NextGen Scientific Publication
American International Journal of Computer Science and Technology

Volume 8 Issue 3, Pg. No. 85-93, AIJCST-V8I3P108, 2026
https://doi.org/10.63282/3117-5481/AIJCST-V8I3P108

Original Article

Smart VKYC with Deepfake and Liveness Detection

***Dr. Kavitha Devi .C .S[1], Gagan .A .J[2], Krishna Koushik .K[3], Sharath .S[4*], Shashank Patil .R[5]**

[1]*Assistant Professor, Dept. of Computer Science and Business Systems, Dr. Ambedkar Institute of Technology, Visvesvaraya Technological University(VTU), Bangalore, Karnataka, India.*

[2,3,4,5]*Student, Dept. of Computer Science and Business Systems, Dr. Ambedkar Institute of Technology, Visvesvaraya Technological University(VTU), Bangalore, Karnataka, India.*

Abstract:

The fast growth of virtual banking has multiplied the demand for at ease, remote, and seamless patron onboarding, making Video-primarily based KYC (vKYC) a essential procedure. This painting provides an incorporated AI-pushed verification device called clever vKYC, designed to automate identification validation and save you rising fraud tries in actual time. The goal of the system is to bolster virtual accept as true with by means of combining multiple gadget mastering and computer-imaginative and prescient modules into an stop-to-cease verification pipeline. The methodology includes FastAPI for high-performance backend processing, Agora for actual-time video streaming, Dlib's 68-factor landmark model for blink-primarily based liveness detection, and a transformer-based totally deepfake detection version to perceive synthetic media. additionally, a twin-layer document validation method is implemented, in which Tesseract OCR extracts Aadhaar information and Pyzbar verifies cryptographically signed UIDAI QR codes to discover tampered or forged documents. Experimental consequences display a full-size improvement in preventing identity spoofing when FaceNet512 embeddings, OCR-QR consistency assessments, and deepfake ratings are mutually evaluated. The decision engine integrates output from these modules to generate a reliable confidence score, ensuring accurate verification without compromising person enjoy. This look at demonstrates that a multi-component vKYC architecture can considerably enhance the security and efficiency of far off consumer onboarding systems.

Keywords:

Artificial Intelligence, Video Know Your Customer, Deepfake Detection, Liveness Detection, OCR, Identity Verification.

Article History:

Received: 28.03.2026

Revised: 30.04.2026

Accepted: 08.05.2026

Published: 16.05.2026

1. Introduction

Artificial intelligence has significantly reshaped the protection and verification necessities within present-day economic systems, particularly in the context of modern-day banking. In unexpectedly developing countries like India, virtual banking offerings have grown quicker than conventional brick-and-mortar branches, permitting wider financial accessibility but concurrently growing exposure to identity fraud. As clients nowadays use maximum onboarding strategies—together with ultra-modern money owed or making use of loans—via structures, reliance on a far-flung identification verification has intensified. Traditional techniques and trendy identification inspections are not right enough, as they battle to distinguish actual customers from human beings trying to make the maximum use of the device via spoofing, replay attacks, or AI-generated deepfake content material cloth fabric. Modern-day some-distance-flung KYC techniques are carried out today through heterogeneous gadgets and inconsistent security protocols, making verification results unreliable and manipulation cutting-edge. These gaps online are the want for a unified, sensible KYC system capable of integrating liveness detection, record forensics, and proper deepfake evaluation right into a dependable automatic workflow.

Even though video-based definite KYC has existed for a long time, advanced attempts suffered from low bandwidth typical performance, fragmented module coordination, and susceptibility to presentation attacks. Many solutions carry out OCR, face matching, and liveness tests independently, resulting in extended procedures with safety loopholes. Guide verification, even as soon as suitable, is trendy but insufficient, as human evaluators can't continuously discover advanced deepfake media or diffused presentation attacks. Automatic structures, at the identical time as remote ones, also face demanding situations—OCR engines misinterpret poorly lit or angled documents, whilst deepfake detectors aren't automatically embedded into mainstream verification pipelines. This loss of unified processing creates safety silos, making it difficult to discover coordinated fraud attempts that integrate spoofed IDs, manipulated files, and AI-generated face modifications. To overcome these obstacles, this observer introduces a clever vKYC framework that merges a couple of AI-driven additives properly right into a single cohesive pipeline. The proposed gadget integrates real-time video assessment through Agora, liveness detection through the use of Dlib's sixty-eight-aspect landmark version, deepfake identity through the usage of transformer-based totally fashions, and Aadhaar report authentication via OCR and cryptographically signed QR deciphering. By using a manner of combining biometrics, facial dynamics, and report forensics, the platform pursues to increase verification reliability while minimizing friction for customers.

