Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/22228
DOI URL: http://dx.doi.org/10.21474/IJAR01/22228

ISSN NO. 2320-5407

RESEARCH ARTICLE

HYBRID REALNESS VERIFICATION SYSTEM

Gonugunta Harshitha[1], Ilindra Krishna Lekha[1], Kandimalla Lakshmi Sruthi Laya[1], Ragham Jitesh[2], Ekabathini Chanchu Suresh[2] and Kommu Pramod[3]

1. UG Student, Department of CSE, QIS College of Engineering and Technology (A), Ongole, Andhra Pradesh, India.
2. UG Student, Department of IoT, QIS College of Engineering and Technology (A), Ongole, Andhra Pradesh, India.
3. UG Student, Department of ECE, QIS College of Engineering and Technology (A), Ongole, Andhra Pradesh, India.

Manuscript Info

Manuscript History
Received: 04 September 2025
Final Accepted: 06 October 2025
Published: November 2025

Key words:-
Deepfake detection, hybrid verification, media authenticity, AI-generated content, multimodal analysis, voice forensics, synthetic media identification.

Abstract

With the rapid advancement of artificial intelligence, the creation of synthetic media, commonly known as deepfakes, has become increasingly realistic and accessible. This growing sophistication poses significant threats to digital security, media authenticity, and online trust. To counter these challenges, this work presents a Hybrid Realness Verification System (HRVS) — a multimodal deepfake detection framework designed to assess the authenticity of image, video, and audio data through domain-specific machine learning techniques.For visual analysis, HRVS integrates a vision transformer-based model to identify subtle inconsistencies such as facial distortion, unnatural texture patterns, and absence of micro-expressions. Audio verification is achieved by transforming voice signals into Mel spectrograms and evaluating them using a convolutional neural network trained to detect artifacts of AI-generated speech, including irregular pitch, rhythm, and tonal energy.The individual predictions from these modules are combined through a fuzzy logic-based decision engine, enabling the system to manage uncertain or borderline cases with adaptive reasoning instead of fixed thresholds. The overall architecture is implemented on a Flask backend, supporting real-time user interaction, scalability, and seamless integration.By fusing multimodal feature analysis with intelligent decision-making, HRVS enhances detection reliability and interpretability. Its applications span security verification, content moderation, digital forensics, and identity authentication, contributing to a safer and more trustworthy digital ecosystem.

"© 2025 by the Author(s). Published by IJAR under CC BY 4.0. Unrestricted use allowed with credit to the author."

Introduction:-

This template, modified in MS Word 2007 and saved as a “Word 97-2003 Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided through The rapid evolution of artificial intelligence has significantly transformed how digital media is created and consumed. Among these advancements, deepfake technology—capable of generating highly convincing synthetic images, videos, and audio—has emerged as both an innovation and a threat. While it has potential applications in entertainment, education, and accessibility, its misuse has raised major concerns regarding privacy, misinformation, and digital trust. Deepfakes blur the line between authentic and fabricated content, challenging the credibility of online information and undermining confidence in visual and auditory evidence.

Traditional deepfake detection approaches often focus on a single modality, such as image or audio analysis. Although these methods can identify basic inconsistencies, they struggle against advanced multi-modal manipulations where generated content maintains coherence across multiple media types. The increasing sophistication of such forgeries demands a more integrated and adaptive detection strategy.

To address this gap, we propose the Hybrid Realness Verification System (HRVS) — a comprehensive multi-modal framework designed to verify the authenticity of visual and auditory content. HRVS employs a vision transformer-based model to identify subtle visual distortions, including unnatural facial expressions and irregular texture patterns. For audio evaluation, it utilizes a convolutional neural network trained on Mel spectrogram representations to detect artifacts indicative of synthetic speech.

The outputs from these models are then combined using a fuzzy logic-based decision engine, which evaluates ambiguous or uncertain cases more effectively than rigid thresholding techniques. The complete system is deployed on a Flask-based web platform, enabling real-time user interaction and scalable deployment for practical use.

By integrating machine learning across multiple data modalities and incorporating adaptive reasoning, HRVS enhances the reliability, accuracy, and interpretability of deepfake detection. The system presents a promising solution for applications in digital forensics, content moderation, cybersecurity, and identity verification.

