International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.035 ISSN: 2581-4621

SOLUTION FOR DETECTION OF FACE-SWAP BASED FAKE VIDEOS

Sangeetha.V, Department of Artificial Intelligence and Data Science, Kamaraj College Of Engineering and Technology, Virudhunagar
Deepika Shri.N, Department of Artificial Intelligence and Data Science, Kamaraj College Of Engineering and Technology, Virudhunagar
Sri Shivanuja.S, Department of Artificial Intelligence and Data Science, Kamaraj College Of Engineering and Technology, Virudhunagar

***Abstract*—**Recent advancements in deep learning have dramatically simplified the creation of highly realistic deepfake (DF) videos, which were traditionally the domain of skilled visual effects professionals. This surge in AI-synthesized media poses challenges in sectors such as politics, security, and entertainment. Detecting deepfakes is increasingly difficult due to the realism of these videos. In this paper, we propose a deepfake detection approach that combines Convolutional Neural Networks (CNNs) for extracting detailed frame-level features with Recurrent Neural Networks (RNNs) for modeling temporal inconsistencies across video sequences. The CNN identifies critical features such as facial landmarks and textures, while the RNN focuses on detecting unnatural transitions between frames, including irregular facial movements or subtle manipulations. Our method is capable of handling videos of varying lengths and has demonstrated competitive performance when tested on a comprehensive dataset of deepfake videos.

Keywords*—*Deepfake detection, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), temporal modeling, video forensics.

I. INTRODUCTION

The increasing sophistication of smartphone cameras, combined with widespread access to reliable internet, has significantly expanded the reach of social media and media- sharing platforms. This has made creating and sharing digital videos easier than ever. Alongside this, the rapid advancement in computational power has made deep learning remarkably powerful, achieving things that would have seemed impossible just a few years ago. However, like any revolutionary technology, this progress has introduced new challenges. One such challenge is the rise of "DeepFakes," which are manipulated video and audio clips generated by deep generative adversarial models. The spread of DeepFakes on social media has become increasingly common, leading to spamming and the dissemination of false information across platforms. These types of DeepFakes can be highly problematic, causing confusion, fear, and misleading the general public.

In order to get out of this kind of circumstance, DF detection is crucial. That being said, we present a novel deep learning approach that efficiently separates artificial intelligence (AI)- generated fake videos (DF Videos) from real videos. The development of technologies capable of identifying fakes is crucial in order to detect the DF and stop it from propagating online.

It is crucial to comprehend how the Generative Adversarial Network (GAN) generates the DF in order to detect it. GAN converts an image and video of a particular person (referred to as the "target") into another video in which the faces of the target are swapped out with those of a different person (referred to as the "source"). Deep adversarial neural networks, which are trained on target videos and face images to automatically map the faces and facial expressions of the source to the target, are the core components of deep face mapping (DF). The video was divided into frames by the GAN, which then changed the input image in each frame. It reconstructs the footage even further. Autoencoders are typically used to accomplish this operation.

By dividing the video into frames, extracting the features with a ResNext Convolutional Neural

DOI:10.46647/ijetms.2025.v09i02.035 ISSN: 2581-4621

Network (CNN), and using a Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) to capture the temporal inconsistencies between frames introduced by GAN during the reconstruction of the DF, our method detects such artifacts. We streamline the procedure by directly modeling the resolution discrepancy in affine face wrappings in order to train the ResNext CNN model.

A. Importance of the work

Our work addresses the growing threat of DeepFake videos, which pose serious challenges to media integrity in sectors like politics, security, and entertainment. By integrating Convolutional Neural Networks (CNNs) for frame-based feature extraction with Recurrent Neural Networks (RNNs) for detecting temporal inconsistencies, we offer a robust approach to accurately identify DeepFake manipulations. This method captures subtle artifacts like unnatural transitions and irregular facial movements, which are often missed by traditional techniques. Tested on extensive datasets, our model demonstrates competitive performance, contributing to the development of scalable DeepFake detection systems crucial for maintaining trust in digital media.

