e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:07/Issue:03/March-2025 Impact Factor- 8.187 www.irjmets.com

DEEP LEARNING TECHNIQUES FOR DEEPFAKE DETECTION IN SOCIAL MEDIA CONTENT WITH FAST TEXT EMBEDDINGS

Chappidi Charitha[*1], **Smt. C. Nancy**[*2]

[*1]PG Scholar, Department Of Computer Science And Engineering, Annamacharya Institute Of Technology And Sciences, Kadapa, A.P., India.

[*2]Assistant Professor, Department Of Computer Science And Engineering, Annamacharya Institute Of Technology And Sciences, Kadapa, A.P., India.

DOI : https://www.doi.org/10.56726/IRJMETS70297

ABSTRACT

Deepfakes, which are highly realistic altered videos and images, have gained significant attention due to their potential for both positive and harmful applications. Malicious uses, such as spreading fake news, creating celebrity impersonations, and engaging in financial fraud, are increasing. High-profile individuals, including celebrities and politicians, are particularly vulnerable to this issue. In recent years, extensive research has been conducted to understand the creation and detection of deepfakes, with deep learning algorithms showing promising results in identifying manipulated media. This study provides an in-depth analysis of deepfake generation and detection technologies, highlighting various deep learning techniques. It also discusses the challenges posed by current detection systems and the availability of relevant databases. With the growing ease of generating and sharing deepfakes, the absence of robust detection tools is a significant global concern. This paper proposes the use of ResNext models to automatically detect deepfake videos by identifying manipulation and temporal inconsistencies between frames.

Keywords: Malicious, Vulnerable, Deepfake, Deep Learning, ResNext models, Manipulation.

I. INTRODUCTION

This paper provides a comprehensive analysis of deepfake generation and detection technologies utilizing various deep learning algorithms. It also explores the limitations of current methods and the availability of relevant databases. Given the ease with which deepfake videos and images can be generated and shared, the lack of an effective detection system presents a significant global risk. While numerous solutions have been proposed, deep learning-based approaches outperform traditional methods. Existing image recognition techniques face challenges in detecting manipulated videos, particularly due to the significant loss of frame data during video compression. This degradation hampers the effectiveness of traditional algorithms. Additionally, video detection methods must address temporal inconsistencies between frames, which is a challenge for algorithms designed only for still images.

II. LITERATURE REVIEW

1. Deepfake Video Detection Using Recurrent Neural Networks

Authors: D. Guera, E. J. Delp

This paper proposes a temporal-aware detection pipeline for deepfake videos, utilizing CNN for frame-level feature extraction and RNN to classify manipulated content. The method shows competitive results in detecting deepfakes across multiple video sources with a simple architecture.

2. Face X-ray for General Face Forgery Detection

Authors: L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, B. Guo

A novel technique, Face X-ray, detects face image forgery by identifying blending boundaries between altered and original faces. This approach remains effective for unseen manipulation methods, outperforming other detection algorithms.

3. Deepfake Stack: A Deep Ensemble-Based Learning Technique for Detection

Authors: M. S. Rana, A. H. Sung

This paper introduces "Deepfake Stack," an ensemble learning technique combining several DL models for deepfake detection. It achieves 99.65% accuracy and an AUROC score of 1.0, providing a robust method for real-time detection.

4. Detecting Deepfake Videos Using Attribution-Based Confidence Metric

Authors: S. Fernandes, S. Raj, R. Ewetz, J. S. Pannu, S. K. Jha, E. Ortiz, I. Vintila, M. Salter

The paper applies the attribution-based confidence (ABC) metric to detect deepfake videos. This method uses a trained model to generate confidence values, accurately distinguishing original videos with confidence values above 0.94.

