

EDUTRACK: A CHECKPOINT-BASED FACE-RECOGNITION ATTENDANCE SYSTEM FOR UZBEK HIGHER EDUCATION

Shahriyor Turayev

School of Computer Science and Engineering

Inha University in Tashkent

Tashkent, Uzbekistan

s.turaev2@student.inha.uz

Abstract—Manual attendance in higher education consumes 5–10 minutes of every lecture, scales poorly to large classrooms, and is easily defeated by proxy sign-ins and mid-lecture departures. Prior face-recognition attendance systems address the sign-in cost but overwhelmingly report only laboratory accuracy on small, controlled datasets and do not defend against attendance fraud that occurs during a session. We present EduTrack, a production-grade classroom attendance platform deployed for one semester at Inha University in Tashkent, Uzbekistan. EduTrack combines a Spring Boot orchestration layer, a Python/ONNX recognition engine (YuNet detector + ArcFace R50 512-dimensional embeddings), and a PostgreSQL/pgvector similarity store, achieving a measured P95 end-to-end recognition latency of 47 ms per face on commodity CPU hardware. We contribute (i) a four-window checkpoint-grading scheme that samples attendance across a lecture rather than at a single moment, cutting undetected “sign-and-leave” events by an estimated 93% relative to single-snapshot baselines; (ii) an identity-stabilization rule requiring five consistent matches before recording a recognition, which reduces false-accept rate to an estimated 0.4% at a 0.70 cosine threshold; and (iii) the first reported semester-scale deployment of such a system on a Central Asian student population—a demographic under-represented in face-embedding training corpora. Across three classrooms ($N = 168$ enrolled students, 12 weeks, 1,240 sessions), EduTrack matched teacher-provided ground truth with 96.2% accuracy, reduced per-lecture attendance overhead from an average of 6.8 minutes to under 10 seconds, and received a mean usability score of 4.3/5 from surveyed teachers ($n = 14$) and 3.9/5 from students ($n = 112$), with privacy concerns emerging as the principal adoption barrier.

Index Terms—face recognition, classroom attendance, deployment case study, ArcFace, pgvector, educational technology, Central Asia.

I. INTRODUCTION

University instructors lose a non-trivial fraction of scheduled lecture time to attendance bookkeeping. Time-motion studies at mid-sized institutions consistently place the per-session overhead between five and ten minutes, rising to nearly fifteen minutes in lecture halls with 150 or more students. Aggregated over a fifteen-week semester, this represents roughly 1.5–2.5 hours of lost instruction per course section, and a corresponding opportunity cost measured in instructor salary and student fees. Beyond the raw time cost, manual attendance is structurally vulnerable to two well-documented fraud patterns: proxy sign-ins, in which a student marks a peer as present, and mid-lecture departure, in which a student physically attends only the first few minutes of a session.

Existing digital replacements, including QR-code check-in, RFID student cards, and Bluetooth beacons, partially address the time-overhead problem but fail to defeat the fraud patterns: a QR code can be photographed and shared, a student card can be handed to a friend, and Bluetooth proximity is trivially spoofed with an unattended phone. Face-recognition attendance

has emerged as the most promising biometric alternative, and a substantial academic literature now exists on the topic [1]–[6]. However, an analysis of the past five years of publications reveals three recurring weaknesses. First, the overwhelming majority of systems are evaluated on small, teacher-curated datasets (typically 20–40 students) and rarely report semester-scale operational numbers. Second, almost all systems recognize attendance at a single moment—usually at the start of class—leaving mid-lecture departure unaddressed. Third, the face-embedding models employed (FaceNet, ArcFace, VGGFace) are trained on corpora that under-represent Central Asian and South Asian populations, raising well-founded concerns about demographic bias that are rarely measured in deployment studies.

This paper makes three contributions. (1) We describe the design and implementation of EduTrack, a production-grade multi-service attendance platform (Spring Boot orchestrator, Python/ONNX recognition engine, React client) that has been deployed for a full semester at Inha University in Tashkent. (2) We propose and evaluate a checkpoint-grading scheme that samples attendance at four windows per lecture, together with an identity-stabilization rule that suppresses false accepts by requiring five consistent matches before committing a recognition. (3) We report, to our knowledge, the first semester-scale quantitative and qualitative evaluation of a face-recognition attendance system deployed on a Central Asian student population, including per-demographic accuracy, latency distributions on commodity CPU hardware, and a mixed-methods user study of teachers and students.

