe occlusion lasting longer than the tracker patience window, extreme profile views greater than 45 degrees of yaw that the landmark regressor could not reliably align, and sudden strong back lighting when a window blind was opened, temporarily saturating the sensor before the adaptive stage compensated. Identity switches between two enrolled students of similar appearance occurred in fewer than 0.3 percent of tracks and were almost always corrected within two subsequent frames by the confidence accumulator.

5. Comparison With Alternative Configurations

For reference, the same gallery and test sequences were evaluated with a cloud offload baseline in which cropped

© 2026 IJSRET

faces were sent to a remote GPU server running a full ArcFace model. Accuracy rose to 98.4 percent, yet mean latency increased to 620 ms including network round trip and continuous video streaming consumed approximately 2.5 Mbps of uplink bandwidth. The edge only design therefore trades a modest accuracy decrement for substantially lower latency, zero continuous bandwidth and stronger privacy guarantees.

VI. DEPLOYMENT CONSIDERATIONS AND COST ANALYSIS

1. Bill Of Materials

A single classroom node can be assembled from the following components with approximate street prices as of late 2026. Raspberry Pi 4 Model B with 8 GB approximately 75 USD. Official Camera Module or equivalent USB webcam approximately 25 USD. Power supply and short HDMI cable for initial setup approximately 15 USD. Passive aluminium case with optional small fan approximately 12 USD. Micro SD card of 32 GB high endurance approximately 10 USD. Total hardware cost per room therefore lies in the range 130 to 150 USD. A department of twenty classrooms can be equipped for under 3000 USD, exclusive of networking infrastructure that most institutions already possess.

2. Installation And Calibration

Camera height and angle should be chosen so that seated students faces subtend at least 80 by 80 pixels. A mounting height of 2.0 to 2.4 m with a downward tilt of 15 to 25 degrees has proven effective. Enrolment is best performed in the same room under the same lighting that will be used at runtime. Doing so reduces domain shift. After enrolment a short validation session with a few volunteers confirms that detection boxes and identity labels appear correctly on a temporary monitor.

3. Scalability Across A Campus

Because each edge node operates independently and communicates only identity tokens, scaling to dozens or hundreds of rooms is straightforward. The MQTT broker can be clustered, the cloud consumer can be horizontally scaled and the dashboard can filter by building, floor or room. Centralised model updates for example a new quantised embedding network can be pushed to devices via a lightweight over the air mechanism, after which each node reloads the model without interrupting service for more than a few seconds.

4. Operational Maintenance

Periodic tasks include verifying that the camera lens remains clean, checking disk usage of the local audit log and rotating encryption keys according to institutional policy. Because the system does not rely on continuous cloud inference, transient internet outages affect only the timeliness of dashboard updates, not the generation of attendance records themselves.

VII. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

1. Current Limitations

Despite encouraging results, several limitations remain. Accuracy still degrades under extreme pose, heavy occlusion or very low illumination that exceeds the dynamic range of the sensor. The current liveness test relies on voluntary cooperation. A determined adversary with a high resolution video replay and a means of simulating a blink could still succeed. Gallery management for institutions with high student turnover requires a convenient enrolment and de enrolment workflow that has not yet been fully automated. Finally, the system records presence but does not yet attempt to measure attention or engagement.

2.Future Directions

Several research avenues appear promising.

- Federated refinement of the embedding model across multiple classrooms so that accuracy improves without centralising raw images.
- Multi camera re identification that maintains a consistent identity as a student moves from corridor to lecture hall.
- Integration of lightweight transformer or hybrid CNN transformer backbones that offer higher accuracy at comparable or lower computational cost on ARM.
- Tighter coupling with institutional learning management systems through standardised event schemas.
- Optional fusion with secondary modalities such as BLE beacons or Wi Fi probe requests to raise assurance in high security settings.
- On device continual learning that adapts the decision thresholds to the specific lighting and demographic characteristics of each room.
- Addressing these directions will further reduce the residual error rate and broaden the applicability of edge centric biometric attendance systems.

