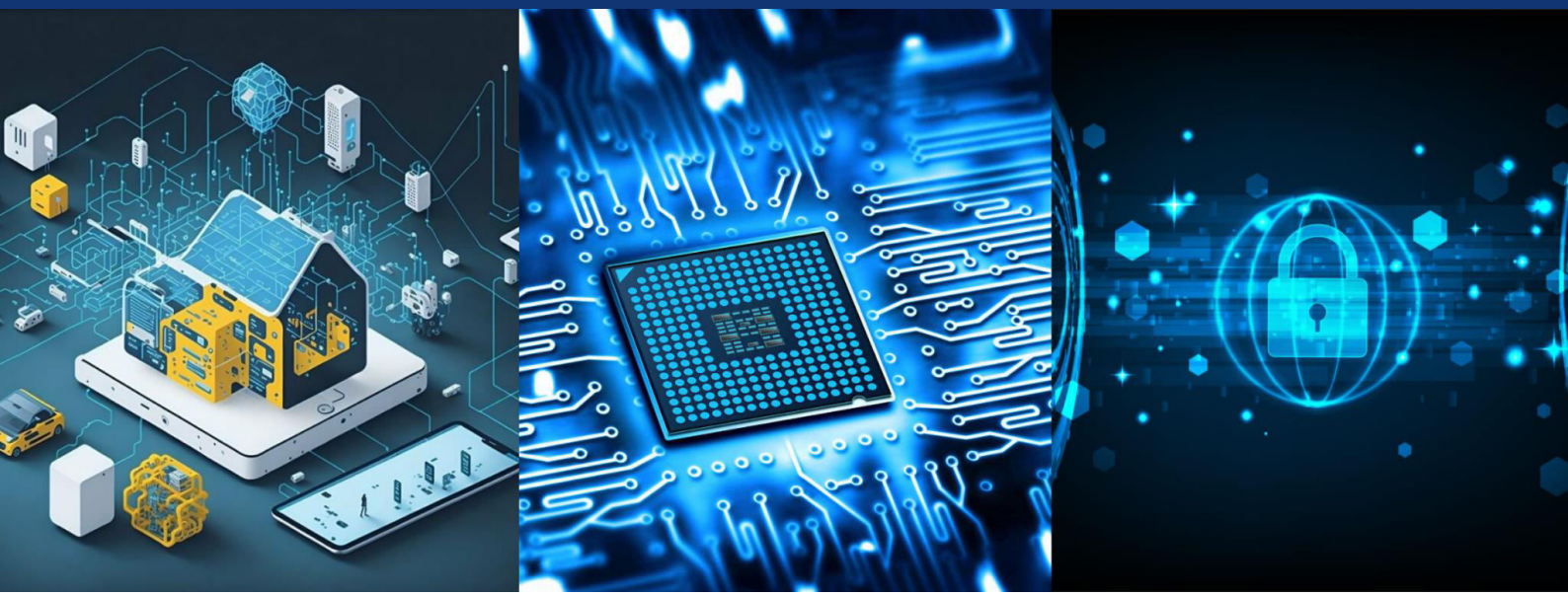
IJIRCCE

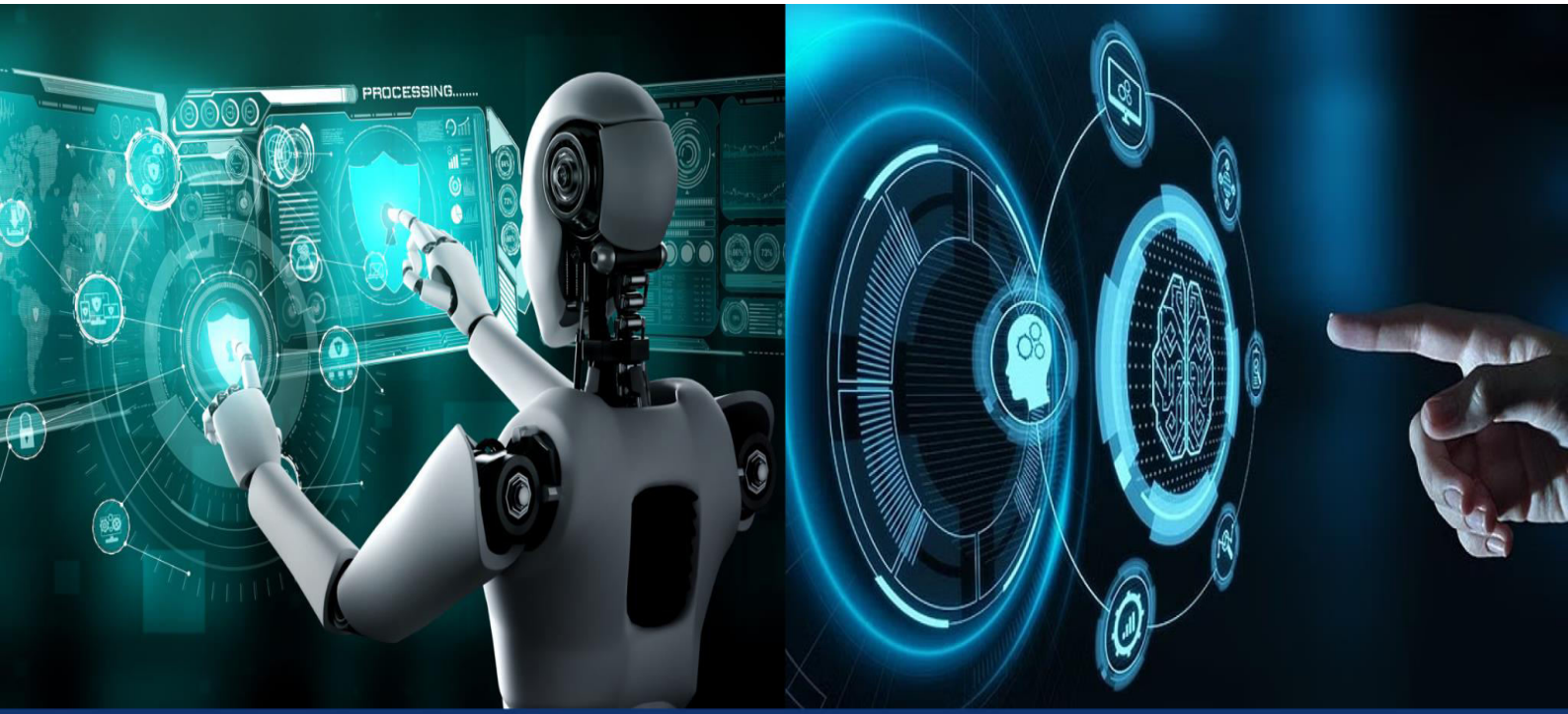
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 5, May 2026

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405072

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Deepfake Video Detection Using ResNet-Based Feature Extraction and Temporal Modeling

Mrs Asha N S, Renuka HS, Ramya S Patil, Varsha P, Tejashwini B Kumbar

Assistant Professor, Department of Information Science and Engineering, Jain Institute of Technology, Davanagere, Karnataka, India

UG Student, Department of Information Science and Engineering, Jain Institute of Technology, Davanagere, Karnataka, India

UG Student, Department of Information Science and Engineering, Jain Institute of Technology, Davanagere, Karnataka, India

UG Student, Department of Information Science and Engineering, Jain Institute of Technology, Davanagere, Karnataka, India

UG Student, Department of Information Science and Engineering, Jain Institute of Technology, Davanagere, Karnataka, India