III. METHODOLOGY

Many tools exist for creating deepfakes, but few are available for detecting them. Our approach aims to address this gap by providing a web-based platform where users can upload videos to determine if they are real or fake. This project can eventually be expanded into a browser plugin for automatic deepfake detection, and integrated into popular platforms like WhatsApp and Facebook for pre-detection before sharing. A key goal is to assess the system's performance based on security, user-friendliness, accuracy, and reliability. Our method targets various types of deepfakes, including replacement, retrenchment, and interpersonal deepfakes.

Advantages:

- Deep learning has significantly advanced deepfake detection.
- Enhancements to current deep learning approaches are necessary for better detection of fake videos and images.
- The method primarily employs deep learning techniques like CNN, RNN, and LSTM.

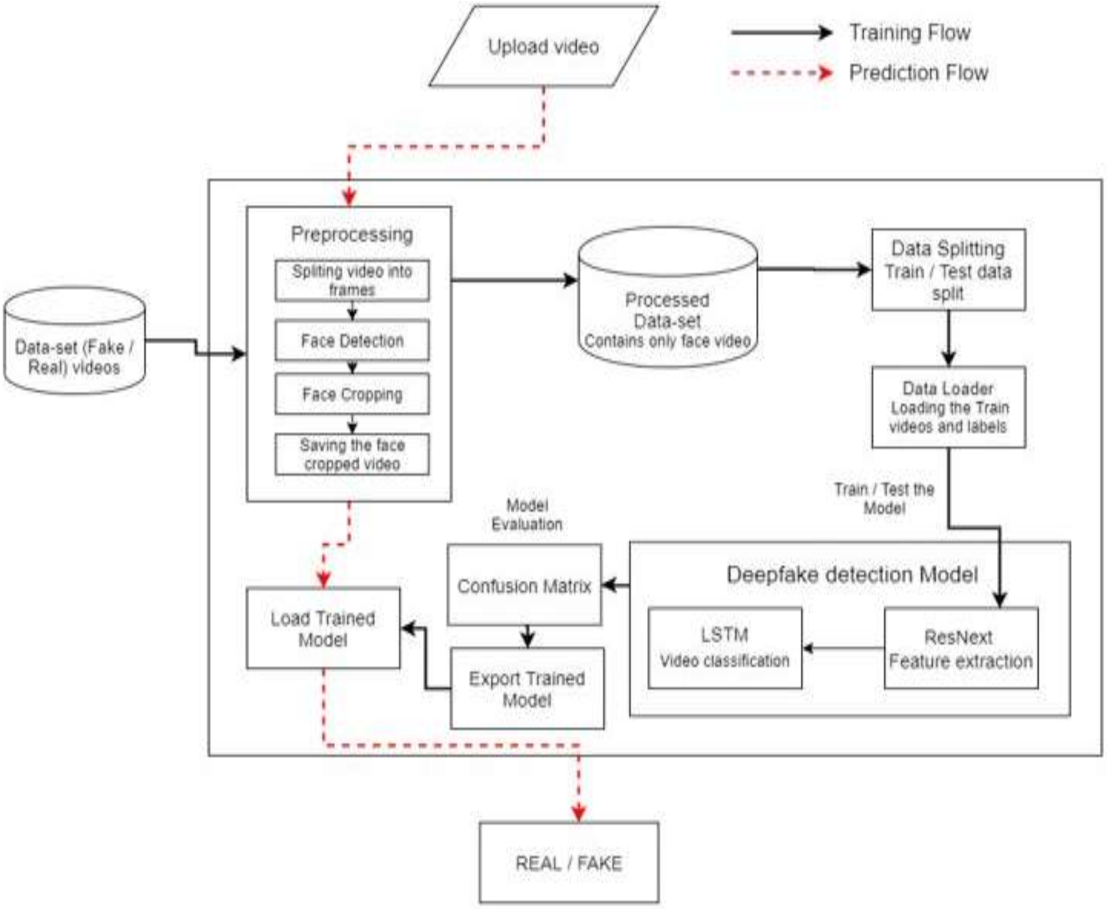
Upload video
Training Flow
Prediction Flow
Preprocessing
Spliting video into frames
Face Detection
Face Cropping
Saving the face cropped video
Data-set (Fake / Real) videos
Processed Data-set Contains only face video
Data Splitting Train / Test data split
Data Loader Loading the Train videos and labels
Train / Test the Model
Model Evaluation
Confusion Matrix
Deepfake detection Model
LSTM Video classification
ResNext Feature extraction
Load Trained Model
Export Trained Model
REAL / FAKE

