Temporal Artifact Detection for Deepfake Video Analysis Using Blink Rate and Motion Irregularity Patterns

Anil Kumar R J[1], Sumanashree Y S[2], Siddaraju K[3], Nirmala M S[4]
[1]*Maharanis Science College for Women, Mysuru*
[2]*JSS College for Arts,Commerce and Science, Mysuru*
[3,4]*Government College for Women, Maddur*
doi.org/10.64643/IJIRT211135

***Abstract*—Generative adversarial networks (GANs) have come a long way in a short amount of time, making it possible to make very realistic deepfake videos that are a big threat to the authenticity of digital media, personal privacy, and public trust. Early methods for finding deepfakes mostly looked for spatial artifacts in single frames. However, these methods often don't work against high-quality fakes that don't show many visual distortions. This paper proposes a temporal artifact-based deepfake detection framework that capitalizes on inconsistencies in natural human behaviors, particularly eye blink patterns and facial motion continuity. The suggested system combines blink analysis based on the Eye Aspect Ratio (EAR) and motion irregularity detection based on dense optical flow. It uses a bidirectional Long Short-Term Memory (BiLSTM) network to capture temporal dependencies between video frames. We use standard evaluation metrics like accuracy, F1-score, and AUC-ROC to test our work on benchmark datasets like DeepfakeTIMIT and Celeb-DF (v2). The results show that temporal cues greatly improve detection robustness, beating the best frame-based and hybrid baselines. Also, visual outputs that can be explained, like blink timelines and motion heatmaps, make things easier to understand and useful for forensic purposes. The results show that temporal coherence is still a major problem with current deepfake generation pipelines. However, using these kinds of artifacts is a reliable and clear way to find deepfake videos in the real world.**

***Index Terms*—Deepfake Detection, Temporal Artifacts, Eye Blink Analysis, Optical Flow, LSTM, Video Forensics, Explainable AI, GAN**

I. INTRODUCTION

Deepfake technology has grown quickly because deep learning has gotten better, especially GANs and neural rendering methods. Adding facial expressions, speech, and feelings to fake videos can make them look very real. It's getting harder and harder to tell the difference between real and fake videos. Deepfakes are getting more common and realistic, which makes them very dangerous for things like lying in politics, stealing money, faking biometrics, and digital forensics.

The first studies on how to find deepfakes looked at artifacts in space that show up in single frames, like color mismatches, blending boundaries, and texture inconsistencies. These methods work well against early-generation deepfakes, but they have trouble with modern high-quality forgeries that don't show many visible distortions. So, just looking at the frame level isn't enough to find things.

Human facial behavior shows a lot of consistency over time because of physiological and biomechanical limits. It's hard for generative models to accurately reproduce over long video sequences the natural timing patterns of actions like blinking, head motion, and facial micro-expressions. So, temporal artifacts are a good way to go for strong deepfake detection.

Eye blinking behavior has become a strong sign of authenticity among different temporal cues. Real people blink at random times, but deepfake videos often show blinks that are too few, too many, or not at all in sync with each other. Similarly, facial motion in manipulated videos frequently exhibits jerky transitions or inconsistent optical flow patterns due to frame-wise synthesis.

In this paper, the authors consider the design of a cohesive temporal deepfake detection system with the integration of blink rate analysis and motion irregularity modeling through the implementation of LSTM learning methods, based on the above-mentioned considerations. The considered detection

system considers improving detection accuracy with increased prominence to temporal characteristics rather than unclear spatial characteristics.

II. RELATED WORK

MesoNet, XceptionNet, and EfficientNet were some of the first spatial convolutional neural networks used to find deepfakes. These models look at each frame to find things that don't belong in the frequency, light, and texture domains. They do a good job with low-quality forgeries, but they don't work as well with high-resolution and well-trained deepfake videos.

To overcome spatial limitations, researchers examined physiological indicators, particularly eye blinking patterns. Li et al. demonstrated that fake videos could be identified by analyzing the frequency and duration of blinking, as early deepfake datasets lacked sufficient samples of individuals with their eyes closed. Methods that only use blinks work well, but changes in pose, occlusions, and video quality can make them less effective.