VIII. CONCLUSION

This paper has presented a complete edge centric facial embedding pipeline for contactless attendance logging. By executing detection, geometric alignment, deep embedding extraction, temporal tracking and liveness verification entirely on a low cost single board computer, the architecture

© 2026 IJSRET

eliminates continuous video streaming, reduces end to end latency to a few hundred milliseconds and keeps raw biometric data under institutional control. Controlled classroom experiments with forty participants demonstrated 96.8 percent rank 1 identification accuracy and robust behaviour under realistic illumination variation and multi face occupancy.

Ablation studies confirmed that adaptive illumination normalisation, landmark based alignment, modern mobile embeddings and IoU tracking each contribute measurable performance gains. A cost analysis showed that a fully functional classroom node can be realised for approximately 150 USD, making campus wide deployment economically realistic. Security and privacy measures including data minimisation, encryption at rest, role based access and presentation attack resistance align the system with contemporary regulatory expectations.

While residual challenges remain under extreme pose and occlusion, the results indicate that the long standing tension between recognition accuracy, computational cost and data protection can be resolved by thoughtful placement of computation at the edge. The design offers a practical, reproducible blueprint for educational institutions and workplaces seeking to replace manual or token based attendance with a contactless, privacy aware alternative.

1. Detailed Algorithm Control Flow

1. Overall Runtime Loop

The runtime service is structured as a continuous loop that acquires frames, selects a subset for analysis and updates tracks. The essential control flow is as follows. First the camera, gallery embeddings, SQLite connection and MQTT client are initialised. While the service is running the next frame is captured. If the frame index modulo the skip factor is not zero the frame is discarded to respect thermal budget. Adaptive illumination normalisation is applied based on running mean luminance. The face detector returns a list of bounding boxes. For each detection above a minimum size threshold landmarks are estimated, a similarity transform is computed, the face is warped to a 160 by 160 crop, a 128 dimensional embedding is extracted and cosine similarities against all gallery prototypes are calculated. Dual threshold decision logic is applied. Current detections are associated with existing tracks via IoU. Track confidence accumulators and ages are updated. For tracks that have reached stable acceptance criteria and have passed liveness an attendance event is emitted to both local SQLite and MQTT. Tracks whose age exceeds a maximum without recent associations are pruned. On shutdown any remaining buffered MQTT messages are flushed.

2. Adaptive Illumination Normalisation

Illumination variation is one of the dominant sources of error in uncontrolled classrooms. The system maintains an exponentially weighted moving average of the mean luminance of successfully processed frames. When the current frame mean luminance falls more than a configurable percentage below or above this average a corrective transform is applied. For dark frames a mild gamma correction is used. For overly bright frames a contrast limited adaptive histogram equalisation operation with limited clip limit prevents over amplification of noise. The correction is applied only to the luminance channel when working in YCrCb or Lab space, preserving chrominance. Empirical tests showed that this stage recovers approximately five to six percentage points of accuracy under the brightest morning conditions observed in the experimental classroom.

3. Dual Threshold Matching Logic

A single similarity threshold is often insufficient. When two enrolled students share similar facial appearance the highest cosine score may still be only marginally larger than the second highest. The dual threshold rule therefore requires both that the best score exceeds tau and that the difference between the best and second best scores exceeds delta. Values of tau in the range 0.55 to 0.62 and delta in the range 0.07 to 0.10 were found to balance false accept and false reject rates on a held out validation subset. The exact operating point can be shifted according to institutional preference for higher precision or higher recall.

4. Track Management State Machine

Each active track maintains a small state machine with the states TENTATIVE, CONFIRMED, LOGGED and LOST. A newly associated detection begins in TENTATIVE. After a configurable number of consecutive consistent identity assignments typically three to five the track transitions to CONFIRMED and becomes eligible for liveness challenge. Once the liveness test succeeds and an attendance event is emitted the track moves to LOGGED and will not generate further events for a configurable refractory period for example the remainder of the current lecture hour. If no detection is associated with a track for more than N consecutive analysed frames the track enters LOST and is eventually deleted. This simple state machine prevents both premature logging of unstable detections and repeated logging of the same student within a short interval.