Objective:-

The primary objective of this research is to develop an intelligent and multimodal framework, termed the Hybrid Realness Verification System (HRVS), for detecting deepfakes and verifying the authenticity of multimedia content, including images, videos, and audio. The system employs advanced machine learning techniques—specifically, vision transformers for visual content and convolutional neural networks for audio analysis—to capture subtle indicators of synthetic manipulation.

Additionally, HRVS integrates a fuzzy logic-based decision engine to combine the results from each modality, enabling more adaptive and reliable interpretation of uncertain or conflicting cases. By overcoming the limitations of single-modality detection, this research aims to deliver a scalable and efficient solution for real-time identity verification, media authenticity assessment, and mitigation of deepfake-driven cybersecurity risks.

Literature Survey:-

A review of recent research underscores the need for multi-modal biometric systems. Numerous works have focused on individual authentication modalities with notable success. For instance, Fatima M. Salman and Samy S. Abu-Naser implemented convolutional neural networks to distinguish between real and fake human faces, achieving over 99% accuracy.

Similarly, Suk-Young Lim and colleagues developed a CNN-LSTM-based framework augmented with Explainable AI techniques for detecting AI-generated voices, demonstrating substantial improvements in recognition reliability. C. Ranjeeth Kumar et al. explored the use of Siamese Networks and CNN-KNN hybrid models to improve facial recognition performance under varying conditions, such as poor lighting or unconventional poses. While these methods offer strong performance within their respective domains, their single-modality nature leaves them vulnerable to cross-domain attacks.

Multi-modal systems, on the other hand, enhance resilience by introducing redundancy and contextual cross-verification. Integrating facial and voice recognition has been shown to significantly reduce false acceptance rates. Behavioral biometrics, though still emerging, add a subtle yet powerful layer of verification by capturing individual usage patterns that are difficult to imitate precisely. The HRVS builds upon these insights, incorporating each of these techniques into a unified, adaptive platform that maximizes security without compromising efficiency.

Existing System:
Current deep fake detection systems typically rely on single-modal analysis, focusing solely on either visual or audio inputs. Image-based systems use convolutional neural networks (CNNs) or similar models to identify facial inconsistencies, such as unnatural textures, distorted features, or irregular lighting, while audio detection tools analyse waveform patterns or spectrograms to capture synthetic voice characteristics. Although effective in controlled conditions, these systems often struggle in real-world scenarios due to limitations like varying quality, compression artifacts, and adversarial attacks designed to bypass detection.

Moreover, many existing tools rely on fixed thresholds for classification, which can lead to false positives or negatives when input data lies in a grey area. This rigid decision-making approach lacks adaptability and fails to account for subtle, ambiguous cases commonly found in sophisticated deep fakes. Additionally, most systems are not designed to integrate results from multiple input types, making them vulnerable when deep fakes span across visual and auditory modalities. As a result, the reliability and accuracy of these standalone systems remain a challenge, especially in high-stakes applications such as biometric authentication, video conferencing, or digital evidence verification.

Proposed Methodology:-
The proposed Hybrid Realness Verification System (HRVS) introduces a multi-modal approach that integrates image, video, and audio analysis to improve the accuracy and robustness of deep fake detection. This methodology begins with a user uploading a media file—either an image, video, or audio—through a web interface. Based on the media type, the system intelligently routes the file to the corresponding detection module.

For video inputs, the system extracts several frames and analyses each using a Vision Transformer-based image classifier trained specifically to detect facial manipulations. This model looks for subtle cues of tampering, such as inconsistent skin textures, unnatural eye movements, and distorted facial geometry. The same model is utilized for standalone images, ensuring uniform detection across static and dynamic visual inputs.

Audio inputs are first transformed into Mel spectrograms, capturing the time-frequency representation of the voice signal. These spectrograms are then fed into a Convolutional Neural Network (CNN) trained to identify signs of synthetic speech. The model focuses on identifying discrepancies in pitch, tone, rhythm, and energy, which are common in AI-generated voices.

