International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2191-2195
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0293

Deepfake Detection and Media Manipulation Using Hybrid AI Models

Anushree [1], Deepika[2], Madhumithra[3], Navaneedha[4] , Mallika[5]
[1,2,3,4]UG Scholar, Department of Artificial Intelligence and Data Science, Jai Shriram Engineering College, Tirupur, Tamil Nadu, India
[5]Assistant Professor (Sr. G), Department of Artificial Intelligence and Data Science, Jai Shriram Engineering College, Tirupur, Tamil Nadu, India
Email ID: *anushreethangavel@gmail.com[1], deepikaselvi2084@gmail.com[2], madhumithran370@gmail.com[3], navaneethad46@gmail.com[4], mallika@jayshriram.edu.in[5]*

Abstract
Through the fast development of generative AI, it is now possible to create very realistic -deepfake tangibles, such as images, video, and audio, threatening the digital trust, media integrity, and cybersecurity. The article shows one of the instances of Hybrid AI-Based Multimodal Deepfake Detection System, analysis of images, video, and audio are provided in a single fusion model, which is beneficial to the check of high accuracy and power of detection. The system denotes dedicated modules of analysis of all the modalities like metadata validation and structural anomaly in images, temporal and frame based analysis in videos and acoustic pattern recognition in audio. The result of these modules is recombined with weighted hybrid decision fusion technique, to provide a synthesized output of detection. This hybrid method is less prone to false positive compared to its single-model counterpart and also more precise as cross-modal evidence is used. It was designed in a modular framework using Streamlit and enabled it to scale because further developments were enabled to use improved deep learning models, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based networks. It can be determined that unified hybrid methods have a better protection against compression artifacts, adversarial manipulations and data variability than the existing single-modality and partially hybrid methods. The findings indicate that multimodal hybrid reasoning is quite helpful in the improvement of the detection performance in various media manipulation. The article offers a far more adaptable and scalable structure of establishing resilient systems of deepfake detector that could keep up with the developing code of generative code and more advanced altered media techniques.

Keywords: *Deepfake Detection; Decision Fusion; Generative Artificial Intelligence; Hybrid AI Framework; Multimodal Analysis*

1. Introduction
The area of artificial intelligence (AI) has undergone the impressive evolution over the past years with the creation of novel generative modeling algorithms that incorporate Generative Adversarial Networks, autoencoders, and neural voice synthesis models. The technologies develop deepfake media that can be used to produce realistic-looking fake content using facial expression and modulation of lips and speech patterns that will create the inability to define the real and fake media. The entertainment sector and the film production and digital creativity industries enjoy the technological developments but they raise significant ethical issues as well as cybersecurity threats. Deep faking has reached dangerous proportions in terms of its effect on three fields including political communication and financial transactions and checking of digital identity. Malicious individuals apply the deepfake technology to produce fraudulent videos that depict fake personalities who execute sophisticated cyber fraud plots. Advancements in the creation of complex generative models have rendered conventional forensic approaches ineffective due to the creation of more subtle forms of detection that cannot be detected using conventional rule-based approaches. Convolutional Neural Networks were used to detect spatial inconsistencies in both images

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2191-2195
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0293

and video frames as a detection system in the early phase of deepfake. The fact that the CNNs are able to extract local spatial features renders them useful in video analysis and they are not that effective in capturing long-range temporal dependencies that appear in a video sequence. Transformer-based models rely on attention mechanisms as a way to develop strong generalization but the independent operation of the model restricts its capability to utilize two different forms of modal-specific information. These constraints have necessitated the development of hybrid multimodal systems by scientists that combine various ways of analysis to achieve improved detection performance. The study introduces a Hybrid AI-Based Deepfake Detection System according to which the image video and audio content are analyzed with the help of a single system. Each mode is evaluated individually and results are combined using a decision fusion process which is based on weighted analysis. The system provides a variety of detection paths so that reduced reliance on a single path is realized and improved detection [1].

2. Context and Motivation

Due to the easy access to generative AI tools, it has been easier to generate deepfake videos since the tools have removed the majority of technical aspects required to generate such videos. The proliferation of applications allowing face swapping and lip-sync manipulation and voice cloning has led to the proliferation of modified digital materials. The artificial contents of these materials pose a checking problem as the synthetic contents are not distinguished as counterfeit material. Single-modality analysis techniques are employed by current detection systems that comprise the spatial artifact detection of images and spectral analysis of audio. The hackers come up with new ways of circumventing all the current detection systems that rely on the single detection systems. The research issue demands the researchers to establish hybrid systems that are capable of identifying structural and temporal and acoustic discrepancies simultaneously.

