Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

LAYERED DETECTION OF TAMPERED IDENTITY DOCUMENTS IN ELECTRONIC KNOW-YOUR-CUSTOMER SYSTEMS: A CONCEPTUAL FRAMEWORK

Cheah Chee Keong[1]
Kong Yin Mei[2]
Tan Kock Lim[3]
Chim Weng Kong[4]
Ong Hock Siong[5]
Dinesh Kumar a/l Saundra Rajan[6]

[1]Universiti Tunku Abdul Rahman, Teh Hong Piow Faculty of Business and Finance Kampar, Perak (Email: cckeong@utar.edu.my)
[2]Universiti Tunku Abdul Rahman, Teh Hong Piow Faculty of Business and Finance Kampar, Perak (Email: kongym@utar.edu.my)
[3] UOW Malaysia KDU Penang University College, Penang (Email: kocklim.tan@kdupg.edu.my)
[4] Universiti Tunku Abdul Rahman, Teh Hong Piow Faculty of Business and Finance Kampar, Perak (Email: chimwk@utar.edu.my)
[5] Universiti Tunku Abdul Rahman, Teh Hong Piow Faculty of Business and Finance Kampar, Perak (Email: onghs@utar.edu.my)
[6] Universiti Tunku Abdul Rahman, Teh Hong Piow Faculty of Business and Finance Kampar, Perak (Email: dineshk@utar.edu.my)
*Corresponding author: Kong Yin Mei (kongym@utar.edu.my)

Article history
Received date : 6-5-2026
Revised date : 6-5-2026
Accepted date : 28-6-2026
Published date : 30-6-2026

To cite this document:
Cheah, C. K., Kong, Y. M., Tan, K. L., Chim, W. K., Ong, H. S., & Saundra Rajan, D. K. (2026). Layered detection of tampered identity documents in electronic know-your-customer systems: A conceptual framework. *Journal of Islamic, Social, Economics and Development (JISED)*, 11 (83), 990-1001.

Abstract: *Identity fraud constitutes one of the most consequential threats confronting digital financial ecosystems, with sophisticated document falsification techniques undermining the foundational trust upon which regulatory compliance and consumer protection depend. This paper examines the effectiveness of electronic Know-Your-Customer (e-KYC) systems in detecting tampered identity documents and proposes an original conceptual framework designed to address persistent detection gaps. The inquiry proceeds through a structured, conceptual and analytical review of extant scholarship across information security, digital identity verification, and financial technology regulation, integrating insights from signal detection theory and risk-based authentication models to construct a theoretically grounded response to an underexplored problem. The proposed framework comprises four interacting layers: document authenticity verification, biometric integrity assessment, contextual risk scoring, and an adaptive decision engine. Three principal propositions are advanced. It is argued, first, that the integration of multimodal forensic analysis within e-KYC pipelines substantially reduces false-negative rates for tampered document detection. Second, that contextual risk signals, including behavioural biometrics, device intelligence, and geolocation anomalies, augment static document verification in a manner that individually neither modality can achieve. Third, that adaptive machine learning architectures capable of continuous*

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

recalibration offer a more durable defence against adversarial manipulation than rule-based heuristic systems. These findings carry significant implications for financial institutions navigating obligations under the Financial Action Task Force recommendations, Basel Committee guidance, and the European General Data Protection Regulation. By synthesising dispersed technical and regulatory literatures into a unified analytical model, this paper contributes to a nascent but critical area of scholarship at the intersection of cybersecurity, fintech governance, and digital trust architecture.

Keywords: *e-KYC, identity document tampering, digital identity verification, biometric authentication, risk-based authentication, anti-money laundering, fintech regulation*

Introduction

The proliferation of digital financial services has fundamentally reconfigured the terrain upon which identity verification operates (Demirgüç-Kunt et al., 2022). Where traditional banking once relied upon the physical presence of customers, notarised documentation, and human judgment to establish identity, the contemporary fintech landscape demands that equivalent, and arguably more rigorous, assurances be achieved remotely, at scale, and in near-real time. This structural transformation carries profound implications not only for consumer experience but for the integrity of the global financial system. Regulatory frameworks, originally conceived to combat money laundering and terrorist financing through in-person processes, have been compelled to evolve, giving rise to electronic Know-Your-Customer (e-KYC) mechanisms as the dominant paradigm for remote identity onboarding (Kubam, 2024).