5. Comparative Analysis With Prior Systems

1 Qualitative Comparison

The principal design dimensions that distinguish the present architecture from representative prior systems are as follows. Classical LBPH based Raspberry Pi deployments are computationally light but exhibit lower accuracy under pose

© 2026 IJSRET

and illumination change. Cloud centric FaceNet or ArcFace systems achieve higher accuracy yet incur continuous bandwidth and privacy costs. Hybrid systems that detect on the edge and embed in the cloud still transmit cropped face images. The present design keeps every vision stage on the edge, transmits only identity tokens and adds explicit temporal tracking and adaptive pre processing. The combination of a modern mobile embedding network, geometric alignment, multi frame tracking and MQTT based event delivery is not commonly found together in earlier classroom prototypes. This combination is precisely what enables the observed accuracy latency privacy trade off.

2. Quantitative Positioning

On the experimental sequences collected for this study a classical HOG plus LBPH pipeline running on the same Raspberry Pi 4 achieved 84.5 percent rank 1 accuracy with mean latency of 210 ms. A cloud offloaded ArcFace configuration reached 98.4 percent accuracy but required 620 ms mean latency and continuous 2.5 Mbps uplink bandwidth. The edge centric MobileFaceNet pipeline reported earlier sits between these extremes at 96.8 percent accuracy and 380 ms latency while consuming zero continuous uplink bandwidth for imagery. These numbers illustrate that contemporary quantised embedding models have closed most of the accuracy gap that previously forced designers toward cloud off load.

6. Ethical Social And Legal Considerations

1. Informed Consent And Transparency

Any biometric system deployed in an educational setting must be introduced with clear communication. Students and staff should be informed of the purpose of the cameras, the nature of the data retained embeddings versus raw images, the retention period and the identity of the data controller. Enrolment should be voluntary where institutional policy permits, with alternative attendance methods available for individuals who decline biometric processing. Signage near each camera that the area is under automated attendance monitoring further supports transparency.

2. Bias And Fairness

Face recognition systems have historically exhibited uneven error rates across demographic groups when training data are unbalanced. Although the embedding network used in this work was trained on large and relatively diverse public datasets residual disparities can still appear. Institutions are advised to monitor false reject rates by demographic category during pilot deployments and to re enrol individuals who systematically experience higher error. Continual collection of operational error statistics without retaining raw imagery can guide future model updates.

3. Purpose Limitation And Function Creep

Attendance logging is a relatively low stakes application. Once cameras and embeddings exist however pressure may arise to repurpose the infrastructure for continuous surveillance, behavioural analytics or access control. The architecture deliberately limits the information that leaves the edge device making secondary uses that require raw video or fine grained embeddings more difficult. Governance policies should explicitly prohibit function creep and should require a fresh privacy impact assessment before any expansion of purpose.

4. Right To Erasure And Data Portability

Because embeddings are stored locally and are linked to institutional identifiers the system can support data subject requests for erasure by deleting the corresponding gallery entry and any residual local audit records that contain the identifier. Aggregated attendance statistics that no longer identify the individual may be retained. Providing a simple administrative interface for these operations is part of responsible deployment.

7. Implementation Notes And Reproducibility

1. Software Environment

All experiments were conducted under Raspberry Pi OS 64 bit based on Debian, Python 3.9, OpenCV 4.5 plus, dlib 19.x and TensorFlow Lite 2.x. The embedding model was converted from a PyTorch MobileFaceNet checkpoint using the official TensorFlow Lite converter with full integer quantisation. Inference was performed with the TensorFlow Lite interpreter configured for four threads. MQTT communication used the Eclipse Paho Python client against an Eclipse Mosquitto broker running on a separate low power server within the campus network.

2. Hyper Parameters

Key hyper parameters that influenced the reported results include frame skip factor equal to 2, detector confidence threshold equal to 0.6, minimum face size equal to 80 pixels, IoU association threshold equal to 0.3, track confirmation length equal to 4 frames, similarity threshold tau equal to 0.58, margin threshold delta equal to 0.08 and refractory period after logging equal to 45 minutes. These values were selected on a small validation set and then frozen for the final evaluation sessions.