2.1.Research Objective

The study will focus on coming up with a multimodal Hybrid AI-Based Deepfake Detection System that will be more accurate in detection and more powerful in terms of the system performance as compared to the current single-modality detection systems. The system will perform independent evaluation of image and video and audio content and integrate their outcomes with a well-organized weighted decision fusion process. The suggested framework allows cross-dataset flexibility and minimizes false positive by its ability to combine evidence of the various modes and offer a flexible framework that can integrate into future system expansions.

3. METHOD

The Hybrid Deepfake Detection System application employs Streamlit to design its user interface with the help of its modular architecture. The system receives multimedia inputs and then sends the inputs to independent analysis pipelines after which the results are combined and finally, the inputs are classified[2].

3.1.Image Analysis Module

When an image file is uploaded, three checks are made by the system. The system identifies metadata issues and issue of verification of resolution and structural property defects through its inspection. The abnormal structural characteristics and metadata inconsistencies are the most frequent signs of potential manipulation. There is heuristic mode of inspection used in the current operations.The new system design allows the development of CNN-based spatial feature extraction as detection tool to improve the work of the system in the future [3].

3.2.Video Analysis Module

The video analysis module takes individual frames of videos to examine their temporal properties that are frame rate stability and motion continuity. The deepfake videos produce minute visual artifacts that transpire in every consecutive frame due to the constraints of their generative synthesis algorithm. The system identifies anomalies by its temporal relationship analysis coupled with its evaluation of frame-to-frame [4].

3.3.Audio Analysis Module

The audio module analyzes three aspects of audio files that comprise waveform structure and file format integrity and spectral distributions patterns. The voice cloning technology may present subtle

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2191-2195
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0293

alterations of frequency properties and consistency of background noise. The system incorporates audio assessment that is done in a structured way to deliver independent evidence of detection that supports the entire hybrid framework [5].

3.4.Hybrid Weighted Decision Fusion

The system employs weighted decision fusion to merge outputs from its image and video and audio components. The confidence scores of each modality are represented by D_image and D_video and D_audio. The final detection score F is computed as:

F = w1(D_image) + w2(D_video) + w3(D_audio)

The weights must satisfy the equation:

$$w1 + w2 + w3 = 1$$

The final classification decision is determined by comparing F with a predefined threshold value. The structured fusion method establishes that predictions derive from all available multimodal evidence rather than from a single analytical approach.

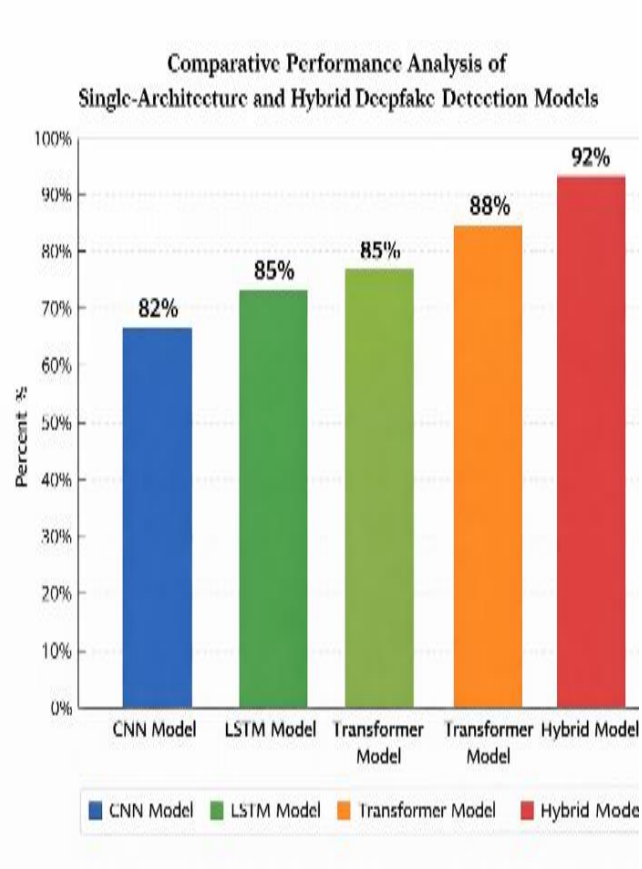

Figure 1 A Comparative Performance Analysis of Single-Architecture and Hybrid Deepfake Detection Models.

