World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Deepfake-Resistant Telehealth: Multi-Factor Voice-Face-Contextive Verification Under Real-Time Constraints

Leeman Takunda Gunzo [1,*], Tendai Nemure [2], Munashe Naphtali Mupa [3] and Japhet Dalokhule Muchenje [4]

[1] *University of Tennessee Knoxville.*
[2] *Munashe Naphtali Mupa.*
[3] *Hult International Business School.*
[4] *Suffolk University.*

World Journal of Advanced Research and Reviews, 2025, 28(01), 152-159

Publication history: Received on 23 August 2025; revised on 27 September 2025; accepted on 30 September 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.1.3387

Abstract

Telehealth's high growth rate has raised the apprehensions that the impersonation attacks, which have been conducted due to the deepfake technologies, are dangerous to patient safety and compliance with regulations. The proposed research proposes a multi-factor verification with voice spectra analysis, facial micro-motion, and contextual metadata (IP, device, timing) with real-time limitations. The system exploits lightweight CNN and transformer encoders that are privacy-preserving by federated learning and differential privacy. Implemented using SOC workflow integration and deployed by using WebRTC middleware, the framework demonstrated an improvement in true positive rate by 295%, a false positive rate of less than 3%, and a latency of less than 250 milliseconds. Findings also identify resistance to adversarial perturbations and multi-strategy deepfakes. Artifacts of evidence, such as an AI intrusion detection prototype and the impact of teaching, support the applicability of the system. The paper concludes with policy implications about HIPAA/GDPR compliance and establishes future directions, such as IoT wearables, federated adversarial training, and zero-trust telehealth architecture.

Keywords: Constraints; Contextive; Deepfake; Multi-Factor; Real-Time; Telehealth

1. Introduction

The past years have seen a rapid increase in the use of telehealth services due to the development of digital connectivity and the necessity to provide easily accessible healthcare services. Video and voice consultations have become a more dependable tool that allows patients to communicate with clinicians, making it convenient, enabling them to continue care, and allowing access to previously underserved areas. While these benefits are undeniable, the rise of artificial intelligence (AI)-generated media has introduced new vulnerabilities. The development of voice and video deepfakes poses a high risk to the assurance of identity since it may allow rogue actors to impersonate patients or clinicians during a remote session. This type of impersonation not only destroys trust but also poses threats to patient safety and regulatory compliance in vulnerable health settings (Cheng et al., 2023).

Deepfake technology has become so sophisticated that nowadays, something can be secured not only with the help of traditional authentication policy. All passwords, fixed face, and simple voice biometrics can be used in syncretic forgery. This is among the cases where the authentication of an individual during a telehealth session is not guaranteed. The problem is more widely used in the healthcare industry, in which the disclosure of the safeguarded health information, or PHI, can take place. The attempted impersonation that has already been achieved may also serve as the basis of the

* Corresponding author: Leeman Takunda Gunzo

violation of the Health Insurance Portability and Accountability Act (HIPAA) (Alvarez et al., 2024). The privacy identity authentication requirement to obtain, secure, and sustain has remained a priority in the telehealth environment.

Students and scientists are shifting toward context-sensitive and multi-modal mechanisms as more competitive defenses against the high invariance. With a combination of behavioral, environmental, and biometric cues, one can build deepfake-invariant verification pipelines. Beyond the correctness, such measures are also required within narrow latency constraints because the telehealth/workflows rely on the radiological decision of the patient and professionals, as well as the capacity of the patients and the professionals to communicate with each other in real-time. According to Gondaliya (2025), security vs. usability implies system design with delicacy as well as lightweight verification schemes and data-in-the-air techniques that retain the confidentiality of the information.