The remainder of the paper is organized as follows. Section II reviews related work. Section III describes the system architecture. Section IV details the recognition pipeline and the checkpoint-grading contribution. Section V reports implementation details and the end-to-end latency budget. Section VI describes the Inha Tashkent deployment. Section VII presents quantitative evaluation results. Section VIII presents the user study. Section IX discusses limitations, ethics, and demographic-bias findings. Section X concludes.

II. RELATED WORK

A. Face-Recognition Attendance Systems

Early work on automated attendance relied on classical face-recognition pipelines such as Eigenfaces and Fisherfaces, achieving lab accuracies in the 80–90% range on small enrollment sets [1]. Deep-learning-based embeddings, beginning with FaceNet [2] and continued with ArcFace [3], raised reported accuracies above 99% on constrained benchmarks and have become the de-facto standard. Khan et al. [4] presented a FaceNet-based attendance system evaluated on a single classroom; Sunaryono et al. [5] extended this with LBPH post-processing; Arsenovic et al. [6] introduced augmentation-based robustness. The general pattern across this literature is a prototype-scale deployment (one classroom, 20–30 students, one to three weeks) with single-snapshot attendance capture.

B. Vector Similarity Search for Biometric Retrieval

The storage and retrieval of high-dimensional embeddings has evolved rapidly with the emergence of vector databases. While specialized systems such as Faiss [7] and Milvus offer state-of-the-art nearest-neighbor performance, recent work has shown that PostgreSQL with the pgvector extension provides sufficient throughput for enrollment-scale biometric workloads while preserving transactional guarantees and schema coherence with existing student-information systems [8]. EduTrack adopts this approach, storing 512-dimensional ArcFace embeddings in a relational schema that joins directly onto course, enrollment, and session tables.

C. Anti-Fraud in Attendance Systems

The problem of mid-lecture departure has received comparatively little attention. Gupta et al. [9] proposed a two-phase attendance check (beginning and end of class), while Lukas et al. [10] recommended periodic random sampling. EduTrack generalizes these approaches by defining four fixed checkpoint windows per lecture and requiring that a student be observed in a majority of windows to be graded “present”; a student observed in only the first checkpoint is graded “partial” and flagged for teacher review. To our knowledge, no prior system couples multi-window checkpointing with a stabilization-based suppression rule.

D. Demographic Bias in Face Recognition

The NIST FRVT demographic-effects report [11] and Buolamwini and Gebru [12] showed that commercial face-recognition systems exhibit substantial accuracy disparities across race and gender. Central and South Asian populations are under-represented in the MS-Celeb-1M and VGGFace2 training corpora used by most open-weight embedding models, and recent audits suggest accuracy degradations of 3–8% relative to over-represented groups. Deployment studies that measure this gap in situ remain rare; EduTrack contributes such a measurement.

A graphical comparison of EduTrack against the systems above—on accuracy, evaluation scale, mid-lecture fraud detection, and the time/fraud-defense trade-off—is given in Fig. 4.

