Continuous Multi-Modal Biometric Authentication for KYC and Fraud Prevention

December 11, 2025

Abstract

Know-Your-Customer (KYC) processes increasingly leverage biometric authentication to verify user identities. However, sophisticated attacks such as deepfakes pose serious threats to remote identity verification. This paper explores a secure multi-modal biometric pipeline (combining face and voice recognition with liveness and deepfake detection techniques) and proposes continuous authentication strategies. By using services including Active Speaker Detection, Audio-Visual Forensics, Visual Speech Recognition, Face Liveness, Face Recognition, and Voice Recognition in tandem, the system can more reliably detect impersonation or spoofing attempts. Furthermore, we discuss how continuous authentication – leveraging historical successful login data and adaptive biometric templates – can improve accuracy and resilience over time. We review related approaches and suggest techniques such as updating facial embeddings from prior authentications (e.g., using running averages) to reduce false rejections while guarding against fraud.

1 Introduction

Financial institutions and online services are adopting biometric-based KYC verification to strengthen security and combat fraud. In recent years, the rise of generative AI *deepfakes* has introduced new risks: attackers can synthetically impersonate a customer's face or voice to bypass biometric checks. These AI-enabled attacks are no longer just theoretical – industry reports indicate that roughly one in ten companies has already encountered deepfake-related fraud, with projected losses in the billions if defenses do not improve [1]. Deepfakes can undermine standard liveness detection by mimicking facial movements and even micro-expressions, eroding trust in traditional one-time verification methods [1].

To counter these threats, there is a growing consensus that multi-modal biometrics and continuous authentication are needed for robust KYC. Multi-modal biometric verification combines multiple independent traits (e.g., face *and* voice) to provide layered security, making it far more difficult for an impostor to spoof all modalities simultaneously [4]. In fact, studies have shown that fusing face and voice recognition results can increase authentication accuracy and reduce error rates significantly compared to using either modality alone [4]. This approach aligns with user expectations as well: a recent survey found that over half of consumers believe biometric checks (especially facial recognition) will become more valuable in the era of deepfakes, and 80% are willing to spend extra time on identity verification if it improves security [2].

Another key enhancement is a shift from point-in-time verification to **continuous authentication**. Instead of authenticating a user only at login, continuous authentication methods monitor the user's identity throughout the session or across repeated logins [3]. This provides ongoing protection: if an unauthorized user hijacks a session or if anomalies are detected after login, the system can re-verify identity or flag fraud in real time. Continuous authentication can be implemented by periodically re-scanning biometric traits or passively analyzing user behavior patterns, ensuring the active user remains the legitimate customer [3]. Moreover, continuous approaches can adapt to legitimate intra-user variations over time, offering a balance between security and usability.

In this paper, we propose a secure biometric KYC pipeline that integrates multiple biometric subsystems with deepfake detection measures, and we explore how leveraging historical authentication data in an adaptive manner can enhance system performance. Section 2 reviews the components of our multi-modal biometric verification approach and how each combats specific fraud vectors. Section 3 discusses continuous authentication and the use of historical embeddings for adaptive biometric matching. We also highlight related work and real-world examples (such as adaptive facial recognition on consumer devices) that inform our approach. Finally, Section 4 concludes with remarks on how these techniques together fortify KYC against evolving fraud threats.

2 Multi-Modal Biometric Verification Pipeline

A multi-modal biometric pipeline can significantly strengthen KYC identity verification by combining several complementary detectors and verifiers. Our system employs the following components in a secure authentication flow:

- **Face Liveness Detection**: Ensures the presented face is from a live person and not a spoof (photo, video replay, or mask). Advanced liveness techniques use 3D facial mapping, texture analysis, challenge-response actions, and even physiological cues (e.g., heartbeat or skin reflectance) to detect fake presentations [1]. These measures go beyond simple blink detection, making it far more difficult for deepfake videos or static forgeries to appear "alive."

- **Active Speaker Detection**: Verifies that the audio speech is coming from the person's face on camera. This involves analyzing the synchronization of lip movements with the spoken words. By checking audio-visual consistency, the system can detect anomalies such as dubbed voices or misaligned mouth movements that often indicate deepfake or tampering [1].