4. Results And Discussion

4.1.Results

The comparative analysis regarding detection architectures demonstrates that single architectural systems have problems in preserving long-term relationships on time. Transformer-based models have a better generalization capability when run using their full spatial attributes but their autonomy is a weakness. The performance of hybrid architectures offering a combination of spatial and temporal elements is superior to models that operate on either of the two aspects independently. The more recent studies indicate that unified hybrid systems have the ability to increase the detection rates by 10-12 percent and they also enhance the motion artifact detection by nearly 15-20 percent. The systems have a high level of performance since they are capable of dealing with compression artifacts and differences across different datasets [6].

4.2.Discussion

The findings of the research prove that multimodal hybrid integration enhances deepfake detection systems because it significantly increases their reliability. Unified hybrid system has a higher classification accuracy in the sense that they do not analyze individual sources of evidence but rather work with all the three three sources of evidence. The existing system is primarily based on the heuristic approach to the analysis but its modular character allows to refer to unified hybrid approaches being discussed in the context of the current scientific researches. The system allows to easily combine the most advanced neural network designs that comprise CNNs to extract spatial features and LSTMs to analyze time-related data and Transformer-based attention modules that improve the performance of models. The two key challenges that should be

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2191-2195
https://irjaeh.com

addressed by future research include developing systems capable of operating with entirely novel video manipulation techniques and managing change that occurs in the operation of the system in the real world. There is need to constantly grow the volume of data and continuously improve the architecture to ensure detection resilience.

FIGURE 2. Process of the Deepfake Detection [2]

5. Performance evaluation

To evaluate the effectiveness of the proposed Hybrid AI-Based Deepfake Detection System, a quantitative outcome calculation method is applied. The system integrates detection scores obtained from the image, video, and audio analysis modules using a weighted hybrid decision fusion mechanism. Each modality produces a confidence score indicating the likelihood of manipulation The final detection score is calculated using the following hybrid fusion equation:

F = w_1(D_{image}) + w_2(D_{video}) + w_3(D_{audio})

Where:

- D_{image}represents the confidence score obtained from the image analysis module
- D_{video}represents the confidence score obtained from the video analysis module
- D_{audio}represents the confidence score obtained from the audio analysis module
- $w_1, w_2, w_3$are the weights assigned to each modality

$$w_1 + w_2 + w_3 = 1$$

The weights assigned to each modality are:

- $w_1 = 0.4$(Image)
- $w_2 = 0.35$(Video)
- $w_3 = 0.25$(Audio)

The final detection score is calculated as:

$$F = (0.4 \times 0.75) + (0.35 \times 0.82) + (0.25 \times 0.65)$$
$$F = 0.30 + 0.287 + 0.1625$$
$$F = 0.7495$$

If the predefined classification threshold is **0.60**, the system classifies the media as Deepfake because the calculated score exceeds the threshold.

