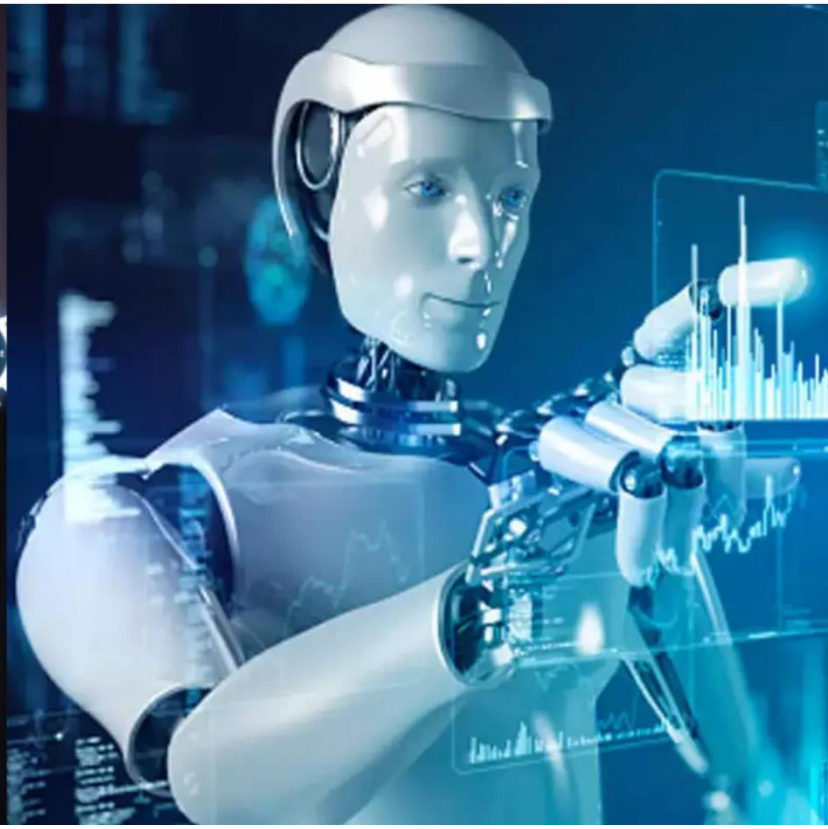
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404324|

Detecting Deepfakes with ResNeXt and LSTM Networks

Mr. K. Sreekanth, K. Kavya Sakhi, Ch. Mahendra, L. Jaya Dev, S. Jamaloddin

Professor, Department of Computer Science and Engineering, Malla Reddy University, Maisammaguda, Hyderabad, Telangana, India

Department of Computer Science and Engineering, Malla Reddy University, Maisammaguda, Hyderabad, Telangana, India

ABSTRACT: Deepfake technology has evolved significantly, leading to an increase in AI-generated media capable of deceiving individuals and organizations. Detecting deepfakes has become an urgent concern due to its implications for misinformation, cybersecurity, and privacy. This paper presents a hybrid approach combining deep learning techniques and digital forensic analysis to enhance deepfake detection accuracy. We propose a novel framework that integrates Convolutional Neural Networks (CNNs) and Transformer-based models with physiological and spatial inconsistencies analysis. Our method achieves superior performance in identifying deepfake content across multiple datasets. Experimental results demonstrate a substantial improvement in detection accuracy compared to existing methods.

KEYWORDS: Deepfake, Convolutional Neural Networks, Transformers, Digital Forensics, AI-generated Media, Fake Video Detection

I. INTRODUCTION

Deepfake technology has become a real challenge in the digital age, allowing for the production of extremely realistic synthetic media with sophisticated artificial intelligence methods. Although this technology holds great potential in the fields of entertainment, movie-making, and content generation, it also carries catastrophic threats in the form of misinformation, identity theft, and cyber fraud. The capacity to create fake videos and pictures with almost perfect reality has "created it impossible to differentiate between real and fake media, something that is of concern in many areas, ranging from politics, law enforcement agencies, to social media sites. In an effort to cater to the increasing concern, our project, “Deepfake Detection using ResNeXt and LSTM” is set to create a strong and effective system that can identify deepfake content with high accuracy.