The work currently mitigates similar issues by the integration of a verification system that is impervious to deepfakes and well-suited to telehealth applications. Unlike typical schemes, the introduced scheme integrates three authentication levels in one: voice spectral pattern, face micro-motion, and session context in terms of IP, device identity, and timing analysis. Such modalities are combined extended structures with an effective encoder that is suitable for online processing. The model is further enriched by identification through adversarial evaluation of robustness, liveness detection, and preservation of privacy processing in order to comply with the regulative HIPAA.

Research Objectives

- To create and deploy a multi-factor identity verification solution, combining voice, facial, and contextual signals, for telehealth security.
- To test the system in real-time constraints, the maximum additional latency should not exceed 250 ms, without compromising user experience.
- To create reference middleware, security operations center (SOC) runbooks, and HIPAA-compliant data flow diagrams that can be used to support adoption in healthcare workflows

2. Literature Review

2.1. Telehealth Security Landscape

Telehealth ushered in the age of transformation in the dynamics of healthcare delivery, and distant healthcare consultation is among the features of routine clinical practice. However, with that came new cybersecurity issues, particularly in authentication and information protection. Part of the main issue is the added complication in the acts of impersonation, such as deepfakes using artificially produced faces and voices in order to get around traditional security measures.

As indicated by Aror and Mupa (2025), there is the application of two conflicting functionalities of AI in the current risk management practice: on the one hand, there is the application of AI for the automation of security surveillance, and on the other hand, the building of adversary capabilities (i.e., deepfakes). The dilemma puts healthcare providers in an untimely position because they have to use the protection through the use of the AI while they are battling against the AI-based assault.

Kalu-Mba et al. (2025a) emphasize the role of AI as a catalyst for innovation in the public sector and that there are opportunities and risks associated with the use of AI in healthcare. Though AI can assist in enhancing efficiency and resiliency in remote care, implementation of the technology must be closely observed to avoid accidental exposure to the patients. This can be linked to the higher policy necessity of the way we strike a balance between the necessity to be innovative in terms of innovation and regulatory protection.

The study by Chitemerere et al. (2025) is also an addition to this debate here, taking place from a strategic angle with a focus on the digital trust that will continue to be the backbone of any successful cross-border or remote service model. When applied in telehealth, it implies that the possibility to provide secure and authentic communication between providers and patients is the key to sustainable adoption.

2.2. Deepfake Detection Methods

Deepfake attack mitigation on telehealth is based on literature on synthetic media detection in face and voice modalities. Of particular interest are deepfakes in the voice modality because there is an opportunity to evade traditional phone- and voice-based authentication. Fazeeha et al. (2025) demonstrate an artificial intelligence-based system to detect

spoofed voices in real time and provide the evidence that the analysis of the features based on the spectrogram is sufficient to detect the slight discrepancies. While this is encouraging, there is the requirement in this solution that high-quality input audio be available—something that can be lost in noisy telehealth environments.

In the aspect of visual, Lei et al. (2025) suggest using a multi-feature decision fusion framework in order to recognize face deepfakes. The authors show experimentally that the robustness to adversarial attacks can be adjusted upwards by having the combination of the texture, motion, and geometric cues, which are better than individual feature models. Uddin et al. (2025), however, utilize a multi-level discrete wavelet transform combined with vision transformers and show good comparisons with state-of-the-art face swap techniques. Though both techniques encourage active participation, there is a strong emphasis on adversarial robustness for Lei et al. (2025) and an emphasis on accuracy on different datasets for Uddin et al. (2025).

Waseem et al. (2023) also cover this point and tackle the problem in a more comprehensive way by combining it with the multi-attention mechanism to identify the face swaps and expression manipulations. This is in contrast to the traditional convolutional methods, which do not enable the model to focus dynamically on the suspicious parts of the face. Not only does this enhance detection rates but also offers interpretability, which is useful in medical environments where explainability is commonly necessary.

In general, the literature demonstrates a tendency toward multi-modal and ensemble-based techniques as the most promising methods of deepfake detection. Nevertheless, the common weakness in the reviewed studies is that most of them studied isolated modalities or fixed conditions, which highlights the necessity of cohesive frameworks that can act effectively in the dynamic telehealth setting.