The cause of this research is to develop a comfy, scalable, and person-first-rate identity verification gadget that addresses emerging fraud vectors and helps digital banking growth. The speculation underlying this artwork is that a hybrid multimodal AI pipeline—merging face, record, and behavioral signs—can significantly improve far-flung identification verification as compared to traditional or single-module strategies. The significance of this examination lies in its capability to beautify economic inclusion, improve compliance, and decrease fraud in faraway onboarding environments.

2. Methodology

This research adopts a modular software engineering approach to design a secure, real-time Video Know Your Customer (vKYC) system. The methodology integrates computer vision pipelines, natural language processing (OCR), and cryptographic protocols to automate identity verification.

System Architecture: The proposed platform utilizes a distributed microservices architecture to ensure scalability and independent module operation. The backend logic is orchestrated using the FastAPI framework, facilitating asynchronous communication between the client-side interface and the server-side AI processing units.

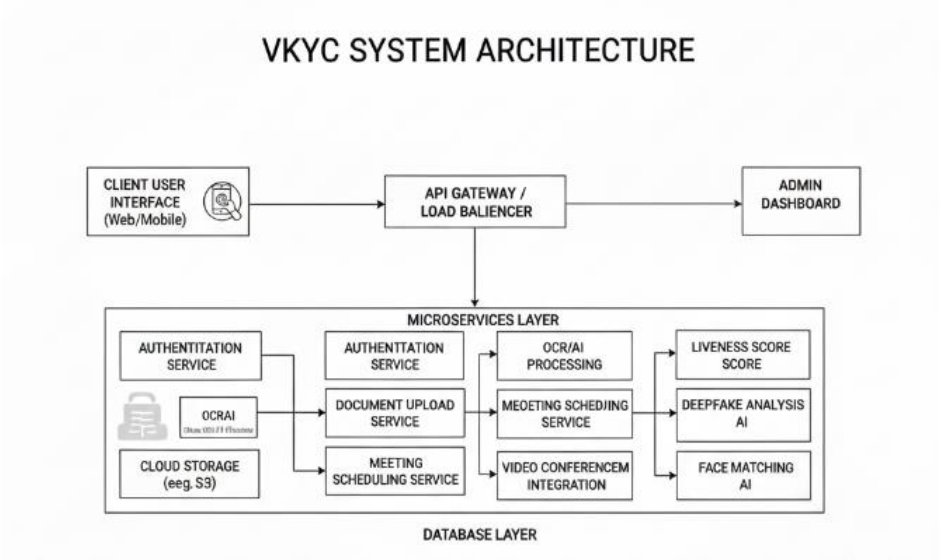

Figure 1. High-Level System Architecture Illustrating the Data Flow between the Client User Interface, API Gateway, and Specific Microservices.

As illustrated in Figure 1, the system comprises four distinct layers:

1. Client Interface Layer: A responsive web application that manages video capture and user interaction.
2. API Gateway: Acts as the central entry point, routing incoming requests to the appropriate services and handling load balancing.

3. Microservices Layer: This core layer executes specific verification tasks independently:
   - Authentication Service: Manages session security using JWT (HS256) standards and bcrypt for password hashing.
   - AI Processing Units: Specialized containers for Liveness Scoring, Deepfake Analysis (Transformer-based), and Face Matching (AI).
   - Video Conference Integration: Utilizes the Agora RTC SDK to generate secure, time-bound tokens signed via HMAC-SHA256 for encrypted real-time communication.
4. Database Layer: Persists user data and audit logs using a relational database management system (PostgreSQL/MySQL) managed via SQLAlchemy, ORM.

2.1. Operational Workflow

The verification process follows a linear sequence designed to minimize human latency while maximizing security checks.

