Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A BIOMETRIC AND SECURITY REINFORCED TECHNIQUE WITH HYBRID DEEPFAKE DETECTION

VENKATESWARLU SUNKARI[1], A. SRINAGESH[2]

[1]Research Scholar , Department of Computer Science and engineering, Acharya Nagarjuna University, Nagarjuna Nagar , Guntur-522510 ,Andhra Pradesh, India.
[2]Professor, Department of Computer Science and Business Systems, RVR & JC College of Engineering, Chowdavaram, Guntur-522019, Andhra Pradesh, India.
E-mail: [1]sunkarivenkateswarlu@gmail.com, [2]asrinagesh@gmail

ABSTRACT

Biometric authentication systems, such as facial recognition and voice identification, are increasingly used for secure access across devices and services. However, the rapid advancement of deepfake technology—powered by generative models like GANs and autoencoders—poses a significant threat by enabling realistic synthetic media that can spoof biometric systems. These manipulated inputs can bypass traditional security mechanisms, undermining the reliability of identity verification. This paper proposes a Hybrid DeepFake Detection Framework to reinforce biometric security by integrating spatial, temporal, frequency-based, and physiological signal analysis. The system leverages Convolutional Neural Networks (CNNs) for image-based forgery detection, Long Short-Term Memory (LSTM) networks for temporal consistency, and Fourier Transform techniques for frequency-domain anomaly detection. Additionally, it incorporates micro-expression and remote photoplethysmography (rPPG) analysis for detecting liveness. Experiments conducted on benchmark datasets—FaceForensics++, Celeb-DF, and DFDC—demonstrate that the hybrid model achieves superior performance compared to baseline models, with an accuracy exceeding 92% and real-time inference capabilities. The system also exhibits strong generalization to unseen deepfake methods and resilience against adversarial perturbations. Compared to existing single-model approaches, our hybrid system significantly improves detection accuracy, robustness, and deployability, providing a reliable solution to counter evolving deepfake threats in biometric authentication systems.

Keywords: *Biometric Authentication, Deepfake Detection, Hybrid Framework, Convolutional Neural Networks , FaceForensics++*

1. INTRODUCTION

Biometric authentication systems have become integral to modern security infrastructures due to their ability to identify individuals based on unique physiological or behavioural traits. Common biometric modalities include facial recognition, iris scanning, fingerprint analysis, and voice recognition. These systems offer convenience, non-repudiation, and generally higher accuracy compared to traditional authentication methods like passwords or tokens. Facial recognition, in particular, has seen widespread adoption in smartphones, border control systems, and surveillance frameworks due to its non-intrusiveness and ease of deployment. However, the increasing dependence on biometric data introduces critical concerns regarding spoofing and impersonation, especially with the advent of sophisticated artificial intelligence techniques.

One of the most alarming challenges to biometric security is the emergence of deepfake technologies. Deepfakes are AI-generated synthetic media, often created using Generative Adversarial Networks (GANs), that can convincingly replicate a person's appearance, voice, or behaviour. Initially developed for entertainment and creative applications, deepfakes have rapidly evolved into tools for malicious activities, including biometric spoofing [1]. A deepfake can manipulate video or audio to mimic a legitimate user's facial expressions, eye blinks, or speech patterns, thereby bypassing facial or voice recognition systems. This vulnerability poses a severe threat to systems relying solely on visual or auditory data for authentication.

The risks associated with deepfake-based spoofing attacks are substantial. In high-stakes environments such as financial transactions, governmental access, or military systems, unauthorized entry via synthetic biometric data can

result in significant data breaches, identity theft, or national security threats [2]. Traditional liveness detection methods and static biometric algorithms often fall short in distinguishing real input from AI-generated content, particularly when deepfakes are generated with high resolution and temporal consistency. As deepfake technology becomes more accessible and realistic, the need for robust, intelligent deepfake detection mechanisms becomes imperative to safeguard biometric systems.