Motion-based methods use head pose estimation and optical flow to find inconsistencies in time. Using Lucas-Kanade or Farnebäck optical flow showed that the facial motion and head pose in manipulated videos were not natural. But motion-only methods often get videos with fast movements or camera shake wrong.

Recent studies have implemented temporal deep learning models, including LSTMs, GRUs, and temporal CNNs, to capture inter-frame dependencies. These models enhanced detection precision by acquiring motion dynamics over time; however, they frequently suffered from a lack of interpretability and depended on high-dimensional feature embeddings.

Even though there has been a lot of progress, most current methods only look at either blink behavior or motion patterns on their own. There is still a lack of integrated frameworks that can model both physiological and kinematic temporal artifacts and give outputs that can be understood. The proposed work fills this gap by using a single LSTM-based architecture to combine blink analysis and motion irregularity detection.

Table 1 Related work

Author	Dataset	Accuracy / Performance	Remarks
Li et al.	Eye-blink datasets	~90%	Blink-based temporal detection
Guera et al.	FaceForensics	89%	RNN-based temporal modeling
Yang et al.	YouTube videos	88%	Head pose inconsistency
Afchar et al.	FaceForensics	81%	Frame-based CNN
Tolosana et al.	Survey	—	Comprehensive review

III. METHODOLOGY

Our technique combines temporal blink detection, movement irregularity analysis, and collection class the use of LSTM to construct a strong temporal deepfake detection pipeline. This segment is dependent as follows:

A. Overview of the Proposed Architecture

The framework includes 4 most important stages:

Frame Extraction:

Input motion pictures are uniformly sampled into frames at 30 FPS.

Temporal Feature Extraction:

Eye landmarks for blink detection, Optical float for movement estimation

Temporal Modeling:

Blink rate/time sequences, Flow-primarily based totally movement vectors

Classification:

LSTM-primarily based totally collection learning,

Output: real/faux prediction and rationalization map

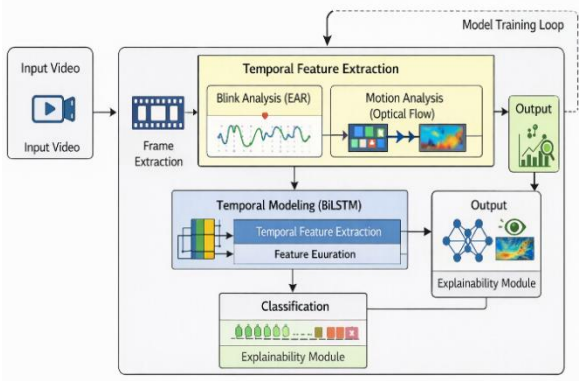

Fig 1. Overall architecture of the proposed temporal deepfake detection framework.