Traditional KYC procedures, long considered a cornerstone of anti-money laundering (AML) compliance, exhibit well-documented limitations in digital contexts. Manual document inspection is inherently prone to human error; paper-based audit trails are difficult to scale; and face-to-face verification is simply incompatible with asynchronous, app-based financial onboarding (Zhang et al., 2025). These limitations have not gone unnoticed by regulators. The Financial Action Task Force has progressively endorsed technology-assisted customer due diligence (Financial Action Task Force, 2022), while jurisdictions across Asia, Europe, and North America have enacted enabling legislation to facilitate digital identity verification under appropriate safeguards. In this environment, e-KYC has emerged not as a convenience but as a regulatory necessity.

Yet the digital migration of identity verification has simultaneously created new attack surfaces. The very technologies that enable remote onboarding: optical character recognition, liveness detection, facial matching, and automated document classification, can be subverted through increasingly sophisticated means. Identity document tampering, once the preserve of well-resourced criminal networks, has become accessible to a broader spectrum of fraudulent actors through the democratisation of image editing tools and artificial intelligence (Zhang et al., 2025). Forged identity cards, manipulated passports, and synthetically generated documents now routinely challenge the discriminatory capacity of automated systems (Moulianitis et al., 2025), generating false negatives that represent systemic risk to financial institutions and their regulators alike.

Despite the urgency of this threat, the academic literature reveals a discernible gap that operates at three distinct levels. At the operational level, technical research on specific detection modalities: metadata analysis, ultraviolet pattern recognition, machine learning classifiers,

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

exists in relative abundance, yet these modalities are rarely integrated into a coherent detection pipeline that practitioners can deploy end-to-end; institutions consequently operate with disaggregated toolkits rather than unified systems (Kang et al., 2024). At the regulatory compliance level, a related but distinct gap persists: the relationship between how an e-KYC system is architected and whether that architecture satisfies supervisory expectations under frameworks such as FATF guidance and Basel operational resilience principles has received comparatively little scholarly attention, leaving compliance officers without theoretically grounded criteria for evaluating vendor claims or justifying design choices to regulators. At the conceptual/theoretical level, the literature lacks a unified framework explaining why and how disaggregated technical modalities should be combined; existing work tends to evaluate individual detection techniques in isolation rather than theorising their interaction, sequencing, or evidential complementarity. These three gaps are related but not reducible to one another: closing the operational gap through better tooling does not by itself close the regulatory gap, and closing the regulatory gap through compliance checklists does not by itself supply the conceptual architecture needed to reason about how detection layers should combine. It is this triad of gaps, taken together, that impairs both the effectiveness of detection and the coherence of regulatory compliance strategies.

This paper seeks to address that gap. We pursue three interconnected objectives: to synthesise the existing literature on e-KYC technologies and document tampering detection; to propose a novel four-layer conceptual framework for tampered identity detection within digital onboarding systems; and to derive theoretical and practical propositions that may guide both future empirical research and institutional deployment decisions. The analysis proceeds as follows. Section 2 outlines the review methodology before surveying relevant scholarship across four thematic areas. Section 3 presents the proposed conceptual framework and its theoretical foundations. Section 4 discusses implications, limitations, and future research directions, while Section 5 offers a synthesising conclusion.

Literature Review

This review adopts a structured search protocol rather than an ad hoc narrative approach. Relevant literature was identified through systematic searches of Scopus, Web of Science, Google Scholar, and arXiv, conducted over the period 2015–2026, using combinations of the following search terms spanning the technical and regulatory dimensions of the problem: "electronic KYC," "identity document tampering," "deepfake detection," "risk-based authentication," and "anti-money laundering compliance technology." Sources were prioritised using a combination of recency and direct relevance to the four layers of the proposed framework, with preference given to peer-reviewed publications; recent preprints were included selectively where they represented the most current evidence available on fast-moving technical subfields such as deepfake generation and detection, notwithstanding their pre-publication status.

