ISSN: 1672 - 6553

JOURNAL OF DYNAMICS AND CONTROL

VOLUME 9 ISSUE 8: P221-231

DEEPFAKE VIDEO DETECTION: A COMPREHENSIVE SURVEY OF TECHNIQUES, CHALLENGES, AND A PROPOSED CNN-LSTM-BASED FRAMEWORK

Sunitha K, Aditi Shenoy, Ayushman Singh, Dhanvi Aithal K, Keerthi A Rao

Department of Information Science and
Engineering, RNS Institute of Technology,
Affiliated to VTU,
Bengaluru, India

DOI: https://doi.org/10.71058/jodac.v9i8015

DEEPFAKE VIDEO DETECTION: A COMPREHENSIVE SURVEY OF TECHNIQUES, CHALLENGES, AND A PROPOSED CNN-LSTM-BASED FRAMEWORK

[1]Sunitha K, [2]Aditi Shenoy, [3]Ayushman Singh, [4]Dhanvi Aithal K, [5]Keerthi A Rao
Department of Information Science and Engineering
RNS Institute of Technology, Affiliated to VTU
Bengaluru, India
[1]sunithakrisnamurthy@gmail.com,[2]aditishenoy35@gmail.com,
[3]ayushmans012@gmail.com, [4]dhanviaithal@gmail.com, [5]kee2rao@gmail.com

Abstract: *Deepfakes—synthetic media generated via advanced deep learning techniques such as Generative Adversarial Networks (GANs)—pose a growing threat to digital authenticity, trust, and security. While these technologies have legitimate applications in entertainment and content creation, they have also enabled the spread of misinformation, fraud, and identity manipulation. This paper presents a comprehensive survey of current deepfake detection techniques, categorizing them into traditional, machine learning-based, and deep learning-based approaches. It explores the key datasets and performance metrics which are used to benchmark detection models and it also discuss technical, ethical, and robustness concerns. This paper proposes a conceptual hybrid CNN-LSTM-based architecture that combines spatial and temporal feature analysis to improve detection accuracy of deep fake videos. The modular design is planned to support future enhancements, including the incorporation of Vision Transformers and attention mechanisms. In this study, we provide a fundamental understanding of deepfake detection and determine promising avenues for further research and real-world applications.*

Keywords: *Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), Vision Transformers (ViT), Spatiotemporal Modeling, Dataset Generalization.*

1. Introduction

The field of digital image forensics, or DIF, has long been crucial for confirming the legitimacy of digital photographs by addressing the manipulators like image enhancers, copy move and splicing. In order to identify tampering, traditional forensic techniques look for anomalies in sensor noise patterns, compression artifacts, or pixel-level features [1,2]. Forensic methods are essential for preserving digital trust and accountability as sophisticated editing tools become more widely available and media circulates quickly on social media platforms.

Deepfakes are synthetic media that are created using deep learning techniques such as Generative Adversial networks (GANs) to appear highly realistic. Unlike conventional image forgeries, deepfakes can realistically alter entire video and audio sequences in addition to images, giving the impression that people are saying or doing things they never did.

The word “deepfake” was first coined by a Reddit user posting under the handle ‘deepfakes’. He integrated "deep learning" and "fake" in his own work. When this user posted AI-based videos which utilized advanced deep learning techniques to swap faces onto previously recorded video footage, it garnered a lot of attention.

As deepfake tools have grown easier to get, their effects on society—particularly regarding politics, finance, and security—have become quite serious. The tool gives people the power to create and believe in media content as a means of spreading misinformation, aiding fraud, and bringing new security risks. Detection of deepfakes is as much a technical problem as it is a societal problem, mandating solid research and innovation. This paper articulates critical gaps and provides perspectives on evolving deepfake detection methodologies across effectiveness, existing limitations, and future research directions.

DOI: https://doi.org/10.71058/jodac.v9i8015

The paper is organized as follows. Section 2 gives background information and technical foundations related to deepfake generation and detection. Section 3 gives an overview of existing deepfake detection techniques, including both traditional methods and those based on deep learning. Key challenges and limitations in current deepfake detection research are addressed in Section 4. Section 5 presents the conceptual proposed CNN-LSTM-based detection model and Section 6 outlines potential future research directions. Conclusion is proposed in Section 7 of the paper.