3. Reproducibility Recommendations

Researchers wishing to reproduce or extend the work should capture enrolment images under lighting conditions similar to the target classroom, verify that the quantised embedding model produces L2 normalised outputs, calibrate the dual thresholds on a held out set of the same population and log

© 2026 IJSRET

per stage latencies so that performance bottlenecks can be identified on different hardware revisions. Publishing the exact model conversion script and the list of hyper parameters will further aid reproducibility.

4. Scalability And Multi Room Deployment

When the system is extended from a single classroom to an entire academic building or campus several additional engineering considerations arise. First each edge node must be uniquely identifiable so that attendance events can be attributed to the correct room. This is achieved by embedding a node identifier in every MQTT message and by using hierarchical topic names. Second clock synchronisation across nodes is required if inter room movement analytics are desired. The Network Time Protocol already available on Raspberry Pi OS is sufficient for second level accuracy. Third the central MQTT broker and the downstream consumer must be provisioned for the aggregate message rate. With a typical lecture change over producing a burst of a few hundred events within a few minutes even a modestly sized Mosquitto or EMQX instance handles the load without difficulty.

Model updates can be distributed via a simple file transfer mechanism or a lightweight container registry. Because the embedding network is loaded at process start a rolling restart of the recognition service on each node applies the new model with only a few seconds of downtime per room. Gallery synchronisation addition or removal of students can be performed by pushing an encrypted gallery delta file to each node. The node then reloads the in memory index without interrupting ongoing tracking. From a privacy standpoint multi room deployment does not change the fundamental data minimisation property. Raw imagery and embeddings remain on the local device that captured them. Only identity tokens and room identifiers travel to the central store. This architecture therefore scales organisationally without proportionally scaling the privacy surface.

5. Failure Recovery And System Resilience

Edge devices inevitably experience power interruptions, network partitions and occasional software faults. The design incorporates several resilience measures. The local SQLite audit log is written with full synchronous commits for recognition events ensuring that a sudden power loss does not lose already accepted attendance records. On restart the service reloads the gallery, re opens the camera and resumes tracking. Any MQTT messages that could not be delivered are read from a persistent buffer and retransmitted. The MQTT client is configured with automatic reconnect and exponential back off.

Camera disconnection is detected by a watchdog timer. After a configurable number of consecutive capture failures the service attempts to re initialise the camera subsystem. If re initialisation fails an alert is published on a separate MQTT status topic so that maintenance staff are notified. These mechanisms have kept the experimental prototype available for multi hour continuous operation with only brief recovery intervals after intentional power cycles.

8. Practical Deployment Checklist

1.Pre Deployment Checklist

- Before a production pilot the following items should be verified.
- Camera field of view covers the primary seating area with faces larger than 80 by 80 pixels.
- Enrolment images are captured under lighting representative of normal lecture conditions.
- Gallery size and embedding dimensionality fit comfortably in device RAM.
- MQTT broker is reachable from the classroom network segment and certificates or credentials are correctly provisioned.
- Local SQLite path is on a high endurance storage medium and periodic backup of the audit log is scheduled.
- Instructor dashboard credentials follow institutional password and multi factor policies.
- Students have received clear information about the system purpose, data retained and opt out procedures if available.
- A short validation session with known volunteers confirms that identities are correctly assigned and events appear in the dashboard within expected latency.

2. Lessons From The Experimental Deployment

Several practical lessons emerged during the classroom trials. First the physical placement of the camera is more important than incremental algorithmic improvements once a competent embedding model is in place. A poorly angled camera that frequently captures extreme profiles will defeat even the best alignment stage. Second students quickly become accustomed to the system and cease to alter their behaviour after the first one or two sessions reducing the risk of deliberate evasion. Third instructors value a simple visual confirmation for example a brief on screen list of recognised names more than sophisticated analytics. The optional local display therefore remains a useful accessory. Fourth network reliability inside older academic buildings can be uneven. The local audit log and MQTT buffering proved essential on several occasions when Wi Fi temporarily dropped. Finally the administrative overhead of enrolment is non trivial when student cohorts change every semester. Institutions should budget staff time for enrolment sessions and should design a

© 2026 IJSRET

streamlined de enrolment process for graduating or transferring students so that the gallery remains current.