B. Objective

- To discover and identify the manipulated elements within deepfake videos in order to reveal the distorted truths of AI-generated forgeries.
- To reduce the risks of abuse and misinformation in order to minimize the potential for deepfakes to deceive and mislead the public on the internet.
- To develop a system capable of accurately distinguishing and classifying videos as either deepfakes or authentic in order to improve the reliability of video content verification.
- To provide an intuitive and user-friendly platform that allows users to upload videos in order to easily determine their authenticity.

C. Social Impact

D. The solution will empower users, organizations, and governments to detect and prevent the spread of deepfake content, safeguarding trust and security in digital media.

E. By curbing the spread of harmful deepfakes, the solution helps protect individuals from misinformation, harassment, and potential identity theft, promoting a safer online environment.

F. Businesses can avoid reputational damage and financial losses caused by deepfakes, while reducing the costs associated with combating disinformation and fraud.

G. Challenges

- The increasing realism of DeepFakes makes it difficult for detection algorithms to identify subtle manipulation artifacts.
- Detection models often struggle to generalize across various datasets and formats, limiting their effectiveness on new, unseen videos.
- Real-time detection poses scalability challenges due to high computational demands, while DeepFake creators continually refine methods to evade detection systems.

H. Limitations

- Our solution currently focuses solely on video, without analyzing audio, making it ineffective at detecting audio deepfakes.
- As a result, audio manipulations will go undetected by our current method.
- In the future, we plan to expand our approach to include audio deepfake detection.

II. LITERATURE SURVEY

The rapid rise in the creation and illegal use of deepfake videos poses a serious threat to democracy, justice, and public trust. As a result, the need for deepfake video analysis, detection, and intervention has significantly increased. Below are some relevant works in the field of deepfake detection:

Exposing DeepFake Videos by Detecting Face Warping Artifacts [1] introduced a method that detects artifacts by comparing generated face regions with surrounding areas, using a specialized

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.035 ISSN: 2581-4621

Convolutional Neural Network (CNN). The approach identifies two types of face artifacts, based on the observation that current deepfake algorithms produce images with limited resolution, which must then be warped to match the faces in the source video.

Exposing AI-Created Fake Videos by Detecting Eye Blinking

[2] describes a technique that reveals deepfake face videos generated by deep neural networks. This method detects the lack of eye blinking, a physiological signal often missing in synthesized videos. The technique was tested on benchmarks for eye-blinking detection and showed promising results in identifying deepfake videos. However, it focuses solely on eye blinking, neglecting other important cues such as facial wrinkles or teeth alignment, which our proposed method aims to address.

Using Capsule Networks to Detect Forged Images and Videos [3] presents an approach utilizing capsule networks to identify manipulated images and videos in scenarios like replay attacks and computer-generated video forgeries. Although their model showed success with their dataset, it was trained with random noise, which may limit its real- world applicability. Our proposed method aims to train on noiseless, real-time datasets for better accuracy.

Detection of Synthetic Portrait Videos Using Biological Signals[5] extracts biological signals from facial regions in both real and fake portrait videos. The approach applies transformations to compute spatial coherence and temporal consistency, capturing signal features and PPG maps, and then uses a probabilistic SVM and a CNN to aggregate authenticity probabilities. While their model, FakeCatcher, performs well across different video content, resolution, and quality, the absence of a discriminator limits the preservation of biological signals, making it challenging to develop a loss function that follows their signal processing steps.

Our proposed method seeks to address some of the limitations in these approaches, integrating a more comprehensive set of parameters for enhanced deepfake detection.