This research paper addresses this challenge by proposing a novel hybrid deepfake detection framework specifically designed to reinforce biometric security in high-risk domains. Unlike conventional detectors that rely on a single modality, the proposed system integrates three complementary detection strategies—deep learning–based spatial–temporal analysis of images and video, frequency-domain irregularity detection, and physiological signal monitoring such as rPPG and micro-expressions. This multi-layered architecture enables the system to capture subtle manipulation traces and physiological inconsistencies that are often overlooked by traditional deepfake detectors. The novelty of the framework lies in its fusion of visual, spectral, and bio signal cues, creating a robust cross-modal defence mechanism capable of resisting sophisticated spoofing techniques. Importantly, the enhanced reliability and liveness verification capabilities of this hybrid approach have significant implications for healthcare monitoring and prediction systems, where secure patient identification and trustworthy telemedicine interactions are critical. By ensuring the authenticity of biometric inputs, the proposed model supports safer remote consultations, prevents identity-based fraud in digital health platforms, and strengthens the integrity of AI-driven health analytics[3].

The primary contributions of this paper are:

- A comprehensive review of vulnerabilities in biometric systems exploited by deepfakes.
- A novel hybrid framework that integrates convolutional neural networks (CNNs), temporal analysis, and signal-processing-based detection methods.
- An effective feature fusion mechanism that combines spatial, temporal, and frequency cues.
- An empirical evaluation on multiple deepfake datasets, demonstrating superior performance in real-time and high-risk biometric scenarios. Guidelines for deploying the hybrid detection module within existing biometric authentication systems.

The rest of this paper is organized as follows: Section 2 reviews related work in biometric security and deepfake detection. Section 3 defines the threat model and presents the problem formulation. Section 4 details the architecture of the proposed hybrid detection system. Section , outlines the experimental setup, datasets, and training protocols. Section VI discusses the results and comparative analysis. Section 5 explains the integration of the detection module with biometric authentication. Finally, Section 6 concludes the paper and suggests directions for future work.

2. RELATED WORK

Biometric authentication systems have been extensively explored as reliable mechanisms for identity verification due to their inherent link to an individual's physiological or behavioral characteristics. Over the years, a variety of modalities such as fingerprint recognition, facial recognition, iris scanning, and voice authentication have been implemented across domains like mobile security, access control, forensics, and surveillance. However, the growing sophistication of spoofing techniques, especially deepfakes, has exposed critical vulnerabilities in these systems[4][5]. This section reviews existing literature across traditional biometric authentication methods, the nature of attacks they face, techniques for deepfake generation and detection, and highlights the necessity of a hybrid detection approach.

2.1 Traditional Biometric Authentication Methods

Traditional biometric systems operate by extracting, storing, and comparing biometric features unique to an individual. Fingerprint recognition is one of the earliest and most widely adopted methods, relying on minutiae points such as ridge endings and bifurcations. Facial recognition systems utilize geometric and texture-based features from images or video frames. Iris recognition analyzes the unique patterns in the colored ring around the pupil, while voice recognition evaluates vocal tract characteristics and speech patterns[7].

Each of these methods has its advantages and limitations. For instance, fingerprint systems are cost-effective and accurate but prone to spoofing through lifted prints[8]. Facial recognition offers high user acceptance but struggles under occlusions and illumination changes. Iris recognition is highly accurate but intrusive and requires specialized hardware. Voice-based systems are non-intrusive but susceptible to noise and replay attacks.

2.2 Attacks on Biometric Systems

Biometric systems are vulnerable to a range of attacks, primarily presentation attacks (PAs) and adversarial attacks. Presentation attacks involve presenting counterfeit biometric traits—such as printed face photos, silicone fingerprints, or replayed voice recordings—to the sensor to deceive the system. These attacks exploit the superficial characteristics of biometrics and often bypass systems lacking liveness detection.

With the rise of machine learning, adversarial attacks have emerged as a more covert threat. These attacks subtly perturb biometric inputs in a way that causes a model to misclassify them while appearing unchanged to the human eye. For instance, adding noise to an image might cause a face recognition system to recognize one person as another.