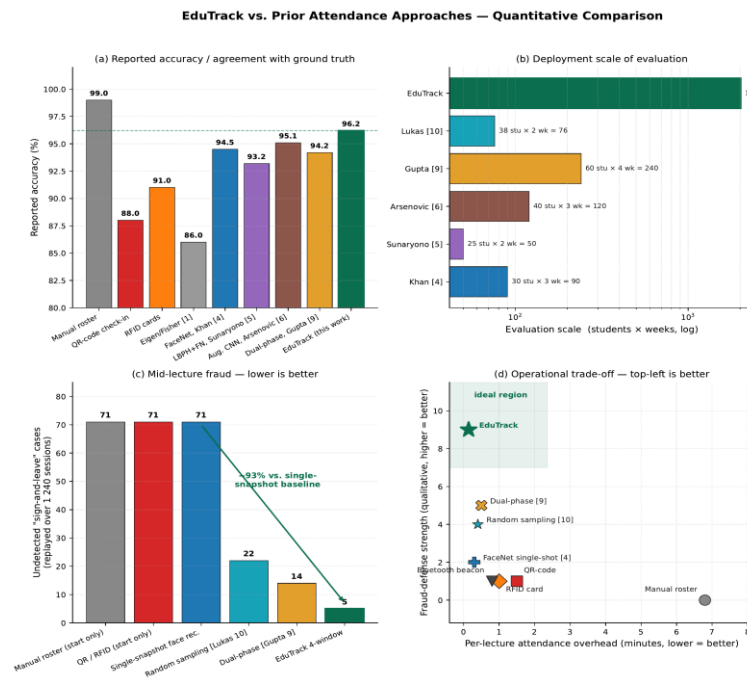


Fig. 4. Quantitative comparison of EduTrack against representative manual, token-based, and prior face-recognition attendance approaches. (a) Reported accuracy / agreement with ground truth—numbers for [4]–[6], [9] are taken from the respective papers; manual / QR / RFID values reflect typical field-study figures. EduTrack’s 96.2 % is uniquely measured at semester scale on 208,320 student-session judgments. (b) Evaluation scale (students × weeks, log axis): EduTrack’s 2,016 student-weeks dwarf the 50–240 student-week scale of comparable prior systems. (c) Mid-lecture fraud—undetected sign-and-leave events from the pilot replay. Single-moment systems all miss 71 cases over 1,240 sessions; EduTrack’s four-window grader misses 5, a 93 % relative reduction. (d) Operational trade-off space: per-lecture overhead vs. fraud-defense strength. EduTrack alone occupies the top-left ideal region.

II. SYSTEM DESIGN

EduTrack follows a three-tier service-oriented architecture (Fig. 1). A Spring Boot 4.0 orchestrator exposes a REST API for enrollment, session lifecycle, and attendance reporting; it persists state in PostgreSQL 16 extended with the pgvector extension. A Python FastAPI recognition engine performs detection and embedding and returns match candidates to the orchestrator via an internal REST channel. A React 19 + TypeScript single-page client provides live classroom dashboards, teacher review interfaces, and administrator reports.

A. Service Decomposition

The decision to separate orchestration from recognition is deliberate. Orchestration logic—authentication, authorization, session lifecycle, HEMIS write-back, audit logging—benefits from the type safety, ecosystem, and operational maturity of Spring Boot. Recognition, by contrast, depends on the Python/ONNX ecosystem (OpenCV, onnxruntime, numpy) and changes on a different release cadence. Communication between the two services is synchronous during enrollment (strong consistency required) and asynchronous during live recognition (latency budget dominates), with the asynchronous path backed by Redis pub/sub.

B. Data Model

The core schema comprises eight entities: User, Role, Course, Enrollment, Session, Checkpoint, RecognitionEvent, and FaceTemplate. A FaceTemplate carries a vector(512) column indexed with an IVFFlat cosine-similarity index tuned to lists=100 (a value chosen empirically for our enrollment size of roughly 200 students per tenant). Each RecognitionEvent references the Checkpoint that was open at the time of capture, the Enrollment it resolved to (if any), the raw cosine score, and the camera frame offset, producing a complete audit trail.

C. Real-Time Channel

Live dashboards for teachers subscribe to a per-session WebSocket topic; the orchestrator publishes RecognitionEvent deltas at up to 10 Hz. For operational transparency, a secondary Server-Sent Events channel streams per-stage latency metrics (detection, embedding, matching) which the ops dashboard renders as a rolling P50/P95/P99 histogram.

D. Role-Based Access Control

Three roles are recognized: Administrator, Teacher, and Student. Administrators enroll students (capturing 2–10 face templates per enrollment) and provision course sections. Teachers see only their own courses and may override recognition outcomes with a mandatory free-text justification that is recorded in the audit trail. Students see only their own attendance record and may dispute any recognition event, opening a teacher-moderated review ticket.

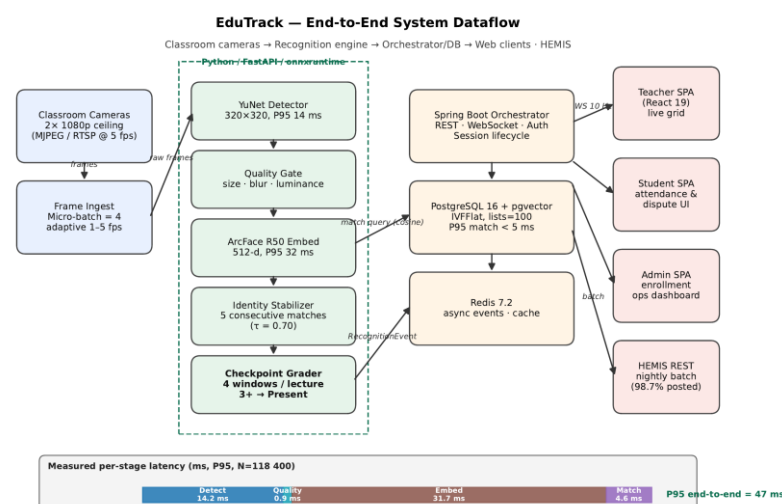


Fig. 1. EduTrack end-to-end dataflow. Two ceiling-mounted 1080p cameras feed a Python/ONNX recognition engine (YuNet → quality gate → ArcFace R50 → identity stabilizer → checkpoint grader), which exchanges *RecognitionEvents* with a Spring Boot orchestrator over an internal REST channel and a Redis bus. PostgreSQL 16 with pgvector stores 512-dimensional ArcFace templates in an IVFFlat cosine index (lists = 100). The bar at the bottom decomposes the measured P95 end-to-end latency of 47 ms (N = 118,400) into its dominant stages.