2. Background and Technical Foundations

2.1 Evolution of Deepfake Technology

In recent years, advancements in deep learning models have played a vital role in the evolution of deepfake technology. Techniques such as autoencoders, Generative Adversarial Networks (GANs), and face-swapping methods are commonly employed to generate synthetic media. The rising ease with which such fabricated videos can be produced, with the availability of open-source platforms like DeepFaceLab [3] and benchmark datasets such as Celeb-DF [4] and FaceForensics++ [5], has further accelerated progress in this domain.

2.2 Key Techniques for Generating Deepfakes

Autoencoders are commonly used in deepfake generation to extract and reconstruct facial features. A shared encoder captures key components from input facial images, while individual decoders rebuild specific faces. This structure allows the model to convincingly map one person’s face onto another in video sequences [6].

Generative Adversarial Networks (GANs) are another class of deep learning models introduced by Goodfellow et al. in 2014 [7]. GANs, operating in an unsupervised manner, consist of two networks: a generator that creates synthetic data, and a discriminator that attempts to differentiate real data from fake. The adversarial training loop enables the generator to gradually produce more realistic outputs, while the discriminator simultaneously becomes more adept at detecting fakes [8].

Face swapping is another well-known technique that entails replacing a subject’s face with that of another while preserving the original facial expressions. As explained in [9], this process usually involves three core steps: selecting the source face, aligning and adjusting facial landmarks (such as the eyes, mouth, nose, and ears) to match the target, and finally, merging the facial features seamlessly to produce a realistic composite.

2.3 Common Deepfake Applications

Deepfake technology is being utilized across a range of sectors. In the entertainment industry, it is used for purposes like digitally de-aging actors or recreating deceased performers, enhancing narratives without extensive reshooting. In political contexts, deepfakes have been exploited to disseminate misinformation—especially through social media platforms like WhatsApp—by fabricating realistic videos of prominent political figures.

In cybersecurity and financial sectors, deepfakes have contributed to various fraudulent activities, including identity theft, impersonation, social engineering attacks, and high-stakes scams such as fraudulent fund transfers and CEO fraud.

3. Deepfake Detection Techniques

The progress in the creation of accurate deepfake detection systems has become crucial due to the increasing refinement of synthetic media generation. Different detection

DOI: https://doi.org/10.71058/jodac.v9i8015

strategies can be categorized into the following classes: traditional methods, machine learning-based approaches, and deep learning-based techniques. Each of these approaches uses unique capabilities to identify and mitigate manipulated media which is illustrated in Figure 1.

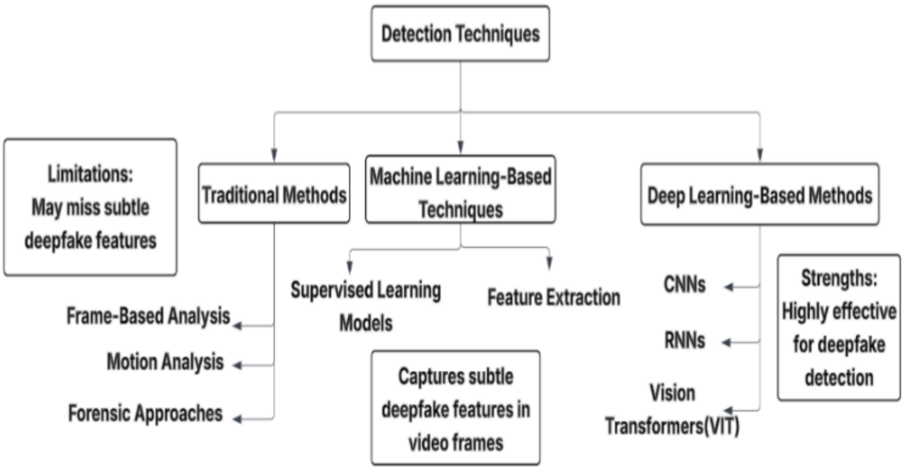
Detection Techniques
Limitations: May miss subtle deepfake features
Traditional Methods
Machine Learning-Based Techniques
Deep Learning-Based Methods
Frame-Based Analysis
Motion Analysis
Forensic Approaches
Supervised Learning Models
Feature Extraction
Captures subtle deepfake features in video frames
CNNs
RNNs
Vision Transformers(VIT)
Strengths: Highly effective for deepfake detection

Figure 1. Comparison of Detection Techniques

3.1 Traditional Methods

Traditional detection methods focus on identifying low-level artifacts and inconsistencies brought on by video manipulation. These techniques incorporate frame-analysis, motion analysis, and forensic approaches.