2.3. Multi-Modal Biometric Fusion

Due to the shortcomings of unimodal deepfake detection, scientists are becoming more proponents of multi-modal biometric fusion. These systems increase resiliency against impersonation by integrating contextual metadata (face, voice, and metadata) with face and voice. Balaji et al. (2025) propose a federated deep learning model that combines facial and eye-blink appearances and provides strong authentication. Their study shows that federated methods can maintain privacy and enhance accuracy, but the use of eye-based signals may not always be feasible in telehealth sessions.

Komarulu (2025) suggests a multi-modal framework that is transformer-based and that can have an adaptive authentication of various biometric inputs. In comparison to Balaji et al. (2025), who focus on privacy, the contribution that Komarlu makes is that of creating an architecture, which enhances its judgment of the authentication process through repeated optimization, thus making it more resilient to changing deepfake tactics. This flexibility comes into use especially in telehealth, where the opposition techniques are likely to become more advanced.

Zen et al. (2025) choose an ensemble-based methodology, which resists multi-strategy deepfake image generation. The significance of redundancy in their findings is emphasized by the fact that a mixture of different types of biometric check-ups is certain to create a case when one of the modalities is compromised; the rest of the systems will maintain the integrity of a system. Zen et al. (2025) emphasize latency less than Komarulu (2025) and accentuate a trade-off that telehealth applications need to address: robustness versus latency.

Additionally, Hu et al. (2025) add to a human-inspired model of multi-face detection a scenario that is more of a group telehealth session with caregivers, specialists, or family members. Their focus on contextually sensitive verification broadens the range of applicability to multi-modal systems to real-world health care interactions, which are frequently not just one-to-one interactions. All these studies collectively demonstrate that multi-modal biometric fusion can provide greater protection than single-modality strategies.

2.4. Challenges in Adversarial Robustness & Real-Time Constraints

Although biometric fusion and detection have advanced, a number of obstacles are yet to be overcome. Latency of healthcare workflow Healthcare is a latency-sensitive area, and any security mechanism should perform at maximum delay in order to prevent consultation interruptions. Nazeri et al. (2024) analyze detection transformers under adversarial conditions and illustrate the accuracy-computational-overhead trade-off. While their models provide high levels of resilience, their inference times could be prohibitive in live telehealth environments.

Liao et al. (2025) propose efficiency-oriented calibration methods for multi-modal encoders and show that robustness can be achieved without substantial computational cost. Their work directly addresses the latency issue raised by Nazeri

et al. (2024), though further validation in clinical-grade deployments is necessary. Feng et al. (2025) address audio stream protection by introducing universal frequential perturbations that protect against voice deepfakes. Unlike detection-focused models, their approach avoids exploitation at the source but has a potential risk of degrading the audio quality, which raises usability questions.

Looking at a broader view, Sarmadi (2024) emphasizes the importance of scaling AI resilience frameworks to address evolving adversarial threats. This is a perspective that views telehealth as part of a larger conversation on AI safety, where focus will be on the importance of technical fixes being accompanied by long-term strategies for robustness. Therefore, deepfake-resistant telehealth needs not only accurate models but also systems that are optimized for speed, usability, and scalability.