ABSTRACT: Continuous progress in AI-based media generation has transformed deepfake videos into a mounting threat against the integrity of digital information. Synthetic video content crafted via generative adversarial networks and associated neural architectures has achieved a degree of visual realism that renders manual inspection unreliable as a standalone verification strategy. The societal risks extend from harm to individual reputations to coordinated disinformation campaigns operating at unprecedented scale. In response to these concerns, this paper introduces a detection framework that merges convolutional neural network (CNN) based spatial feature analysis with recurrent neural network (RNN) based temporal reasoning, equipping the system to identify manipulation evidence both at the individual frame level and across sequential video segments. The resulting tool is engineered to serve practical deployment needs within cybersecurity operations and fact-verification environments where scalable content authenticity assessment is required.

KEYWORDS: Deepfake Detection, Machine Learning, CNN, RNN, Video Forensics, Face Manipulation, Deep Learning

I. INTRODUCTION

Over the past decade, the landscape of synthetic media creation has shifted from a niche research topic to a widespread practical capability accessible to non-expert users with modest hardware. Tools leveraging GAN-based video synthesis can now be operated without deep technical expertise, fueling a significant increase in manipulated video content circulating across online platforms.

The consequences of this trend are well established. Fabricated video footage has been weaponized in political interference campaigns, financial fraud, online harassment, and impersonation attacks. As the visual quality of AI-synthesized content continues to improve, the perceptual boundary separating authentic recordings from artificial constructs diminishes—placing both ordinary viewers and professional content moderators in an increasingly untenable verification position.

Algorithmic detection methods present a scalable solution to this problem. Frame-level CNN examination effectively surfaces pixel-domain irregularities linked to facial synthesis operations, while LSTM-driven temporal analysis can expose inconsistencies within physiological motion patterns—such as irregular eye movements or asynchronous facial

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

dynamics—that only become apparent when evaluating a contiguous sequence of frames. This work integrates both approaches into a unified, deployment-ready detection system.

II. LITERATURE REVIEW

[1] Afchar et al. (2018) introduced MesoNet, a compact convolutional architecture purpose-built for identifying manipulated facial video. The design targets medium-granularity texture patterns and blending artifacts characteristic of GAN-derived content, achieving computational efficiency suitable for near real-time operation.
[2] Güera and Delp (2018) proposed a two-stage detection approach that pairs CNN-based spatial encoding with LSTM temporal sequence modeling. Their formulation treated detection as an inherently time-dependent problem, recognizing that motion continuity anomalies across consecutive frames carry detection signals on par with per-frame visual artifacts.
[3] Rossler et al. (2019) introduced FaceForensics++, a benchmark dataset that has since become a standard reference for detection algorithm evaluation. Their experiments using XceptionNet showed that both the scale of training data and the degree of video compression variation significantly affect a model's ability to generalize across previously unseen forgery categories.
[4] Korshunov and Marcel (2019) demonstrated that deepfake content can exploit vulnerabilities in biometric verification systems, and advocated for behavioral signal analysis—including micro-expression tracking and involuntary head motion patterns—as a complementary detection layer alongside appearance-focused techniques.
[5] Li and Lyu (2018) identified geometric distortions produced by affine transformations in face-replacement pipelines as a reproducible detection artifact. Their system was engineered to locate resolution mismatches along the boundary separating the swapped facial area from the native video background.
[6] Li, Chang, and Lyu (2018) noted that early synthesis systems failed to accurately replicate natural eye-blink rhythms. An RNN model trained specifically on blink pattern data established that physiological behavioral cues can function as a highly targeted and effective detection mechanism.
[7] Tariq et al. (2018) broadened the detection problem to encompass manually composited counterfeit images alongside digitally generated fakes. Training across multiple heterogeneous forgery types enhanced cross-domain classification performance of their CNN-based detector.
[8] Dolhansky et al. (2020) published the Deepfake Detection Challenge (DFDC) dataset, which at its release represented the most comprehensive publicly available benchmark, featuring diverse subjects recorded under varying illumination conditions and emotional states to rigorously assess detection robustness.
[9] Nguyen, Yamagishi, and Echizen (2019) applied capsule network architectures to the challenge of media forensics. Unlike standard CNN pooling operations, capsule-based routing mechanisms preserve the spatial relationships among facial sub-regions, providing enhanced sensitivity to subtle structural anomalies.
[10] Mirsky and Lee (2021) produced a broad-ranging survey covering both generation and detection methodologies. Their findings on adversarial robustness weaknesses and cross-dataset generalization gaps continue to define active research priorities across the deepfake detection community.