Frame based techniques involve examination of individual frames within a video to detect visual artifacts. Deepfake methods, mainly those that employ Generative Adversarial Networks (GANs), frequently introduce detectable anomalies such as image blurriness, edge irregularities, and illumination mismatches. These variations in facial texture, geometry, and shading between frames are extracted and analyzed using specialized algorithms.

Motion analysis target temporal inconsistencies across video sequences. Authentic human expressions, eye movements and head movements typically follow biometric rules, while deepfake content frequently doesn't. For example, the approach proposed by Demir and Çiftçi [10] employs motion magnification techniques to amplify minor distortions in facial movement, allowing the detection of synthetic manipulations that may not be perceptible to the human eye.

Forensic methods focus on spatial and temporal inconsistencies introduced during the synthesis and compression stages. These techniques examine irregularities in blinking patterns, eye and mouth textures, or distorted facial landmarks. The difference between head movements and facial landmark trajectories is a crucial sign of manipulation. In order to differentiate real from fake content, forensic analysis also assesses artifacts brought about by encoding and compression pipelines, particularly in multi-stage manipulations.

3.2 Machine Learning-Based Approaches

Machine learning methods have played a vital role in the development of deepfake detection, especially through supervised learning and effective feature extraction techniques. Supervised learning algorithms, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have proven to be highly effective in

DOI: https://doi.org/10.71058/jodac.v9i8015

spotting manipulated media. He et al. [11] introduced a gaze-guided spatial inconsistency learning technique based on CNNs, which significantly boosts detection performance by focusing on fine-level spatial inconsistencies.

Extracting distinctive features through feature extraction methods is crucial for differentiating genuine media from synthetic media. Studies have explored various physiological indicators such as blinking frequency and head motion. Sharma and Dwivedi [12] presented a hybrid architecture that integrates a Multi-Layer Perceptron (MLP)-CNN with LSTM networks to analyze blinking patterns, thereby increasing the reliability of fake media detection. The efficiency of detection models is generally measured using evaluation metrics such as accuracy, precision, and recall. Sharma and Dwivedi's hybrid model was validated on benchmark datasets like the World Leader Dataset (WLDR) and DeepfakeTIMIT, where it outperformed several existing methods [12].

3.3 Deep Learning-Based Approaches

Deep learning techniques have made significant progress in the field by identifying intricate temporal and spatial patterns that strongly suggest artificial manipulation. Several approaches as shown in Table 1 have been developed to enhance the effectiveness of deepfake video detection, with convolutional, recurrent, and transformer-based architectures showing particular promise.

Table 1. Summary of Deepfake Detection Methods

Method	Dataset(s)	Key Features	Limitation
Motion Magnification [10]	Multi-source DF datasets	Captures micro-motion via deep + phase magnification	Sensitive to quality, alignment
GazeForensics [11]	DF benchmarks	3D gaze + deep features	Depends on gaze accuracy
Multi-scale Conv+Transformer [13]	Celeb-DF v2, others	EfficientNet + multi-scale transformers	High computation
MeST-Former [14]	Large-scale benchmarks	Swin Transformer, ID-decoupled features	Poor generalization
Xception [15]	FF++, DFD, Celeb-DF, others	Depthwise conv., ensemble, stable features	High cost, dataset bias
CNN+Landmark [16]	DFDC (242 films, 318 samples)	Landmark-based CNN with augmentation	Landmark/dataset dependent
Systematic Review [17]	108 papers (CNN, RNN, GAN, etc.)	Highlights gaps in generalization & standards	Overfitting, no benchmarks