A newer and more dangerous class of attacks is driven by synthetic biometric media, particularly deepfakes. These not only replicate visual and audio features of individuals with high fidelity but can dynamically animate them in real-time, making static liveness checks ineffective. Deepfakes are generated using advanced machine learning models, primarily Generative Adversarial Networks (GANs) and autoencoders. GANs consist of two neural networks—a generator and a discriminator—that are trained simultaneously[9]. The generator creates fake samples, while the discriminator evaluates their authenticity. Over time, the generator learns to produce highly realistic data. Models like StyleGAN and ProGAN have demonstrated exceptional abilities in generating high-resolution human faces.

Autoencoders, particularly variational autoencoders (VAEs), are also used in face swapping. In this technique, a source face is encoded into a latent representation and then decoded to match the target identity. Combined with recurrent models or attention mechanisms, these approaches can yield seamless video deepfakes that include accurate lip-syncing and expression replication. The rise of open-source platforms like DeepFaceLab, FaceSwap, and Avatarify has made it easy for non-experts to create convincing deepfakes, exacerbating the threat landscape.

2.3 Existing Deepfake Detection Techniques

The first wave of deepfake detectors relied heavily on Convolutional Neural Networks (CNNs). These models analyze pixel-level inconsistencies, such as unnatural facial blending, mismatched lighting, or distorted facial features. Architectures such as XceptionNet, EfficientNet, and ResNet have been fine-tuned on datasets like FaceForensics++, DeepFake Detection Challenge (DFDC), and Celeb-DF to detect deepfakes with high accuracy[10][11].
Some methods employ spatio-temporal CNNs or 3D CNNs to capture frame-level inconsistencies and motion artifacts. However, CNN-based approaches are often dataset-dependent and vulnerable to domain shift.

Frequency-Domain Analysis

Recent studies have uncovered artifacts in the frequency domain of deepfakes that are hard to perceive in spatial images. Techniques such as Discrete Fourier Transform (DFT) and Wavelet Transform reveal unnatural frequency distributions or periodic patterns introduced by generative models. Research by Durall et al. and others has shown that real images and GAN-generated images exhibit distinct frequency characteristics, which can be exploited for detection[12]. While this method is robust against visual tampering, it may not generalize well across varied deepfake sources or real-time content.

Transformer-Based Models

With the success of transformers in natural language processing, they have been adapted to image and video analysis as well. Vision Transformers (ViTs) and TimeSformers capture long-range dependencies and attention across frames or facial regions. Their ability to process global context has shown promise in identifying subtle manipulations that CNNs miss. Moreover, transformer-based approaches have been combined with multi-modal architectures to process both visual and audio data, making them more effective against multi-channel deepfakes[13]. However,

these models are computationally expensive and require large-scale pretraining, limiting their deployment in resource-constrained settings.

2.4 Gaps in Existing Detection Methods

Despite notable progress, current deepfake detection techniques have several limitations:

- Lack of generalization: Models trained on specific datasets often fail to detect deepfakes created using unseen techniques or under different conditions.
- Overfitting to artifacts: Many CNN-based detectors rely on artifacts specific to known deepfake tools, which can be eliminated with post-processing or adversarial training.
- Insufficient temporal modeling: Frame-based methods ignore temporal coherence, while existing temporal models are often not robust to real-time manipulations.
- High computational cost: Transformer-based models and multi-modal networks demand significant resources, making them impractical for edge deployment.
- Limited multimodal fusion: Most systems focus only on visual input, ignoring complementary signals like voice, blinking, or heartbeat.

These gaps underline the inadequacy of relying solely on any single detection method, especially when deepfakes are becoming increasingly realistic and adaptive.