From KYC to e-KYC: Regulatory Drivers and Digital Transformation

The evolution from paper-based KYC to fully digitised onboarding processes is best understood as a regulatory response to commercial pressure rather than a technologically driven disruption per se. Financial institutions operating across multiple jurisdictions faced mounting compliance costs under frameworks such as the European Fourth and Fifth Anti-Money Laundering Directives (European Parliament and Council, 2018, Financial Action Task Force, 2022) and analogous instruments in Southeast Asia and North America. The FATF Recommendation 10, which mandates customer due diligence as a prerequisite for account establishment (Financial

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

Action Task Force, 2023), provided the normative architecture within which digital alternatives were gradually accommodated. What emerged was not a uniform global standard but a mosaic of nationally inflected e-KYC regimes, each calibrating the permissible scope of remote verification against domestic risk appetites. This regulatory fragmentation, while creating complexity for multinational institutions, paradoxically spurred technical innovation as vendors competed to develop solutions compatible with multiple compliance environments simultaneously.

Core Technologies Underpinning e-KYC

The technical substrate of e-KYC draws upon a convergent set of capabilities that, individually, predated the fintech boom but whose integration represents a meaningful qualitative shift. Optical character recognition enables the automated extraction of data fields from identity documents, reducing both processing time and transcription error (Ingle et al., 2020). Biometric authentication, most commonly facial recognition augmented by liveness detection, addresses the fundamental verification challenge: establishing that the person presenting a document is its legitimate holder. Liveness detection, in particular, has attracted substantial research attention given its susceptibility to presentation attacks using printed photographs or video replays (Sabaghi et al., 2021; Kawana et al., 2024). Neural network architectures, including convolutional models trained on large document corpora, have substantially improved classification accuracy across document types and issuing jurisdictions (Harley et al., 2015; Omar et al., 2024; Petmezas et al., 2025). Machine learning has also enabled the development of anomaly detection systems capable of identifying statistical deviations from baseline document templates, a capability with direct relevance to forgery detection (Kubam, 2024). Nevertheless, the empirical performance of these systems under adversarial conditions, where attackers possess knowledge of the detection algorithms, remains an area requiring further investigation. Recent advances have begun to address this gap: Omar et al. (2024) demonstrated that a bagging ensemble of CoAtNet classifiers, combining depth-wise convolution and self-attention, improves generalisation across multiple deepfake generation methods by capturing both local features and long-range spatial dependencies. Petmezas et al. (2025) further showed that integrating three-dimensional morphable models with a hybrid CNN-LSTM-Transformer architecture substantially improves detection accuracy for identity verification tasks, with the Transformer component capturing temporal inconsistencies that purely spatial models overlook.

These performance gains, however, warrant a critical qualification rather than unqualified endorsement. The generalisation improvements reported by Omar et al. (2024) and Petmezas et al. (2025) are measured against specific benchmark datasets and known deepfake generation methods; neither study establishes performance under a genuinely adaptive adversary who has observed and adjusted to the detection model's decision boundary, which is the threat scenario most relevant to a deployed e-KYC pipeline. More broadly, the technical literature reviewed above evaluates individual modalities largely in isolation, using benchmark datasets that may not reflect the demographic, document-format, or image-quality diversity encountered in live onboarding traffic; reported accuracy figures should therefore be read as upper bounds on real-world performance rather than as direct estimates of it.

Typologies of Identity Document Tampering

Document tampering encompasses a heterogeneous range of manipulation techniques that resist reduction to a single threat model. At one end of the spectrum lie relatively crude physical alterations: laminate peeling, photograph substitution, and ink overwriting, that remain

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

detectable through optical and tactile examination. More technically sophisticated attacks exploit digital editing workflows to manipulate scanned document images before submission to e-KYC pipelines. These include pixel-level manipulations that leave no visible artefacts, metadata injection to simulate authentic document provenance (Rafique et al., 2023; Rawat et al., 2026), and the generation of entirely synthetic identity documents using generative adversarial networks (Carrere, 2025). The last category is particularly significant: deepfake identity documents, produced at a fidelity that challenges even experienced human reviewers (Moulianitis et al., 2025; Felouat et al., 2024; Kawana et al., 2024), represent a qualitative escalation in the threat environment that many existing e-KYC systems are poorly equipped to address. Empirical evidence supports this concern: Felouat et al. (2024) documented that eKYC systems lacking deepfake detection capability can be readily compromised, and assembled a corpus of over 228,000 synthetic videos specifically designed to reflect the motion and facial verification protocols required in real-world eKYC deployment. Similarly, Kawana et al. (2024) evaluated publicly available deepfake detection countermeasures against both FaceSwap and Reenactment attack types, finding that the two categories of manipulation present distinct and complementary detection challenges for eKYC pipelines. Detection challenges are compounded by the diversity of document formats across issuing authorities, and the structural heterogeneity of identity documents across even a single national system means that training robust classifiers requires document corpora of considerable breadth and representativeness (Kang et al., 2024).