Convolutional Neural Networks (CNNs) are widely adopted for their ability to capture spatial correlations in images. Lin et al. [13] proposed a hybrid approach that combines multi-scale convolutions with Vision Transformers (ViTs), utilizing EfficientNet as a feature extractor to enhance detection accuracy. In another study, Gura et al. [16] introduced a customized CNN architecture that integrates facial landmark predictors, resulting in improved performance in detecting subtle facial manipulations. Similarly, Alkurdi et al. [15] employed the XceptionNet architecture, which leverages depth-wise separable convolutions, demonstrating robust performance across multiple benchmark datasets. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have been applied to analyze temporal inconsistencies in video sequences. These

DOI: https://doi.org/10.71058/jodac.v9i8015

models capture long-range frame dependencies, enabling the detection of irregular motion patterns and unnatural facial expressions that typically indicate manipulated content.

More recently, transformer-based models have shown significant improvements in both spatial and temporal modeling for deepfake detection. Vision Transformers (ViTs), in particular, have been successfully included into hybrid designs. For example, ViXNet combines ViTs with XceptionNet to capture subtle facial forgeries [17]. Additionally, MeST-Former, proposed by Liu et al. [14], incorporates motion-enhancement modules and spatiotemporal attention mechanisms to identify inconsistencies in facial movements and head pose dynamics. These transformer-based methods have demonstrated high efficiency, especially when applied to high-fidelity datasets.

3.4 Performance Metrics

To measure the effectiveness of deepfake detection models, several standard performance metrics are employed. These metrics, Equations (1)–(5), provide a thorough understanding of classification accuracy, reliability, and the balance between false positives and false negatives, while also emphasizing the correct identification of true positives.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

The overall proportion of correctly classified instances, combining both real and fake predictions gives us the accuracy.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Precision measures how many of the videos classified as fake are truly fake, indicating the reliability of positive predictions.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Recall, also known as sensitivity, reflects the capability of the model to correctly identify fake videos out of all actual fake instances.

$$F1 - Score = \frac{2*(Precision*Recall)}{Precision+Recall} \tag{4}$$

The F1-score offers a balanced assessment in situations where there is a class imbalance by providing a harmonic mean of Precision and Recall.

$$AUC = \int TPR(FPR)\, d(FPR) \tag{5}$$

Area Under the ROC Curve (AUC) measures the model's discriminative power across changing thresholds, with higher values indicating superior separability between real and fake videos.

The combined use of accuracy, precision, recall, F1-score, and AUC-ROC provides a holistic evaluation of deepfake detection models. Accuracy offers an overall measure of correctness, but on its own may be misleading in imbalanced datasets. Precision reflects the reliability of positive detections, while recall ensures that most manipulated content is successfully identified. Together, these metrics ensure a balanced assessment of correctness, reliability, sensitivity, and robustness, making them well-suited for real-world deepfake detection scenarios.

3.5 Datasets and Benchmarking

Benchmark datasets play a key role in evaluating the generalization and performance of detection models. The most prominent datasets are as follows.

DOI: https://doi.org/10.71058/jodac.v9i8015

- DeepFake Detection Challenge (DFDC): A large-scale benchmark introduced for the DFDC competition, offering a diverse set of manipulated videos for research and evaluation [18].
- FaceForensics++: Comprises video sequences generated using various manipulation methods, enabling multi-condition evaluation of detection algorithms. Table 2 provides a comparitive evaluation of different deepfake detection techniques on different datasets [19].
- Celeb-DF: Contains high-quality deepfake videos of celebrities, closely resembling real-world manipulated content and posing significant challenges due to the visual realism of the forgeries.

Table 2. Deepfake Detection Results on Different Datasets

Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score(%)	AUC
FaceForensics++ [5]	Xception	98.5	97.8	98.2	98.0	99.1
DFD [20]	EfficientNet-B4	96.3	95.1	96.5	95.8	97.4
Celeb-DF v2 [4]	MesoNet	89.4	87.6	88.9	88.2	90.1
WildDeepfake [21]	ViT	90.6	89.8	90.1	89.9	91.2
DeeperForensics [22]	Xception+	95.2	94.5	95.0	94.7	96.3

4. Challenges and Limitations

4.1 Technical Challenges

The rapid evolution of generative models has greatly complicated deepfake detection. Contemporary systems can synthesize videos that faithfully reproduce a person’s voice, micro-expressions, and ambient lighting, leaving detectors with very few exploitable artifacts. Table 3 summarizes the core capabilities and shortcomings reported in the literature.