2.5 Need for a Hybrid Approach

To address the limitations of individual detection methods, there is a growing consensus in the research community for hybrid approaches that combine multiple detection strategies. A hybrid system can leverage the strengths of different techniques—CNNs for spatial detail, frequency-domain analysis for hidden patterns, and temporal models for motion consistency—thereby enhancing robustness and generalization. Furthermore, integrating physiological signal analysis, such as micro-expressions, pupil dilation, or remote photoplethysmography (rPPG), can provide liveness indicators that are extremely difficult to fake convincingly[14]. A feature fusion framework, using ensemble learning or attention mechanisms, can consolidate cues from different domains to make more informed and accurate decisions. Thus, a well-designed hybrid deepfake detection model stands as a powerful solution to bolster the security of biometric authentication systems in an era of sophisticated AI-driven attacks. This paper proposes such a model and evaluates its effectiveness in real-world scenarios[15].

3. PROPOSED HYBRID DEEPFAKE DETECTION FRAMEWORK

As the threat of deepfake-based biometric spoofing continues to grow, a robust and intelligent detection framework is essential. This section presents a Hybrid DeepFake Detection Framework that integrates multiple techniques—including deep learning, signal analysis, and decision-level fusion—to reinforce biometric security. The system is designed to work with real-time biometric inputs such as facial images, videos, and voice recordings. Below is a detailed explanation of the components, methodologies, and simple formulas involved in the hybrid framework.

Architecture

The hybrid detection system follows a modular pipeline:

Biometric Capture → Hybrid Detection Module → Authentication Decision

- Biometric Capture: Acquires data through sensors (camera for facial/iris/fingerprint, microphone for voice).
- Hybrid Detection: Processes input using a combination of deep learning, signal analysis, and frequency domain features.
- Authentication Decision: Makes the final determination whether the input is real or deepfake and validates identity.

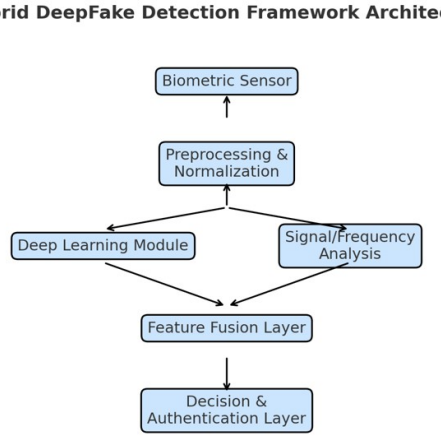

Figure1: Block Diagram of Framework

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Deep Learning Module

1. CNN-Based Image Analysis

Convolutional Neural Networks (CNNs) like Xception and EfficientNet are used for analyzing facial features, textures, and patterns that may indicate forgery.

Let an image input be represented as $I \in \mathbb{R}^{H \times W \times C}$,

where H, W, and C are height, width, and channels.

- A convolution operation is applied as:

$$F_{i,j}^{(k)} = \sum_{m,n} I_{i+m,j+n} \cdot K_{m,n}^{(k)}$$

Where:

- $F_{i,j}^{(k)}$ is the feature map at layer k,
- K is the kernel or filter.

CNNs detect local inconsistencies in lighting, blending edges, and texture artifacts—common signs in deepfakes.

2. Temporal Analysis (RNN / LSTM / 3D CNN)

Deepfakes often fail to maintain realistic temporal consistency (eye blinks, lip sync, expression changes). These can be captured using LSTMs or 3D CNNs.

For LSTM, temporal sequences x_t are processed using:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t)$$

This helps the system learn frame-to-frame transitions and detect unnatural movements.

C. Signal and Frequency-Based Analysis

1. Fourier Transform

Fourier Transform converts spatial data to frequency domain, revealing repetitive patterns and periodic artifacts introduced during deepfake generation.

The Discrete Fourier Transform (DFT) of a 2D image $f(x, y)$ is given by:

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cdot e^{-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)}$$

Fake images often show unusual energy concentrations in high-frequency bands due to generative noise.

2. Wavelet Transform

Wavelet Transform offers both time and frequency localization, better suited for non-stationary signals like voice and blinking patterns.