Gaps in the Existing Research Landscape

A careful survey of the literature reveals several persistent lacunae. First, the preponderance of published work addresses discrete technical problems: improving liveness detection recall (Sabaghi et al., 2021), reducing OCR error rates under suboptimal lighting (Ingle et al., 2020), rather than the systemic question of how multiple verification modalities should be integrated and sequenced within a coherent detection architecture. Second, the relationship between e-KYC system design and regulatory compliance outcomes has received comparatively little scholarly attention, leaving practitioners without theoretically grounded guidance for satisfying supervisory expectations. Third, research on adversarial attacks against e-KYC pipelines, while growing (Yuan et al., 2019), has yet to converge on a shared threat taxonomy, making cross-study comparison difficult. Finally, questions of jurisdictional variability: how detection frameworks should be calibrated to account for different national document standards, regulatory regimes, and fraud epidemiology — are almost entirely absent from the technical literature. These gaps collectively motivate the conceptual synthesis attempted in the present paper.

Two further observations sharpen this account of the literature. First, the evidence base reviewed above is uneven in maturity: foundational contributions such as signal detection theory (Green & Swets, 1966), convolutional document classification (Harley et al., 2015), and risk-based authentication (Wiefling et al., 2020) are well-established in the peer-reviewed literature, whereas several of the findings most directly relevant to deepfake identity document detection (Kang et al., 2024; Carrere, 2025; Moulianitis et al., 2025; Kubam, 2024; Rawat et al., 2026) remain preprints that have not yet undergone formal peer review. This paper treats the two evidence bases accordingly: the peer-reviewed strand is used to ground the framework's theoretical architecture, while the preprint strand is treated as indicative of the current state of a fast-moving field and flagged, where relevant, as awaiting independent replication. Second, the operational and regulatory-compliance gaps identified above are, in part, a symptom of a literature that has developed the technical and regulatory strands largely in parallel: technical

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

papers evaluate detection accuracy against machine-learning benchmarks without reference to specific supervisory expectations, while regulatory and compliance literature (e.g., Financial Action Task Force, 2022, 2023; Basel Committee on Banking Supervision, 2021) specifies outcomes required of institutions without engaging the technical means by which those outcomes are achieved. The conceptual synthesis attempted in this paper is designed explicitly to bridge that separation, treating regulatory requirements not as an external constraint layered on top of a technical system but as a design input shaping the architecture from the outset.

Conceptual Framework

The framework proposed here draws on two theoretical traditions that, taken in isolation, address only partial dimensions of the detection problem. Signal Detection Theory, originally developed to characterise the performance of sensory discrimination tasks (Green & Swets, 1966), provides a rigorous vocabulary for understanding the trade-offs inherent in binary classification decisions: as detection sensitivity increases, so too does the rate of false positives, and any operational system must navigate this trade-off in light of the asymmetric costs of different error types (Green & Swets, 1966). Risk-Based Authentication, developed within the information security literature (Wiefling et al., 2020), offers a complementary perspective by emphasising that verification decisions should be calibrated not only to the characteristics of the presented identity evidence but to the broader contextual signals that indicate the probability of fraudulent intent. Integrating these perspectives yields a framework that is simultaneously attentive to the intrinsic properties of identity documents and to the situational context within which they are presented.

Layer 1: Document Authenticity Verification

The first layer constitutes the forensic foundation of the framework and is concerned exclusively with the intrinsic properties of the presented identity document. Verification processes at this layer include metadata consistency analysis, examining whether embedded file properties conform to the expected provenance of the claimed issuing authority, alongside pixel-level integrity checking using error level analysis and noise residue analysis (Rafique et al., 2023). Rafique et al. (2023) demonstrate that error level analysis, when combined with deep learning classifiers, achieves strong discrimination between authentic and manipulated images by exploiting the characteristic compression artefact signatures introduced during digital editing. These are, techniques that identify regions of an image that have been digitally manipulated. Template matching against certified reference images, where available, provides a further discriminatory signal (Kang et al., 2024). The layer also incorporates spectral analysis capabilities, relevant where physical documents are captured under controlled lighting conditions. Crucially, the output of Layer 1 is not a binary accept/reject decision but a calibrated authenticity confidence score, expressed as a probability distribution over the space of possible document states (Green & Swets, 1966). This probabilistic output is then forwarded to subsequent layers, preserving uncertainty information that would be discarded by a hard threshold classifier.