4.2 Ethical and Legal Concerns

Deepfakes give rise to significant concerns across multiple industries. According to the data presented in [23], the rapid increase in deepfake activity highlights the growing threats associated with their creation and misuse. Malicious uses already include fabricating political speeches, impersonating executives in financial scams, and manipulating public sentiment during elections.

4.3 Limitations in Robustness Strategies

Although many deepfake detection models can be validated for real-world applications, current research still lacks a comprehensive model that effectively addresses all the major challenges as stated by [19] which are - transferability across datasets, interpretability of results, and resistance against adversarial attacks. These limitations make it extremely difficult to create and use detection frameworks in practical settings.

Table 3. Deepfake Detection Results on Different Datasets

Challenge	Description	Example Limitation	Solution

Generalization [16]	Poor cross-dataset performance	Xception on FF++ fails on Celeb-DF	Domain adaptation, transfer learning
Adversarial Robustness [9][24]	Susceptible to perturbations	Noise injection fools CNNs	Adversarial training, robust loss
Computational Cost [13][17]	High complexity, slow	Transformers need GPUs	Lightweight models (e.g., MobileNet+LSTM)
Interpretability [25][12]	Black-box predictions	CNN decisions unclear	Grad-CAM, attention maps
Multi-modal Manipulations [26][27]	Audio-visual fakes	Lip-sync bypasses visual detectors	Multi-modal fusion

For detection models to be helpful in real-world scenarios, their robustness has to be improved. A proposed method is to include a noise layer in the model, which enables it to learn from simulated data distortions and improves its resistance to quality degradations and adversarial attacks [26].

We can observe that the common limitations include poor generalization across datasets, sensitivity to data quality and compression, lack of explainability, and limited use of temporal relationships. Many models focus on frame-level analysis without capturing sequential inconsistencies present in the videos. Additionally, most of these result in high-computational costs and hinder real-world deployment.

To overcome these issues, we propose a lightweight yet effective hybrid architecture: a CNN backbone extracts spatial features, while a unidirectional LSTM captures temporal dynamics. This design leverages the strengths of both spatial and temporal modeling while maintaining computational efficiency. Our approach enables frame-level interpretability and supports sequence-aware predictions.

5. Proposed Implementation

We model deepfake video detection as a spatiotemporal classification task. A CNN extracts frame-level spatial features, which are aggregated by an LSTM to capture temporal dependencies. The overall process is formalized as follows from Equations (6)-(8).

5.1 Problem Formulation

Let $X = \{x_1, x_2, \ldots, x_T\}$ be a sequence of $T = 20$ face-centric video frames, where $x_t \in \mathbb{R}.^{112 \times 112 \times 3}$ each frame is encoded using a CNN:

$$h_t = \phi_{CNN}(x_t) \qquad (6)$$

Here, h_t is a feature vector that captures spatial information such as facial textures, lighting inconsistencies, and artifacts from the t- frame.

The temporal sequence $H = \{h1, h2, \ldots, h_T\}$ is then passed through an LSTM to obtain a video-level feature representation:

$$z = \phi_{LSTM}\ (H) \qquad (7)$$

The LSTM captures temporal dependencies and motion patterns across the video frames. The resulting vector z is a fixed-length representation summarizing the video's spatiotemporal characteristics

Finally, the probability is estimated via a fully connected layer with sigmoid activation:

$$\hat{y} = \sigma(Wz + b) \tag{8}$$

The predicted probability $\hat{y}$ indicates the likelihood that the input video is real. **Algorithm 1** presents the classification pipeline implementing the above formulation.