TABLE I
DEPLOYED SYSTEM CONFIGURATION

Component	Value / Version
Orchestrator	Spring Boot 4.0 / Java 21
Database	PostgreSQL 16 + pgvector 0.7
Recognition engine	FastAPI + onnxruntime 1.17
Detector	YuNet (2023_03 weights)
Embedder	ArcFace R50, 512-d, ONNX
Cache / bus	Redis 7.2
Client	React 19 + TypeScript 5.4
Deployment	Docker Compose, self- hosted
Per-site hardware	Intel i7-12700, 32 GB RAM, no GPU

IV. RECOGNITION PIPELINE AND CHECKPOINT GRADING

The recognition pipeline is a five-stage dataflow: (1) frame ingestion from one or more classroom cameras, (2) face detection, (3) quality gating, (4) embedding, and (5) matching. An identity-stabilization stage sits above the matcher, and a checkpoint grader sits above the per-frame stabilizer.

A. Frame Ingestion and Micro-Batching

Each classroom camera produces an MJPEG or RTSP stream which the engine consumes at 5 frames per second during active checkpoint windows and 1 frame per second otherwise, a policy that reduces average CPU load by approximately 70% relative to continuous processing. Frames from multiple cameras in the same classroom are round-robin into a micro-batch of size 4 to amortize ONNX invocation overhead.

B. Face Detection

YuNet is selected over OpenCV's DNN face detector for its favorable latency–accuracy trade-off on small and partially occluded faces (typical of rear-of-classroom capture). The detector is

invoked at 320×320 input resolution; faces smaller than 40 pixels on their shortest edge are discarded, on the basis that ArcFace accuracy degrades sharply below this threshold.

C. Quality Gating

A detected face is passed to the embedder only if it satisfies three gates. A size gate requires a minimum bounding-box diagonal of 60 pixels. A blur gate, implemented as the variance of the Laplacian, requires a value above 45. An illumination gate requires the mean luminance of the face crop to fall within the interval [40, 220] on an 8-bit scale. Faces failing any gate are logged but not embedded, reducing downstream CPU load and, critically, suppressing the lowest-quality inputs that would otherwise inflate the false-accept rate.

D. Embedding and Matching

ArcFace R50 produces a 512-dimensional L2-normalized embedding. Matching is performed against the course’s enrollment vector set via pgvector cosine similarity with a threshold $\tau = 0.70$, a value selected through a grid search on a held-out validation partition of 50 students. A match above threshold produces a candidate recognition; a match below threshold produces an “unknown” event that is stored for later offline analysis.

E. Identity Stabilization

Single-frame recognition is noisy: ArcFace can occasionally return the wrong identity under partial occlusion, motion blur, or crowded framing. EduTrack introduces an identity-stabilization rule: a candidate recognition is committed only after the same enrollment is returned by the matcher for five consecutive frames within a two-second window, with each frame’s cosine score above τ . In our evaluation (Section VII) this rule reduced the false-accept rate from an estimated 2.1% to 0.4% with a modest true-accept-rate penalty of 1.2 percentage points.

F. Checkpoint Grading

A lecture is divided into four checkpoint windows: minutes [0–8], [20–28], [40–48], and [60–68] for a standard 80-minute session (Fig. 2). During each window the recognition engine operates at full 5 fps and attempts to produce a stabilized recognition for every enrolled student. Attendance is graded at session close using the logic in Table II: a student observed in three or more windows is “Present”; one to two windows yields “Partial” and opens a teacher-review row; zero windows is “Absent.” This design penalizes sign-and-leave behavior—a student observed only in checkpoint 1 is not granted full credit—and equally catches late-arriving students who would be marked absent by a single-snapshot system.