Layer 2: Biometric Integrity Assessment

The second layer addresses the verification of the correspondence between the document holder and the identity document presented. Facial recognition comparing the live capture against the document photograph constitutes the core verification act at this layer, but the framework conceives of this comparison as insufficient in isolation. Passive liveness detection, inferring three-dimensional facial geometry from monocular video sequences, guards against photograph and video replay attacks (Sabaghi et al., 2021; Kawana et al., 2024; Petmezas et al., 2025).

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

Behavioural biometric signals, including typing cadence, device orientation patterns, and scrolling velocity captured during the onboarding session (Kasprowski et al., 2022), contribute an additional verification dimension that is exceptionally difficult to spoof consistently across an entire interaction. The integration of voice biometrics, where regulatory context permits audio capture, further enriches the signal set. Again, the output of this layer is a probabilistic biometric integrity score rather than a deterministic verdict, ensuring that marginal cases are flagged for additional scrutiny rather than summarily accepted or rejected.

Layer 3: Contextual Risk Scoring

It is at the third layer that the framework most clearly operationalises the risk-based authentication paradigm (Wiefling et al., 2020). Contextual risk scoring integrates signals exogenous to the document itself: device fingerprint analysis, IP geolocation consistency, velocity checks against prior authentication events, and network reputation scoring, into a composite risk index. The intuition here is that a document that passes forensic examination with high confidence is nonetheless deserving of additional scrutiny if it is being presented from a device and network profile associated with known fraudulent activity, or if the geolocation of the presenting device is inconsistent with the claimed residential address on the document. Conversely, an application exhibiting marginal document scores may be appropriately approved if the contextual risk profile is strongly consistent with legitimate behaviour. This layer thus introduces a dynamic and relational dimension to verification that purely document-centric approaches foreclose. Formally, the risk score at this layer is computed as a weighted function of individual contextual signals, with weights calibrated empirically and updated as new fraud patterns emerge.

Layer 4: Adaptive Decision Engine

The fourth layer synthesises the probabilistic outputs of the preceding three layers into an actionable verification decision. Rather than applying fixed thresholds, the adaptive decision engine employs a machine learning model, specifically an ensemble architecture combining gradient-boosted trees and neural network components (Xu et al., 2023; Omar et al., 2024; Petmezas et al., 2025). The selection of an ensemble architecture at this layer is informed by empirical evidence that ensemble methods outperform single-model classifiers across diverse deepfake generation techniques (Omar et al., 2024), and that temporal modelling via recurrent and transformer components substantially improves detection fidelity in video-based identity verification scenarios (Petmezas et al., 2025), to learn the optimal decision boundary from historical verification outcomes. The adaptive character of this layer is its most consequential feature: as adversarial manipulation techniques evolve, the decision engine can be retrained on new labelled examples without requiring structural modifications to the layers below. The engine produces three possible outputs: approve, reject, and escalate for human review, with the last category explicitly acknowledging that some cases cannot be resolved algorithmically with sufficient confidence to satisfy regulatory expectations. Proposition P1 holds that this layered architecture, through its accumulation of complementary evidential streams, achieves materially lower false-negative rates than single-modality verification systems. Proposition P2 contends that the contextual risk scoring layer, by conditioning decisions on behavioural and environmental signals, reduces the susceptibility of the system to spoofing attacks that successfully defeat document-level forensic checks (Wiefling et al., 2020). Proposition P3 asserts that the adaptive decision engine sustains detection performance over time by learning from emerging fraud typologies (Xu et al., 2023), in contrast to static rule-based systems whose discriminatory capacity degrades as adversarial techniques evolve to exploit known decision boundaries (Yuan et al., 2019).