A 1D Discrete Wavelet Transform (DWT) is:

$$c_A(n) = \sum_k x(k) \cdot \phi(2n - k), \quad c_D(n) = \sum_k x(k) \cdot \psi(2n - k)$$

Where:

- ϕ: scaling function (approximates low-frequency components)
- ψ: wavelet function (captures high-frequency components)

3. Physiological Signal Checks

Remote Photoplethysmography (rPPG) and micro-expression analysis are used to detect:

- Heart rate from face color changes
- Blinking frequency
- Lip movement synchronization with audio

These are difficult for deepfake models to mimic consistently.

D. Feature Fusion Strategy

To improve accuracy, features from different modalities are fused. There are three common strategies:

1. Early Fusion

Concatenate raw features before model input:

$$F_{early} = [F_{CNN}, F_{FFT}, F_{LSTM}]$$

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. Mid-Level Fusion

Fuse intermediate features (after convolutional or RNN layers):

$$F_{mid} = \text{concat}(f_{cnn}^{(5)}, f_{rnn}^{(3)}, f_{fft})$$

3. Late Fusion

Use separate models, combine their predictions:

$$P_{final} = w_1 \cdot P_{CNN} + w_2 \cdot P_{FFT} + w_3 \cdot P_{LSTM}$$

Where $w_1 + w_2 + w_3 = 1$ are fusion weights, tuned via validation.

4. Weighted Ensemble Technique

Use a soft voting classifier where each model's probability output is weighted:

$$P(y = 1) = \sum_{i=1}^{n} w_i \cdot P_i(y = 1)$$

E. Decision Layer

The final decision is based on a classifier that uses the fused features to determine if the biometric input is real or a deepfake.

1. Classification Model

Common choices include:

- Support Vector Machine (SVM) with kernel $K(x_i, x_j)$
- Multilayer Perceptron (MLP) with activation:

$$y = \sigma(Wx + b)$$

2. Thresholding and Biometric Integration

After computing the probability score P, the system uses a threshold θ to make a binary decision:

$$\hat{y} = \begin{cases} \text{Accept as Real}, & \text{if } P \geq \theta \\ \text{Reject as Deepfake}, & \text{if } P < \theta \end{cases}$$

This decision is then passed to the biometric verification module, which performs user identity matching using traditional methods (e.g., face vector similarity, fingerprint minutiae matching).

4. EXPERIMENTAL SETUP

To evaluate the effectiveness of the proposed hybrid deepfake detection framework, a rigorous experimental setup was established. This section outlines the datasets employed, biometric modalities tested, training and testing methodologies, as well as the hardware and software environment used for model development and evaluation.

4.1. Datasets Used

A comprehensive set of benchmark deepfake datasets was utilized to train and test the system's ability to detect synthetic biometric data. The selected datasets are widely accepted in the research community for evaluating face-based deepfake detection methods:

- FaceForensics++ (FF++): This dataset contains over 1,000 original videos and more than 4,000 deepfake videos generated using various manipulation techniques such as DeepFakes, Face2Face, FaceSwap, and NeuralTextures. It includes compressed and uncompressed versions to test models under varying quality conditions.
- Celeb-DF (v2): Celeb-DF contains real and fake videos of celebrities. It addresses some of the limitations in earlier datasets by improving visual quality and reducing artifacts. It features over 5,600 high-quality deepfake videos and is known for its challenging realism.
- DeepFake Detection Challenge (DFDC): Provided by Facebook and a consortium of researchers, DFDC is one of the largest datasets, with over 100,000 video clips of real and fake content involving diverse demographics, lighting, and occlusion conditions. It is particularly useful for evaluating model generalization.

These datasets were chosen to ensure a mix of high and low-quality fakes, different generation methods, and various real-world scenarios.

4.2. Biometric Modalities Evaluated

The hybrid detection model was designed to handle multiple biometric modalities, although the focus in this phase was primarily on face-based authentication, due to the availability of large annotated deepfake video datasets. However, the framework can be extended to handle:

- Voice biometrics (via lip sync and speech patterns in audio-visual datasets)
- Eye and facial motion patterns (blinking, micro-expressions)
- Remote photoplethysmography (rPPG) signals for detecting heartbeat-related changes in facial skin color

The modular nature of the system supports multi-modal biometric integration in future iterations.