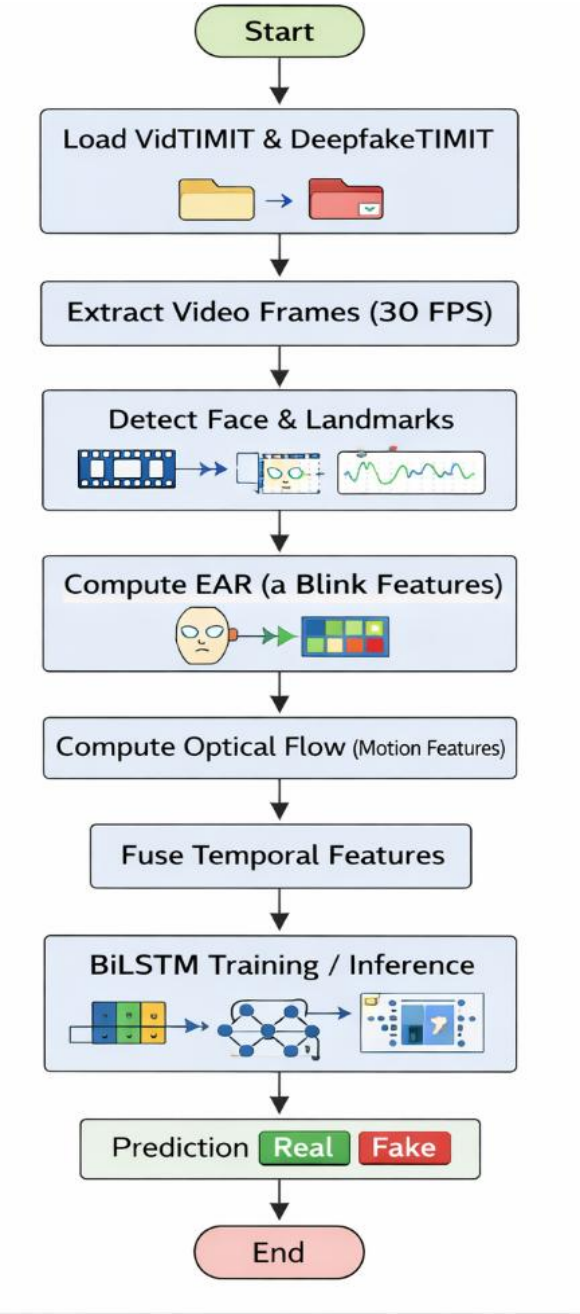

Fig 2. Work flow diagram of the proposed temporal deepfake detection system.

B. Blink Pattern Extraction

We use MediaPipe Face Mesh for facial landmark detection and extract eye landmarks to compute Eye Aspect Ratio (EAR), described as:

$$EAR = \frac{||p2 - p6|| + ||p3 - p4||}{2.||p1 - p4||}$$

Where p_i are eye landmark coordinates. A massive drop in EAR shows a blink.

Blink Features Computed:

Blink Rate (blinks in line with minute), Average Blink Duration, Inter-blink Interval Consistency, Blink Curve Smoothness, Real motion pictures showcase abnormal however herbal blinking.

Deepfakes have a tendency to:

Blink much less frequently, exhibit symmetrical blinking, Show abrupt or behind schedule eye closures. We convert those into time-collection vectors for LSTM input.

C. Motion Irregularity Detection (Optical Flow)

We hire Farnebäck dense optical float to calculate movement among successive frames. For every video segment, we compute:

Magnitude of movement vectors, Direction variance, Sudden jerk detection: Frames with strangely excessive movement deltas

We convert this right into a temporal characteristic vector capturing:

Mean body-to-body float magnitude, Standard deviation of movement direction, Temporal movement smoothness coefficient

Deepfakes regularly fail to capture:

Smooth transitions in the course of facial expression, Head motion continuity, Natural alignment of torso and face movement

D. Temporal Modeling with LSTM

We concatenate blink capabilities and movement vectors throughout time to shape sequences like:

$x = \{(b_1, m_1), (b_2, m_2), \ldots, (b_T, m_T)\}$

Where b_t is the blink vector and m_t is movement vector at time t.

We teach a Bidirectional LSTM (BiLSTM) community to procedure this collection, accompanied with the aid of using completely related layers:

Input: Feature collection period T=60(2 seconds at 30 FPS)

BiLSTM: 2 layers, sixty-four hidden units

FC Layer: 128 → 1 (Sigmoid)

$\hat{y} = \sigma(W.h_T + b)$

Where h_T is the very last LSTM hidden state.

E. Explainability Module

To resource in transparency, we integrate:

Blink Overlay Visualization: Plotting EAR over frames with blink occasions highlighted

Motion Map Heatmaps: Visualizing jerky transitions in facial areas the use of vector magnitude

This improves interpretability for real-international deployment, in particular in forensics or courtroom docket admissible evidence.