Each detection module generates a binary classification—real or fake—alongside a confidence score. These outputs are then processed by a fuzzy logic-based decision engine. Unlike rigid threshold systems, the fuzzy logic engine interprets borderline or conflicting results with a degree of flexibility, allowing for more nuanced and context-aware decisions. For instance, if both visual modules indicate high realness but the audio module shows uncertainty, the system can still confidently classify the media using logical inference and adaptive weighting.

The entire backend is developed using Flask to manage user-uploaded files and production of results. By combining specialized models for each modality and incorporating a flexible decision-making layer, the HRVS significantly enhances the accuracy and reliability of media authenticity verification.

Implementation:-
The implementation of the Hybrid Realness Verification System (HRVS) involves a modular architecture where each media input type—image, video, or audio—is handled by specialized detection components. The backend of the system is built using the Flask web framework, offering a lightweight yet scalable structure for managing API calls, user interactions, and routing of uploaded files.

Upon uploading, the system first determines the type of input. For video files, multiple frames are extracted using frame sampling techniques at fixed intervals. These frames are processed through a Vision Transformer (ViT)-based classifier. This model has been trained on a large dataset of real and manipulated faces, enabling it to detect subtle tampering features like abnormal skin texture, unnatural facial movements, or loss of micro-expressions.

For image inputs, the same ViT-based model is directly used. The consistency in model usage ensures standardization in detection, regardless of whether the input is a still image or a frame from a video.

Audio files are processed by first converting them into mel spectrograms, a visual representation of the voice signal over time. These spectrograms are passed through a pre-trained Convolutional Neural Network (CNN) that has been specifically tuned to identify synthetic speech.

The model focuses on anomalies such as robotic tonal patterns, unnatural pauses, or energy spikes that often exist in AI-generated audio. Each of these modules provides a binary classification (real or fake) along with a confidence score. These results are then forwarded to a Fuzzy Logic Decision Engine. The fuzzy logic system is designed using rule-based inferences that evaluate the degree of realness from each module. This allows the system to handle uncertainty and conflicting results better than a traditional threshold-based approach.

The frontend of the system is developed with HTML, CSS, and JavaScript, offering a user-friendly interface to upload files and view results. As soon as the decision engine finalizes the classification, the result is displayed to the user along with confidence percentages from each module. This modular and intelligent implementation makes HRVS a powerful and practical solution for detecting deep fakes in real-time, enhancing trust in digital media verification systems

Results:-

The performance of the Hybrid Realness Verification System (HRVS) was evaluated using a comprehensive dataset of both real and synthetic media. The system was tested with a range of inputs, including images, videos, and audio files, to assess its effectiveness in detecting deep fakes across multiple modalities.

For image and video inputs, the Vision Transformer (ViT)-based classifier demonstrated high accuracy in identifying manipulated faces and detecting anomalies. The system was able to recognize subtle artifacts such as inconsistent skin texture, unnatural facial expressions, and facial alignment distortions.

In initial tests, the image classification module achieved an accuracy of approximately 92% in detecting real and fake faces. Video frame extraction and subsequent analysis resulted in similarly high accuracy, with the system successfully identifying deep fake videos in over 90% of the cases, even in challenging scenarios such as low-resolution inputs or slight alterations in facial features.

For audio inputs, the Convolutional Neural Network (CNN) trained on mel spectrograms showed impressive performance in identifying synthetic voices. The system detected discrepancies in pitch, rhythm, and tonal energy associated with AI-generated speech.

The CNN classifier achieved an accuracy rate of 88% for differentiating between real and fake audio, with a notable ability to flag audio that exhibited unnatural speaking patterns or robotic-sounding intonations. Despite some ambiguity in cases involving high-quality voice clones, the system maintained an overall reliability rate of 85% in identifying synthetic voices.

The fuzzy logic decision engine played a pivotal role in integrating the results from all three detection modules (image, video, and audio). This component was critical for resolving conflicts between modules and improving overall decision-making, especially in cases where one module gave a slightly ambiguous result.

The fuzzy logic system effectively balanced the individual confidence scores from each module, ensuring that even borderline cases were handled with high accuracy. Overall, the system's final classification of media as "real" or "fake" had a reliability rate of around 90%.

