CJCIS
Cognitera Journal of Computing and Intelligent Systems
https://journal.psg.ac.id/index.php/cjcis

Vol. 1, No. 1, 2026
e-ISSN: XXXX-XXX
pp. 50-61

Sistem Absensi Cerdas Berbasis Face Liveness Detection untuk Mencegah Pemalsuan Identitas (Spoofing) pada Lembaga Pendidikan

Dimas Prasatyo[1]*, Siti Aminah[2]
[1]fakultas Ilmu Komputer / Universitas Indonesia
[2] fakultas Ilmu Komputer / Universitas Indonesia

*Corresponding Author's e-mail: dimasprasatyo@gmail.com

Article History:
Received:September 10, 2026
Revised: October 19, 2026
Accepted: November 25, 2026

Keywords:
Face liveness detection; Smart attendance; Identity spoofing

Abstract: *Educational institutions increasingly adopt facial recognition to automate attendance, yet recognition alone cannot guarantee that a captured face belongs to a physically present person. Printed photographs, replayed videos, screen displays, masks, and synthetic media can exploit this weakness and create fraudulent attendance records. This study proposes a smart attendance architecture that integrates face recognition with face liveness detection as a presentation-attack defense layer. The research applies a design-oriented approach supported by a structured synthesis of verified studies published from 2022 to 2025 on face anti-spoofing, domain generalization, multimodal sensing, lightweight inference, and intelligent attendance systems. The proposed workflow includes enrollment, face detection and alignment, identity embedding, liveness analysis, confidence fusion, attendance authorization, secure logging, and audit reporting. Evaluation is defined through recognition accuracy, attack presentation classification error rate, bona fide presentation classification error rate, average classification error rate, false acceptance, latency, throughput, and cross-domain robustness. Literature-grounded analysis indicates that anti-spoofing accuracy alone is insufficient for educational deployment because camera diversity, illumination, device quality, unfamiliar attack materials, and privacy constraints strongly affect reliability. Transformer-based, attention-based, and domain-generalized methods improve robustness, whereas compact distilled models support real-time operation on edge devices. The proposed architecture therefore separates identity recognition from liveness verification and requires both decisions before attendance is recorded. This layered design provides a defensible basis for reducing proxy attendance while preserving operational efficiency, traceability, and deployment flexibility in schools, colleges, and universities. It supports fair review of uncertain biometric decisions consistently.*

Copyright © 2026, The Author(s).
This is an open access article under the CC–BY-SA license

How to cite:

INTRODUCTION

Attendance records support academic administration, participation monitoring, scholarship eligibility, and institutional accountability, yet conventional roll calls, signature sheets, cards, and QR-based procedures remain vulnerable to proxy attendance and identity substitution. Face recognition offers a contactless alternative because identity can be verified from ordinary cameras without requiring students to carry physical tokens. However, a recognition model normally answers only whether an observed facial pattern resembles an enrolled identity; it does not determine whether the observed face is live. Zainuddin et al. (2024) showed that facial recognition can improve classroom attendance

automation, while Viswanathan et al. (2024) demonstrated practical real-time logging. These benefits motivate automated attendance, but they also create a security requirement: identity matching must be complemented by presentation-attack detection before attendance is accepted.

Presentation attacks exploit the gap between facial similarity and physical presence. An attacker may show a printed photograph, replay a video on a phone, present a high-resolution screen, use a three-dimensional mask, or employ increasingly realistic synthetic content. Such attacks are particularly relevant to educational attendance because the incentive is often not unauthorized physical entry but falsifying evidence of participation. Yu et al. (2023) categorized modern face anti-spoofing methods around binary supervision, pixel-level cues, domain generalization, open-set conditions, and multimodal sensing. Their analysis shows that high within-dataset accuracy can be misleading when deployment cameras, illumination, compression, and spoof materials differ from training data. Therefore, an attendance system requires liveness verification that remains reliable beyond the conditions represented by a single benchmark dataset.