III. PROBLEM STATEMENT

Deepfake video generation has accelerated to a pace that far outstrips the capacity of human reviewers to evaluate individual pieces of content. These forgeries replicate facial movements, lip sync behavior, and personal appearance with sufficient fidelity that even specialist reviewers cannot achieve reliable, consistent identification. The consequent imbalance between production capability and detection capacity represents a fundamental crisis for digital information trust. Addressing this gap requires an automated classification engine capable of processing high-dimensional visual and temporal data to deliver authenticity judgments at scale—preferably intercepting manipulated content before it spreads and inflicts measurable damage. This research addresses that requirement by constructing an ML-driven pipeline that accommodates the variety of manipulation techniques currently observed in real-world contexts.

IV. OBJECTIVES

The research is structured around the following goals:
- Develop an ML classification model capable of dependably distinguishing authentic video recordings from manipulated counterparts.

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405072

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Identify and characterize the discriminative spatial and temporal features that separate genuine from synthetically produced facial footage.
- Verify detection performance against established deepfake evaluation benchmarks using conventional quantitative metrics.
- Execute comparative analysis across multiple detection approaches to isolate the strategy with superior generalization across diverse manipulation methods.
- Architect the system to maintain effectiveness as novel generation techniques emerge and evolve.

V. SYSTEM METHODOLOGY

The proposed system receives raw video as input and produces a binary authenticity decision through a structured sequence of processing stages. In the preprocessing phase, the input video is decomposed frame by frame. A face localization module identifies and extracts facial regions from each frame, which are then cropped and geometrically normalized to eliminate orientation-related variance.

Each normalized face patch passes through a CNN backbone—either ResNet or ResNeXt—which produces a compact feature vector encoding skin texture characteristics, boundary transitions, and visual appearance signatures. These per-frame feature vectors are then forwarded sequentially to an LSTM-based RNN component, which builds a temporal representation of how facial attributes change throughout the video. Temporal anomalies including sudden appearance shifts, motion discontinuities, or unnatural transformations—artifacts that cannot be reliably captured through spatial analysis in isolation—are identified at this stage.

Classification is completed by averaging per-frame prediction scores and comparing the result against a fixed decision threshold. Inputs exceeding the threshold receive a FAKE classification; those falling below are labeled REAL. The full system was built in Python 3.12, leveraging TensorFlow and PyTorch as the primary machine learning libraries, with a Flask-based web service managing inference requests and returning results through a browser front end.

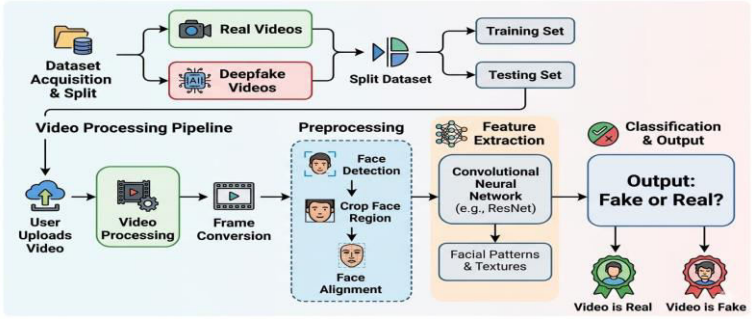


Figure 5.1 illustrates the complete pipeline from data preparation and training set construction through face preprocessing, feature extraction, temporal modeling, and final classification output.