Algorithm 1: *Deepfake Video Classification Pipeline*

```
Input: Video $V$ with $N$ frames
Output: Prediction $y \in \{0,1\}$
Extract face frames $F = \{f_1, \dots, f_k\}$ from $V$
Resize each $f_i$ to $112 \times 112$, normalize
If $|F| < 20$:
    Pad to length 20
Augment frames (flip, color jitter)
for each frame $f_i$:
    $h_i \leftarrow \phi_{CNN}(f_i)$
Form sequence $H \leftarrow \{h_1, \dots, h_{20}\}$
Pass sequence through LSTM: $z = \phi_{LSTM}(H)$
Compute predicted probability: $\hat{y} = \sigma(Wz + b)$
Threshold output: $y = 1$ if $y > 0.5$, else $y = 0$
return y
```

5.2 Dataset Utilization

The proposed framework will be trained and evaluated using publicly available deepfake video datasets, including FaceForensics++, the Deepfake Detection Challenge Dataset (DFDC), and Celeb-DF. These datasets contain a variety of real and manipulated videos, varying in quality, compression levels, and generation techniques, thereby enabling robust model training and comprehensive evaluation across diverse scenarios.

5.3 Frame-Level Interpretation and Visualization

To enhance transparency and understand model decision-making, we incorporate both spatial and temporal interpretability mechanisms into our architecture. Temporal analysis is performed using the LSTM outputs, which provide per-frame predictions, enabling identification of video segments most likely to contain manipulations. For spatial interpretability, Grad-CAM is applied to the final layer of the CNN backbone. The resulting heatmaps highlight discriminative regions influencing the model's predictions. These heatmaps are further combined with facial landmarks and bounding boxes to localize artifacts such as distorted eyes, unnatural textures, or facial inconsistencies.

This interpretability framework validates model and provides visual evidence of the forgery, thereby improving trustworthiness and aiding future research directions.

5.4 Architectural Flexibility

The proposed framework, although initially based on a CNN-LSTM architecture, is intentionally designed to be modular and extensible. This design choice allows for future enhancements as the field evolves. Potential extensions include the integration of 3D Convolutional Neural Networks (3D CNNs) or Vision Transformers (ViTs) to enable more

DOI: https://doi.org/10.71058/jodac.v9i8015

effective spatiotemporal modeling. Additionally, attention mechanisms [24] may be incorporated to improve the model's ability to focus on critical temporal features across video frames. The framework also allows for the exploration of different strategies or hybrid techniques that combine deep learning techniques with other techniques. This architectural flexibility ensures long-term adaptability and relevance in the context of swiftly evolving deepfake generating techniques.

5.5 Expected Effectiveness

The proposed framework is expected to perform well due to its alignment with the core challenges of deepfake detection. CNN-based spatial encoding captures subtle pixel-level inconsistencies, while the LSTM aggregates these features across frames to detect irregular motion and temporal discrepancies. Evidence from earlier applications of CNN–LSTM hybrids in tasks such as face-anti-spoofing [28] has shown competitive performance, indicating that the same architecture is well suited for deepfake detection. Because the network learns both spatial cues and temporal dynamics, it should be more resilient than detectors that rely only on spatial features and should generalize across a wide range of manipulation techniques.

6. Future Directions

The swift progress of deepfake technology demands expanded research into dependable detection methods, real-time scalability, and solid policy structures. Hybrid strategies that blend deep learning with forensic analysis can improve accuracy, with explainable hierarchical designs showing encouraging potential [25]. Additionally, transfer learning boosts generalization, enabling CNN-based models to handle previously unseen datasets more effectively [16]. Lightweight configurations such as a MobileNet backbone paired with an LSTM and supplemented by synthetic data can achieve a practical trade-off among speed, accuracy, and robustness for real-time use while reducing memory and latency demands [29].

Beyond technical measures, a robust legal framework is vital to deter misuse [27] and regular collaboration between social media platforms and policy makers [30] is likewise needed to limit the spread of deepfake content.

7. Conclusion

Deepfake technology has evolved as an artistic innovation and a devastating security threat, setting the top agenda for researchers, industry players, and policymakers. This survey reviews a wide spectrum of detection strategies, stretching from classic forensic techniques to the latest deep-learning models built on CNNs, RNNs, and Vision Transformers. It stresses the importance of benchmark datasets and evaluation metrics, while pointing out ongoing challenges such as cross-dataset generalisation, adversarial robustness, explainability, and real-time deployment. To narrow these gaps, a hybrid CNN-LSTM framework is proposed, pairing spatial feature extraction with temporal sequence modelling to raise accuracy. The design is modular and transparent, making it straightforward to plug in newer components such as Vision Transformers or attention-based modules so the system can keep pace with the rapid evolution of deepfake generation methods.