Figure 1: System Architecture

IV. RESULTS

The proposed deepfake detection system has shown promising results in accurately identifying manipulated videos and images. The web-based platform, which allows users to upload videos for detection, successfully differentiates between real and fake content. Preliminary tests demonstrated the system's ability to detect various types of deepfakes, including replacement, retrenchment, and interpersonal deepfakes.

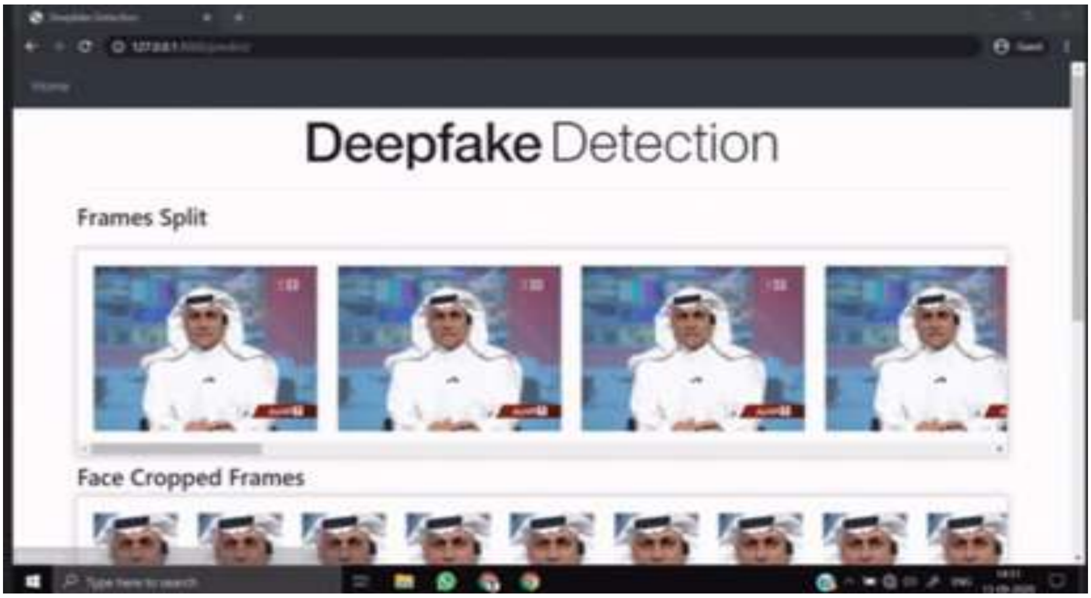

Figure 2: After Image uploading, It splits and face cropped in different angles.