Active speaker verification thwarts attacks where an impostor might play a victim's voice recording over a different video.

- **Audio-Visual Forensics**: Uses AI-based analysis to spot artifacts or inconsistencies in the video and audio streams that could suggest deepfake generation or other manipulation. For example, forensic algorithms may detect unnatural blending on facial boundaries, abnormal motion between frames, or audio artifacts. By assessing both modalities, these detectors can identify subtle signs of synthetic media that single-modality checks might miss [1].

- **Visual Speech Recognition (Lip Reading)**: Analyzes the video of the user's lip movements to determine the speech content, which can be compared to the actual audio. This cross-check helps ensure that the spoken content matches the lip movements. In a KYC scenario, the user might be asked to read a random phrase or numbers aloud; visual speech recognition can confirm that the correct phrase was spoken, adding another layer of verification against deepfake audio insertion or pre-recorded video.

- **Face Recognition (Identity Verification)**: Performs facial recognition by comparing the live face image (or its computed embedding) to the stored reference of the claimed identity. Modern face recognition models generate a high-dimensional embedding of the face and measure similarity to the enrolled template. Using a well-trained deep network (e.g., based on ArcFace or similar) ensures high accuracy. In our context, face recognition confirms that the person in front of the camera matches the user who originally onboarded (or the photo ID on file), thus preventing impostors from using a different face [5].

- **Voice Recognition (Speaker Verification)**: Similarly, the system captures the user's voice and extracts a voiceprint (using features like MFCCs fed into a speaker recognition model). It then compares this voiceprint to the enrolled voice template for the user. This verifies identity through the vocal biometric. Voice recognition adds another factor of authentication, which is especially useful if the user had provided a voice sample during KYC onboarding. Even if the face were somehow mimicked, a mismatched voice can expose the fraud.

Each of these components addresses specific attack vectors. Face and voice recognition verify the identity claim, while liveness, active speaker, and forensic checks verify the authenticity of the biometric data itself. By operating these in parallel, the system achieves a layered defense (sometimes called defense-in-depth). For instance, even if an attacker manages to create a highly realistic deepfake face video of the victim, the voice might not match or the lip synchronization might be slightly off; the active speaker check and voice match would fail, blocking the attempt. Conversely, an impostor using a stolen audio clip of the victim's voice would be caught by the face mismatch or liveness checks. This multimodal approach thus dramatically raises the bar for fraudsters.

Not only does multi-modal fusion improve security, it also enhances accuracy under normal conditions. If one modality is temporarily unreliable (e.g., poor lighting affecting face image quality, or background noise affecting voice), the other modality can compensate, leading to more robust overall performance [4]. Empirical research supports this: Gofman *et al.* demonstrated that combining face and voice biometrics with a quality-weighted score fusion improved the true acceptance rate by several percentage points compared to single-biometric verification [4]. By intelligently weighting the modalities based on their reliability (for example, down-weighting a noisy voice sample and relying more on the face, or vice versa), the system can adapt to environmental variability while still ensuring the user is correctly authenticated.

It is important, however, to maintain a good user experience. Our pipeline is designed to be as seamless as possible: all checks can occur near-instantaneously during a user login or verification session. Users simply follow a prompt to look at the camera, speak a short phrase, and possibly perform a quick instructed motion (for liveness). The system processes these inputs in a few seconds or less. Thanks to advances in optimization and hardware, even resource-constrained devices like smartphones can handle such multi-modal processing efficiently [3]. The payoff is a highly secure KYC verification that is still user-friendly and fast, balancing the requirements of fraud prevention and customer experience.

3 Continuous Authentication and Adaptive Learning

While the multi-modal pipeline significantly secures the initial login or onboarding verification, another layer of defense is *continuous authentication*. Continuous authentication refers to the practice of continually validating the user's identity during the entire session or across recurring interactions, rather than relying on a single checkpoint at login [3]. This approach is particularly valuable for fraud detection because it can catch "session takeover" scenarios or detect gradual changes that might indicate an impostor has replaced the legitimate user.