The system was also evaluated in terms of its processing time. For typical media uploads, including videos of up to 5 minutes in length, the HRVS processed and classified the content in under 10 seconds. Audio files were processed slightly faster, with classification times averaging around 4 seconds.

These performance results ensure that HRVS can be deployed in real-time applications, such as social media platforms and digital forensics, without significant delays. The user interface was found to be intuitive, with users able to easily upload media, view results, and understand the reasoning behind the classification through confidence scores. The system's feedback mechanism was also praised for its transparency, providing users with clear insights into why media was flagged as real or fake.

In summary, the HRVS achieved high accuracy in detecting both deep fake images, videos, and audio, with solid results across all modules. The integration of fuzzy logic significantly enhanced the overall decision-making process, allowing the system to handle complex and uncertain cases. These results demonstrate the potential of HRVS as a reliable tool for real-time deep fake detection and media authenticity verification.

Conclusion:-

The Hybrid Realness Verification System (HRVS) presents a comprehensive and innovative approach to detecting deep fakes across multiple media types, including images, videos, and audio. By leveraging state-of-the-art machine learning models, such as Vision.

Transformers for image and video analysis and Convolutional Neural Networks for audio classification, the system is able to accurately identify manipulated content. The integration of fuzzy logic for decision-making enhances the system's flexibility, allowing it to resolve conflicts and handle ambiguous cases with high reliability.

The results demonstrate that HRVS is highly effective in distinguishing between real and fake media, achieving impressive accuracy rates across all modalities. The system's ability to process and analyse media in real-time makes it suitable for deployment in applications that require immediate verification of content, such as social media platforms, digital forensics, and security systems. Additionally, the user-friendly interface ensures that users can easily interact with the system and obtain valuable insights into the authenticity of the media.

While the system shows significant promise, there are areas for future improvement, such as enhancing the audio classification module to better handle high-quality synthetic voices and expanding the system's ability to detect new forms of deep fakes. Nonetheless, HRVS serves as a valuable tool in the ongoing battle against digital manipulation, offering a robust and scalable solution for ensuring the authenticity of multimedia content.

In conclusion, the Hybrid Realness Verification System contributes significantly to the field of deep fake detection and media verification. Its combination of cutting-edge machine learning techniques and intelligent decision-making ensures high accuracy and reliability, positioning it as a powerful tool for safeguarding digital media and combating the growing issue of synthetic content.

References:-

1. P. Yaswanthram and B. A. Sabarish. (2022) "Face Recognition Using Machine Learning Models - Comparative Analysis and impact of dimensionality reduction," IEEE Fourth International Conference on Advances in Electronics, Computers, and Communications (ICAC): pp. 1-4,.
2. M. Nehru and S. Padmavathi. (2017)"Illumination invariant face detection using viola jones algorithm," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS): pp 1-4.
3. A.S. L. Sravya, K. S. Chinmayi, L. N. Yoshita, S. V and P. V. Manitha,(2019) "A Smart Aid for Visually Challenged in Face Detection and Recognition," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT): pp. 704-708.
4. Shahroz Tariq, Sangyup Lee, Hoyoung Kim, Youjin Shin, Simon S. Woo, Detecting Both Machine and Human Created Fake Face Images In the Wild, Proceedings of the 2nd International Workshop on Multimedia Privacy and Security pages 81-87, October 15-19, 2018, New York, NY, USA. Retrieved from (https://dl.acm.org/citation.cfm?id=3267367).
5. Yaojie Liu, Amin Jourabloo, Xiaoming Liu; Learning Deep Models for face Anti-Spoofing: Binary of Auxiliary Supervision; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp.