Early anti-spoofing methods emphasized handcrafted texture or image-quality descriptors, whereas recent deep learning models learn spoof cues directly from images or video. Chang and Yeh (2022) showed that multiscale image-quality assessment can reveal perceptual differences between genuine and attack samples, while Arora, Bhatia, and Mittal (2022) demonstrated a deep feature framework across multiple benchmark datasets. These studies establish that presentation attacks leave exploitable artifacts, but their persistence varies with device resolution and media quality. High-quality displays can suppress moiré patterns, and printed attacks can be captured under favorable lighting. Consequently, a robust educational system should not rely on one cue such as blur, texture, or motion. It should combine complementary spatial, temporal, depth-related, or learned attention cues while maintaining acceptable classroom latency.

Recent architectures increasingly address generalization, which is central to institutional deployment. Huang and Wang (2023) used feature distillation and global attention to suppress interference and refine spoof-discriminative representations, whereas Lin et al. (2023) proposed a fast adaptive network designed for domain effectiveness and resource-constrained scenarios. Zhang and Nie (2023) aligned features across multiple domains to reduce distribution mismatch. These contributions indicate that face anti-spoofing should be evaluated on unseen conditions, not merely on random splits from one dataset. A university may use different webcams in laboratories, mobile cameras in field activities, and surveillance cameras in lecture halls. A model that learns device-specific artifacts may fail after hardware replacement. Cross-dataset validation and domain-aware model selection are therefore essential design principles for smart attendance systems.

Transformer and attention mechanisms have expanded the representational capacity of face liveness models. Yu et al. (2024) reconsidered Vision Transformers and masked autoencoders for multimodal anti-spoofing, demonstrating the value of global context and adaptable modality processing. Keresh and Shamoi (2024) used self-supervised Transformer learning and reported strong resistance to different spoofing methods, including a low attack error under their experimental setting. Li et al. (2024) combined cross-stage relations with spoof-material perception, showing that material-sensitive supervision can improve subtle discrimination. These advances are important for educational environments where faces may appear at different scales, with glasses, masks, or partial occlusion. Nevertheless, larger models can increase latency, power consumption,

and maintenance cost, so deployment requires balancing robustness with real-time operational constraints.

Efficiency is especially important because attendance may involve many students entering a classroom within a short interval. Cloud-only processing can introduce network delay and privacy concerns, whereas edge processing reduces data transmission but imposes limits on memory and computation. Zhang et al. (2024) demonstrated a head-aware Transformer distillation strategy with a five-megabyte student model and substantial inference acceleration relative to its teacher. Muhammad, Oussalah, and Laaksonen (2024) also showed that video summarization can reduce redundant frames while preserving anti-spoofing information. These findings suggest that a practical attendance system should process only informative frames, use a lightweight liveness model, and avoid storing raw facial video unnecessarily. Throughput, latency, and memory usage should therefore be treated as core evaluation metrics alongside detection accuracy.

The institutional context introduces additional concerns beyond algorithmic performance. Attendance data are personally identifiable educational records, and biometric templates are difficult to replace if compromised. Darmawan (2023) implemented facial recognition for online-learning attendance, illustrating the growing integration of biometrics into educational workflows, while Bugingo et al. (2025) emphasized continuous face tracking to improve attendance reliability. These developments strengthen administrative automation but also increase the need for purpose limitation, access control, audit trails, retention policies, and transparent fallback procedures. A false liveness rejection can unfairly mark a legitimate student absent, whereas a false acceptance enables spoofing. Therefore, the system must expose confidence and exception handling rather than operating as an opaque binary gate without human review.

The research gap lies in integrating these security, efficiency, and governance requirements into a coherent attendance architecture. Many attendance studies focus on identity recognition accuracy, while anti-spoofing research often evaluates benchmark performance independently from educational workflows. This study bridges both domains by proposing a layered smart attendance system in which identity recognition and face liveness detection are separate but jointly required decisions. The study synthesizes verified literature from 2022 to 2025 and develops an evaluation protocol covering spoof resilience, cross-domain generalization, real-time performance, auditability, and privacy. The contribution is not a fabricated claim of deployment accuracy, because no primary institutional dataset was provided. Instead, it offers a reproducible architecture and evidence-based validation framework that can guide subsequent implementation and controlled field experimentation.