In a continuous authentication regime, the system might periodically re-authenticate the user in the background. For example, if a user remains logged in for an extended period, the webcam and microphone could intermittently re-capture face and voice data to confirm the same person is still present. If any of the biometric comparisons fail (or if liveness checks start failing), the system can lock the session or trigger an alert. This real-time monitoring is crucial for high-security applications (e.g., preventing an attacker from accessing sensitive financial data if the real user steps away). Researchers Ayeswarya & Singh [3] have shown that context-aware continuous biometric fusion (even using modalities like gait or keystroke patterns) can greatly enhance security with minimal user inconvenience, adapting the authentication requirements to the current risk context.

Beyond within-session monitoring, we propose leveraging the **history of successful authentications** to improve the biometric matching accuracy over the long term. Biometric traits naturally exhibit some variation over time and across capture conditions. A single enrollment template (say, one face photo or one voice sample) might not fully represent the user's appearance or voice in all scenarios. Therefore, our system can perform **adaptive template updates** using the embeddings or features extracted from each successful login.

One straightforward technique is to maintain a running average of the user's face embedding. For instance, during initial KYC onboarding, the user provides a face image which generates an embedding vector E_0. Each time the user logs in thereafter and passes all checks (indicating a genuine acceptance), the new face embedding E_i could be integrated into the user's reference template. If E_{avg} is the current stored template (average embedding), we can update it as:

$$E_{\text{avg}} \leftarrow (1-\alpha)\, E_{\text{avg}} + \alpha\, E_i,$$

with a learning rate α (e.g., $\alpha = 0.1$) giving more weight to recent samples. Over time, this running mean embedding will incorporate various legitimate appearances of the user – different lighting, angles, hairstyles, glasses, etc. – making the face recognition more robust. Franco *et al.* demonstrated the viability of such incremental template updating for face recognition; their video-based template update method showed improved accuracy in recognizing subjects in home environments as new authentic samples were gradually added [5]. An adaptive biometric system can thus "learn" the intra-class variations of the user, reducing false rejections (for example, the system is less likely to mistakenly reject a user who grew a beard or changed hairstyle, because the updated template has absorbed those variations). Modern commercial systems like Apple's Face ID are known to use adaptive machine learning to accommodate changes in a user's appearance over time, continuously updating the stored face data as the user successfully unlocks their device [6]. This ensures that facial biometric authentication remains accurate even as the user ages or alters their look, without requiring re-enrollment.

For voice recognition, a similar approach can be taken. The system can update the voice print by averaging voice feature vectors from successful authentications, accounting for variations in the user's accent, tone (which might change due to illness or noise conditions), and recording device differences. By maintaining an up-to-date voice model, the system avoids performance degradation that could occur if only an old enrollment sample is used as reference.

While adaptive updating is powerful, it must be done cautiously to avoid introducing fraud into the template (a known risk called template poisoning). Our strategy is to update templates only on *high-confidence* authentications. In practice, this means that all anti-spoofing checks (liveness, speaker verification, etc.) must be firmly passed and the face/voice match scores must be above a stringent threshold before considering an update. Additionally, multi-modal cross-checks provide assurance; for instance, if the face matches strongly and the voice matches as well, it is very likely the true user, so we can safely learn from that sample. If any doubt exists (e.g., borderline match score or slight liveness irregularity), the system would refrain from updating the profile and may even prompt for additional verification. By setting these guardrails, we ensure that impostors cannot gradually corrupt a user's template by mimicking them poorly and getting partial updates. This concept follows the guidelines of adaptive biometric systems as surveyed by Rattani *et al.*, which emphasize supervised (controlled) template updates to handle intra-user variability while maintaining security [5].

Another aspect of using historical data is analyzing behavioral patterns of logins. Although our focus is on biometric factors, continuous authentication can also include non-biometric context features, such as typical login time, geolocation, device characteristics, and user interaction patterns. Incorporating these into a risk engine can further strengthen fraud detection. For example, if a login occurs from an unusual location or at an odd time inconsistent with the user's history, the system could heighten the security requirements (triggering more frequent biometric checks or requiring an additional factor). This kind of adaptive risk-based authentication is recommended in modern fraud prevention frameworks [1], and it complements the biometric pipeline by handling cases that pure biometrics cannot (such as an attacker using stolen biometrics from a different country).