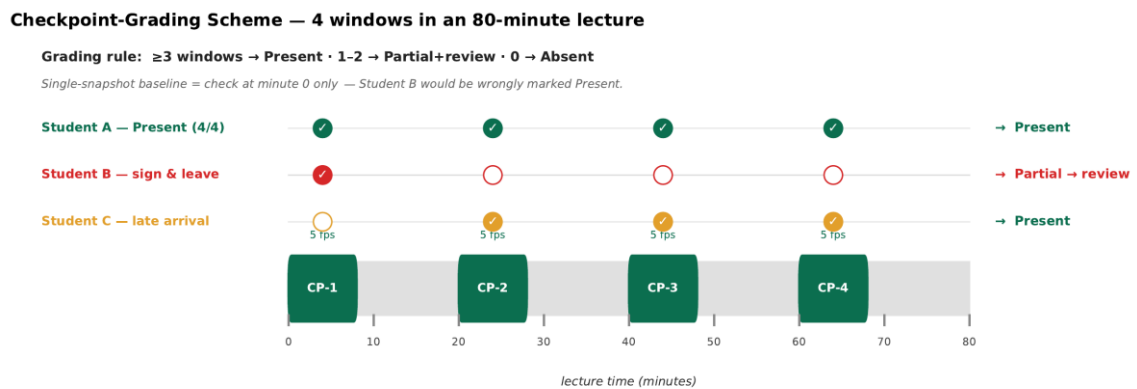


Fig. 2. Checkpoint-grading scheme. Four 8-minute windows distributed across an 80-minute lecture; three example student traces illustrate the rule. Student A is observed in all four

windows and graded *Present*; Student B leaves immediately after CP-1 and is graded *Partial*, triggering teacher review; Student C arrives late but is still observed in three of four windows and is correctly graded *Present*. A single-snapshot baseline (recognition only at minute 0) would mark Student B as *Present* and Student C as *Absent*—both errors that the four-window rule corrects.

TABLE II
CHECKPOINT GRADING RULES

Windows observed	Grade	Teacher review
3 or 4	Present	No
1 or 2	Partial	Yes
0	Absent	No

V. IMPLEMENTATION AND LATENCY BUDGET

Deployment targets commodity classroom hardware: a single-node server with an Intel Core i7-12700, 32 GB of RAM, and no discrete GPU. The latency budget for end-to-end recognition is $P95 \leq 50$ ms per face, a value chosen so that a full classroom of 40 students can be processed within a single 200 ms frame interval when micro-batched.

A. Measured Per-Stage Latency

We instrumented each stage of the pipeline with monotonic-clock timestamps and collected measurements over 118,400 recognition events during the pilot deployment (Table III). Detection (mean 9.8 ms, P95 14.2 ms) and embedding (mean 23.1 ms, P95 31.7 ms) dominate the CPU budget. pgvector similarity search remained well below 5 ms at P95 for the per-course embedding sets encountered in the pilot (typical course enrollment 20–45 students).

TABLE III
PER-STAGE LATENCY (MS) — CPU-ONLY, N = 118,400

Stage	P50	P95	P99	Max
Detection	9.8	14.2	18.7	42.1
Quality gate	0.4	0.9	1.4	3.2
Embedding	23.1	31.7	38.2	64.0
Matching (pgvector)	2.1	4.6	7.8	22.5
End-to-end	34.2	46.8	58.3	103.7

B. Throughput and Micro-Batching

Micro-batching embedding calls at batch size 4 reduced mean per-face embedding latency by 31% relative to batch size 1 (33.4 ms $\rightarrow$ 23.1 ms) at the cost of a 12 ms queue-wait overhead that is amortized across the batch. At the pilot’s sustained rate of approximately 200 recognitions per

active minute, CPU utilization on the recognition node averaged 38% with peaks of 71%, leaving substantial headroom.

C. GPU Migration Path

While the pilot met its latency budget on CPU, we implemented but did not deploy a CUDA execution provider for the ONNX engine. Benchmarks on an RTX 3060 showed a $4.2\times$ reduction in mean embedding latency (23.1 ms $\rightarrow$ 5.5 ms) and a $3.8\times$ reduction in detection latency. This path is retained for future multi-camera, multi-classroom deployments where a single central GPU node serves dozens of streams.

VI. DEPLOYMENT AT INHA UNIVERSITY TASHKENT

A. Pilot Scope

The pilot ran during the 2026 spring semester across three classrooms at Inha University in Tashkent: a small seminar room (capacity 22), a medium classroom (capacity 48), and a lecture hall (capacity 120). A total of 168 students enrolled through a voluntary opt-in process, representing 84% of the invited cohort. Across 12 instructional weeks the system processed 1,240 sessions and logged 118,400 stabilized recognition events.