References

[1] *Sunitha, K., Krishna, A.N. & Prasad, B.G. Copy-move tampering detection using keypoint based hybrid feature extraction and improved transformation model.*

DOI: https://doi.org/10.71058/jodac.v9i8015

Applied Intelligence, 2022, Volume 52, pp. 15405-15416. DOI: https://doi.org/10.1007/s10489-022-03207-x

[2] *Sunitha K, Krishna A N, "Efficient Keypoint Based Copy Move Forgery Detection Method using Hybrid Feature Extraction," Proceedings of 2nd IEEE International Conference on Innovative Mechanisms for Industry Applications, Bengaluru, March , pp.542-547,2020*

[3] *K. Liu, I. Perov, D. Gao, N. Chervoniy, W. Zhou, and W. Zhang, "DeepFaceLab: Integrated, flexible and extensible face-swapping framework," Pattern Recognition, vol. 141, p. 109628, 2023. [Online]. Available: https://doi.org/10.1016/j.patcog.2023.109628*

[4] *Y. Li, X. Yang, P. Sun, H. Qi and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 3204-3213, doi: 10.1109/CVPR42600.2020.00327.*

[5] *Rossler, A., et al.: Faceforensics++: Learning to detect manipulated facial images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1–11 (2019)*

[6] *R. Khan, M. Sohail, I. Usman, M. Sandhu, M. Raza, M. A. Yaqub, and A. Liotta, "Comparative study of deep learning techniques for DeepFake video detection," Computers & Security, vol. 118, p. 102731, 2022, doi: 10.1016/j.cose.2022.102731.*

[7] *I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Advances in neural information processing systems, vol. 27, 2014.*

[8] *M. Vashistha, S. Jain, S. Pandey, A. Pradhan, and S. Tarwani, "A Comparative Analysis of Machine Learning and Deep Learning Approaches in Deepfake Detection," in Proceedings of the 2024 IEEE Region 10 Symposium (TENSYMP), 2024, pp. 1–8. doi: 10.1109/TENSYMP61132.2024.10752209.*

[9] *F. Ding, G. Zhu, Y. Li, X. Zhang, P. K. Atrey, and S. Lyu, "Anti-forensics for face swapping videos via adversarial training," IEEE Transactions on Multimedia, vol. 24, pp. 3429–3441, 2022, doi: 10.1109/TMM.2021.3098422.*

[10] *Demir and U. A. Çiftçi, "How Do Deepfakes Move? Motion Magnification for Deepfake Source Detection," 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2024, pp. 4768-4778, doi: 10.1109/WACV57701.2024.00471*

[11] *Qinlin He, Chunlei Peng, Decheng Liu, Nannan Wang, Xinbo Gao, GazeForensics: DeepFake detection via gaze-guided spatial inconsistency learning, Neural Networks, Volume 180, 2024, 106636, ISSN 0893-6080, doi: 10.1016/j.neunet.2024.106636*

[12] *Ruchika Sharma, Rudresh Dwivedi, Unmasking deepfakes: Eye blink pattern analysis using a hybrid LSTM and MLP-CNN model,Image and Vision Computing,Volume 154, 2025, 105370, ISSN 0262-8856, doi: 10.1016/j.imavis.2024.105370*

[13] *Hao Lin, Wenmin Huang, Weiqi Luo, Wei Lu, DeepFake detection with multi-scale convolution and vision transformer, Digital Signal Processing, Volume 134, 2023, 103895, ISSN 1051-2004, doi: 10.1016/j.dsp.2022.103895*

[14] *Baoping Liu, Bo Liu, Ming Ding, Tianqing Zhu, MeST-Former: Motion-enhanced Spatiotemporal Transformer for generalizable Deepfake detection, Neurocomputing, Volume 610, 2024, 128588, ISSN 0925-2312, doi: 10.1016/j.neucom.2024.128588.*

[15] *Dunya Ahmed Alkurdi, Mesut Cevik, Abdurrahim Akgundogdu, Advancing Deepfake Detection Using Xception Architecture: A Robust Approach for Safeguarding against Fabricated News on Social Media, Computers, Materials*