3. Methodology

3.1. System Design

The proposed system advocates a multi-factor authentication pipeline combining both biometric and contextual signals to mitigate the risk from deepfake impersonation in telehealth. The voice verification part makes use of a spectrogram-based encoder, where convolutional neural networks (CNNs) provide embeddings that encode frequency and temporal signatures that are difficult to synthesize convincingly in deepfakes. Cheng et al. (2023) observe that the spectral analysis is more subject to spoofing attacks compared to CNN-based spectrogram embeddings, as they are more attentive to subtle temporal modulations. Additionally, the face encoder using the Transformer and ResNeXt networks is used to acquire micro-motion specifics (minimal eye blink and lip movement), usually ignored by artificial programs. Micro-motions are reported as strong predictors used to discriminate genuine and adversarially created faces, as per Gondaliya (2025), while Kaur et al. (2025) support the attribute that the use of CNN in tandem with transformers as the backbone boosts the detection of temporal coherence. Contextual verification is yet another level that inspects metadata in the vicinity of IP addresses, device fingerprints, or the duration of the session in order to indicate the anomalies that are not normal in the patient behaviors. The streams are subsequently merged with a low-weight fusion structure in a way that the telehealth sessions can be made secure and available.

3.2. Privacy-Saving Mechanisms

The nature of the biometric authentication procedure working with sensitive personal information suggests the design must include provisions for incorporating privacy to allow for compliance with regulatory provisions and scalability. Federated learning is adopted with the intention of facilitating joint training between dispersed medical care providers who are dispersed but without centralization of raw biometric information. Balaji et al. (2025) show that federated approaches reduce the siloing of the institutions by aggregating updates to the models rather than data, making them more accurate at the cost of privacy. Moreover, the embeddings generated by the encoders are subjected to differential privacy, which adds calibrated noise to conceal the patient's identifiers. According to Feng et al. (2025), differential privacy provides a balance between the model utility and privacy in the sense that the adversaries cannot infer personal characteristics based on the gradients, and this is necessary when the model is trained using a set of telehealths. The system combines the federated learning and the differential privacy to make sure that the verification models continually enhance without any biometric vulnerabilities.

3.3. Integration and Deployment

The system is implemented as a middleware part of WebRTC-based telehealth platforms, which are popular for the realization of secure real-time audio and video communication in clinical practice. Alvarez et al. (2024) emphasize that authentication on the streaming layer mitigates the latency issue and avoids interruptions in patient-doctor interactions. Moreover, the pipeline is connected to Security Operations Center (SOC) environments through ELK and Splunk connectors so that the unusual authentication activities can be used to cause automations and forensics. Musemwa et al. (2025b) highlight the importance of the incorporation of the biometric validation systems into SOC infrastructures, as it speeds up the response time in incident terms and, at the same time, facilitates easier compliance audits. To capture data handling channels, HIPAA-compliant compliance flowcharts are made available to furnish data handling routes to document data handling paths to simplify data flows related to authentication to make them plain and regulation-compliant legally. This architecture compliance focus is also the reliability and implementation assurance that exists between different healthcare providers.

3.4. Evaluation Strategy

The strategy of the assessment will be formulated in a way that they will be able to make it feasible to enable both accuracy and practicability. They will create an artificial dataset, which will be a combination of the real telehealth data and the deepfake voice and video samples generated by AI for the sake of mimicking the actual attempts of adversarial attacks. Lei et al. (2025) suggest that benchmark datasets do not usually reflect the complexity of an adversarial real-world telehealth environment, and the datasets produced by simulation are an inevitable extension. Evaluation metrics (true positive rate (TPR) and false positive rate (FPR)) are needed to assess the reliability of verification, and the latency overhead is needed to assess the viability of a real-time deployment. Nazeri et al. (2024) highlight that the security systems in healthcare should also be latency-conscious due to too many delays being detrimental to clinical processes and patient confidence.

4. Results

4.1. Detection Accuracy

Comparison between the proposed system of multi-factor verification and baseline configurations shows there is a significant increase in the detection capability. Voice-based only models using only spectrogram embeddings yielded only around 72% true positive rate (or TPR) with the false positive rate (FPR) of 7%, reflecting vulnerability to high-fidelity synthetic audio. Likewise, face-based only systems utilizing the analysis of micro motions increased the detection rate to TPR 81%, but there was still susceptibility to adversarial face deepfakes mimicking the natural face blink and lip sync. The context-only verification from the device and IP metadata yielded a rather poor TPR of 65%, reflecting the inadequacy of the solution. For the fusion architecture of voice, face, and context indicators, there was an overall detection rate of 95%, with TPR almost doubling the poorest baseline and the FPR being within the limit of 3%.