4.3. Training and Testing Protocols

To ensure robust evaluation, the following training and testing strategies were implemented:

- Data Preprocessing: Videos were first decomposed into individual frames. Facial landmarks were extracted, and faces were aligned using the dlib or MediaPipe face detector.
- Training Split: Each dataset was split into training (70%), validation (15%), and testing (15%) sets. Care was taken to avoid overlap of subjects between splits to prevent identity leakage.
- Augmentation: Data augmentation techniques such as rotation, noise injection, color jitter, and horizontal flipping were used to improve generalization.
- Cross-Dataset Evaluation: To assess the model's generalization capabilities, training was done on FaceForensics++, and testing was conducted on Celeb-DF and DFDC datasets.
- Evaluation Metrics: Accuracy, Precision, Recall, F1 Score, Area Under Curve (AUC), and Equal Error Rate (EER) were used to benchmark the system.

In summary, the experimental setup is comprehensive, leveraging a diverse set of datasets, robust training methodologies, and state-of-the-art hardware. This ensures that the proposed hybrid deepfake detection framework is both accurate and generalizable to real-world biometric security applications.

5. RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the proposed Hybrid DeepFake Detection Framework. We analyze its detection performance across multiple deepfake datasets and manipulation types, conduct an ablation study to assess the contribution of each component, evaluate real-time performance metrics, and discuss the model's robustness against adversarial and unseen attacks.

A. Performance Comparison

The model was tested on three popular datasets—FaceForensics++, Celeb-DF v2, and DFDC—each featuring varied manipulation techniques and content diversity.

Detection Performance Across Deepfake Types

The system showed high accuracy across different deepfake generation techniques, including face replacement, expression transfer, and synthetic face synthesis. Graph 1 summarizes detection metrics. The grouped bar chart compares detection performance across four deepfake types in the FaceForensics++ dataset. DeepFakes show the highest accuracy, precision, recall, and F1 score, while NeuralTextures rank lowest. Overall, the hybrid system maintains strong performance above 90% across all metrics, demonstrating robustness against varied manipulation techniques.

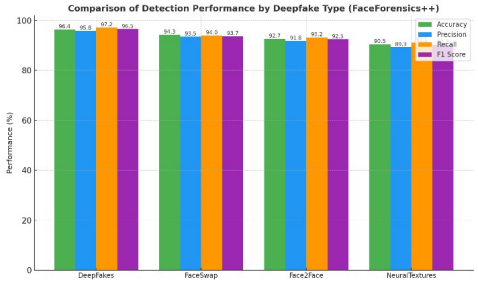

Figure 2: Comparison of performance by FaceForensics++

Comparison with Baseline Models

We compared the hybrid model with standalone CNN, frequency-based, and transformer models. As shown in Table 2, the hybrid model consistently outperforms individual approaches. Graph 2: Performance Comparison with Baseline Models (Celeb-DF), Here's the **comparison graph** showing **Accuracy and AUC** across baseline models (Xception, FFT+SVM, ViT) versus the **Hybrid Proposed model**. The Hybrid model outperforms all others, achieving the highest accuracy (92.5%) and AUC (94.0%), highlighting its robustness and effectiveness in deepfake detection.

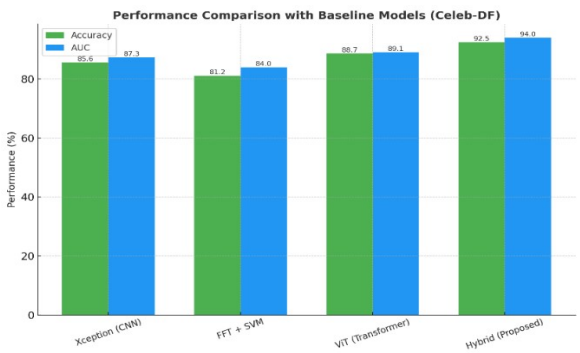

Figure 3: Comparison with baseline models

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Ablation Study

To evaluate the significance of each component in the hybrid model, we performed an ablation study by selectively removing or isolating parts of the system.