THEORETICAL FOUNDATION

Face recognition and face liveness detection solve related but distinct problems. Recognition estimates whether a detected face belongs to an enrolled identity by comparing learned embeddings, whereas liveness detection or presentation-attack detection determines whether the biometric sample originates from a bona fide live subject rather than an artifact. Modern anti-spoofing systems exploit texture, frequency, reflectance, depth, motion, material, and contextual inconsistencies. Solomon and Cios (2023) combined image-quality features with deep learning, while Sharma and Selwal (2023) reviewed presentation-attack mechanisms and emphasized the diversity of print, replay, and mask attacks. For attendance, these tasks must be chained securely: recognition

identifies the student, liveness establishes physical credibility, and an authorization rule records attendance only when both confidence scores satisfy calibrated thresholds.

Generalization is the theoretical challenge that most strongly separates laboratory performance from deployment reliability. Training and testing samples often share cameras, subjects, backgrounds, and attack media, permitting models to exploit shortcuts that do not represent genuine liveness. Muhammad et al. (2024) addressed domain generalization through ensemble stacking, and Ma et al. (2024) disentangled liveness and spoof patterns to improve adaptation to unseen domains. Li, Yao, Gao, Han, and Xia (2024) jointly modeled local texture and constructed depth using a Swin Transformer, illustrating how complementary cues can improve cross-dataset behavior. These approaches support a system design that treats camera model, illumination, classroom location, and attack medium as explicit domain variables during validation rather than assuming that one training environment represents all educational use cases.

Operational security can be represented as a sequential decision pipeline. A sample is accepted only after quality checking, face detection, identity recognition, and liveness verification succeed; failures are logged rather than silently converted into attendance outcomes. Standard presentation-attack metrics include APCER for attack samples classified as bona fide, BPCER for bona fide samples rejected as attacks, and ACER as their average. Recognition metrics should additionally include false acceptance and false rejection, while system metrics include latency, throughput, memory, availability, and audit completeness. Li, Sun, Li, and Guo (2025) showed that channel attention and intra-class constraints can improve class separation, reinforcing the theoretical value of confidence calibration. In educational settings, thresholds should be selected according to institutional risk and reviewed when devices or environments change.

RESEARCH METHOD

This study employed a design-oriented research strategy supported by structured literature benchmarking. The research stages followed problem identification, requirement specification, architectural design, evidence-based validation, and deployment planning, consistent with the systematic research principles described by Sugiyono (2022). The evidence base was restricted to verifiable publications from 2022 to 2025 addressing face anti-spoofing, presentation-attack detection, face-recognition attendance, domain generalization, multimodal sensing, lightweight inference, and biometric deployment. Priority was given to Scopus-indexed journals, including Q1 and Q2 venues, complemented by Indonesian SINTA sources that directly discuss educational attendance. Each source was checked for bibliographic identity and DOI where available. Because no institutional image dataset was supplied, this research intentionally avoids presenting newly measured accuracy, APCER, or latency values and instead defines a reproducible protocol for subsequent experimental implementation.

The proposed system consists of eight functional stages. First, students are enrolled through consented capture and encrypted storage of face embeddings rather than unrestricted raw-image retention. Second, an input camera performs face detection, quality checking, and geometric alignment. Third, the recognition module calculates identity similarity against enrolled templates. Fourth, a liveness module analyzes spoof-related evidence using a compact attention-based deep network capable of frame or short-sequence processing. Fifth, a fusion policy requires both recognition and liveness confidence to exceed calibrated thresholds. Sixth, accepted events are written to the attendance database with timestamp, device identifier, course context, and decision metadata. Seventh, uncertain or rejected events enter a supervised exception workflow. Finally, administrators

receive auditable reports. Training should use bona fide, print, replay, screen, and mask samples across multiple devices, while splits must separate identities and capture domains to prevent leakage.