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

Discussion

Theoretical Implications

The framework proposed in this paper extends existing models of e-KYC verification in several theoretically significant ways. Most contemporary treatments of digital identity verification conceive of the verification task as a succession of independent checks: document scan, biometric match, liveness test, each producing a binary verdict that cascades toward a final decision. This sequential, modular architecture is epistemically impoverished: it discards the probabilistic information carried by marginal results at each stage and fails to leverage the evidential complementarity that exists between different verification modalities (Kubam, 2024). By contrast, the four-layer architecture advanced here operationalises a coherent probabilistic inference chain, preserving and accumulating uncertainty across layers in a manner consistent with Bayesian principles. The incorporation of Signal Detection Theory (Green & Swets, 1966), moreover, makes explicit the cost asymmetry between false positives and false negatives in the e-KYC context — a distinction that has operational significance given that the regulatory consequences of approving a fraudulent identity typically far exceed the commercial cost of incorrectly rejecting a legitimate applicant. The framework thus bridges a gap between formal statistical theory and practical system design that the existing literature has largely left unaddressed.

Practical Implications

For financial institutions, the practical implications of the framework are considerable. Deployment of a four-layer architecture requires non-trivial investment in both technical infrastructure and data governance, and this paper does not minimise the implementation challenges involved. The document reference template databases required for Layer 1 demand ongoing maintenance as issuing authorities update document designs; biometric models at Layer 2 require training data that reflects the demographic diversity of the customer population being served (Kang et al., 2024); and the adaptive decision engine at Layer 4 requires a labelled historical dataset of sufficient volume and quality to avoid spurious model fits. Notwithstanding these demands, the framework aligns closely with supervisory expectations expressed across several major regulatory instruments. The FATF Guidance on Digital Identity acknowledges the legitimacy of technology-assisted CDD while emphasising the importance of assurance frameworks that assess reliability and independence of identity evidence (Financial Action Task Force, 2022): a requirement the multi-layer architecture directly addresses. The Basel Committee on Banking Supervision has similarly underscored the need for robust customer onboarding controls within operational risk frameworks (Basel Committee on Banking Supervision, 2021). The GDPR introduces a further set of constraints, particularly regarding the processing of biometric data (Abomhara & Yayilgan, 2021; Rawat et al., 2026). Rawat et al. (2026) propose one practical approach to this challenge: their federated micro-expression and multi-modal metadata fusion framework processes biometric data on-device within a federated learning architecture, achieving high detection accuracy while preserving personally identifiable information, and financial institutions implementing the framework must ensure that data minimisation and purpose limitation principles are observed at each layer, with privacy-by-design principles embedded in system architecture rather than appended as compliance afterthoughts.

Limitations and Critical Reflections

This paper does not advance empirically validated claims, and its propositions remain at the level of theoretically grounded hypotheses awaiting experimental substantiation. The absence of real-world performance data is a limitation that honesty requires to be stated plainly. Beyond

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

this, three categories of limitation merit specific attention. First, the framework as presented assumes an adversary with limited knowledge of the detection system; in practice, sophisticated fraud networks engage in systematic probing of e-KYC pipelines to identify exploitable weaknesses, and adversarial machine learning techniques capable of generating input perturbations specifically designed to evade neural network classifiers (Yuan et al., 2019) represent a credible threat that the adaptive decision engine alone cannot fully neutralise. Second, data privacy concerns, already noted in relation to GDPR, are particularly acute for the biometric and behavioural data processed across Layers 2 and 3; the collection of such data at scale creates concentrated repositories whose breach or misuse could cause substantial harm to data subjects (Abomhara & Yayilgan, 2021). Third, jurisdictional variability in document standards, regulatory requirements, and fraud typologies means that a framework calibrated in one national context may perform poorly when transplanted without adaptation; the claim to universality implicit in a conceptual framework must be tempered by recognition of the structural heterogeneity of the global identity landscape.

Future Research Directions

Several directions for future inquiry present themselves naturally from the foregoing analysis. Most urgently, empirical testing of the framework across representative e-KYC deployment environments is required to move the propositions advanced here from theoretical conjecture to evidential standing. Such testing should incorporate adversarial stress-testing methodologies: red-team exercises using state-of-the-art document falsification and deepfake generation tools (Carrere, 2025; Felouat et al., 2024; Kawana et al., 2024), to probe the framework's resilience under realistic attack conditions. Cross-border harmonisation of identity document standards represents another underexplored frontier: to the extent that the detection architecture at Layer 1 depends on reference template databases, the development of international standards for cryptographically authenticated document templates would substantially improve detection reliability in cross-border digital onboarding contexts. Finally, the integration of deepfake video detection into the liveness assessment component of Layer 2 warrants dedicated investigation (Sabaghi et al., 2021; Petmezas et al., 2025), given the rapidly improving fidelity of generative video models and their potential application to real-time biometric spoofing attacks.