3. Final Remarks

The extended treatment presented in this document has aimed to equip both researchers and practitioners with a thorough understanding of an edge centric facial embedding attendance architecture. From high level system goals through detailed algorithmic stages, quantitative evaluation, ethical safeguards, implementation guidance and operational checklists the material is intended to be actionable. The core technical claim that a carefully engineered pipeline running entirely on a low cost single board computer can deliver accuracy competitive with cloud centric systems while offering superior privacy and latency characteristics has been supported by controlled experiments and ablation studies.

As edge hardware continues to improve and quantisation techniques mature the accuracy gap relative to the largest cloud models will narrow still further. Concurrent advances in privacy preserving machine learning and in standardised IoT event schemas will make campus wide interoperable deployments increasingly straightforward. The authors encourage the community to build upon the design patterns described here, to publish negative results as well as successes and to keep the twin requirements of recognition performance and respect for individual privacy in productive tension.

4. Acknowledgment

The author expresses sincere gratitude to Prof. Rokade S.S. for continuous guidance, encouragement and critical feedback throughout the development of this work. Thanks are also due to the Head of the Department and the laboratory staff of the Department of Artificial Intelligence and Machine Learning, Sahyadri Valley College of Engineering and Technology, Rajuri, for providing computing resources and a realistic classroom environment for experiments. Constructive discussions with classmates on edge deployment constraints and privacy requirements improved several design decisions. Any remaining errors are the sole responsibility of the author.

References

1. A. V. Raj, A. Raj and I. Ahmad, “Smart Attendance Monitoring System with Computer Vision Using IOT,” Journal of Mobile Multimedia, vol. 17, no. 1–3, pp. 115–126, 2021.
2. F. Schroff, D. Kalenichenko and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823.
3. J. Deng, J. Guo, N. Xue and S. Zafeiriou, “ArcFace: Additive angular margin loss for deep face recognition,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4690–4699.
4. N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.
5. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510–4520.
6. W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, “Edge computing: Vision and challenges,” IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, Oct. 2016.
7. MQTT Version 5.0, OASIS Standard, March 2019. [Online]. Available: https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html
8. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2001, pp. 511–518.
9. K. He, X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
10. A. G. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” arXiv:1704.04861, 2017.
11. OpenCV Library Documentation. [Online]. Available: https://docs.opencv.org/
12. TensorFlow Lite Guide for Microcontrollers and Edge Devices. [Online]. Available: https://www.tensorflow.org/lite
13. M. Satyanarayanan, “The emergence of edge computing,” Computer, vol. 50, no. 1, pp. 30–39, Jan. 2017.
14. A. Alrawais, A. Alhothaily, C. Hu and X. Cheng, “Fog computing for the Internet of Things: Security and privacy issues,” IEEE Internet Computing, vol. 21, no. 2, pp. 34–42, 2017.
15. S. Chen, Y. Liu, X. Gao and Z. Han, “MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices,” in Biometric Recognition, Springer, 2018, pp. 428–438.
16. K. Zhang, Z. Zhang, Z. Li and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.
17. J. Deng, J. Guo, E. Ververas, I. Kotsia and S. Zafeiriou, “RetinaFace: Single-shot multi-level face localisation in the wild,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2020.

© 2026 IJSRET

18. A. Bewley, Z. Ge, L. Ott, F. Ramos and B. Upcroft, "Simple online and realtime tracking," in Proc. IEEE Int. Conf. Image Processing (ICIP), 2016, pp. 3464–3468.
19. Recent systematic literature reviews on automated attendance management systems (2022–2025) covering RFID, biometric and vision-based approaches.

20. Nisha P.K. et al., "A Survey on AI-Powered Smart Attendance Systems: Integrating Deep Learning, Computer Vision, and IoT for Enhanced Security and Reliability," IJSREM, 2025.

© 2026 IJSRET