DOI: https://doi.org/10.71058/jodac.v9i8015

and Continua, Volume 81, Issue 3, 2024, Pages 4285-4305, ISSN 1546-2218, doi: 10.32604/cmc.2024.057029

[16] *Dmitry Gura, Bo Dong, Duaa Mehiar, Nidal Al Said, Customized Convolutional Neural Network for Accurate Detection of Deep Fake Images in Video Collections, Computers, Materials and Continua, Volume 79, Issue 2, 2024, Pages 1995-2014, ISSN 1546-2218, doi: 10.32604/cmc.2024.048238.*

[17] *Ramcharan Ramanaharan, Deepani B. Guruge, Johnson I. Agbinya, DeepFake video detection: Insights into model generalisation — A Systematic review, Data and Information Management,2025, 100099, ISSN 2543-9251, doi: 10.1016/j.dim.2025.100099.*

[18] *Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., & Ferrer, C. C. (2020). The DeepFake Detection Challenge (DFDC) dataset. arXiv preprint arXiv:2006.07397.*

[19] *Wang, Tianyi, Xin Liao, Kam Pui Chow, Xiaodong Lin, and Yinglong Wang. "Deepfake detection: A comprehensive survey from the reliability perspective." ACM Computing Surveys 57, no. 3 (2024): 1-35.*

[20] *Pokroy, Artem A. and Alexey Egorov. "EfficientNets for DeepFake Detection: Comparison of Pretrained Models." 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus) (2021): 598-600.*

[21] *Nguyen, Dat, et al. "Fakeformer: Efficient vulnerability-driven transformers for generalisable deepfake detection." arXiv preprint arXiv:2410.21964 (2024).*

[22] *Jiang, Liming, et al. "Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.*

[23] *S. P. Koritala, M. Chimata, S. N. Polavarapu, B. S. Vangapandu, T. K. Gogineni and V. M. Manikandan, "A Deepfake detection technique using Recurrent Neural Network and EfficientNet," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-6, doi: 10.1109/ICCCNT61001.2024.10723875.*

[24] *W. Lu et al., "Detection of Deepfake Videos Using Long-Distance Attention," in IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 7, pp. 9366-9379, July 2024, doi: 10.1109/TNNLS.2022.3233063*

[25] *Samuel Henrique Silva, Mazal Bethany, Alexis Megan Votto, Ian Henry Scarff, Nicole Beebe, Peyman Najafirad, Deepfake forensics analysis: An explainable hierarchical ensemble of weakly supervised models, Forensic Science International: Synergy, Volume 4, 2022, 100217, ISSN 2589-871X,*

[26] *P. Yu, Z. Xia, J. Fei, and Y. Lu, "A Survey on Deepfake Video Detection," IET Biometrics, vol. 10, no. 6, pp. 607–624, Nov. 2021, doi: 10.1049/bme2.12031. https://doi.org/10.1016/j.fsisyn.2022.100217.*

[27] *Gueltoum Bendiab, Houda Haiouni, Isidoros Moulas, Stavros Shiaeles, Deepfakes in digital media forensics: Generation, AI-based detection and challenges, Journal of Information Security and Applications, Volume 88, 2025,103935, ISSN 2214-2126,*

[28] *H. Li, P. He, S. Wang, A. Rocha, X. Jiang and A. C. Kot, "Learning Generalized Deep Feature Representation for Face Anti-Spoofing," in IEEE Transactions on Information Forensics and Security, vol. 13, no. 10, pp. 2639-2652, Oct. 2018, doi: 10.1109/TIFS.2018.2825949.*

[29] *Abdelwahab Almestekawy, Hala H. Zayed, Ahmed Taha, Deepfake detection: Enhancing performance with spatiotemporal texture and deep learning feature fusion, Egyptian Informatics Journal, Volume 27, 2024, 100535, ISSN 1110-8665*

[30] *Walter Matli, Extending the theory of information poverty to deepfake technology, International Journal of Information Management Data Insights, Volume 4, Issue 2, 2024, 100286, ISSN 2667-0968, https://doi.org/10.1016/j.jjimei.2024.100286.*