III. Proposed System

There are numerous tools available for creating DeepFakes (DF), but there is a scarcity of effective tools for detecting them. Our approach to DF detection will be a significant contribution in preventing the spread of DeepFakes across the internet. We aim to develop a web-based platform where users can upload videos to classify them as either real or fake. This project has the potential to scale, from a web platform to a browser plugin for automatic DF detection. Major platforms like WhatsApp and Facebook could even integrate this project into their systems to enable DF detection before videos are shared. A key goal is to evaluate the solution's performance in terms of security, user-friendliness, accuracy, and reliability.

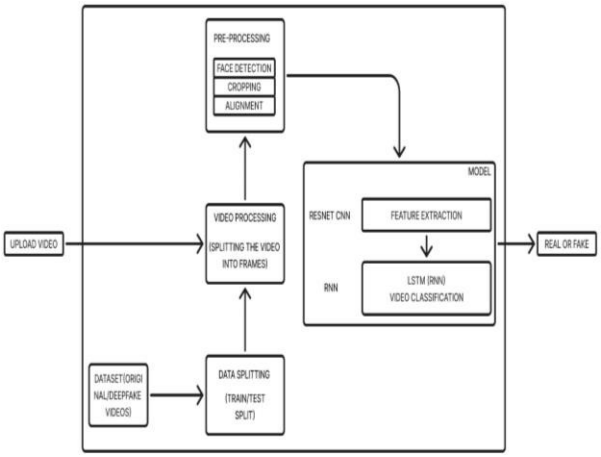

Fig. 1: outlines the basic system architecture of the proposed system.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.035 ISSN: 2581-4621

A. Software Requirements

Programming Language: Python, JavaScript Programming framework: Pytorch, Django Version control: Git Cloud Services: Google cloud platform

B. Dataset Specification

We are utilizing a mixed dataset that includes an equal number of videos from various sources such as FaceForensics++ [14], the Deepfake Detection Challenge dataset [13], and YouTube. Our newly created dataset consists of 50% resized deepfake videos and 50% original, unaltered videos. The dataset is divided into two parts, with 70% of the videos designated for training and the remaining 30% set aside for testing purposes.

C. Pre Processing

In the preprocessing stage, videos are split into individual frames. We then detect and crop the faces from each frame. To maintain uniformity, we calculate the average number of frames in the dataset videos, and the new processed dataset contains a consistent number of face-cropped frames. Frames without detectable faces are discarded. For experimental purposes, due to computational constraints, we propose using only the first 100 frames from each video, instead of the full 300 frames for a 10-second video.

D. Model

Our model consists of a ResNext50_32x4d architecture followed by a Long Short-Term Memory (LSTM) layer. The Data Loader loads the preprocessed face-cropped videos and splits them into training and testing sets. The processed frames are then passed to the model for training and testing in mini-batches.

E. ResNext CNN for Feature Extraction:

Rather than designing a custom classifier, we propose using the ResNext CNN for feature extraction to accurately capture frame-level features. We will fine-tune the network by adding the necessary layers and selecting an optimal learning rate to ensure proper convergence of the model's gradient descent. The 2048-dimensional feature vectors produced after the final pooling layers are used as input for the LSTM sequence processor.

F. LSTM for Sequence Processing:

We assume that the sequence of ResNext CNN feature vectors represents the input frames, and we use a 2-node neural network to classify whether the sequence belongs to a DeepFake or authentic video. The challenge is designing a model that can process sequences meaningfully. For this, we propose using a 2048-unit LSTM with a 0.4 dropout rate, which will help achieve our goal. The LSTM analyzes the frames in sequence to perform temporal analysis, comparing the frame at time 't' with a frame at 't-n', where n represents a number of frames prior to t.