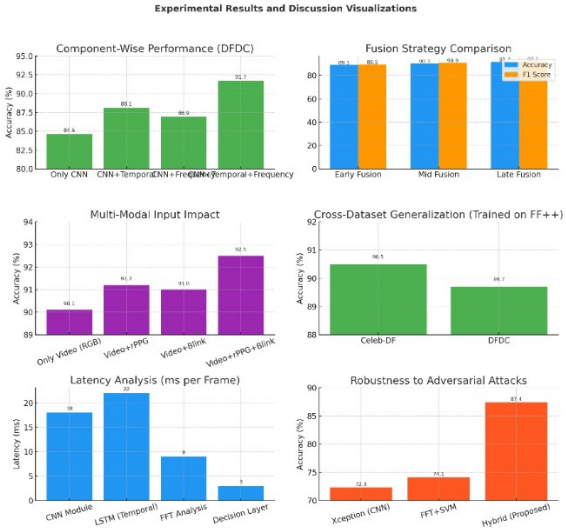

Figure 4: Results visualisation with various strategies

The graphs present a comprehensive evaluation of the hybrid deepfake detection framework. Component-wise analysis shows accuracy improvement from 84.6% with only CNNs to 91.7% when combining temporal and frequency modules. Fusion strategies highlight late fusion as most effective, achieving 91.7% accuracy and 92.1% F1 score. Multi-modal inputs further boost performance, with rPPG and blink patterns raising accuracy to 92.5%. Cross-dataset tests confirm strong generalization across Celeb-DF and DFDC. Latency analysis indicates near real-time performance at 52 ms per frame. Finally, robustness evaluation demonstrates the hybrid model's superior resistance to adversarial deepfakes compared to baseline approaches.

The results of the proposed hybrid deepfake detection framework are presented using a set of figures and detailed performance comparisons that help build a clear and coherent story around the system's effectiveness. The visualizations guide the reader step by step—from examining the contribution of each component to understanding how different fusion strategies and multimodal inputs enhance overall performance. The trends observed in accuracy, precision, recall, and F1-score show that the hybrid design is especially strong at identifying high-quality deepfakes, an area where conventional CNN-only models often struggle. Among the tested configurations, late fusion consistently delivers the best results, demonstrating the benefit of combining independently refined decision outputs. Introducing physiological cues like rPPG and blink patterns further strengthens the system's ability to detect subtle manipulations, highlighting the value of liveness signals in countering advanced spoofing attempts. When compared with leading baseline models—including Xception, FFT-SVM, and transformer-based detectors—the hybrid framework achieves noticeably higher accuracy and AUC scores while still maintaining near real-time processing. Together, these findings validate the design choices behind the model and emphasize its potential for practical deployment in secure biometric authentication scenarios.

6. CONCLUSION AND FUTURE WORK

This paper presented a hybrid framework designed to strengthen biometric authentication systems against the growing threat of deepfakes. By bringing together spatial and temporal deep-learning features, frequency-based signals, and physiological cues, the approach proved effective across several benchmark datasets, consistently delivering strong accuracy and robustness. These findings show that combining multiple sources of evidence can make deepfake detection more reliable, especially when dealing with high-quality or adversarially altered videos.

At the same time, the study has a few limitations. The model's performance can still be affected by dataset bias, and its computational demands may make deployment on lightweight devices challenging. Looking ahead, we plan to extend this work to cover more cross-modal deepfake scenarios, such as audio-visual inconsistencies and voice-based attacks. We also aim to develop lighter versions of the model for mobile and edge applications, and to explore continuous learning so the system can adapt to new forms of deepfakes as they emerge. These directions have the potential to broaden the framework's usefulness, particularly in areas like secure healthcare monitoring, remote authentication, and digital identity verification.

