Deep Fake Detection

Mr. SANJAY M [1], P B SANDEEP [2], RACHANA G C [3], SOORYA PRAKASH S [4], SRUJAN S R[5]

[1] Assistant. Professor, Dept. of Information Science & Engineering, Rajeev Institute of Technology, Hassan
[2] Information Science & Engineering, Rajeev Institute of Technology, Hassan[3]
Information Science & Engineering, Rajeev Institute of Technology, Hassan[4]
Information Science & Engineering, Rajeev Institute of Technology, Hassan[5]
Information Science & Engineering, Rajeev Institute of Technology, Hassan

---

Abstract - Authenticity of Smart Media A method called Deep Fake identification With Machine Learning uses deep learning approaches to enhance the identification of AI-manipulated media. Artificial intelligence (AI) produces incredibly lifelike synthetic movies known as "deep fakes," which can cause political instability, disinformation, and harm to one's reputation. This project uses preprocessing methods like face cropping and frame extraction to analyse video material. While LSTM is used for temporal sequence modelling to categorise movies as real or deepfake, ResNeXt CNN is employed for feature extraction. Real-time detection and increased accuracy are the outcomes of the system's automation of video forensics. It guarantees dependable results and offers users an easy-to-use online interface by utilising deep learning.

Key words: LSTM, deepfake detection, facial recognition, video forensics, computer vision, deep learning, ResNeXt, and media authenticity.

1. INTRODUCTION

Deepfake technology, in which artificial intelligence (AI) algorithms are used to make synthetic yet remarkably realistic videos, has emerged as a result of the quick development of deep learning and processing capacity. Because they allow media manipulation for nefarious objectives like identity theft, political propaganda, extortion, and disinformation, these deepfakes represent a serious threat. In order to differentiate authentic information from fake media, the method presented in this paper uses sophisticated machine learning algorithms to analyse and categorise videos. The suggested method improves the accuracy of deepfake detection by utilising CNN and LSTM models for frame-level feature extraction and temporal pattern identification.

Grade I:In today's media-driven environment, deepfakes have emerged as a serious digital danger. Every year, millions of edited movies are disseminated online, impacting people, organisations, and governments. Traditional detection techniques rely on inefficient and error-prone manual inspection. Deepfake creation tools and platforms are becoming more and more common, hence automated detection techniques utilising artificial intelligence are crucial. Large video datasets can be analysed using deep learning to find minute irregularities that indicate manipulation, which helps stop the spread of false information.

Grade II: Using deep learning models like ResNeXt and LSTM greatly improves the precision and effectiveness of deepfake detection. In order to identify face-swapped information, these models examine individual video frames and their temporal correlations. The suggested solution uses CNNs to extract significant features from videos and LSTM to classify the sequential frame data. This method increases identification accuracy across a range of video formats and situations, automates the deepfake detection process, and lessens the need for human verification.

Grade III:AI-powered deepfake detection systems are essential for preserving digital integrity and confidence in media communications, even after they have first identified manipulated media. These tools help content platforms, journalists, and investigators confirm the authenticity of videos by detecting fake content. Ensuring the validity of video evidence and recorded encounters can also be advantageous for the legal, healthcare, and educational sectors. A larger ecosystem of safe information exchange and digital responsibility is supported by the use of deep learning into media forensics.

Grade IV: Deepfake detection using machine learning also offers wider social and financial benefits. Real-time content verification and broad accessibility are made possible by incorporating detection techniques into mobile applications and web platforms. This makes digital safety technologies more accessible to everyone, especially in underserved or distant regions. Transparency and dependability in digital media are improved by the automation and digitisation of detection systems, which also expedites legal and regulatory procedures. Integrating these solutions into communication platforms as technology advances guarantees a safe online environment for coming generations.

Media security systems must incorporate machine learning to counter the growing risks of synthetic material. For precise, real-time deepfake identification, this study offers an automated and user-friendly method based on ResNeXt and LSTM. By reducing human error and improving media verification, the system encourages safe communication and well-informed choices. Cross-platform scalability, smartphone integration, and browser plug-ins are possible future developments that will expand the use and significance of deepfake detection technologies.