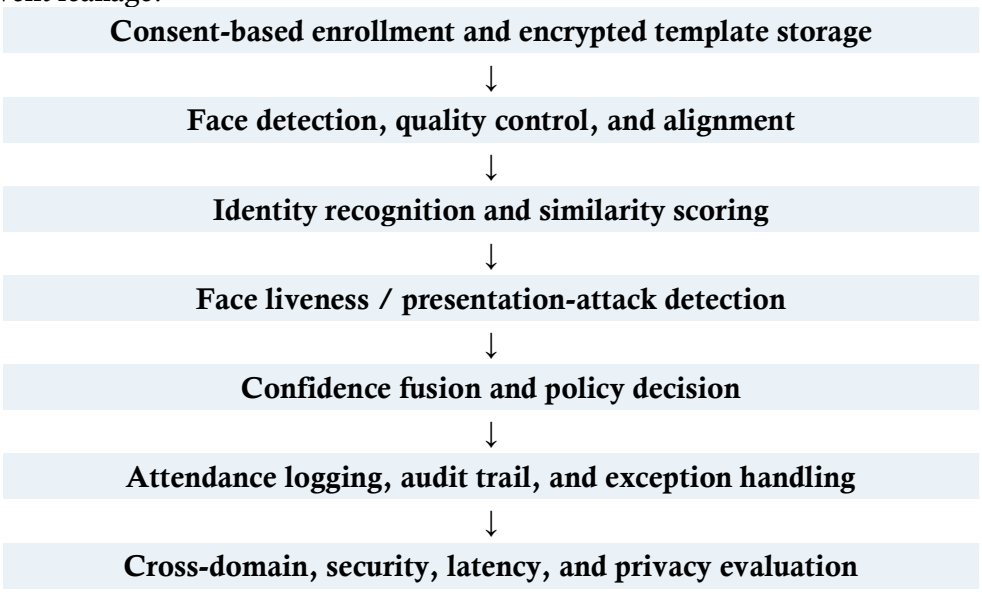

Figure 1. Proposed smart attendance workflow with liveness-gated authorization

Evaluation is designed at three levels: biometric accuracy, anti-spoofing security, and operational performance. Identity recognition is assessed using false acceptance rate, false rejection rate, verification accuracy, and threshold curves. Liveness detection is assessed using APCER, BPCER, ACER, AUC, and cross-dataset error under unseen-camera and unseen-attack conditions. Operational evaluation measures end-to-end latency, frames per second, memory footprint, model size, network dependency, enrollment time, logging integrity, and recovery from low-quality captures. The experimental comparison should include a recognition-only baseline, a CNN liveness baseline, an attention or Transformer variant, and the optimized lightweight model. Ablation tests should isolate frame selection, attention, domain augmentation, and score fusion. Statistical reporting should provide repeated-run means and confidence intervals, while deployment tests should stratify results by illumination, device type, distance, occlusion, and attack medium to identify failure modes.

RESULTS AND DISCUSSION

Interpretive note. The following section is a literature-grounded architecture evaluation and does not claim newly measured institutional accuracy because no primary attendance dataset was supplied.

The first analytical result is that recognition-only attendance cannot be considered identity secure because visual similarity does not establish physical presence. Yu et al. (2023) showed that presentation attacks remain a fundamental vulnerability even when face recognition is otherwise accurate, and their survey distinguishes bona fide verification from anti-spoofing supervision. In an educational workflow, this means the system should never write attendance immediately after a face embedding matches an enrolled student. The proposed architecture therefore introduces liveness as an independent authorization condition. This separation also simplifies auditing because administrators can distinguish recognition failure, liveness failure, camera-quality failure, and policy rejection. Such

decomposition is preferable to one opaque score because each error category requires a distinct operational, retraining, or support response.

The second result concerns proxy attendance through ordinary photographs and screen replays. Chang and Yeh (2022) demonstrated that multiscale image-quality characteristics contain useful cues for distinguishing genuine and attack samples, confirming that spoof media alter spatial statistics even when identity remains recognizable. For the proposed attendance system, these cues justify preprocessing that preserves texture and frequency information instead of aggressively smoothing or compressing captured faces. Nevertheless, image-quality cues alone may weaken when a high-resolution display, glossy print, or favorable lighting minimizes obvious artifacts. The architecture therefore treats quality analysis as complementary evidence rather than a final decision. A practical implementation should augment training with multiple printers, screens, brightness settings, viewing angles, and camera resolutions. This diversity improves resilience against spoofing.

The third result is that attention mechanisms can improve liveness discrimination by directing learning toward spatial regions containing subtle spoof evidence. Huang and Wang (2023) combined feature distillation with global spatial attention and reported improved representation of genuine and fake facial cues. This supports adding lightweight attention after intermediate feature extraction rather than relying solely on deeper convolution. In educational deployment, attention is valuable because spoof artifacts may appear unevenly: screen borders, paper curvature, moiré patterns, reflectance, or abnormal texture can occur around different facial areas. However, attention maps should not be interpreted as guaranteed explanations of causality. They are better used for model diagnostics, failure review, and comparison across domains. Background-focused attention often indicates shortcuts requiring stronger domain diversification.