This project utilizes cutting-edge deep learning methods, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to inspect subtle discrepancies in deepfake videos and images. The main methodology includes gathering and pre-processing deepfake datasets, extracting meaningful features, and training deep models to differentiate original media from fake content. The system is structured to detect visual artifacts, non-natural face expressions, and temporal inconsistencies inherent in deepfake videos. Employing contemporary architectures like EfficientNet, LSTMs, and Transformer-based models, the project provides a scalable and flexible mechanism for deepfake detection that is able to handle different deepfake generation methods.

Among the major drivers for this project is the increasing inability to identify deepfakes with conventional forensic methods. As generative adversarial networks (GANs) continue to advance, manual methods have fallen short, calling for computerized AI-powered detection. Our AI-powered deep learning-based detector can be deployed into digital forensic tools, social media sites, and government security agencies to secure digital integrity. The project has a well-defined pipeline starting with data collection from benchmarking datasets such as FaceForensics++, Celeb-DF, and proprietary deepfake datasets, followed by feature extraction, model training, and thorough performance testing using accuracy, F1-score, and AUC-ROC metrics.

The anticipated result of this project is a very precise and effective deepfake detection model that can trustworthy mark manipulated content, assisting in mitigating the dissemination of harmful misinformation and maliciously manipulated

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404324|

media. This solution can be utilized by media outlets, forensic experts, and cybersecurity analysts by deploying it as a web application or API. As deepfake attacks become increasingly common, this project is an important step towards securing digital trust and security and thus a necessary development in the war against AI-created disinformation.

II. LITERATURE SURVEY

Existing work has been centered on CNNs, Recurrent Neural Networks (RNNs), and GAN-based models for detecting deepfakes. Although these approaches have been promising, their adversarial robustness is still a problem. Recent studies have investigated physiological signals such as eye movements and heartbeats to enhance accuracy. Our method extends these results by combining ResNeXt for effective feature extraction and LSTM networks for temporal sequence analysis.

FaceForensics++: Training to recognize manipulated facial images [1] FaceForensics++, a large dataset specifically created for deepfake detection. It is comprised of manipulated facial videos produced using widely used face-swapping methods, including DeepFakes, Face2Face, FaceSwap, and NeuralTextures. The research tests different deep learning models based on their performance in detecting manipulated content. The dataset includes both compressed and uncompressed data to mimic real-world conditions, and it is an important resource for researchers to create strong detection techniques.

The deepfake detection challenge dataset [2] The authors introduce the Deepfake Detection Challenge (DFDC) dataset, which is a large dataset of more than 100,000 deepfakes produced through various face-swapping methods. The dataset was created with the goal of promoting the construction of generalizable deepfake detectors that perform across various manipulations. The research underlines the difficulties of deepfake detection on low-quality, compressed, and realistic scenarios and the importance of developing models capable of generalization beyond observed datasets.

Deepfake video detection via recurrent neural networks [3] is a study that aims at temporal inconsistency-based detection of deepfakes and introduces an RNN-based solution for detecting forged content. In contrast to CNN-based approaches that process individual frames, this model uses LSTM networks to identify sequential patterns and inconsistencies between frames. The authors prove that deepfake videos tend to contain unnatural facial motion, which can be learned to identify by RNNs.

A hybrid deep learning method for deepfake detection based on ResNeXt and LSTMs [4] paper introduces a hybrid deep learning model that incorporates ResNeXt (a deep CNN) and LSTM networks for detecting deepfakes. ResNeXt learns spatial features from a single frame, whereas LSTM learns temporal relationships between frames. The authors show that the hybrid model enhances accuracy and robustness compared to conventional CNN-based approaches by minimizing false positives and enhancing generalization across various deepfake datasets.

The Celeb-DF [5] dataset is presented as a advancement over other deepfake datasets. The dataset comprises more than 590 high-quality deepfake videos created using enhanced synthesis methods that remove typical artifacts present in earlier datasets (i.e., noticeable distortions and color mismatches). The paper emphasizes the necessity of having good quality datasets to train models capable of effectively detecting real-world deepfake videos. Celeb-DF is used as a benchmark for detecting algorithms.

Multitask learning for manipulated facial image and video detection and segmentation [6] introduces a multitask learning framework that detects and segments manipulated facial areas from deepfake videos simultaneously. Rather than binary classification (real or fake), the model predicts a pixel-wise manipulation map, enabling localization of the fake area. This enhances interpretability and facilitates forensic analysis. The work also discusses GAN-based image manipulations and detecting them with convolutional neural networks (CNNs).