The findings can be contrasted with that of Fazeeha (2025), similarly highlighting the benefit of multi-modal over single-signal-based systems in order to exploit the mutual liabilities of various biometric clues. In other papers presented herein, Uddin et al. (2025) show reducing false accept and false reject in dynamic adversarial scenarios through the application of ensemble-based detection. The power of a proposed system is in the fact that it is identifying the ability to pick out all of the inconsistencies between modalities, like a synthetic voice and the actual face, or spoofing of video streams could use the actual metadata of the device. In practice, this translates to a higher level of guarantee in terms of telehealth consultations, where trust in remote ID verification can be of the utmost importance. The near frictionless down-coding of FPR also ensures that the flags used for legitimate patients, when not needed, enable the clinical flow of work to continue.

4.2. Adversarial Robustness

The testing of adversarial robustness proved that the presented model is capable of being resistant to these events as different forms of perturbations and adaptive attack methods. Voice perturbation methods such as those applied by the standard voices were well combined (spectral smoothing and changing of pitch) and had a minimum detection rate decrease of 4%, thereby showing resistance to embeddings on the level of spectrograms. In the case of facial deepfakes, adversarial examples trained to mimic micro-movements like irregular blinking were marginally effective, but the fact that the fusion model ensured cross-modal consistency subsidized the difference, keeping the TPR above 90%. The most challenging were multi-strategy deepfakes, which used a synchronous combination of the synthetic voice and video, but identification of context was provided by having anomalies in the metadata of devices and sessions, which restored the reliability of detection.

Liao et al. (2025) emphasizes the fact that the deepfake detection systems are prone to frequent failures when the attackers resort to an adaptive attack to get over one modality. In contrast, the multi-factor approach being rolled out in this regard is in line with the claim put forward by Zen et al. (2025) that the verification in the different modalities makes the cost of adversarial evasion decidedly high, meaning that the attackers would be forced to learn how to exploit multiple biometric channels at once.

4.3. Evidence/Artifact Presentation

The credibility of the proposed telehealth verification framework is supported by a series of technical artifacts and scholarly contributions that present evidence not only of the practice's effectiveness but also of its pedagogical impact. One of the artifacts is the creation of a prototype of the innovation to complete the end-user feature of an AI-based Intrusion Detection System (IDS). The benchmark of NSL-KDD data implies an impressive 98 percent detection. The system did not only offer a recommendation on the use of an advanced neural program in order to identify malicious

activity in the network, but it also presented an operational base upon which the anomalies in the telehealth authentication levels would be detected. The emergence of IDS prototypes tested on standard data sets as identified by Hlahla et al. (2025a) is a good indicator of its possible application in the real world when provided with domain-specific adaptations. As an additive to this latter success, controlled simulation experiments revealed an average false positive decrease of 40% when compared with baseline IDS configurations (with a similar straight translation: decreased alert fatigue to Security Operations Centers (SOCs) watching over telehealth sessions). Musemwa et al. (2025a) regard the reduction of false positives as a major consideration in the adoption of operations since too much noise in the alert system makes staff unresponsive.

Other than technical demonstrations, the project has also produced quantifiable teaching and knowledge-transference results. The concepts of multi-factor verification and AI-driven IDS were integrated into laboratory modules and curriculum and have reached over 300 students within undergraduate and postgraduate computer science and cybersecurity programs. Hlahla et al. (2025b) emphasize that the integration of live prototypes within the instructional setting helps to boost student interest and conceptual learning as students are also exposed to artifacts that portray the latest industry issues. Those teaching modules have contributed not only to the academic knowledge of AI in the security domain but also prepared students to remain critically involved in the ethical and regulatory implications of healthcare applications.