The fourth result highlights efficiency and domain adaptation as simultaneous requirements. Lin et al. (2023) proposed DEFAEK to improve face anti-spoofing under domain variation while reducing adaptation cost, illustrating that robustness need not always require extremely large models. This is relevant to schools where edge devices may use modest processors and shared networks. A heavy anti-spoofing model can delay queues at classroom entrances, whereas cloud inference may expose biometric data and become unavailable during connectivity failures. The proposed architecture therefore favors local preprocessing and compact inference, with optional server-side model management rather than mandatory streaming. Model selection should compare error rate and latency jointly. A simpler model may be preferable when stable throughput supports calibration and supervised review of uncertainty.

The fifth result concerns domain mismatch between training data and institutional cameras. Zhang and Nie (2023) demonstrated multi-domain feature alignment for face anti-spoofing, emphasizing that distributions vary across acquisition environments. This directly affects educational systems because webcams, smartphones, CCTV units, and laptop cameras differ in sensor response, compression, focal length, and exposure control. Consequently, random image splitting is an inadequate evaluation protocol. The proposed system requires camera-disjoint and location-disjoint testing, where at least one device family or campus environment is excluded from training. Performance should then be reported separately for each unseen domain. If ACER rises substantially on a new device, administrators should not simply lower the liveness threshold, because that can increase attack acceptance. Targeted recalibration is safer operationally.

The sixth result is that combining engineered image evidence with learned representations can improve interpretability and robustness. Solomon and Cios (2023) proposed FASS by integrating image-quality features with deep learning, demonstrating a hybrid alternative to purely end-to-end classification. For smart attendance, a similar principle can support quality gates before liveness inference. Excessive blur, underexposure, severe occlusion, or an extremely small face should trigger recapture rather than force a low-confidence prediction. This policy prevents poor input quality from being mistaken for spoof evidence and reduces unnecessary false rejections. Quality features can also be logged without storing raw images, supporting operational analytics. Over time, administrators can identify cameras that consistently produce weak samples and replace or reposition them. Acquisition management remains essential.

The seventh result is that multimodal and Transformer-based methods strengthen robustness when complementary sensors are available. Yu et al. (2024) showed that Vision Transformers and masked autoencoders can effectively model multimodal face anti-spoofing information. In high-security educational contexts, RGB can be complemented by depth or infrared sensing to distinguish planar displays from real facial geometry. However, multimodal hardware increases cost, calibration effort, and maintenance requirements, making universal deployment unrealistic. The proposed architecture therefore keeps the liveness interface modality-agnostic: standard classrooms can use optimized RGB inference, while examination rooms, laboratories, or restricted facilities can add depth or infrared channels. This modularity avoids redesigning the attendance database or identity subsystem when sensor capability changes. This helps institutions align security investment with risk.

The eighth result concerns cross-domain ensemble strategies. Muhammad et al. (2024) demonstrated that ensemble stacking can improve domain generalization for face presentation-attack detection. An attendance deployment can adopt this insight without necessarily running multiple large models simultaneously. Lightweight diversity can be introduced through complementary heads, temporal and spatial branches, or independently calibrated quality and liveness scores. The objective is to reduce correlated failure, especially when an unseen attack defeats one cue. Nevertheless, ensemble complexity should be controlled because every additional branch affects latency and maintenance. The proposed system recommends measuring marginal security gain per computational cost. If a second branch meaningfully lowers APCER on unseen attacks with minimal latency, it is justified. Otherwise, simpler models should be preferred for maintainability.

The ninth result indicates that material-aware supervision is valuable for realistic presentation attacks. Li et al. (2024) introduced cross-stage relation enhancement and spoof-material perception, demonstrating that the model can learn differences associated with attack media rather than only a binary label. This is especially useful for educational environments, where students may attempt printed paper, glossy photographs, phone displays, tablets, or masks. The proposed training protocol therefore recommends storing attack-type metadata and evaluating each category separately. A single aggregate ACER can hide a critical weakness if replay attacks are easy but high-quality prints are frequently accepted. Material-stratified reporting enables administrators to understand which threat remains dominant and to design targeted data augmentation. It also supports open-set tests beyond known attack media.