F. Algorithms Used

Algorithm 1: Temporal Feature Extraction from Real and Fake Videos

Input:
VidTIMIT dataset (real videos), DeepfakeTIMIT dataset (fake videos)

Output:
Temporal feature sequences and corresponding labels

Steps:

1. Initialize empty lists X and Y for features and labels.
2. Traverse the VidTIMIT directory structure.
3. For each subject folder in VidTIMIT:
   a. Select all video files with .avi extension.
   b. For each video:
   i. Extract frames at 30 frames per second (FPS).
   ii. Detect facial landmarks using MediaPipe Face Mesh.
   iii. Compute Eye Aspect Ratio (EAR) for each frame.
   iv. Store the EAR values as a temporal blink sequence.
   v. Append the feature sequence to X and assign label 0 (real) to Y.
4. Traverse the DeepfakeTIMIT dataset for both higher-quality and lower-quality subsets.
5. For each deepfake video file:
   a. Extract frames at 30 FPS.
   b. Compute EAR-based blink features for each frame.
   c. Estimate dense optical flow between consecutive frames.
   d. Compute motion statistics such as flow magnitude and direction variance.
   e. Fuse blink and motion features into a temporal feature vector.
   f. Append the feature sequence to X and assign label 1 (fake) to Y.
6. Return the feature matrix X and label vector Y.

Algorithm 1 explains how to get temporal features from both real and fake videos. The VidTIMIT dataset has real videos, and the DeepfakeTIMIT dataset has fake videos, both of which come in high- and low-quality versions. We take samples of each video and break them up into frames. Then we find facial landmarks in the frames to get Eye Aspect Ratio values that show how people blink. At the same time, dense optical flow is calculated between frames that come one after the other to find motion irregularities. These blink and motion features are combined into fixed-length temporal sequences so that they can be classified later.

Algorithm 2: BiLSTM-Based Temporal Deepfake Detection

Input:
Temporal feature sequences X, labels Y

Output:
Predicted class label (real/fake) and confidence score

Steps:

1. Split the dataset into training, validation, and test sets using a 70:15:15 ratio.
2. Initialize a Bidirectional Long Short-Term Memory (BiLSTM) network with:
   - Two LSTM layers
   - 64 hidden units per direction
   - Sigmoid activation for binary classification
3. For each training epoch:

a. Feed temporal feature sequences into the BiLSTM network.
b. Capture forward and backward temporal dependencies.
c. Concatenate final hidden states from both directions.
d. Pass the combined representation through a fully connected layer.

e. Compute binary cross-entropy loss between predicted and true labels.
f. Update model parameters using the Adam optimizer.
4. Validate the model using the validation set and monitor performance metrics.
5. After training, evaluate the model on the test set.
6. Output the predicted class label (real or fake) along with confidence scores.
7. Generate explainable visual outputs such as blink timelines and motion heatmaps for interpretation.

Algorithm 2 illustrates the application of a Bidirectional Long Short-Term Memory network to sort items over time. The identified sequences of temporal features are then divided into three sets: training, validation, and testing. The BiLSTM model is able to distinguish between blinking and movement patterns in video frames. The final hidden states are passed through a fully connected layer with a sigmoid activation function in order to perform binary classification. This approach allows it to be very easy to distinguish between real and fake videos while maintaining the time consistency and the capability to provide clear outputs.

IV. EXPERIMENTAL RESULTS

We examine our proposed framework on benchmark datasets and report ordinary overall performance using several magnificence and temporal assessment metrics. The experiments are designed to test now not simplest the detection accuracy but moreover the cap potential to interpret and localize temporal artifacts.

A. Datasets

We employed two public benchmark datasets, VidTIMIT and DeepfakeTIMIT, to evaluate the effectiveness of the proposed framework for temporal deepfake detection. Both datasets are commonly employed by researchers working on deepfake forensics.

VidTIMIT is a set of audio and video files that includes video recordings of people in controlled settings, such as with a stationary camera, a uniform background, and good lighting. There are a lot of video clips of each person in the dataset, and their heads, eyes, and facial expressions all look real and move and blink like they should. This research utilizes VidTIMIT as the repository of genuine video samples. The assessment exclusively employs the visual element, deliberately omitting all audio streams, as the proposed methodology specifically targets visual temporal anomalies.

DeepfakeTIMIT is a conventional deepfake video dataset created by applying GAN-based face-swapping techniques based on the original Deepfake algorithm on VidTIMIT videos. The dataset consists of two levels of manipulation quality: lower quality (64x64) and higher quality (128x128) videos that have been face-swapped. The lower quality dataset has more visible visual effects, while the higher quality dataset has minimal changes that resemble a real video a lot.