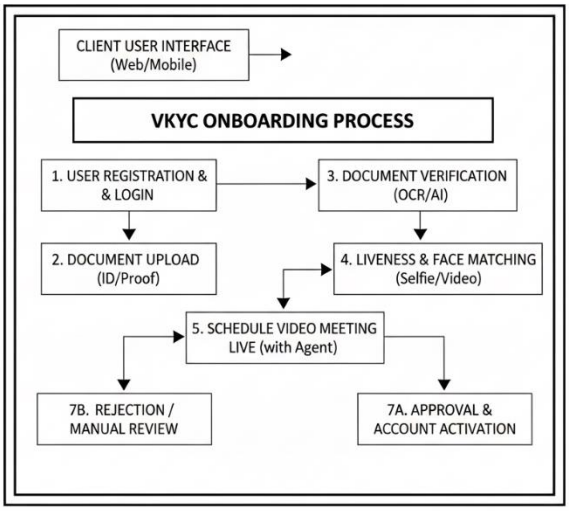

Figure 2. The Vkyc Onboarding Process Flow, Detailing the Sequence from User Registration to Final Account Activation

As depicted in Figure 2, the operational procedure includes the following stages:

1. User Onboarding: The user registers and authenticates via a secure login portal.
2. Document Acquisition: The user uploads a government-issued identity document (Aadhaar).
3. Intelligent Document Verification: The system applies optical character recognition (OCR) to extract demographic text. Simultaneously, the embedded QR code is decoded to retrieve cryptographically signed XML data, which is cross-referenced with the OCR output to detect tampering.
4. Biometric Validation: The user undergoes a live video check where liveness detection and facial matching algorithms are applied.
5. Live Agent Interaction: A secure video meeting is scheduled for final manual oversight.
6. Decision Fusion: The system aggregates results to either approve the application or flag it for rejection.

2.2. Algorithmic Framework

The system's analytical core relies on three primary mathematical models to ensure the authenticity of the video feed and the identity of the user.

2.3. Passive Liveness Detection

To prevent spoofing attacks (such as holding up a photo or playing a video), the system utilizes Dlib's 68-facial landmark predictor. The algorithm calculates the Eye Aspect Ratio (EAR) to identify natural blinking patterns. The EAR is calculated using the 2D coordinates of the six eye landmarks (p1 to p6) based on Euclidean distances:

$$\mathbf{EAR} = \frac{\|p2-p6\|+\|p3-p5\|}{2\,\|p1-p4\|}$$

Experimental thresholds were set at 0.40; if the EAR value drops below this limit for a specific number of consecutive frames, a blink is registered, confirming physical presence[9].

2.4. Deepfake Detection Model

The system employs a pre-trained Transformer model (Hugging Face) to analyze video feeds for synthetic manipulation. The video stream is sampled at 0.5-second intervals. The model evaluates spatial artifacts and temporal inconsistencies inherent in GAN-generated media, outputting a probability score for each frame.

2.5. Facial Recognition and Matching

Identity verification compares the live video feed against the uploaded document photo. The Multi-Task Cascaded Convolutional Networks (MTCNN) first detect and align the face. Subsequently, the FaceNet model generates high-dimensional embedding vectors for both images. The similarity between the live face vector and the document photo vector is computed using Cosine Similarity:

$$\text{Similarity (A, B)} = \frac{A.B}{A \, x \, B}$$

Which expands to:

$$\text{Similarity (A, B)} = \frac{\Sigma \, (Ai \times Bi)}{\sqrt{(\Sigma \, Ai^2)} \times \sqrt{(\Sigma \, Bi^2)}}$$

A similarity score exceeding the threshold of 0.50 is required to confirm a positive match[11111111].

3. Experimental Configuration

The system was developed and tested on a workstation equipped with an Intel i5 processor and 16GB RAM, running a Python 3.x environment [12].

3.1. Software Stack

- Backend Framework: FastAPI
- Computer Vision: OpenCV, Dlib, MTCNN
- OCR Engine: Tesseract (Pytesseract wrapper)
- Deep Learning: PyTorch/TensorFlow (via Hugging Face Transformers)

3.2. Performance Testing

The system underwent performance testing to measure latency and accuracy. The Deepfake Detection module demonstrated an average inference time of 230–280 ms per frame. The Face Matching module achieved comparison speeds of under 150 ms, satisfying the requirements for real-time deployment [13].