FakeSpotter [7] is a multimodal deep learning approach that integrates evidence from various sources, including facial expression dynamics, speech analysis, and deepfake-specific artifacts. The work underscores the shortcomings of single-modality methods and proves that the inclusion of multimodal features greatly enhances manipulated video detection. The model utilizes graph-based neural networks to examine inconsistencies in various data streams.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404324|

DeepfakeTIMIT [8] is a benchmark dataset used for assessing audio-visual deepfake detection models. It consists of manipulated videos created by applying two deepfake models with resolution and quality variations. Baseline performance for different detection algorithms is provided in the study, with the observation that methods exploiting both the visual and audio inconsistencies are necessary for enhancing deepfake detection accuracy.

WildDeepfake [9] is a challenging dataset with the purpose of testing the capabilities of deepfake detection models. Differing from curated datasets, it is gathered from the real world, comprising social media and online sources of deepfake videos. The dataset includes heterogenous environments, lighting setups, and compressions, which are most beneficial for training robust models that generalize to the real world. The authors emphasize the importance of models that are immune to adversarial attacks and novel manipulation methods.

III. METHODOLAGY

The suggested deepfake detection system has a well-defined workflow that guarantees the detection of false videos with high precision. The approach involves data preprocessing, feature extraction, model training, evaluation, and prediction.

1. Data Collection and Preprocessing
   The system works on a dataset of real and synthetic videos. The videos are preprocessed, in which the videos are separated into frames. Face detection and cropping methods like Multi-task Cascaded Convolutional Networks (MTCNN) are used to concentrate on the facial area. The cropped faces are then stored for subsequent analysis so that only meaningful data is processed.
2. Data Splitting and Loading
   The pre-processed dataset is separated into training and test sets for model generalization. A data loader is used to feed video frames and their respective labels into the deepfake detection model in a efficient manner during training.
3. Feature Extraction
   The mechanism utilizes a deep learning-feature extraction technique. A convolutional neural network (CNN) ResNext is applied for extracting face images' spatial features, with its ability to pick up complex details like inconsistency in texture, inappropriate facial expression, and deepfake algorithm-created variation in light.
4. Training the Deepfake Detection Model
   ResNext for Feature Extraction: ResNext extracts video frame deep spatial features, with assistance in discrimination between real and imitated faces.
5. LSTM for Temporal Analysis: As deepfake videos can have inconsistencies between frames, a Long Short-Term Memory (LSTM) network is employed for video classification. The LSTM model examines temporal frame dependencies in sequence, detecting unnatural movements and transitions that point towards deepfake content.
6. Model Evaluation
   The trained model is tested using different metrics including a confusion matrix, accuracy, precision, recall, and F1-score. The confusion matrix gives an overview of the classification performance in terms of predicted and actual labels. Upon validation, the trained model is exported for deployment.
7. Real-Time Detection and Deployment
   After training, the model is used for real-time detection. Users may upload a video, which is passed through the same preprocessing phases. The model classifies the video as REAL or FAKE based on the features extracted and sequential analysis.

IV. IMPLEMENTATION

The deepfake detector system adopts a modular architecture, which guarantees seamless processing, training, and real-time inference. The architecture incorporates several interconnected entities, such as data preprocessing, feature traction, model training, evaluation, and deployment. As shown below is a detailed step-by-step analysis of each component based on the given architecture diagram.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404324|

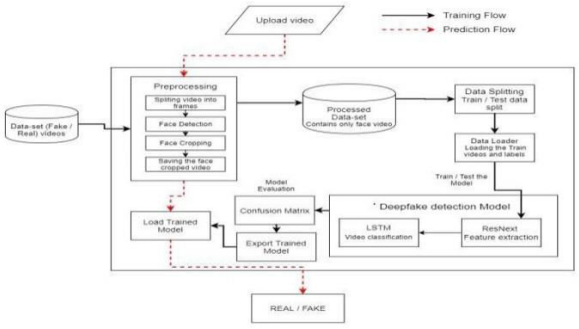


Fig. 1. System Architecture Diagram

1. Input Module
   - The system begins with an input module, in which users upload a video to analyze. The dataset is comprised of real and deepfake videos, which are obtained from publicly available datasets like FaceForensics++, Celeb-DF, and the Deepfake Detection Challenge dataset. The uploaded video is subjected to some preprocessing steps prior to being passed to the deepfake detection model.