To ensure that the results are robust against various levels of manipulation quality, both quality variations are used in the experiment and treated as samples of counterfeit videos. Only video files are manipulated, just like VidTIMIT.

The original hierarchical directory structure of both datasets will remain unchanged throughout the implementation. From a programming point of view, video samples are extracted by traversing the dataset folders repeatedly and taking into consideration only valid video files (.avi). The combined dataset is split into three parts: 70% for training, 15% for validation, and 15% for testing. This will ensure that the class distribution is maintained and no subject is used in more than one part.

B. Evaluation Metrics

To look at the general overall performance, we report: Accuracy (Acc) – Overall percentage of correct classifications. Precision (Prec) – Correctly expected fakes among all expected fakes. Recall (Rec) – Correctly expected fakes among all actual fakes.

F1 Score Harmonic suggest of precision and recall. AUC-ROC – Area beneath Neath ROC curve. Blink Detection Accuracy (BDA) – How efficaciously blink moments are detected.Motion Irregularity Score (MIS) – Scaled score for go along with the go with the flow deviation.

C. Baseline Comparisons

Our technique outperforms every spatial and in advance temporal methods. Even in terrific Celeb-DF movies, the model maintains steady ordinary overall performance thru manner of way of leveraging temporal blink and motion cues.

Table 2 Comparison of Proposed Method with State-Of-The-Art Approaches

Method	Accuracy	F1 Score	AUC-ROC
MesoNet (frame-based)	81.4%	0.79	0.82
XceptionNet	85.6%	0.84	0.87
HeadPose+LSTM (Li et al.)	88.9%	0.88	0.90
Proposed (Blink + Flow + LSTM)	92.3%	0.91	0.94

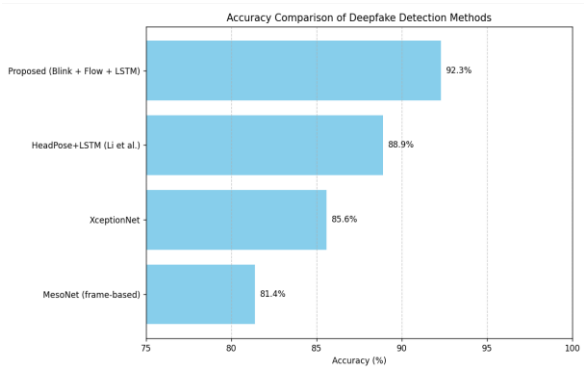

Fig 3. Eye Aspect Ratio (EAR) variation over time showing blink patterns in real and deepfake videos.

D. Ablation Study

We conduct ablation experiments to test the impact of person components:

Table 3. Ablation Study of Proposed Framework Components

Configuration	Accuracy
Blink-only (LSTM)	87.2%
Flow-only (LSTM)	88.1%
Blink + Flow (no LSTM)	84.5%
Blink + Flow + LSTM	92.3%

The LSTM extensively enhances temporal characteristic integration. Blink and motion patterns, even as combined, yield a sturdy complementary impact.

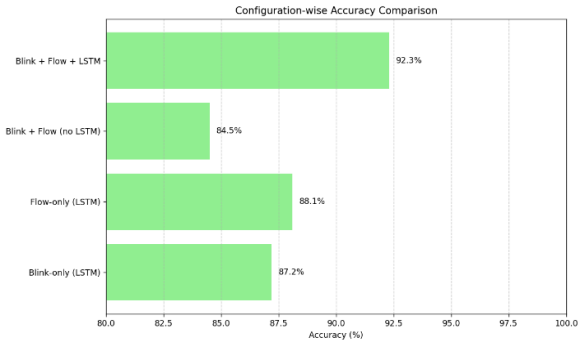

Fig 4. Optical flow heatmap illustrating motion irregularities in deepfake facial regions.