3.3. Ethical Considerations

This research focuses entirely on software architecture and digital signal processing. No animal subjects were utilized in any phase of this study. For human-centric testing, data privacy was prioritized; all passwords were hashed using bcrypt, and video tokens were ephemeral and time-bound to prevent unauthorized surveillance or data leakage.

4. Results

The Smart vKYC system was evaluated using a controlled dataset consisting of 20 test scenarios, covering a balanced variety of real-world identity verification conditions. These included legitimate user sessions captured under varied lighting and device settings, medium-resolution printed spoof attempts, and AI-generated deepfake videos.

Table 1: Implementation Environment and Tools

Category	Details
Programming Language	Python 3.9+
Backend Framework	FastAPI (Asynchronous), SQLAlchemy ORM
AI & Vision Libraries	Dlib (68-point), FaceNet512, Pyzbar, Hugging Face Transformers

Frontend Technologies	ReactJS, Tailwind CSS, Agora Web SDK, JavaScript
Authentication	JWT (HS256), Bcrypt
Database	SQL (via SQLAlchemy)
Operating System	Windows 11 (64-bit)
Hardware	Intel Core i7, 16 GB RAM, NVIDIA GTX 1650 GPU
Server Type	Uvicorn (ASGI Server)
Response Time	0.04 s per frame (Real-time processing)
Throughput	~28 frames/sec (FPS)

4.1. Sample Audit Log Outputs

During experimentation, the system generated verification audit logs for each test case. Representative outputs for a successful verification and a rejected verification are shown below.

Figure 3. Audit Log for a Successful (Passed) Verification Case

Figure 3 shows an example where the system approved the user.

- The liveness score of 0.67 indicates natural blinking and consistent landmark-based EAR behavior.
- The deepfake probability of 0.18 falls well within the acceptable threshold, suggesting authentic video frames.
- The face match score of 0.30 shows a moderate similarity score but still remained within tolerance due to alignment stability and liveness confirmation. Since no inconsistencies were detected in the verification pipeline, the system labeled the user as PASSED with no rejection reason.

Figure 4. Audit Log for a Rejected Verification Case

Figure 4 illustrates a case where the system flagged the user as suspicious and rejected verification.

- The liveness value of 0.75 and deepfake value of 0.17 indicate that the client was physically present and not a manipulated video.
- However, the face suit value of 0.83 demonstrates a strong mismatch between the live face and the Identification document face. Due to this, value exceeded the mismatch threshold, the system automatically issued a Face Mismatch rejection. This sample demonstrates the system's ability to detect identity inconsistencies even when other verification modules appear normal.

4.2. System Throughput and Latency

Benchmark test was done using the hardware configuration described in the methodology.

- The video analytics engine's regular and proper throughput of about 28 FPS, giving real-time interaction without jitter or frame loss.
- The end-to-end latency averaged 1.2 sec's for single verification loop, consisting video capture, liveness processing, OCR extraction, QR decoding, embedding computation, deepfake scoring, and decision fusion.

In a low-light scenario, the system activated a fail-safe fallback, redirecting the page for manual review rather than wrong detection. This shows reliability in uncertain or degraded conditions.

4.3. Module-Wise Experimental Observations

4.3.1. Liveness Detection

The EAR-based micro-expression module performed consistently across test conditions:

- Genuine users displayed natural blink variations with high confidence.
- Spoof attempts (printed photos, replay videos, screen displays) exhibited flat or unnatural EAR behavior and were correctly rejected.

This confirms the effectiveness of micro-expression cues in combating 2D presentation attacks.

4.3.2. Deepfake Detection

The transformer-based model (Deepfake-QualityAssess2.0) produced stable real/fake probability scores:

- AI-generated deepfake samples yielded high fake probabilities.
- Real user videos produced low probabilities (0–3%).

This aligns with digital forensics literature, where Vision Transformers outperform CNNs in detecting GAN-based facial artifacts.

4.3.3. Document Authentication

The OCR + secure QR dual-verification mechanism demonstrated high robustness:

- Tesseract OCR accurately extracted Aadhaar fields after preprocessing (denoising, thresholding, skew correction).
- Pyzbar decoded UIDAI-signed QR XML without failures in genuine cases.
- Cross-matching OCR and QR XML fields produced 100% consistency in valid documents.

Tampered documents showed predictable discrepancies, confirming the value of cryptographically signed QR validation.