2. LITERATURE REVIEW

A literature review is an essential step in the software development process since it provides valuable insights and improvements for existing methods. This section highlights the key papers that have influenced the suggested work on deepfake detection utilising machine learning and deep learning approaches.

A technique for identifying face warping artefacts in deepfake videos was presented by Li and Lyu (2018). To find discrepancies

between the created face regions and the surrounding areas, they employed a convolutional neural network (CNN). The study highlighted how resolution constraints in current deepfake algorithms frequently create artefacts, but it lacked temporal frame analysis, which is essential for thorough identification.

Since many synthetic videos were unable to replicate realistic blink frequencies, Li, Chang, and Lyu (2018) conducted another significant study that concentrated on identifying deepfakes using eye-blinking patterns. Using Long-term Recurrent Convolutional Networks (LRCN), the model investigated temporal discrepancies in eye behavior. But as deepfake generation techniques improved, eye-blinking by itself was no longer enough for accurate detection.

In order to detect manipulated images and videos in a variety of contexts, including replay attack detection and synthetic face recognition, Nguyen et al. (2019) investigated capsule networks. Although the method produced encouraging results, its practicality was restricted by the use of random noise in training. In order to achieve robustness and generalisability, our approach trains on clean, realistic datasets.

In order to detect deepfake videos, Güera and Delp (2018) suggested using Recurrent Neural Networks (RNNs) in conjunction with CNN-extracted features and temporal sequence learning. A tiny, homogeneous dataset limited the model's performance despite its positive outcomes. For real-time relevance, on the other hand, our system is trained on a vast and varied dataset that combines many sources.

The FaceForensics++ dataset was first presented by Rossler et al. (2019) and served as the basis for numerous detection methods. Their research highlighted the value of carefully selected, superior datasets for creating reliable classifiers and showed how well deep learning works to identify edited video.

3. SYSTEM DESIGN

Existing system:

Conventional methods for identifying deepfake movies mostly depend on forensic video analysis and manual observation, where professionals spot irregularities like abnormal lighting and reflections, improper lip-syncing, or strange facial motions. This approach is subjective, time-consuming, and prone to human mistake, even if it can be helpful in certain situations. This is particularly true when dealing with extremely realistic AI-generated content. Furthermore, temporal irregularities across frames are frequently missed by current methods that rely on simple frame-based comparisons or shallow machine learning models, which lowers detection accuracy.

Several modern techniques that use semi-automated ultrasonic analysis but lack advanced image processing and feature extraction reduce accuracy. The absence of machine learning-based predictive models in existing diagnostic methods further restricts personalised risk assessment. Therefore, an AI-driven system that integrates deep learning, ultrasound segmentation, and clinical data analysis is needed to increase accuracy, efficiency, and early PCOS identification.

Proposed system:

The suggested method improves the precision and effectiveness of deepfake video detection by utilising machine learning and deep learning. It combines a Long Short-Term Memory (LSTM) network to examine temporal inconsistencies and sequential patterns across video frames with a pre-trained ResNeXt Convolutional Neural Network (CNN) for frame-level feature extraction. To make sure that only pertinent visual information is handled, the system uses a structured pipeline that begins with data preprocessing, which includes face detection, frame extraction, and noise reduction.
By spotting minute irregularities and artefacts that are not evident to the naked eye, this hybrid architecture enables precise categorisation of films as either real or deepfake. To enhance generalisation across various deepfake creation techniques, the model is trained on a balanced, varied dataset that incorporates information from several public sources. Users can upload films and get real-time forecasts with confidence scores using a web-based front-end. Future improvements will include real-time detection and alert systems that integrate with web browsers and messaging platforms, guaranteeing greater accessibility and defence against online disinformation.