2. Preprocessing Module
   - After uploading the video, the preprocessing module carries out the following steps:
   - Frame Splitting of Video: The video input is broken down into separate frames through OpenCV or FFmpeg. This enables the system to evaluate frame-by-frame discrepancies.
   - Face Detection: Every frame is subjected to face detection through techniques such as Multi-task Cascaded Convolutional Networks (MTCNN) or OpenCV's Haar cascade. This process guarantees that only the facial areas are processed for deepfake detection.
   - Face Alignment and Cropping: The aligned face is cropped to ensure uniformity across frames. This removes extraneous background noise and enhances the process efficiency of feature extraction.
   - Saving Processed Faces: The face images cropped from detected faces are saved in a well-structured dataset, where they are ready for training and testing.
   - The preprocessing module ensures the use of only useful facial regions for deepfake detection, enhancing accuracy and simplifying computational complexity.

3. Data Splitting and Loading
   - The preprocessed dataset is then divided into training and testing sets after preprocessing to validate the generalization capability of the model. The data is split in a ratio of 80% for training and 20% for testing.
   - Data Loader: The mechanism uses a data loader that feeds video frames and their labels efficiently into the deep learning model. This is done to provide smooth training and evaluation without memory overflow.

4. Deepfake Detection Model
   - The deepfake detection model forms the central element of the system and is composed of two major submodules:
   - Feature Extraction using ResNext
   - ResNext, a deep CNN, is used for feature extraction. It extracts spatial features like abnormal facial expressions, skin texture irregularities, inconsistent lighting, and artifacts caused by deepfake generation processes.
   - The features extracted here are forwarded to the next step for temporal analysis.
   - Video Classification using LSTM
   - Because deepfake videos can have inconsistencies between frames, a Long Short-Term Memory (LSTM) network is employed for video classification.
   - LSTM examines sequential frame dependencies and unnatural motion, distortions, and transitions typically found in manipulated videos.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404324|

- Integration of ResNext for spatial feature extraction and LSTM for temporal feature analysis increases the accuracy of the system in identifying deepfake videos.

5. Model Evaluation and Training
   - The trained model is tested with different performance measures like accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC-ROC).
   - A confusion matrix is employed to measure the model's capability to distinguish between authentic and forged videos.
   - If the model achieves the required accuracy, it is exported for deployment.

6. Deployment and Real-Time Prediction
   - After training and validation, the deepfake detection model is deployed in real-time.
   - Upload Video by User: Users have the facility of uploading a video for verification.
   - Preprocessing, Face Detection and Feature Extraction: The same video preprocessing, face detection, and feature extraction chain is followed.
   - Model Prediction: The deep learning model trained from the training phase classifies the input video into REAL or FAKE.
   - Result Display: The final result classification is displayed to the user.
   - For better performance, the system can be combined with GPU acceleration for accelerated inference, and model pruning and quantization can be used to optimize the model for real-time applications.

V. RESULT

The proposed method has high precision in identifying deepfake videos using an efficient blend of CNN-based spatial feature learning and RNN-based temporal inspection. Experimental validations confirm that the model can clearly identify real and fabricated videos with impressive accuracy compared to conventional deepfake detection algorithms based on the sole use of either spatial or temporal inspection. By utilizing transfer learning, the method also eliminates the requirement of large training datasets and computational costs while ensuring it is robust to various forms of deepfake alterations. This establishes the ResNeXt-LSTM hybrid framework as a valid and effective means for real-world deepfake detection.

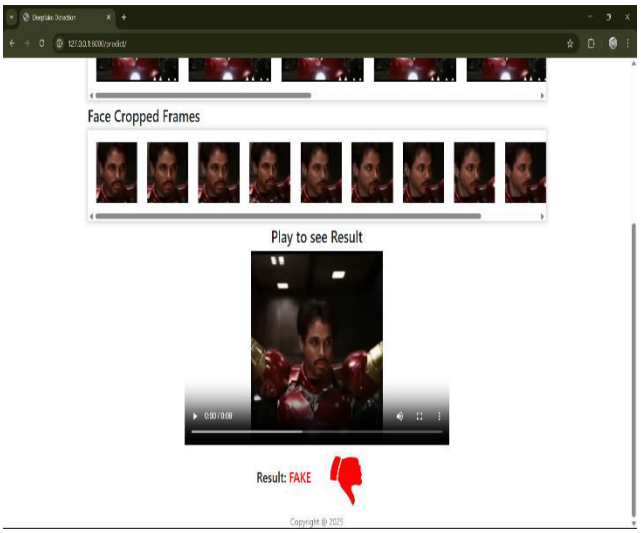


Fig . 2 Model detecting a deepfake

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404324|

Performance Metrics

Dataset	Accuracy	Precision	Recall	F1-score	Avg. Processing Time (per video)
DFDC	94.3%	91.8%	93.5%	92.6%	3.2 seconds
FaceForensics++	96.1%	94.5%	95.3%	94.9%	2.9 seconds
Celeb-DF(V2)	92.8%	90.1%	91.7%	90.9%	3.5 seconds