In summary, continuous authentication and historical data utilization provide a proactive defense: the system not only verifies identity at the start, but continues to *learn and monitor* thereafter. By continuously refining the user's biometric templates with genuine data, the system becomes more accurate (reducing user friction from false negatives) and more resilient to slow-evolving attacks (like an impostor who tries to slowly train a deepfake model—such attempts would consistently fail one of the checks and never be incorporated). Meanwhile, by constantly watching for discrepancies during a session, the system can terminate suspicious activity in real-time, limiting the damage window of any breach.

4 Conclusion

As fraudsters turn to AI-driven impersonation techniques, KYC verification systems must evolve beyond traditional single-modal, one-time authentications. In this paper, we outlined an approach for a multi-modal biometric authentication pipeline enhanced with continuous authentication and adaptive learning. By combining face recognition, voice recognition, and stringent liveness and deepfake detection (active speaker verification, audio-visual consistency checks, etc.), the system achieves a robust defense-in-depth against identity fraud. We also discussed how leveraging the history of user authentications to update biometric templates can yield more accurate and user-tolerant security over time, all while carefully safeguarding against template poisoning.

Our proposed methods align with industry observations that multi-layered verification and ongoing adaptation are crucial to keep pace with evolving threats [1]. The integration of multiple biometrics means that an attacker must defeat several independent checks at once, a substantially harder endeavor than fooling any single biometric in isolation. Continuous monitoring ensures that even after login, any anomalous behavior or biometric mismatch can be caught swiftly, limiting the opportunity for fraudulent access. Adaptive template updates ensure that the system remains calibrated to the legitimate user, reducing false alarms and enhancing the customer experience, which is vital for user acceptance of security measures.

Future work will involve evaluating this integrated system on real-world data, including simulated attack scenarios with deepfake videos and voices. It will be important to quantify the false acceptance rate when facing sophisticated forgeries, and to tune the fusion and update algorithms for an optimal balance between security and convenience. Another avenue is exploring advanced machine learning models that can jointly embed face and voice data for each user, potentially allowing even more robust multi-modal matching and anomaly detection. Additionally, privacy considerations must be addressed: storing and updating biometric data requires strong encryption and compliance with data protection regulations, and continuous authentication should be done in a user-transparent and privacy-preserving manner.

In conclusion, a KYC process fortified with multi-modal biometrics and continuous authentication stands as a promising solution to current fraud challenges. By learning from each genuine interaction and vigilantly checking for impostors, such a system can significantly raise the trust and safety in digital customer onboarding and authentication, ensuring that *“you are who you claim to be”* at all times.

References

[1] G. Regan, *Securing Financial KYC Verification Against Deepfake Breaches.* Reality Defender Insights, Jan. 2025. [Online]. Available: `https://www.realitydefender.com/`

[2] Jumio, *2023 Online Identity Study.* Jumio Global Consumer Research Report, 2023. [Online]. Available: `https://www.jumio.com/2023-identity-study/`

[3] S. Ayeswarya and K. J. Singh, *Enhancing security and usability with context aware multi-biometric fusion for continuous user authentication. Scientific Reports*, vol. 15, art. 30627, 2025.

[4] M. Gofman, S. Mitra, K. Cheng, and N. Smith, "Quality-Based Score-level Fusion for Secure and Robust Multimodal Biometrics-based Authentication on Consumer Mobile Devices," in *Proc. 10th Intl. Conf. on Software Engineering Advances (ICSEA)*, 2015, pp. 274–279.

[5] A. Franco, D. Maio, and D. Maltoni, "Incremental template updating for face recognition in home environments," *Pattern Recognition*, vol. 43, no. 8, pp. 2891–2903, 2010.

[6] REPART Team, *Unlocking Face ID Secrets: A Comprehensive Analysis.* REPART Blog, Oct. 2023. [Online]. Available: `https://irepart.com/blogs/articles/unlocking-face-id-secrets`