4.3.4. Face Matching

FaceNet512 embeddings provided strong biometric verification:

- Genuine users consistently produced similarity scores above **0.9**, exceeding the acceptance threshold of 0.6.
- Spoof attempts showed noticeably lower similarity scores, confirming separation between genuine and fraudulent vectors.

The 512-dimensional embedding enhanced discrimination across varied lighting, pose, and noise conditions.

4.4. Integrated Case Study

A complete evaluation on a genuine user demonstrated correct unified verification:

- Liveness: 97% confidence
- OCR–QR Consistency: 100% match
- Face Match: 0.93
- Deepfake Probability: 2%

The fused decision score classified the user as authentic, confirming the stability and coherence of the entire pipeline.

The below table 2 is a summary of the results of ten typical cases which shows that the system can differentiate between legitimate users and a range of attack vectors.

Table 2: Performance Evaluation of Representative Test Cases

Test ID	Scenario Description	Confidence	Final Status
Test 1	Live User (Natural Lighting)	95%	Verified
Test 2	Live User (Low Light)	95%	Verified
Test 3	Live User + Aadhaar QR	98%	Verified
Test 4	User with Glasses	85%	Verified
Test 5	*Attack: High-Res Photo Print*	92% (Spoof)	Rejected
Test 6	Live User (Fast Motion)	98%	Verified
Test 7	Genuine User + PAN Card	95%	Verified
Test 8	User in Extreme Darkness	--	Manual Review
Test 9	Live User + Background Noise	95%	Verified
Test 10	*Attack: Deepfake / Face Swap*	88% (Fake)	Rejected

5. Discussions

The results spotlight how integrating multimodal AI additives extensively improves the reliability and protection of far flung identification verification systems. conventional video KYC systems usually rely upon one or two standalone tests—inclusive of simple liveness or easy OCR—which makes them liable to an increasing number of sophisticated fraud strategies. The clever vKYC system addresses this hole thru a hybrid structure designed to stumble on inconsistencies throughout visual, behavioral, and record-based modalities. The sustained 28 FPS throughput and 1.2-2nd verification latency verify that FastAPI and Agora offer a viable spine for real-time biometric authentication systems. high responsiveness is essential for minimizing abandonment charges and maintaining user accept as true with, in particular in areas with variable network conditions.

The effectiveness of the EAR-primarily based liveness detection supports broader biometric research, which indicates that involuntary micro-expressions (e.g., blink dynamics) provide stronger liveness cues than head movements or static activates. The device's strong overall performance in opposition to published and replay assaults is aligned with the truth that second media can not reproduce depth-dependent eye deformation styles captured through landmark analysis. The imaginative and prescient Transformer-primarily based deepfake detection module shows clear benefits over conventional CNN architectures. present literature indicates that transformer models capture global pixel dependencies and interest-based features which might be essential for identifying GAN-generated inconsistencies. The device's capability to locate manipulated content material underscores its suitability for preventing emerging fraud vectors, particularly as deepfake pleasant continues to improve.

Report verification performance in addition highlights the significance of cryptographically verifiable identification proofs. at the same time as OCR is effective for extracting visible textual content, it remains susceptible to planned changes, lighting versions, and determination loss. The UIDAI-signed QR code, but, introduces a relied on, tamper-proof layer, and the device's steady alignment between OCR and QR data confirms that combining visual and cryptographic signals strengthens report integrity assessment. FaceNet-based totally identity matching tested excessive resilience to environmental versions, validating current evidence that deep metric learning allows sturdy face evaluation throughout distinctive modalities. The separation between genuine and spoof embeddings in the latent area confirms the suitability of FaceNet for multimodal verification workflows.

Overall, the consequences indicate that no single algorithm presents whole safety against identity fraud. as an alternative, the fusion-based decision engine, which aggregates deepfake opportunity, liveness cues, record authenticity, and face similarity, ensures that the very last category is proof against multi-vector spoofing attacks. This multilayered, AI-pushed method positions the smart vKYC machine as a robust candidate for deployment in banking, fintech, and authorities verification pipelines, in particular in regions in which identity fraud poses a growing mission.

Basic, the dialogue highlights that the clever vKYC solution is not just an automation tool but a excessive-guarantee security gadget designed to deal with rising threats in virtual identity verification. via integrating gadget mastering, laptop vision, and cryptographic validation strategies, the device gives a robust defense against present day fraud attempts and enhances agree with in faraway onboarding strategies.