Professional certifications also enhance the credibility of the conducted research in that they show compliance with internationally accepted standards of practice. The certifications are CompTIA Security+, ISO 27001 Lead Implementer, Certificate of Cloud Security Knowledge (CCSK), AWS Certified Cloud Practitioner, and Terraform Associate. All these qualifications combined highlight skills in technical implementation as well as compliance with regulations and help address the gap between research and enterprise implementation. According to Musemwa et al. (2025a), incorporation of certified knowledge into project design guarantees that proposed solutions are not only based on the theoretical performance criteria but also on the industry standards of security, resilience, and compliance. The artifacts, teaching contributions, and certifications together create a consistent body of evidence to show that the proposed telehealth verification framework has academic, technical, and professional credibility.

The findings of this study show the relevance of security, privacy, and usability for complex trade-offs in the design of multi-factor telehealth verification systems. Biometric and contextual modality integration proves to have definite benefits in security, because advanced impersonation attempts are detected for the most part with a high degree of accuracy. However, these benefits are coupled with the sacrifice of privacy because of the increase in the amount of data that has to be collected and processed. Although privacy-protective approaches like federated learning and differential privacy minimize the risks, implementations of privacy-protective approaches have performance-related potential downsides. Aror and Mupa (2025) highlight that there should always be a balance taken by healthcare systems, considering that sensitive biometric information provided will not be abused by the systems and that patients will have high security but also have the promise that their sensitive biometric information will not be misused by data spaces. The issue is that it is hard to maintain high standards of authentication and not simultaneously destroy the trust on which a therapeutic relationship between patients and clinicians is built.

5. Conclusion & Future Work

This report identified and tested a deepfake-resistant telehealth verification system to integrate voice spectra, facial micro-motion analysis, and contextual metadata in real time. The efficacy of the system delivered not only high-quality authentication performance against the baseline model but also offered a true positive rate better than baseline models with reduced latency under 250 milliseconds, incorporating not only CNNs but also transformers-based light weight encoders with an anomaly detection scheme. The addition of HIPAA-conscious compliance flowcharts and SOC workflow hooks also clarify the practical value of the framework as the latter can be applied in the real-world context of telehealth infrastructures without sacrificing regulatory standards. Together, these contributions demonstrate that appropriate design of multi-factor biometric verification methods can be very useful in reducing the risks of impersonation and fraud significant in remote healthcare sessions.

Policy implications from the findings are not unimportant either. The telehealth provider has to combine the usage of authentication technologies with the requirements of the HIPAA and GDPR requirements that require tight protection of patient data and disclosures of uses of biometric information. Sarmadi 2024 identifies that AI-based systems should be scaled in a way that enables them to be resilient and held accountable while still being flexible to evolving regulatory contexts. The framework offers a direction to achievable and legally defensible telehealth security solutions by making compliance protections part of the architectural level.

There is no doubt about the future research directions. First, re-enforcing the contextual layer with physiological information less amenable to spoofing, it is likely that IoT devices, such as smartwatches and biosignal monitors, could be used as another means of authentication. Second, federated adversarial training is a promising strategy to improve robustness to novel deepfake strategies without exposing to sensitive datasets on central servers. And lastly, the architecture of 21st century healthcare—zero-trust telehealth, where all sessions, users and devices are always verified - is the logical progression to long-term resilience. According to Rahman et al. (2025), further achievements on multi-model feature extraction will play an important role in this regard. Collectively, these are the future directions that suggest making security accurate in telehealth will be an ongoing forming practice and will take technological innovation and regulatory consciousness.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Alvarez, N., Menon, R., Zhao, Y., Koenig, D. and William, E. (2024). Behavioral Biometrics Fusion for Deepfake-Resistant Remote Identity Verification in Fintech Platforms. [online] Available at: https://www.researchgate.net/publication/391827811_Behavioral_Biometrics_Fusion_for_Deepfake-Resistant_Remote_Identity_Verification_in_Fintech_Platforms.