4. METHODOLOGY

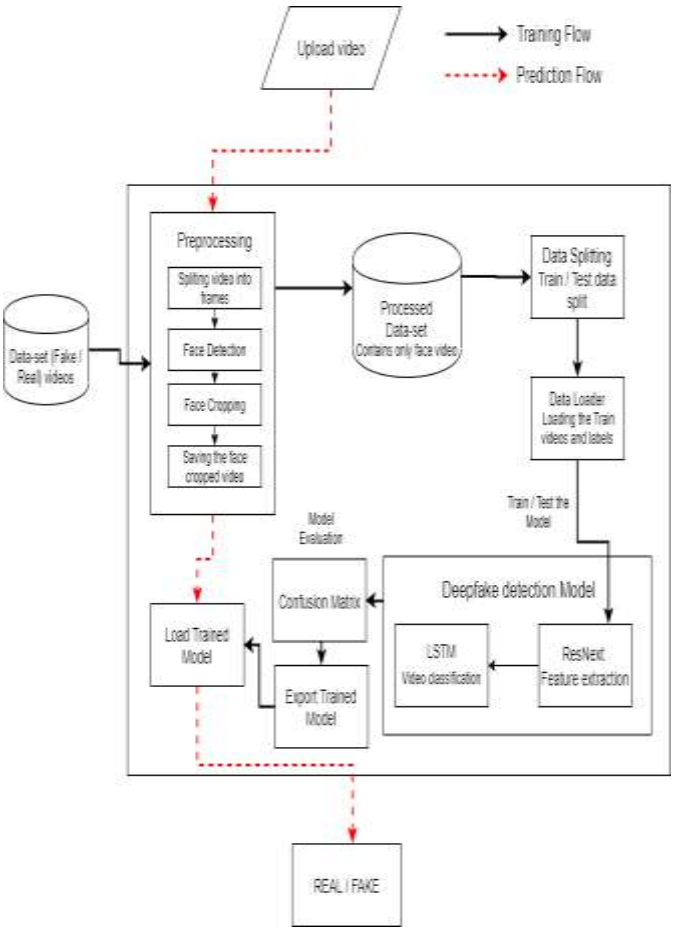

Figure 1: Architectural Diagram

The architecture of a deepfake video detection system is depicted in this flow diagram. A video is first uploaded, and then preprocessing operations including trimming, facial detection, and frame extraction are performed. Preprocessed data is divided into

training and testing sets during the training phase, and ResNeXt is used to extract features. LSTM is then utilised for temporal video classification. A confusion matrix is used to assess the model before it is exported for real-time forecasting. The trained algorithm uses learnt patterns to determine whether submitted movies are real or false during prediction.

5. CONCLUSIONS

The Deepfake Detection System is a cutting-edge tool that precisely detects AI-generated fake videos by applying deep learning and machine learning algorithms. Using a hybrid architecture that incorporates Long Short-Term Memory (LSTM) networks for temporal video analysis and ResNeXt for frame-level feature extraction, the system successfully identifies movies as real or fraudulent. The trained model's durability in identifying slight modifications is demonstrated by its excellent accuracy of up to 97.76% on benchmark datasets. The solution improves the integrity of digital content, guarantees prompt intervention, and slows the spread of false information by automating the detection process.

6. FUTURE DIRECTIONS

Future developments could greatly increase the Deepfake Detection System's accuracy, scalability, and accessibility on a variety of platforms. Real-time video authentication for social media users and content producers can be made possible by creating small, mobile, and browser-based plug-ins. Before it spreads, altered content can be identified and flagged with the aid of integration with messaging apps and streaming services. For legal and investigative reasons, tele-forensics capabilities can facilitate remote collaboration and verification. Furthermore, adding multimodal analysis—like consistency checks for text and audio—can increase the system's functionality. Future developments might potentially include real-time detection of synthetic voice and full-body deepfakes, guaranteeing complete defence against manipulation of digital media.

.

REFERENCES

[1] Li, Y., & Lyu, S. (2018). Exposing DeepFake Videos by Detecting Face Warping Artifacts. *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, USA, 46-52.

[2] Li, Y., Chang, M.-C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, Hong Kong, China, 1-7.

[3] Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Using Capsule Networks to Detect Forged Images and Videos. *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2307-2311.

[4] Güera, D., & Delp, E. J. (2018). Deepfake Video Detection Using Recurrent Neural Networks. *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Auckland, New Zealand, 1-6.

[5] Rossler, 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*, arXiv:1901.08971.

[6] Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, 3207–3216.