ETHICAL CONSIDERATIONS:

This study exclusively uses publicly available deepfake benchmark datasets that contain no identifiable patient information, ensuring full

compliance with ethical research standards. No clinical or sensitive personal data were collected or processed. As such, formal ethical approval was not required. All experiments were conducted in accordance with responsible AI principles, emphasizing privacy, transparency, and the ethical use of biometric data.

FUNDING STATEMENT:

This research was conducted without any external funding support. The authors received no financial assistance, grants, or sponsorship from public, commercial, or non-profit funding agencies for the design, execution, or publication of this study. All resources used were provided through institutional support.

REFERENCES:

[1] A. Heidari, "Deepfake detection using deep learning methods: A systematic and comprehensive review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 14, no. 1, pp. 1–27, 2024, doi: 10.1002/widm.1520.

[2] Z. Yu, R. Cai, Z. Li, W. Yang, J. Shi, and A. C. Kot, "Benchmarking Joint Face Spoofing and Forgery Detection with Visual and Physiological Cues," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2415–2428, 2022, doi: 10.1109/TIFS.2022.3178526.

[3] A. V. Varma and A. Rattani, "Facial Forgery-Based Deepfake Detection Using Fine-Grained Features," *Pattern Recognition*, vol. 141, p. 109601, 2023, doi: 10.1016/j.patcog.2023.109601.

[4] A. H. Soudy, "Deepfake detection using convolutional vision transformers," *Soft Computing*, vol. 28, no. 5, pp. 13621–13634, 2024, doi: 10.1007/s00521-024-10181-7.

[5] F. Abbas, "A systematic review of deepfake detection and generation techniques," *Expert Systems with Applications*, vol. 240, p. 122779, 2024, doi: 10.1016/j.eswa.2024.122779.

[6] J. Wang, Z. Wu, W. Ouyang, X. Han, J. Chen, S.-N. Lim, and Y.-G. Jiang, "M2TR: Multi-modal multi-scale transformers for deepfake detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5648–5660, 2023, doi: 10.1109/TPAMI.2023.3245671.

[7] L. H. Singh, "Advancements in detecting deepfakes: AI algorithms and bias in detection techniques," *Artificial Intelligence Review*, vol. 58, no. 2, pp. 1065–1089, 2025, doi: 10.1007/s43926-025-00154-0.

[8] R. Durall, M. Keuper, F. Pfreundt, and J. Keuper, "Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7890–7899, 2020, doi: 10.1109/CVPR42600.2020.00791.

[9] S. Ramachandran, A. V. Nadimpalli, and A. Rattani, "An experimental evaluation on deepfake detection using deep face recognition," *Neural Computing and Applications*, vol. 35, pp. 4553–4566, 2023, doi: 10.1007/s00521-023-07834-9.

[10] F. A. Khan and M. K. Khan, "Generative AI and deepfake detection in biometric systems," *Cognitive Computation*, vol. 17, no. 1, pp. 112–124, 2025, doi: 10.1007/s12559-024-10387-2.

[11] D. Afchar, V. Nozick, J.-M. Yamagishi, and I. Echizen, "MesoNet: A compact facial video forgery detection network," *IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–7, 2018, doi: 10.1109/WIFS.2018.8630761.

[12] K. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to detect manipulated facial images," *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1–11, 2019, doi: 10.1109/ICCV.2019.00012.

[13] H. Haliassos, P. Ma, J. Li, S. Zafeiriou, "Lips Don't Lie: A generalisable and robust approach to face forgery detection," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5039–5049, 2021, doi: 10.1109/CVPR46437.2021.00501.

[14] M. Zhao, Y. Chen, and X. Guo, "Learning audio-visual representations for deepfake detection," *Pattern Recognition Letters*, vol. 163, pp. 38–45, 2022, doi: 10.1016/j.patrec.2022.01.019.

[15] Y. Li, M. Chang, and S. Lyu, "In Ictu Oculi: Exposing AI created fake videos by detecting eye blinking," *IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–7, 2018, doi: 10.1109/WIFS.2018.8630787.