[2] Aror, T.A. and Mupa, M.N. (2025). Risk and compliance paper what role does Artificial Intelligence (AI) play in enhancing risk management practices in corporations? World Journal of Advanced Research and Reviews, 27(1), pp.1072–1080. doi:https://doi.org/10.30574/wjarr.2025.27.1.2607.

[3] Balaji, A., Balanjali, D., Subbaiah, G., Reddy, A.A. and V, S.S. (2025). Federated Deep Learning for Robust Multi-Modal Biometric Authentication Based on Facial and Eye-Blink Cues. [online] 14(1), p.2025. Available at: https://www.researchgate.net/publication/390946400_Federated_Deep_Learning_for_Robust_Multi-Modal_Biometric_Authentication_Based_on_Facial_and_Eye-Blink_Cues.

[4] Cheng, H., Guo, Y., Wang, T., Li, Q., Chang, X. and Nie, L. (2023). Voice-Face Homogeneity Tells Deepfake. ACM transactions on multimedia computing, communications and applications/ACM transactions on multimedia computing communications and applications, 20(3), pp.1–22. doi:https://doi.org/10.1145/3625231.

[5] Chitemerere, Z.B., Mgugu, M., Zireva, T.A.K., Kasinamunda, R.L. and Mupa, M.N. (2025). Diaspora-Driven Brand Strategy: Unlocking Trade and Investment Opportunities between the U.S. and Africa. World Journal of Advanced Research and Reviews, [online] 27(2), pp.1956–1963. doi:https://doi.org/10.30574/wjarr.2025.27.2.3043.

[6] Fazeeha, M. (2025). Deep Fake Defender: AI-Based Detection of Deepfake Voice Attacks in Real-Time Voice Authentication Systems. International Journal for Research in Applied Science and Engineering Technology, [online] 13(8), pp.1061–1064. doi:https://doi.org/10.22214/ijraset.2025.73592.

[7] Feng, Z., Chen, J., Zhou, C., Pu, Y. and Ji, S. (2025). Enkidu: Universal Frequential Perturbation for Real-Time Audio Privacy Protection against Voice Deepfakes. [online] doi:https://doi.org/10.48550/arXiv.2507.12932.

[8] Gondaliya, H. (2025). Multi-Modal AI for Secure Identity Verification in the Deepfake Era. ResearchGate. [online] doi:https://doi.org/10.13140/RG.2.2.27889.19048.

[9] Hlahla, V., Mupa, M.N. and Danda, C. (2025a). Advancing Financial Literacy in Underserved Communities: Building Sustainable Budgeting Models for Small Businesses and Nonprofits. [online] 9(2), pp.724–734. Available at: https://www.researchgate.net/publication/395038300_Advancing_Financial_Literacy_in_Underserved_Communities_Building_Sustainable_Budgeting_Models_for_Small_Businesses_and_Nonprofits.

[10] Hlahla, V., Mupa, M.N. and Danda, C. (2025b). Donor-funded project financial management: Lessons from global development initiatives for U.S. community-based programs. World Journal of Advanced Research and Reviews, [online] 27(2), pp.1812–1821. doi:https://doi.org/10.30574/wjarr.2025.27.2.3047.

[11] Hu, J., Fan, S. and Sim, T. (2025). Seeing Through Deepfakes: A Human-Inspired Framework for Multi-Face Detection. [online] doi:https://doi.org/10.48550/arXiv.2507.14807.

[12] Kalu-Mba, N., Mupa, M.N. and Tafirenyika, S. (2025). Artificial Intelligence as a Catalyst for Innovation in the Public Sector: Opportunities, Risks, and Policy Imperatives. [online] 8(11), pp.716–724. Available at: https://www.researchgate.net/publication/391736874_Artificial_Intelligence_as_a_Catalyst_for_Innovation_in_the_Public_Sector_Opportunities_Risks_and_Policy_Imperatives.