VI. PSEUDOCODE

Algorithm: DetectDeepfakeVideo

```
Input:
  V → Input video
  T → Classification threshold
Output: "Fake Video" or "Real Video"
Begin
  frames ← ExtractFrames(V)
  predictions ← empty list
  For each frame Fi in frames do
    face_region ← DetectFace(Fi)
    If face_region = NULL then Continue End If
```

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405072

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
  cropped_face ← CropImage(Fi, face_region)
  resized_face ← ResizeImage(cropped_face, 224, 224)
  Pi ← NormalizePixels(resized_face)
  fi ← CNN_ExtractFeatures(Pi)
  Li ← TrainedClassifier.Predict(fi)
  Append Li to predictions
 End For
 If Length(predictions) = 0 then Return "Real Video" End If
 S ← Average(predictions)
 If S ≥ T then Return "Fake Video" Else Return "Real Video" End If
End
```

VII. RESULTS

System testing was conducted using a locally deployed Flask inference server linked to a browser-based interface. This configuration permitted direct video file submission and on-demand analysis without requiring additional external software or specialized client tooling. The user interaction model is straightforward: a video file is selected and submitted, after which the system returns a classification verdict accompanied by a confidence percentage rendered as a circular progress gauge.

In validation experiments, the system accurately flagged a known synthetic video as FAKE with 82% confidence, and correctly authenticated a genuine recording as REAL with 84% confidence. The system also handled edge-case inputs appropriately—footage with no identifiable facial content, as well as files in unsupported formats, triggered clear diagnostic output rather than producing silent misclassification errors.

Figures 7.1 through 7.3 document the operational interface:

Figure 7.1 — System home screen displaying the video upload interface.

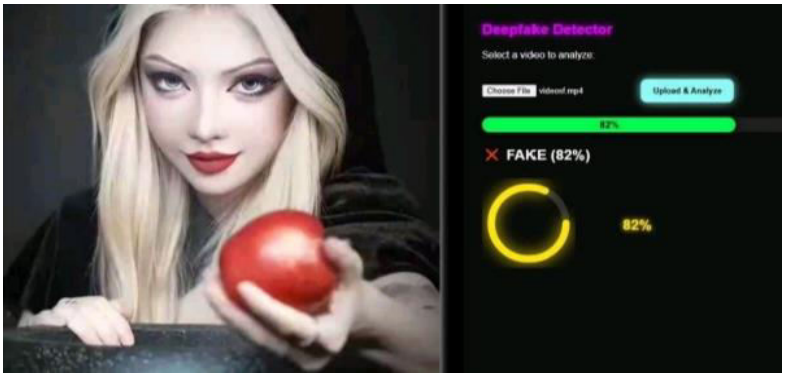


Figure 7.2 — FAKE classification result with an 82% confidence gauge display.

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405072

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

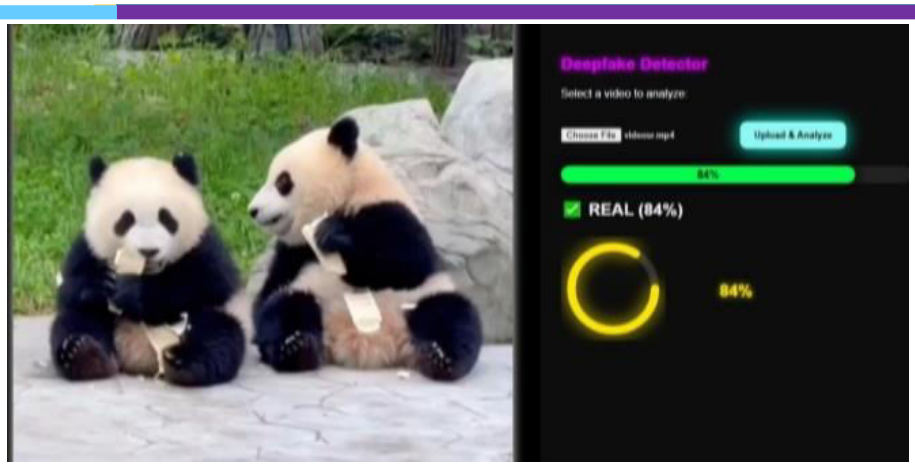


Figure 7.3 — REAL verdict rendered at 84% confidence for authentic video content.

VIII. CONCLUSION

This work developed a machine learning framework targeting the growing challenge posed by AI-generated synthetic video content. The CNN-LSTM detection pipeline demonstrates the advantage of integrating complementary analytical techniques: spatial processing handles artifact identification at the individual frame level, while temporal modeling reveals sequential irregularities in facial dynamics that would remain undetected by frame-only analysis.