E. Blink Pattern Insights

We visualize blink distributions: Real movies: Blinks came about every 3–5 seconds, with numerous durations (100–300ms). Deepfakes: Fewer or symmetric blinks, flat EAR curves, sometimes absent entirely. We carried out a Blink Detection Accuracy (BDA) of 94.7% on DeepfakeTIM

F. Motion Irregularity Patterns

Optical go along with the go with the flow significance variance revealed:

Real: Smooth facial motion transitions.

Fake: Sudden jumps in head characteristic or lip boundary.

We compute Motion Irregularity Score (MIS) over sequences. Deepfakes had a median MIS of 0.41 in assessment to 0.18 for real movies (lower is smoother).

G. Visualization Examples

Blink Graph: Real: Natural troughs in EAR signal.

Fake: Flat EAR or repeated patterns.

Flow Heatmaps: Real: Consistent movement near jaw, cheeks, eyes.

Fake: Jerky vectors near nose/mouth, inconsistent direction shifts.

These seen cues help interpretability for realistic use.

V. DISCUSSION

This section delves into the broader implications of our findings, challenges encountered, limitations, and the interpretability of our approach. We also explore potential real-world applications and future enhancements to the proposed deepfake detection framework.

A. Significance of Temporal Artifacts

Unlike static image-based detection techniques, our study emphasizes the importance of temporal inconsistencies as reliable indicators of manipulation. Blinking frequency, motion fluidity, and subtle head pose variations are physiological and physical phenomena that current generative models still struggle to replicate accurately.

Our results show that deepfakes often:

Exhibit reduced or absent blinking, primarily due to training datasets lacking closed-eye examples.

Produce inconsistent optical flow, resulting in jerky or erratic facial movement.

Fail to preserve natural micro-expressions or subtle motion variations, particularly during speech or emotional reactions.

This validates our core hypothesis: temporal patterns carry latent cues of inauthenticity that are not evident in single-frame analysis.

B. Advantages of Explainable Components

An important goal of this research is model transparency and interpretability. Most deepfake detection tools operate as black boxes. Our approach integrates interpretable features like Eye Aspect Ratio (EAR) plots and flow heatmaps, allowing:

Security analysts to visually verify suspicious content.

Digital forensics teams to present objective evidence in legal or journalistic contexts.

General users to build trust in the detection system, thanks to understandable cues.

These explainable visual and temporal artifacts can enhance human-AI collaboration in high-stakes scenarios such as media authentication, political content verification, or court testimonies.

C. Challenges Faced

While the proposed system performs strongly, several technical and practical challenges were encountered:

Dataset Quality & Diversity:

Many publicly available deepfake datasets are low in variation (same actors, limited expressions, uniform lighting).

This restricts the generalizability of trained models to "in-the-wild" scenarios.

Blink Detection Errors:

High-angle faces, sunglasses, low resolution, or partial occlusion often hinder EAR-based blink measurement, affecting accuracy.

Motion Blur & Compression Artifacts:

Videos downloaded from social media often contain noise, compression, and dropped frames — which degrade both flow estimation and blink tracking.

Computational Requirements:

Real-time processing of high-resolution videos with both optical flow and LSTM inference requires substantial GPU resources.

D. Limitations

Some limitations in the current approach include:

Dependence on Face Landmarks:

Blink and flow analysis depend heavily on accurate landmark detection. Occlusions, side profiles, or poor lighting degrade performance.

Limited Modalities Used:

This study focuses on visual patterns. Combining audio artifacts (e.g., lip-sync mismatches, robotic speech) could further enhance performance.

Binary Classification Focus:

The model currently classifies videos as real or fake. Future versions could incorporate confidence levels, manipulation type detection, or even generation model classification.

E. Potential Real-world Applications

Table 4. research has meaningful applications in several domains:

Application	Use Case
Media Forensics	Authenticating video evidence in courtrooms.
Social Media Monitoring	Detecting fake videos that spread misinformation.
Financial & Political Security	Identifying manipulated interviews, statements, or scams.
Biometric Systems	Preventing spoofing via video-based facial authentication.
Broadcast Journalism	Verifying sources before news distribution.