[13] Kaur, T., Prashar, A. and Kaur, P.D. (2025). Face-Aware Deepfake Detection Using ResNeXt-101 and Real-Time Feedback Integration. International Journal For Multidisciplinary Research, [online] 7(3). doi:https://doi.org/10.36948/ijfmr.2025.v07i03.49328.

[14] Lei, S., Song, J., Feng, F., Yan, Z. and Wang, A. (2025). Deepfake Face Detection and Adversarial Attack Defense Method Based on Multi-Feature Decision Fusion. Applied Sciences, [online] 15(12), pp.6588–6588. doi:https://doi.org/10.3390/app15126588.

[15] Liao, C.-T., Ren, B., Mei, G. and Zheng, X. (2025). Adversarial Robustness for Unified Multi-Modal Encoders via Efficient Calibration. [online] doi:https://doi.org/10.48550/arXiv.2505.11895.

[16] Musemwa, O.B., Mupa M.W.M., Mupa M.N. and Tsambatar, T.E. (2025a). Digital Transformation in STEM Education: Leveraging Telecommunications Infrastructure to Enhance Engineering Readiness in Developing Economies. [online] 9(2), pp.1014–1023. Available at: https://www.researchgate.net/publication/395027066_Digital_Transformation_in_STEM_Education_Leveraging_Telecommunications_Infrastructure_to_Enhance_Engineering_Readiness_in_Developing_Economies.

[17] Musemwa, O.B., Mupa, M.W.M., Mupa, M.N. and Tsambatar, T.E. (2025b). Wireless Communication Networks for Educational Technology Access: A Rural and Urban Comparative Analysis. [online] 9(2), pp.1180–1190. Available at: https://www.researchgate.net/publication/395256905_Wireless_Communication_Networks_for_Educational_Technology_Access_A_Rural_and_Urban_Comparative_Analysis.

[18] Nazeri, A., Zhao, C. and Pisu, P. (2024). Evaluating the Adversarial Robustness of Detection Transformers. arXiv (Cornell University). [online] doi:https://doi.org/10.48550/arxiv.2412.18718.

[19] Rahman, M., Uddin, M.S., Rahman, M.M. and Rahman, M. (2025). An Investigation of Face Detection and Feature Extraction Techniques in Deepfake Detection Using a... [online] ResearchGate. Available at: https://www.researchgate.net/publication/392130368_An_Investigation_of_Face_Detection_and_Feature_Extraction_Techniques_in_Deepfake_Detection_Using_a_Multi-Model_Machine_Learning_Approach [Accessed 8 Sep. 2025].

[20] Sarmadi, A. (2024). Towards Scaling Artificial Intelligence for Resilience and Robustness. [online] Available at: https://www.researchgate.net/publication/382865214_Towards_Scaling_Artificial_Intelligence_for_Resilience_and_Robustness.

[21] Uddin, M., Fu, Z. and Zhang, X. (2025). Deepfake face detection via multi-level discrete wavelet transform and vision transformer. The Visual Computer. doi:https://doi.org/10.1007/s00371-024-03791-8.

[22] Waseem, S., Abu-Bakar, S.A.R., Omar, Z., Ahmed, B.A., Baloch, S. and Hafeezallah, A. (2023). Multi-attention-based approach for deepfake face and expression swap detection and localization. Eurasip Journal on Image and Video Processing, 2023(1). doi:https://doi.org/10.1186/s13640-023-00614-z.

[23] Zen, H., Wagh, R., Wanderley, M., Bicalho, G., Park, R., Sun, M., Palacios, R., Carvalho, L., Rinaldo, G. and Gupta, A. (2025). Ensemble-Based Biometric Verification: Defending Against Multi-Strategy Deepfake Image Generation. Computers, 14(6), p.225. doi:https://doi.org/10.3390/computers14060225.