G. Prediction:

When a new video is input for prediction, it undergoes the same preprocessing as the training data. The video is split into frames, faces are cropped, and the frames are passed directly to the trained model for detection, without storing them in local storage.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.035 ISSN: 2581-4621

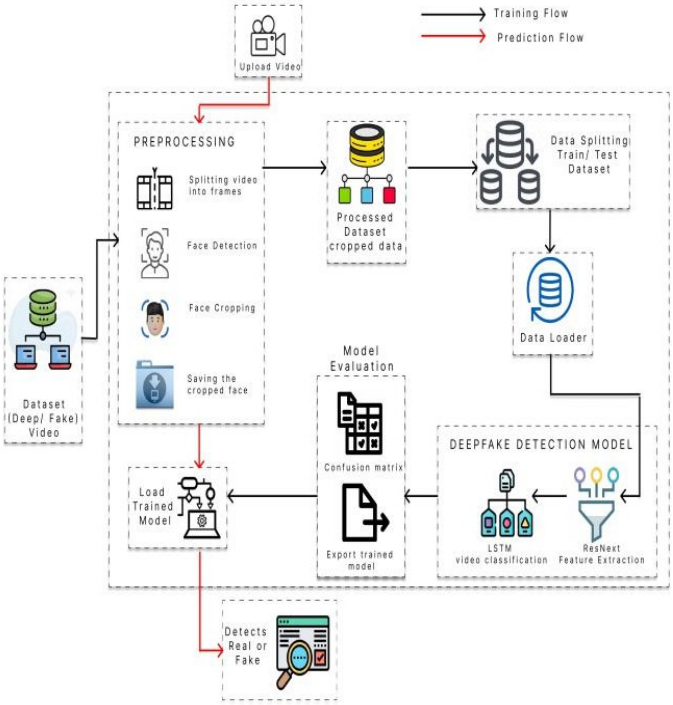

Fig 2: Flow chart

IV. CONCLUSION

In conclusion, we introduced a neural network-based approach for classifying videos as either DeepFake or authentic, while also providing the confidence level of the proposed model. Our method draws inspiration from the creation process of DeepFakes using Generative Adversarial Networks (GANs) and Autoencoders. We utilize ResNext CNN for frame-level detection and an RNN with LSTM for video classification. This approach enables the detection of videos as either DeepFake or real based on the parameters outlined in this paper. We anticipate that it will achieve high accuracy when applied to real-time data.

V. REFERENCES

[1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1-135.

[2] Huang, L., Soong, F. K., & Juang, B. H. (2001). Spoken language processing: A guide to theory, algorithm, and system development. Prentice Hall.

[3] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning,
C. D., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the conference on empirical methods in natural language processing (EMNLP), 1631-1642.

[4] Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., ... & Weninger, F. (2013). The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. Proceedings INTERSPEECH 2013, 148-152.

[5] Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. Proceedings of the 45th annual meeting of the association of computational linguistics, 440-447.

[6] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

[7] Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting of the Association for

DOI:10.46647/ijetms.2025.v09i02.035 ISSN: 2581-4621

Computational Linguistics (ACL), 417-424.
[8] Yuezun Li, Siwei Lyu, “ExposingDF Videos By Detecting Face Warping Artifacts,” in arXiv:1811.00656v3.
[9] Yuezun Li, Ming-Ching Chang and Siwei Lyu “Exposing AI Created Fake Videos by Detecting Eye Blinking” in arXiv:1806.02877v2.
[10] Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen “ Using capsule networks to detect forged images and videos ” in arXiv:1810.11215.
[11] D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6.
[12] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, June 2008.
Anchorage, AK
[13] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin “Detection of Synthetic Portrait Videos using Biological Signals” in arXiv:1901.02212v2
[14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv:1412.6980, Dec. 014.
[15] ResNext Model: https://pytorch.org/hub/pytorch_vision_resnext/ accessed on 06 April 2020
[16] https://www.geeksforgeeks.org/software-engineering-cocomo-model/ Accessed on 15 April 2020
[17] Deepfake Video Detection using Neural Networks http://www.ijsrd.com/articles/IJSRDV8I10860.pdf
[18] International Journal for Scientific Research and Development http://ijsrd.com/