Conclusion

This paper has pursued a specific and, it is argued, consequential contribution: the development of a theoretically grounded, four-layer conceptual framework for detecting tampered identity documents within electronic Know-Your-Customer systems. The framework was constructed at the intersection of signal detection theory (Green & Swets, 1966) and risk-based authentication (Wiefling et al., 2020), integrating document forensics, biometric assessment, contextual risk intelligence, and adaptive machine learning into a unified evidential architecture. Three propositions were advanced in support of the framework's claims: that multimodal integration reduces false-negative rates relative to single-modality systems; that contextual risk signals augment document-level verification in ways that reduce susceptibility to sophisticated spoofing; and that adaptive learning mechanisms sustain detection efficacy over time in the face of evolving adversarial techniques (Xu et al., 2023).

The broader significance of this inquiry extends beyond the technical domain. Digital identity verification is rapidly becoming the primary mechanism through which financial inclusion and regulatory compliance are jointly pursued in economies undergoing digital transformation (Demirgüç-Kunt et al., 2022). Failures in e-KYC systems are not merely operational embarrassments; they represent entry points for money laundering, terrorist financing, and large-scale consumer fraud that impose real costs on both financial systems and the individuals

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

they serve (Zhang et al., 2025). The development of theoretically coherent detection architectures is therefore not an abstract academic exercise but a contribution to the infrastructure of digital trust upon which modern financial markets increasingly depend.

We recognise that the framework presented here is necessarily provisional. Its propositions await empirical validation; its performance characteristics under adversarial conditions remain untested at scale; and its applicability across diverse jurisdictional contexts is subject to the caveats articulated in the discussion. These are not grounds for diminishing the contribution but for contextualising it appropriately: as a starting point for a programme of systematic empirical and normative inquiry that the field both needs and, judging by recent regulatory attention to digital identity risk, is beginning to mobilise resources to conduct.

The trajectory of identity fraud is, arguably, one of progressive technical sophistication (Zhang et al., 2025). As generative artificial intelligence continues to lower the cost and raise the quality of synthetic document production (Carrere, 2025), the adequacy of today's detection systems will be continuously tested. Meeting that challenge requires not only better algorithms but better conceptual frameworks within which algorithms are deployed, evaluated, and iteratively improved. It is to that longer-term project of framework development that this paper offers its contribution.

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

References

Abomhara, M., & Yayilgan, S. Y. (2021). An example of privacy and data protection best practices for biometrics data processing in border control: Lesson learned from SMILE. In S. Y. Yildirim Yayilgan, I. S. Bajwa, & F. Sanfilippo (Eds.), *Intelligent Technologies and Applications. INTAP 2020.* Springer. https://doi.org/10.1007/978-3-030-71711-7_29

Basel Committee on Banking Supervision. (2021). *Principles for operational resilience.* Bank for International Settlements. https://www.bis.org/bcbs/publ/d516.htm

Carrere, B. (2025). Can generative models actually forge realistic identity documents*?* arXiv. https://arxiv.org/abs/2601.00829

Demirgüç-Kunt, A., Klapper, L., Singer, D., & Ansar, S. (2022). *The Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19.* World Bank. https://doi.org/10.1596/978-1-4648-1897-4

European Parliament and Council. (2018). Directive (EU) 2018/843 of the European Parliament and of the Council (5th Anti-Money Laundering Directive). *Official Journal of the European Union.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L0843

Felouat, H., Nguyen, H. H., Le, T. N., Yamagishi, J., & Echizen, I. (2024). eKYC-DF: A large-scale deepfake dataset for developing and evaluating eKYC systems. *IEEE Access, 12*, 30876–30892. https://doi.org/10.1109/ACCESS.2024.3369187

Financial Action Task Force. (2022). *Guidance on digital identity.* FATF-GAFI. https://www.fatf-gafi.org/content/dam/fatf-gafi/guidance/Guidance-on-Digital-Identity.pdf

Financial Action Task Force. (2023). *International standards on combating money laundering and the financing of terrorism & proliferation: The FATF Recommendations (updated October 2023).* FATF-GAFI. https://www.fatf-gafi.org/en/topics/fatf-recommendations.html

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* Wiley.