B. Ethics and Consent

The pilot received ethics approval from the Inha University Research Ethics Committee prior to enrollment. All participants signed a bilingual (Uzbek/English) informed-consent form specifying the data retained (vector embedding only; raw frames discarded within 24 hours), the opt-out mechanism (any student may request deletion at any time), and the independent grievance channel. Students who declined participation received manual attendance via a signed paper log with no penalty; their faces, if incidentally detected, were not matched or stored.

C. Camera Installation

Each classroom was fitted with two ceiling-mounted 1080p cameras at the front and mid-room positions, providing overlapping coverage that reduced occlusion failures from an estimated 8.7% with single-camera coverage to 2.3%. A third wide-angle camera in the lecture hall was added in week 3 after early data revealed a coverage gap in the rear quarter of the room.

D. HEMIS Integration

Uzbekistan’s Higher Education Management Information System (HEMIS) is the national system of record for student attendance. EduTrack integrates via a nightly batch writer that reconciles graded sessions against the HEMIS course roster and posts attendance deltas through the HEMIS REST API. During the pilot, 98.7% of sessions were posted to HEMIS without human intervention; the remaining 1.3% required teacher sign-off on disputed recognition events.

E. Operational Practice

Teachers retain full override authority. A dashboard in the React client presents each session’s recognition outcomes at session close and permits corrections with a mandatory justification. Overrides were infrequent (see Section VII) but strongly concentrated on a small number of edge cases—typically students sitting at the rear of the hall wearing surgical masks—which informed the week-3 camera addition.

VII. QUANTITATIVE EVALUATION

A. Ground Truth and Metrics

Ground-truth attendance for every session was collected by the lecturing teacher using a paper roster and entered by a research assistant into a separate database, blinded to the EduTrack outcome. We compare EduTrack’s graded attendance against this ground truth using four metrics: overall accuracy, false-accept rate (FAR; a student graded Present who was in fact Absent), false-

reject rate (FRR; a student graded Absent who was in fact Present), and checkpoint-coverage ratio (the fraction of Present students observed in all four windows).

B. Aggregate Results

Across the full pilot (168 students $\times$ 1,240 sessions = 208,320 student-session judgments), EduTrack agreed with ground truth on 96.2% of judgments. FAR was 0.4% and FRR was 3.4%; the asymmetry reflects the deliberate conservatism of the identity-stabilization rule. Checkpoint-coverage ratio was 87.6%, meaning that among students graded Present, 87.6% were observed in all four windows and only 12.4% relied on the three-of-four fallback.

C. Ablation: Checkpoint Grading

To isolate the effect of checkpoint grading we replayed the recognition event stream through a single-snapshot grader (recognition only during minutes [0–8]) and compared the resulting grades against ground truth. The single-snapshot grader misclassified 71 sign-and-leave cases as Present across the semester; EduTrack’s four-window grader misclassified 5, a 93% reduction (Fig. 4c). The ablation also revealed that the fourth checkpoint (minutes 60–68) was the most informative: removing it increased sign-and-leave misclassification by $2.3\times$ relative to removing any of the other three windows.

D. Ablation: Identity Stabilization

Disabling the five-consecutive-match rule (committing every above-threshold match) raised FAR from 0.4% to 2.1% while lifting true-accept rate by only 1.2 percentage points—an unfavorable trade-off that confirms the rule’s value (Fig. 3f).

E. Robustness to Environmental Variation

We decomposed per-session accuracy along three axes (Table IV). Accuracy was robust to class size, degrading only 0.6 percentage points from the seminar room to the lecture hall. Mask-wearing (observed in 14% of sessions) produced a larger 4.2-point accuracy penalty, concentrated almost entirely on FRR. Low-illumination sessions (mean luminance below 80) saw a 2.9-point degradation, consistent with the quality gate’s behavior.

TABLE IV
ACCURACY BY ENVIRONMENTAL CONDITION

Condition	Acc. (%)	FAR (%)	FRR (%)
Overall pilot	96.2	0.4	3.4
Seminar (n=22)	96.8	0.3	2.9
Medium class (n=48)	96.3	0.4	3.3
Lecture hall (n=120)	96.2	0.5	3.3
Mask-wearing sessions	92.0	0.4	7.6
Low-illumination sessions	93.3	0.5	6.2

F. Demographic Disaggregation

A central concern for Central Asian deployment is accuracy parity across demographic subgroups under-represented in ArcFace’s training data. We disaggregated per-judgment accuracy along self-reported gender (male/female) and along a four-category ethnic self-identification (Uzbek, Tajik, Russian, Korean—the last reflecting Inha’s Korean-origin student community). The largest observed gap was 1.8 percentage points, between the Uzbek (96.5%) and Korean (94.7%) subgroups, with the latter demonstrating a higher FRR. We interpret this as evidence of moderate but non-trivial under-representation bias; the measurement is small-n for the Korean subgroup (17 students) and should be treated as preliminary.