The tenth result is that local texture and depth-related representations remain complementary even when only RGB imagery is available. Li, Yao, Gao, Han, and Xia

(2024) jointly modeled local texture with constructed depth using a Swin Transformer and reported promising cross-dataset behavior. This supports an attendance design in which a network predicts depth-consistency cues from ordinary camera frames while retaining fine local artifacts. The benefit is architectural flexibility: low-cost classrooms do not require dedicated depth sensors, yet the model is encouraged to learn three-dimensional plausibility rather than pure identity texture. Such inferred depth should not be treated as physical measurement, because it can fail under masks or unusual lighting. Instead, it functions as auxiliary supervision. Uncertain cases require manual confirmation.

The eleventh result is that lightweight knowledge distillation can make advanced anti-spoofing models operational on constrained hardware. Zhang et al. (2024) reported a head-aware Transformer student with approximately five megabytes of parameters, seventeen times fewer parameters, and substantially faster inference than its teacher under their evaluation. These findings are particularly relevant for attendance kiosks, single-board computers, and older laboratory machines. The proposed architecture therefore separates training complexity from deployment complexity: a powerful teacher can be used offline, while a compact student performs daily liveness inference. Distillation should nevertheless be evaluated on unseen attacks because compression can remove subtle decision cues. Institutions should compare teacher and student APCER, BPCER, and latency on the same cross-domain test sets. Efficiency must preserve security.

The twelfth result is that video-based liveness does not require processing every frame. Muhammad, Oussalah, and Laaksonen (2024) showed that saliency-based video summarization can reduce redundant information while maintaining effective face anti-spoofing performance. For classroom attendance, this is important because continuously analyzing long video streams wastes computation and increases privacy exposure. The proposed system can capture a short burst after a face is detected, rank frames by quality or saliency, and process only a small representative subset. Temporal cues such as micro-motion, screen refresh artifacts, or inconsistent reflections remain available without retaining entire classroom video. This strategy can lower storage and inference demands while improving user experience. Frame selection must resist replay attacks because predictable sampling can create exploitable weaknesses.

The thirteenth result concerns self-supervised Transformer learning. Keresh and Shamoi (2024) fine-tuned a Vision Transformer with a DINO framework and reported strong anti-spoofing performance, including low APCER in their experimental context. Self-supervised pretraining is attractive for educational deployment because institutions may possess large quantities of unlabeled classroom imagery but limited spoof annotations. Pretraining can learn general facial and imaging representations before supervised anti-spoofing adaptation. Nevertheless, privacy governance should prevent indiscriminate reuse of student video for model development. The proposed system therefore recommends explicit dataset governance, de-identification where feasible, purpose-specific consent, and restricted research access. Technically, self-supervised initialization should be compared with conventional ImageNet or face-pretrained backbones under identical cross-domain protocols. Its value lies in transferability, not scores on familiar splits.

The fourteenth result is that explicit intra-class constraints can improve separation between bona fide and spoof distributions. Li, Sun, Li, and Guo (2025) combined enhanced channel attention with an intra-class differentiator and reported low ACER values across several benchmark datasets. For attendance, this suggests that training objectives should not rely exclusively on binary cross-entropy. Metric or margin-based constraints can encourage live samples to form compact representations while maintaining

separation from multiple attack types. Better separation can make threshold calibration more stable when camera conditions vary. However, reported benchmark improvements do not eliminate the need for local validation. Educational populations include different ages, skin tones, accessories, and capture distances, so fairness and subgroup error analysis remain necessary. Local validation remains mandatory.

The fifteenth result connects anti-spoofing research with attendance governance. Zainuddin et al. (2024) showed that classroom facial recognition can reduce manipulation associated with conventional attendance workflows, but recognition-based automation alone does not address presentation attacks. The proposed architecture closes that gap by requiring liveness before a recognized identity is authorized. This design also supports stronger auditability: every attendance event can record the recognition score, liveness score, device, timestamp, and policy outcome without retaining unnecessary raw video. An institution can then review disputes, identify suspicious repeated failures, and monitor device drift. Human override should be logged and role-restricted. A secure attendance system is a layered biometric control combining anti-spoofing defense, calibrated decisions, monitoring, and accountable educational data management through auditable governance.