Experimental findings are encouraging, though the pace of advances in generative techniques requires sustained attention. Adversarial training approaches that deliberately exploit detector vulnerabilities represent an open and consequential challenge for current detection systems. Future work should pursue improvements in real-time inference throughput, expanded evaluation across a wider range of manipulation strategies, and the development of architectures with stronger resistance to adaptive attacks designed to circumvent detection.

From a broader perspective, the capacity to verify the authenticity of video content is not merely a technical achievement—it is a prerequisite for sustaining public confidence in digital media. This project contributes a concrete step toward that goal while recognizing that the field remains an active and rapidly developing research frontier.

REFERENCES

[1]. Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking. ArXiv preprint arXiv:1806.02877v2, 2018.
[2]. B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "Wilddeepfake: A challenging real-world dataset for deepfake detection," in Proc. 28th ACM Int. Conf. Multimedia, 2020, pp. 2382–2390.
[3]. H. A. Khalil and S. A. Maged, "Deepfakes creation and detection using deep learning," in 2021 Int. Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), IEEE, 2021, pp. 1–.
[4]. P. Yang, R. Ni, and Y. Zhao, "Recapture image forensics based on laplacian convolutional neural networks," in Int. Workshop on Digital Watermarking, Springer, 2016, pp. 119–128.
[5]. N.-T. Do, L.-S. Na, and S.-H. Kim, "Forensics face detection from GANs using convolutional neural network," ISITC, vol. 2018, pp. 376–379, 2018.
[6]. S. Tariq, S. Lee, H. Kim, Y. Shin, and S. S. Woo, "Detecting both machine and human created fake face images in the wild," in Proc. 2nd Int. Workshop on Multimedia Privacy and Security, 2018, pp. 81–87.
[7]. Yuezun Li and Siwei Lyu, "Exposing DF Videos by Detecting Face Warping Artifacts," arXiv:1811.00656v3.
[8]. Hyeongwoo Kim, Pablo Garrido, Ayush Tewari, and Weipeng Xu, "Deep Video Portraits," arXiv:1901.02212v2.
[9]. Umur Aybars Ciftci, İlke Demir, and Lijun Yin, "Detection of Synthetic Portrait Videos using Biological Signals," arXiv:1901.02212v2.
[10]. A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in Proc. ICCV, 2019, pp. 1–11.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[11]. D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," in 2018 IEEE Int. Workshop on Information Forensics and Security (WIFS), 2018, pp. 1–7.
[12]. D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in 2018 15th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1–6.
[13]. Brian Dolhansky et al., "The Deepfake Detection Challenge (DFDC) Dataset," arXiv preprint arXiv:2006.07397, 2020.
[14]. Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen, "Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos," in ICASSP 2019, pp. 2307–2311.
[15]. Yisroel Mirsky and Wenke Lee, "The Creation and Detection of Deepfakes: A Survey," ACM Computing Surveys, vol. 54, no. 1, pp. 1–41, 2021.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details