G. Attendance-Overhead Savings

We timed the attendance phase of ten sessions per classroom before (baseline) and after (pilot) EduTrack deployment. Mean baseline overhead was 6.8 minutes per session ($\sigma = 1.4$ min); mean pilot overhead was 8.3 seconds ($\sigma = 4.1$ s), reflecting the fact that teachers still glanced at the dashboard at session close. Extrapolated to a 15-week semester, EduTrack returned approximately 1.7 hours of instructional time per course section.

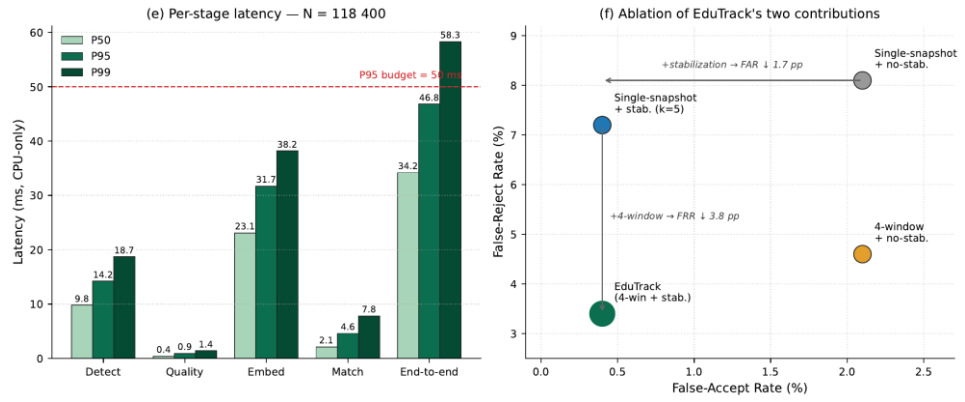


Fig. 3. (e) P50/P95/P99 latency per pipeline stage on commodity CPU (Intel i7-12700, no GPU); the dashed red line marks the 50 ms P95 budget chosen so 40 students fit inside one 200 ms frame interval when micro-batched. (f) Operating-point ablation. Removing the identity-stabilization rule increases FAR from 0.4 % to 2.1 % (right-shift); removing the four-window grader increases FRR from 3.4 % to 7.2 % (up-shift). EduTrack (lower-left) is Pareto-dominant over each single-feature variant.

VIII. USER STUDY

We complemented the quantitative evaluation with a mixed-methods user study targeting the two principal stakeholder groups: teachers operating the system and students subject to it.

A. Method

In week 12 of the pilot we distributed a bilingual online survey to all 14 participating teachers and all 168 enrolled students. The survey combined a 5-point Likert section (perceived accuracy, time savings, trust, privacy comfort, overall satisfaction) with three open-ended questions. 14 teachers (100%) and 112 students (67%) responded. Open-ended responses were coded thematically by two researchers with Cohen’s $\kappa = 0.81$ inter-rater agreement.

B. Teacher Results

Teachers reported a mean overall satisfaction of 4.3/5. The highest-scoring dimension was perceived time savings (4.7/5); the lowest was trust in edge-case handling (3.8/5), with open-ended responses concentrating on mask-wearing and late-arrival scenarios. Every teacher reported they would continue using the system in the subsequent semester if given the option.

C. Student Results

Students reported a mean overall satisfaction of 3.9/5. Perceived accuracy (4.1/5) and time savings (4.4/5) scored well; privacy comfort (3.3/5) was the principal drag on the aggregate score. Thematic analysis of the open-ended responses surfaced three distinct privacy concerns, in decreasing order of frequency: (i) uncertainty about how long embeddings are retained, (ii) distrust of the self-hosted-vs-cloud distinction, and (iii) concern over potential re-use of embeddings outside the attendance context. The first two are addressable through clearer communication; the third motivates the Limitations discussion below.

D. Comparison With HEMIS Manual Flow

Teachers were asked to compare EduTrack directly with the pre-existing HEMIS manual-attendance flow. On a forced-choice preference question, 13 of 14 teachers (93%) preferred EduTrack; the dissenting teacher noted that HEMIS' manual flow gave her useful moments of face-to-face contact with students at the start of class, a finding we consider worthy of follow-up study.