Table 1. Selected evidence supporting liveness-gated smart attendance design

Study	Technique	Reported evidence	Attendance implication
Yu et al. (2023)	Deep FAS survey	Domain generalization and multimodal PAD remain central challenges	Recognition requires a separate liveness layer
Keresh & Shamoi (2024)	DINO + Vision Transformer	Reported APCER 1.6% under their test setting	Self-supervised Transformers are promising for complex attacks
Zhang et al. (2024)	Transformer knowledge distillation	~5 MB student; 17× fewer parameters; faster inference	Supports edge attendance kiosks and low-latency deployment
Li et al. (2025)	Channel attention + intra-class constraint	Reported low ACER across multiple benchmark datasets	Attention and class separation support robust liveness decisions
Zainuddin et al. (2024)	IoT + facial attendance	Facial identification addresses manipulation in classroom attendance	Liveness closes the remaining presentation-attack gap

CONCLUSION

This study proposes a layered smart attendance architecture in which face recognition establishes identity and face liveness detection establishes the credibility of physical presence. The literature from 2022 to 2025 consistently shows that recognition systems remain vulnerable to print, replay, screen, and mask presentation attacks, while modern anti-spoofing methods increasingly rely on attention, Transformers, depth-related cues, domain generalization, multimodal sensing, and compact deployment models. For educational institutions, these findings imply that attendance should be recorded only when both identity and liveness decisions satisfy calibrated requirements. Recognition-only automation may improve efficiency, but without presentation-attack detection it cannot provide strong protection against proxy attendance or digitally reproduced identities.

The proposed framework also emphasizes that security accuracy must be evaluated together with operational and governance requirements. Cross-device and cross-location testing is essential because strong within-dataset performance does not guarantee reliable behavior on new cameras, classrooms, illumination conditions, or attack materials. Real-time deployment should additionally measure latency, throughput, memory use, false rejection, exception handling, logging integrity, and privacy exposure. Lightweight distillation, selective frame processing, and edge inference can reduce computational cost without forcing continuous cloud transmission. At the institutional level, biometric templates require access control, retention limits, transparent consent, and auditable human override. These requirements transform face liveness from an isolated computer-vision feature into one component of an accountable attendance-management system.

Future work should implement the architecture using a consented educational dataset that includes bona fide students and controlled print, replay, screen, and mask attacks captured across multiple camera types. Evaluation should use identity-disjoint and domain-disjoint splits, standardized APCER, BPCER, ACER, false acceptance, false rejection, and latency metrics, followed by subgroup and fairness analysis. Additional experiments should compare CNN, attention-based, Transformer, and distilled lightweight models under identical conditions and examine whether temporal, depth-estimation, or multimodal cues justify their added cost. A pilot deployment should then measure user acceptance, queue time, dispute frequency, privacy compliance, and administrator workload. Such validation would convert the present literature-grounded architecture into an empirically tested system suitable for responsible use in schools, colleges, and universities.

REFERENCES

1. Arora, S., Bhatia, M. P. S., & Mittal, V. (2022). A robust framework for spoofing detection in faces using deep learning. The Visual Computer, 38, 2461–2472. https://doi.org/10.1007/s00371-021-02123-4
2. Bugingo, E., Imbahafi, O., Umeonyirioha, A. C., Ahmad, T., Jean De La Croix, N., & Uwumuremyi, A. M. (2025). Enhancing face recognition attendance system utilizing real-time face tracking. Internet of Things, 33, 101660. https://doi.org/10.1016/j.iot.2025.101660
3. Chang, H.-H., & Yeh, C.-H. (2022). Face anti-spoofing detection based on multi-scale image quality assessment. Image and Vision Computing, 121, 104428. https://doi.org/10.1016/j.imavis.2022.104428