Harley, A. W., Ufkes, A., & Derpanis, K. G. (2015). Evaluation of deep convolutional nets for document image classification and retrieval. *Proceedings of the 13th IAPR International Conference on Document Analysis and Recognition (ICDAR).* https://arxiv.org/abs/1502.07058

Ingle, N., Rajput, U., Bhatt, R., Pimpale, N., Phapale, A., & Shelke, D. (2020). A survey of deep learning approaches for OCR and document understanding. *arXiv.* https://arxiv.org/abs/2011.13534

Kang, J., Doyle, B., Kim, T., & Boult, T. E. (2024). *IDNet: A novel dataset for identity document analysis and fraud detection.* arXiv. https://arxiv.org/abs/2408.01690

Kasprowski, P., Borowska, Z., & Harezlak, K. (2022). Biometric identification based on keystroke dynamics. *Sensors, 22*(9), 3158. https://doi.org/10.3390/s22093158

Kawana, N., Ooshima, K., Suzuki, A., & Yoshino, M. (2024). Detection methods of spoofing attacks for eKYC using deepfake. In *Proceedings of the 38th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI2024)*. JSAI. https://confit.atlas.jp/guide/event/jsai2024/subject/4M3-GS-10-03/detail

Kubam, C. S. (2024). Agentic AI microservice framework for deepfake and document fraud detection in KYC pipelines. *Journal of Intelligent Systems and Emerging Methodologies, 9*(1). https://arxiv.org/abs/2601.06241

Moulianitis, V., Ioannidis, D., Kostavelis, I., & Tzovaras, D. (2025). TwoHead-SwinFPN: A unified DL architecture for synthetic manipulation, detection and localization in identity documents. *arXiv.* https://arxiv.org/abs/2601.12895

Volume: 11 Issues: 83 [June, 2026] pp. 990 - 1001
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: academicinspired.com/jised
DOI: 10.55573/JISED.118381

Omar, K., Sakr, R. H., & Alrahmawy, M. F. (2024). An ensemble of CNNs with self-attention mechanism for DeepFake video detection. *Neural Computing and Applications, 36*(6), 2749–2765. https://doi.org/10.1007/s00521-023-09196-3

Petmezas, G., Vanian, V., Konstantoudakis, K., Almaloglou, E. E. I., & Zarpalas, D. (2025). Video deepfake detection using a hybrid CNN-LSTM-Transformer model for identity verification. *Multimedia Tools and Applications, 84*, 40617–40636. https://doi.org/10.1007/s11042-024-20548-6

Rafique, R., Gantassi, R., Amin, R., Frnda, J., Mustapha, A., & Alshehri, A. H. (2023). Deep fake detection and classification using error-level analysis and deep learning. *Scientific Reports, 13*(1), 7422. https://doi.org/10.1038/s41598-023-34629-3

Rawat, R., Dibouliya, A., Chaturvedi, N., Rawat, A., & Dangi, C. S. (2026). A data analytics based federated biometrics for deepfake fraud detection in financial multimedia systems. *Discover Computing, 29*, 215. https://doi.org/10.1007/s10791-026-10073-5

Sabaghi, A., Oghbaie, M., Hashemifard, K., & Akbari, M. (2021). *Deep learning meets liveness detection: Recent advancements and challenges.* arXiv. https://arxiv.org/abs/2112.14796

Wiefling, S., Dürmuth, M., & Lo Iacono, L. (2020). More than just good passwords? A study on usability and security perceptions of risk-based authentication. In *Proceedings of the Annual Computer Security Applications Conference (ACSAC 2020)* (pp. 203–218). ACM. https://doi.org/10.1145/3427228.3427243

Xu, K., Shi, Z., Chen, J., & Zhang, W. (2023). *Efficient fraud detection using deep boosting decision trees.* arXiv. https://arxiv.org/abs/2302.05918

Yuan, X., He, P., Zhu, Q., & Li, X. (2019). Adversarial examples: Attacks and defenses for deep learning. *IEEE Transactions on Neural Networks and Learning Systems, 30*(9), 2805–2824. https://doi.org/10.1109/TNNLS.2018.2886017

Zhang, C. J., Gill, A. Q., Liu, B., & Anwar, M. (2025). AI-based identity fraud detection: A systematic review. arXiv. https://arxiv.org/abs/2501.09239