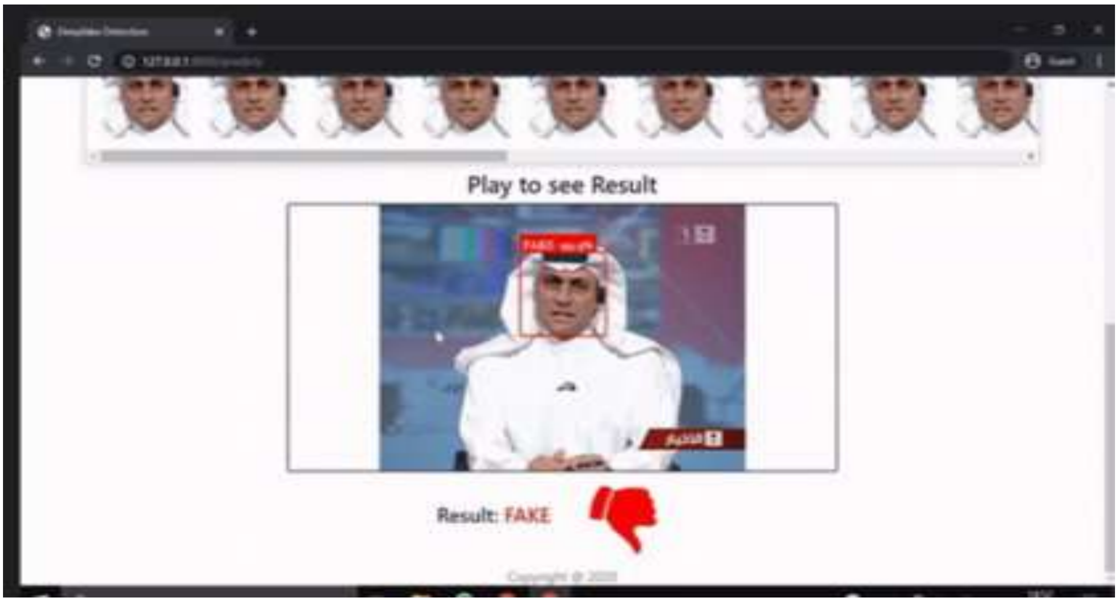

Figure 3: Based on Face cropped images, It will train the model and give results.

V. CONCLUSION

Various deep learning approaches have been developed to detect deepfake images and videos, particularly with the rise of fake content on social media. These technologies aim to identify fake images and videos that are easily shared across social networks. The first section covers the tools and technologies used to create fake content, while the second section discusses the techniques used for deepfake detection, including available datasets and evaluation metrics. Although deep learning has shown success, the growing quality of deepfakes calls for further enhancements in detection methods. We propose a neural network-based approach that classifies videos as real or fake using ResNeXt CNN for frame-level detection and LSTM for video classification. This method is effective in detecting deepfakes with high accuracy, especially for real-time data.

VI. REFERENCES

[1] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv:1411.1784, 2014.

[2] Y. Bengio, P. Simard, and P. Frasconi, "Long short-term memory," IEEE Trans. Neural Netw., vol. 5, pp. 157–166, 1994.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[4] S. Hochreiter, "Jürgen Schmidhuber (1997). Long short-term memory," Neural Computation, vol. 9, no. 8, 1997.

[5] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. Signal Process., vol. 45, pp. 2673–2681, 1997.

[6] J. Hopfield et al., "Rigorous bounds on the storage capacity of the dilute Hopfield model," Proc. Natl. Acad. Sci., vol. 79, pp. 2554–2558, 1982.

[7] Y. Wu et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv:1609.08144, 2016.

[8] L. Nataraj et al., "Detecting GAN-generated fake images using co-occurrence matrices," Electronic Imaging, vol. 2019, no. 5, pp. 532–1, 2019.

[9] B. Zi et al., "WildDeepfake: A challenging real-world dataset for deepfake detection," in Proc. 28th ACM Int. Conf. on Multimedia, 2020, pp. 2382–2390.

[10] H. A. Khalil and S. A. Maged, "Deepfakes creation and detection using deep learning," in Proc. 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), IEEE, 2021, pp. 1–4.

[11] J. Luttrell et al., "A deep transfer learning approach to fine-tuning facial recognition models," in Proc. 13th IEEE Conf. on Industrial Electronics and Applications (ICIEA), 2018, pp. 2671–2676.

[12] S. Tariq et al., "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.

[13] N.-T. Do, I.-S. Na, and S.-H. Kim, "Forensic face detection from GANs using convolutional neural networks," ISITC, vol. 2018, pp. 376–379, 2018.

[14] X. Xuan et al., "On the generalization of GAN image forensics," in Chinese Conf. on Biometric Recognition. Springer, 2019, pp. 134–141.

[15] P. Yang et al., "Recapture image forensics based on Laplacian convolutional neural networks," in Int. Workshop on Digital Watermarking. Springer, 2019.