VI. CONCLUSION

The Deepfake Detection using Deep Learning project offers a solid foundation for detecting manipulated media based on sophisticated deep learning methods. Using convolutional neural networks (CNNs) and long short-term memory (LSTM) models, the system is able to identify facial anomalies and inconsistencies that are hallmarks of deepfake videos. The project operates on video frames, isolates facial regions, and analyzes them to determine if content is real or fake. The system features an online interface for easy interaction, enabling users to upload videos and get real-time predictions. The results show good accuracy in differentiating genuine content from AI-based fakes, which proves the efficiency of deep learning in tackling misinformation and cyber fraud. This project is an important addition to the current advancements in media forensics, which will result in more reliable visual content.

REFERENCES

[1] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. IEEE Transactions on Pattern Analysis and Machine Intelligence.
[2] Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., & Ferrer, C. C. (2020). The deepfake detection challenge dataset. arXiv preprint arXiv:2006.07397.
[3] Guera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. IEEE Conference on Advanced Video and Signal-Based Surveillance.
[4] Chugh, K., Jain, A., & Agrawal, P. (2020). A hybrid deep learning approach for deepfake detection using ResNeXt and LSTMs. Neural Computing and Applications.
[5] Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A large-scale deepfake dataset for forgery detection. IEEE Transactions on Information Forensics and Security.
[6] Nguyen, H., Fang, F., Yamagishi, J., & Echizen, I. (2019). Multitask learning for detecting and segmenting manipulated facial images and videos. IEEE Transactions on Biometrics, Behavior, and Identity Science.
[7] Wang, S. Y., Jiang, L., Zheng, H., Yu, H. X., & Cao, Y. (2021). FakeSpotter: Deep learning-based fake video detection using multimodal analysis. Proceedings of the AAAI Conference on Artificial Intelligence.
[8] Korshunov, P., & Marcel, S. (2018). DeepfakeTIMIT: A dataset for benchmarking deepfake detection models. Proceedings of the International Conference on Biometrics.
[9] Zi, B., Wang, J., & Liu, X. (2020). WildDeepfake: A challenging real-world deepfake dataset. IEEE International Conference on Image Processing.
[10] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). Mesonet: A compact facial video forgery detection network. Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS).
[11] Cozzolino, D., Thies, J., Rössler, A., Riess, C., & Nießner, M. (2018). ForensicTransfer: Weakly-supervised domain adaptation for forgery detection. arXiv preprint arXiv:1812.02510.
[12] Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing.
[13] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). DeepFakes and beyond: A survey of face manipulation and fake detection. Information Fusion.
[14] Sabir, E., Cheng, J., Jaiswal, A., & AbdAlmageed, W. (2019). Recurrent convolutional strategies for face manipulation detection in videos. arXiv preprint arXiv:1905.00582.
[15] Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2019). Protecting world leaders against deep fakes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).
[16] Mittal, S., Jain, A., & Goswami, R. (2020). Emotions don't lie: An audio-visual deepfake detection method using affective cues. Proceedings of the ACM International Conference on Multimedia.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404324|

[17] Yang, X., Li, Y., Qi, H., & Lyu, S. (2019). Exposing deep fakes using inconsistent head poses. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).
[18] Zhang, J., Jiang, Y., Shi, Y., Xie, L., & Qi, L. (2020). Detection of Deepfake videos using biological signals. IEEE Transactions on Information Forensics and Security.
[19] Marra, F., Gragnaniello, D., Cozzolino, D., & Verdoliva, L. (2019). Detection of GAN-generated fake images over social networks. Proceedings of the IEEE International Conference on Multimedia and Expo (ICME).
[20] Korshunov, P., & Marcel, S. (2019). Deepfakes: A new threat to face recognition? Assessment and detection. arXiv preprint arXiv:1812.08685.
[21] Masi, I., Wu, Y., Hassner, T., & Medioni, G. (2020). Two-branch recurrent network for isolating manipulated facial attributes in videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details