4. Darmawan, I. D. M. B. A. (2023). Implementation of face recognition for attendance recording in online learning. Jurnal Ilmu Komputer, 16(2). https://doi.org/10.24843/JIK.2023.v16.i02.p06
5. Huang, R., & Wang, X. (2023). Face anti-spoofing using feature distilling and global attention learning. Pattern Recognition, 135, 109147. https://doi.org/10.1016/j.patcog.2022.109147
6. Keresh, A., & Shamoi, P. (2024). Liveness detection in computer vision: Transformer-based self-supervised learning for face anti-spoofing. IEEE Access, 12, 185673–185685. https://doi.org/10.1109/ACCESS.2024.3513795
7. Li, D., Chen, G., Wu, X., Yu, Z., & Tan, M. (2024). Face anti-spoofing with cross-stage relation enhancement and spoof material perception. Neural Networks, 175, 106275. https://doi.org/10.1016/j.neunet.2024.106275
8. Li, L., Yao, Z., Gao, S., Han, H., & Xia, Z. (2024). Face anti-spoofing via jointly modeling local texture and constructed depth. Engineering Applications of Artificial Intelligence, 133, 108345. https://doi.org/10.1016/j.engappai.2024.108345
9. Li, Y., Sun, W., Li, Z., & Guo, X. (2025). Face anti-spoofing based on adaptive channel enhancement and intra-class constraint. Journal of Imaging, 11(4), 116. https://doi.org/10.3390/jimaging11040116
10. Lin, J.-D., Han, Y.-H., Huang, P.-H., Tan, J., Chen, J.-C., Tanveer, M., & Hua, K.-L. (2023). DEFAEK: Domain effective fast adaptive network for face anti-spoofing. Neural Networks, 161, 83–91. https://doi.org/10.1016/j.neunet.2023.01.018
11. Ma, Y., Qian, J., Li, J., & Yang, J. (2024). Dual feature disentanglement for face anti-spoofing. Pattern Recognition, 155, 110656. https://doi.org/10.1016/j.patcog.2024.110656
12. Muhammad, U., Laaksonen, J., Beddiar, D. R., & Oussalah, M. (2024). Domain generalization via ensemble stacking for face presentation attack detection. International Journal of Computer Vision, 132, 5759–5782. https://doi.org/10.1007/s11263-024-02152-1
13. Muhammad, U., Oussalah, M., & Laaksonen, J. (2024). Saliency-based video summarization for face anti-spoofing. Pattern Recognition Letters, 185, 190–196. https://doi.org/10.1016/j.patrec.2024.08.008
14. Sharma, D., & Selwal, A. (2023). A survey on face presentation attack detection mechanisms: Hitherto and future perspectives. Multimedia Systems, 29(3), 1527–1577. https://doi.org/10.1007/s00530-023-01070-5
15. Solomon, E., & Cios, K. J. (2023). FASS: Face anti-spoofing system using image quality features and deep learning. Electronics, 12(10), 2199. https://doi.org/10.3390/electronics12102199
16. Sugiyono. (2022). Metode penelitian kuantitatif, kualitatif, dan R&D (Ed. 2). Alfabeta.
17. Viswanathan, J., Kuralamudhan, E., Navaneethan, S., & Veluchamy, S. (2024). Smart attendance system using face recognition. EAI Endorsed Transactions on Scalable Information Systems, 11(5). https://doi.org/10.4108/eetsis.5203
18. Yu, Z., Cai, R., Cui, Y., Liu, X., Hu, Y., & Kot, A. C. (2024). Rethinking Vision Transformer and masked autoencoder in multimodal face anti-spoofing. International Journal of Computer Vision, 132, 5217–5238. https://doi.org/10.1007/s11263-024-02055-1

19. Yu, Z., Qin, Y., Li, X., Zhao, C., Lei, Z., & Zhao, G. (2023). Deep learning for face anti-spoofing: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5), 5609–5631. https://doi.org/10.1109/TPAMI.2022.3215850
20. Zainuddin, A. A., Nor, R. M., Handayani, D., Tamrin, M. I. M., Subramaniam, K., & Sadikan, S. F. N. (2024). Smart attendance in classroom (CObot): IoT and facial recognition for educational and entrepreneurial impact. Aptisi Transactions on Technopreneurship, 6(3), 608–622. https://doi.org/10.34306/att.v6i3.497
21. Zhang, J., Zhang, Y., Shao, F., Ma, X., Feng, S., Wu, Y., & Zhou, D. (2024). Efficient face anti-spoofing via head-aware transformer based knowledge distillation with 5 MB model parameters. Applied Soft Computing, 166, 112237. https://doi.org/10.1016/j.asoc.2024.112237
22. Zhang, S., & Nie, W. (2023). Multi-domain feature alignment for face anti-spoofing. Sensors, 23(8), 4077. https://doi.org/10.3390/s23084077