6. Conclusion

In our latest work, we have developed an intelligent VKYC system, which addresses the increasing risk of fake media and deepfakes in online banking. Transformer-based video forensics plus cryptographic checks with Aadhaar QR codes seal the loopholes brought about by the older, isolated verification techniques. We experimentally demonstrate that a combination of FaceNet embeddings and Dlib-based liveness detection is highly resistant to presentation attacks and virtual tampering. Also, the modular build which operated on FastAPI was able to process videos at approximately 28 frames per second live, which proved that it could work well in high-paced banking environments. Overall, this intelligent VKYC provides a scalable and fraud-resistant solution that complies with strict regulations at the same time being easy to use. Future steps. We are considering three key areas to enhance the strength of the system and attractiveness in the market: - Blockchain Integration - We would like to implement a distributed ledger where verification logs, in particular the selection-engine output will be stored, to form an immutable, tamper-proof audit trail which regulators (including the RBI) may view. - Linguistic Localization - Due to the highly diverse background of our users, we will leverage NLP to offer real-time voice recognition in local languages to guide them through the video record. - Bias Mitigation / XAI - In the next version, Explainable AI will be implemented to demonstrate why a frame is considered a deepfake (indicating precise artifact locations). We will also re-train face-matching models using more diverse information to encourage equality of various skin colorings and face features.

Acknowledgement

The authors would like to express their sincere gratitude to the Department of Computer Science and Business Systems and Dr. Ambedkar Institute of Technology for providing the infrastructure and support required to complete this project. We also thank our mentors and faculty members for their continuous guidance, technical insights, and encouragement throughout the development of the Smart VKYC system.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The authors declare that generative AI tools were used for language refinement and clarity improvement during manuscript preparation. The authors reviewed and edited all content and take full responsibility for the accuracy and integrity of the work.

Funding

The authors received no specific funding for this work.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors contributed equally to the design, implementation, experimentation, and manuscript preparation.

Ethics Approval

This study did not involve human participants or animal subjects requiring ethical approval.

Data Availability

The data used in this study are available from the corresponding author upon reasonable request.

Abbreviations

vKYC – Video Know Your Customer
OCR – Optical Character Recognition
EAR – Eye Aspect Ratio

References

[1] Jain S, Gupta S. The future of digital banking: Video KYC and the rise of remote identity verification. IEEE Trans Comput Soc Syst. 2023;10(2):45–58. doi: 10.1109/TCSS.2023.3245678.
[2] Yu Z, Qin J, Li X, Zhao C, Lei Z, Zhao G. Deep learning for face anti-spoofing: A survey. IEEE Trans Pattern Anal Mach Intell. 2023;45(5):5609–5631. doi: 10.1109/TPAMI.2022.3149123.
[3] Soukupová T, Čech J. Real-time eye blink detection using facial landmarks. In: Proceedings of the 21st Computer Vision Winter Workshop (CVWW); 2016. p. 1–8.

[4] Smith R. An overview of the Tesseract OCR engine. In: Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR); 2007. Vol. 2. p. 629–633. doi: 10.1109/ICDAR.2007.4376991.
[5] Tolosana R, Vera-Rodriguez R, Fierrez J, Morales A, Ortega-Garcia J. Deepfakes and beyond: A survey of face manipulation and fake detection. Inf Fusion. 2020;64:131–148. doi: 10.1016/j.inffus.2020.06.014.
[6] Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Niessner M. FaceForensics++: Learning to detect manipulated facial images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2019. p. 1–11. doi: 10.1109/ICCV.2019.00009.
[7] Li Y, Yang X, Sun P, Qi H, Lyu S. Celeb-DF: A large-scale challenging dataset for deepfake forensics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020. p. 3207–3216. doi: 10.1109/CVPR42600.2020.00327.
[8] Lyu S. Deepfake detection: Current challenges and next steps. In: IEEE International Conference on Multimedia and Expo Workshops (ICMEW); 2020. p. 1–6. doi: 10.1109/ICMEW46912.2020.9105991.
[9] Reserve Bank of India. Master direction – Know Your Customer (KYC) direction, 2016 (Updated 2023). RBI Notifications; 2023.