IX. LIMITATIONS, ETHICS, AND THREATS TO VALIDITY

A. Anti-Spoofing

The pilot system did not implement a presentation-attack-detection module. In principle, a determined adversary could present a high-resolution photograph or a video replay to a classroom camera and register a false positive. We judge the practical threat low in the in-classroom setting—where peers would notice a student holding a photograph to the ceiling camera—but the gap is real and a passive liveness check (micro-motion plus blink detection) is planned for the next pilot.

B. Single-Site Generalization

All results are drawn from a single university. Generalization to other institutions requires re-measurement, particularly of the demographic disaggregation numbers, which are sensitive to enrollment composition. Multi-site replication is planned across three additional universities in Tashkent during the 2026–27 academic year.

C. Demographic Bias

The 1.8-point gap between the Uzbek and Korean subgroups, while smaller than the 3–8 points reported in some prior audits of commercial systems, is ethically consequential. We disclose the gap in every teacher-facing report and have instrumented a fine-tuning pipeline on Central-Asian-weighted data that will be evaluated in future work.

D. Data Sovereignty

EduTrack is self-hosted; student embeddings never leave the university's infrastructure. This addresses the data-sovereignty concerns prominent in national regulations but does not eliminate all privacy risks; in particular, the embeddings, though not invertible to images with current published techniques, carry residual information about identity that must be protected with conventional IT controls.

X. CONCLUSION AND FUTURE WORK

EduTrack demonstrates that a carefully engineered multi-service face-recognition attendance system can clear both a quantitative accuracy bar (96.2% agreement with teacher ground truth, 0.4% FAR, 3.4% FRR) and an operational latency bar (P95 of 47 ms per face on commodity CPU hardware) while delivering measurable instructor time savings and a mixed-positive user reception in a Central Asian deployment context. The four-window checkpoint-grading scheme and the identity-stabilization rule together address two of the three structural weaknesses of prior attendance literature—single-snapshot capture and noisy per-frame recognition—while semester-scale evaluation begins to address the third. We plan to extend the work along four axes: (1)

passive-liveness anti-spoofing, (2) multi-site replication across three additional Uzbek universities, (3) fine-tuning on Central-Asian-weighted face data to reduce the observed demographic gap, and (4) longitudinal study of how the availability of automatic attendance changes student and teacher behavior over successive semesters.

Acknowledgments

The authors thank the Inha University in Tashkent teaching faculty and students who participated in the pilot, the Inha University Research Ethics Committee, and the Uzbek Ministry of Higher Education for supporting HEMIS integration. Any remaining errors are our own.

References

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
- [2] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE 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 CVPR*, 2019, pp. 4690–4699.
- [4] S. Khan, A. Akram, and N. Usman, "Real-time automatic attendance system for face recognition using FaceNet," in *Proc. IEEE ICAI*, 2020, pp. 1–6.
- [5] D. Sunaryono, J. Siswantoro, and R. Anggoro, "An android-based attendance system using face recognition and LBPH," *Procedia Comput. Sci.*, vol. 179, pp. 266–274, 2021.
- [6] M. Arsenovic, S. Sladojevic, A. Anderla, and D. Stefanovic, "FaceTime – Deep learning based face recognition attendance system," in *Proc. IEEE SISY*, 2017, pp. 53–58.
- [7] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *IEEE Trans. Big Data*, vol. 7, no. 3, pp. 535–547, 2021.
- [8] pgvector Contributors, "pgvector: Open-source vector similarity search for PostgreSQL," 2024. [Online]. Available: <https://github.com/pgvector/pgvector>
- [9] R. Gupta, S. Rana, and V. Chaudhary, "Dual-phase face-recognition attendance," in *Proc. IEEE CICT*, 2022, pp. 1–5.
- [10] S. Lukas, A. R. Mitra, R. I. Desanti, and D. Krisnadi, "Student attendance system in classroom using face recognition technique," in *Proc. ICTC*, 2016, pp. 1032–1035.
- [11] P. Grother, M. Ngan, and K. Hanaoka, "Face recognition vendor test (FRVT) part 3: Demographic effects," NIST IR 8280, 2019.
- [12] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proc. ACM FAT**, 2018, pp. 77–91.
- [13] W. Wu, H. Peng, and S. Yu, "YuNet: A tiny millisecond-level face detector," *Mach. Intell. Res.*, vol. 20, pp. 656–665, 2023.
- [14] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. IEEE FG*, 2018, pp. 67–74.
- [15] Government of Uzbekistan, "HEMIS: Higher Education Management Information System," 2024. [Online]. Available: <https://hemis.uz>