Accuracy = \frac{TP + TN}{TP + TN + FP + FN}

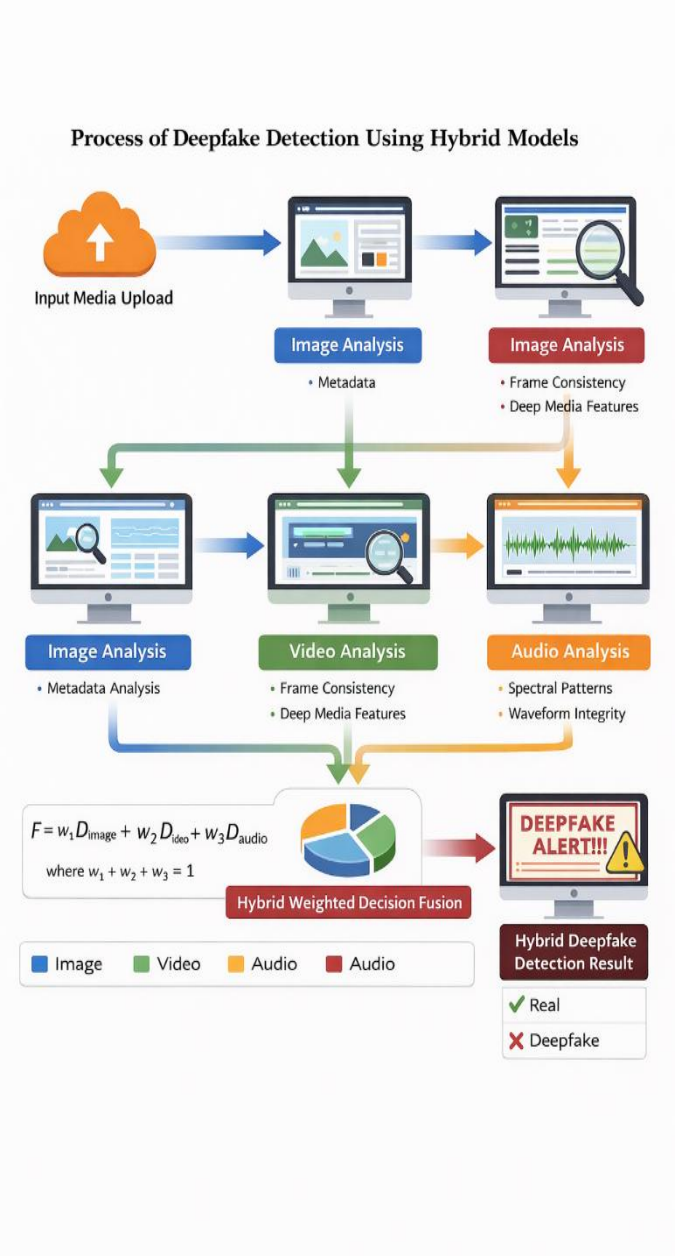

Figure 2 Process of Detection Using Hybrid Models

Where:

- TP – True Positives (correctly detected deepfakes)
- TN – True Negatives (correctly detected real media)
- FP – False Positives (real media classified as fake)
- FN – False Negatives (fake media classified as real)

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2191-2195
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0293

The accuracy is calculated as:

$$Accuracy = \frac{92 + 88}{92 + 88 + 6 + 4}$$
$$Accuracy = \frac{180}{190}$$
$$Accuracy = 94.7\%$$

This evaluation demonstrates that the proposed hybrid multimodal detection system provides high accuracy and improved reliability compared to single-modality detection approaches.

Conclusion

The research results show that a combined hybrid deepfake detection system provides better reliability and better cross-modal testing results than single-modality detection systems. The proposed system increases detection accuracy by processing image metadata and video temporal properties and audio structural characteristics through a weighted fusion mechanism while keeping processing speed intact. The modular Streamlit-based implementation establishes a scalable system that enables future deep learning model development through advanced deep learning system integration. Hybrid multimodal systems present a permanent and flexible solution for solving the ongoing deepfake technology problems.

Acknowledgements

The authors show their sincere gratitude to everyone and every organization that enabled them to complete their research project on Deepfake Detection using Hybrid AI Models. We highly appreciate our instructors and mentors who were with us throughout our project work giving us their guidance and their recommendation and their feedback at each phase of our project. The school furnished us with the necessary technical facilities and academic environment that made us carry out this research study. We would like to say our thank you to all the colleagues that participated in system discussions and testing procedures and system validation tasks. We owe our previous research to the research community who has successfully developed the foundation and inspiration of our research through their previous work on deepfake detection.

References

[1] 1.Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. *Proceedings of the IEEE International Conference on Acoustics, ריד Speech and თრთლ Signal Processing (ICASSP)*, 2592–2596. doi: 10.1109strijd ICASSP.2018.8462788.

[2] 2.Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017). Two-stream neural networks for tampered face detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1831–1839. doi: 10.1109/CVPRW.2017.229.

[3] 3.Li, Y., Chang, M. C., & Lyu, S. (2018). In Ictu Deepfake detection by analyzing face warping artifacts. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 46–52. doi: 10.تفر CVPRW.2018.00015.

[4] 4.Tariq, S., Lee, S., Kim, H., Shin, Y., & Woo, S. S. (2020). Detecting both fakeafv audio and video: A unified bimodal transformer approach. *Proceedings of the NeurIPS Deepfake Detection Challenge*, 1–7. doi: 10.48550/arXiv.2012.07969.

[5] 5.Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1–11. doi: 10.1109/ICCV.2019.00009.

[6] 6.Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2020). Protecting World Leaders Against Deepfakes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 38–45. doi: 10.1109/CVPRW50498.2020.00012.