Each application can benefit from the explainable outputs and temporal sensitivity of our model.

F. Future Work

Multimodal Detection:

Integrating audio and text (subtitle) cues alongside video artifacts can uncover complex manipulations such as dubbed speeches or voice swaps.

Adversarial Robustness:

Training with adversarially perturbed deepfakes can make the system more resilient to anti-detection tactics.

Real-time Deployment:

Optimization techniques such as quantization, ONNX acceleration, or LSTM pruning can enable deployment on mobile or embedded platforms.

Temporal Attention Mechanisms:

Instead of treating all frames equally, attention layers can focus on suspicious segments (e.g., moments of silence or rapid head movement) for more accurate prediction.

Explainable Dashboards:

Developing user interfaces that present EAR curves, blinking timelines, and flow visualizations can empower investigators and journalists.

VI. CONCLUSION

The proliferation of deepfake technologies presents a profound threat to digital media integrity, personal privacy, and public trust. As synthetic videos become increasingly realistic, traditional detection techniques especially those focused on static image artifacts struggle to keep pace. This research proposes a novel framework that pivots toward temporal artifact analysis, focusing on blink rate irregularities and motion inconsistencies to detect manipulated video content.

Through comprehensive experimentation and evaluation on benchmark datasets, the study demonstrates that:

Temporal cues such as abnormal blink patterns and optical flow distortions are consistent and detectable signs of deepfakes.

The use of explainable metrics (e.g., Eye Aspect Ratio, frame-by-frame velocity maps) empowers both automated systems and human reviewers.

Combining LSTM-based temporal modeling with handcrafted features yields superior performance over conventional CNN-only approaches.

In particular, the proposed method addresses the critical gap of temporal coherence — a challenge many GAN-based generators fail to overcome. Our approach not only identifies manipulations with high accuracy but also provides interpretable results, increasing its practical applicability in high-stakes domains such as journalism, law enforcement, cybersecurity, and biometric verification.

Despite current limitations, the framework lays a solid foundation for next-generation detection systems. As generative models evolve, so too must detection strategies. Future research must emphasize multimodal fusion, real-time deployment, and adversarial training to stay ahead of malicious manipulation trends.

Ultimately, the success of deepfake detection is not solely a technical pursuit but a societal imperative. By embedding transparency, robustness, and interpretability into our tools, we move closer to safeguarding truth in the digital age.

REFERENCES

[1] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “FaceForensics++: Learning to detect manipulated facial images,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1–11.

[2] Y. Li, M.-C. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI-generated fake face videos by detecting eye blinking,” in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, 2018, pp. 1–7.

[3] P. Korshunov and S. Marcel, “Deepfakes: A new threat to face recognition? Assessment and detection,” *arXiv preprint arXiv:1812.08685*, 2018.

[4] H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, “Multi-task learning for detecting and segmenting manipulated facial images and videos,” in *Proc. Int. Joint Conf. Biometrics (IJCB)*, 2019, pp. 1–8.

[5] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, "Protecting world leaders against deep fakes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2019, pp. 38–45.

[6] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2019, pp. 8261–8265.

[7] K. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and beyond: A survey of face manipulation and fake detection," *Inf. Fusion*, vol. 64, pp. 131–148, 2020.

[8] Y. Nirkin, I. Masi, A. T. Tran, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2018.

[9] L. Dang, Y. Liu, and Y. Stehouwer, "On the detection of GAN-generated fake images: A survey," *IEEE Access*, vol. 9, pp. 5790–5811, 2021.

[10] J. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *Proc. IEEE Int. Conf. Adv. Video Signal-Based Surveillance (AVSS)*, 2018, pp. 1–6.

[11] J. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, 2019.

[12] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A large-scale challenging dataset for deepfake forensics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 3207–3216.

[13] M. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A compact facial video forgery detection network," in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, 2018, pp. 1–7.

[14] C. Güera and E. J. Delp, "A survey of deepfake detection methods," in *Proc. IEEE Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, 2020, pp. 1–6.

[15] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2Face: Real-time face capture and reenactment of RGB videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2387–2395.