389 398.Retrieved from (http://openaccess.thecvf.com/content_cvpr_2018/html/Liu_Learning_Deep_Models_CVPR_2018_paper.html).
6. Youngshin Kim, Jaekeun Na, Seongbeak Yoon, and Juneho Yi. Masked aments. JOSA A, 2009. Retrieved from(https://www.researchgate.net/publication/24251875_Masked_fake_face_detection_using_radiance_measurements).
7. X. Lv, M. Li, Application and research of the intelligent management system based on Internet of things technology in the era of big data, Mobile Inf. Syst. 2021 (16) (2021) 1–6.
8. M.S. Maggu, Image-based face detection and recognition, Int.J.Mod.Trend.Sci. Technol. 6 (12) (2021) 466–470.
9. T.T. Nguyen, L.E. Son-Thai, L.E. Van-Thuy, Adaptive hyperparameter for face recognition, Int. J. Innovative Technol. Explor. Eng. 10 (2) (2021) 116–119.
10. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 815-823).
11. Grother, P., Ngan, M., & Hanaoka, K. (2019). Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. National Institute of Standards and Technology.
12. Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1701-1708).
13. Abhilash Bhise, Radhika Khichi, Amol Korde and Prof. Deepali Lokare, 'Atten- dance System Using NFC Technology with Embedded Camera on Mobile Device ,' B hise, Khichi, Korde,Lokare, 2015.
14. Aparna Trivedi*1, Chandan Mani Tripathi2, Dr. Yusuf Perwej3, Ashish Kumar Srivastava4, Neha Kulshrestha5, 'Face Recognition Based Automated Attendance Management System' International Journal of Scientific Research in Science and Technology Print ISSN: 2395-6011.
15. Devendra Kumar Yadav , Sumit Singh , Prof. Shashank Pujari , Pragyan Mishra, 'Fingerprint Based Attendance System Using Microcontroller and LabView,' (Ku- mar Yadav, Singh, Pujari, Mishra, 2015).
16. Hoque, M. A., Islam, T., Ahmed, T., & Amin, A. (2020, March). Autonomous face detection system from realtime video streaming for ensuring the intelligence security system. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 261-265). IEEE.
17. Das, A., Ansari, M. W., & Basak, R. (2020, December). Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV. In 2020 IEEE 17th India Council International Conference (INDICON) (pp. 1-5). IEEE.
18. Khan, M., Chakraborty, S., Astya, R., & Khepra, S. (2019, October). Face Detection and Recognition Using OpenCV. In 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 116-119). IEEE.
19. Abadi, M., Barham, P., Chen, J., & Chen, Z. (2016). TensorFlow: A system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016.
20. Chakraborty, S., & Das, D. (2014). An Overview Of Face Liveness Detection. International Journal on Information Theory (IJIT), Vol.3.
21. Pratap, K., Priya, A., & Mani, G. (2019). Technologies to overcome spoofing attack in facial recognition. International Journal of Recent Technology and Engineering, 3.
22. Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Ferrer, C.C. The deep fake detection challenge (dfdc) dataset. arXiv 2020, arXiv:2006.07397.
23. Meng, H.; Yan, T.; Yuan, F.; Wei, H. Speech emotion recognition from 3D log-mel spectrograms with deep learning network. IEEE Access 2019, 7, 125868–125881.
24. Variani, E.; Lei, X.; McDermott, E.; Moreno, I.L.; Gonzalez-Dominguez, J. Deep neural networks for small footprint text-dependent speaker verification. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 4052–4056.
25. M.Murtaza, M. Sharif, M. Raza, and J. Shah, "Face recognition using adaptive margin Fisher's criterion and linear discriminant analysis," Me International Arab Journal of Information Technology, vol. 11, no. 2, pp. 1–11, 2014.
26. S. Vignesh, K. V. S. N. L. Manasa Priya, and S. S. Channappayya, "Face image quality assessment for face selection in surveillance video using convolutional neural networks," in Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, Orlando, FL, USA, December 2015.
27. S. Umer, B. Chandra Dhara, and B. Chanda, "Face recognition using fusion of feature learning techniques," Measurement, vol. 146, 2019.
28. S. G. Beale and V. H. Mankar, "A Review Paper on Face Recognition Techniques," Int. J. Adv. Res. Comput. Eng. Technol., vol. 1, no. 8, pp. 2278–1323, 2012.

29. M. Naveenkumar and A. Vadivel, "OpenCV for Computer Vision Applications", Proceedings of National Conference on Big Data and Cloud Computing (NCBDC'15), March 20, 2015.
30. D. N. Parmar and B. B. Mehta, "Face Recognition Methods & Applications," Int. J. Comput. Technol. Appl.,vol. 4, no. 1, pp. 84–86, 2013.