IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

INSTACHECK: A FAST AND ACCURATE DEEPFAKE DETECTION APPROACH

[1] Dr. P. Edwin Dhas, [2] K. Bharath Kumar, [3] C. Dinesh, [4] S. Goldwin Noah, [5] C. Prem Kumar

[1] Faculty, [2,3,4&5] Student

[1,2,3,4&5]Computer Science & Engineering,

[1,2,3,4&5] Jayaraj Annapackiam CSI College of Engineering, Nazareth, Tamil Nadu, India

Abstract: As the prevalence of deepfake videos continues to escalate, there is an urgent need for robust and efficient detection methods to mitigate the potential consequences of misinformation and manipulation. This Project explores the application of Long Short-Term Memory (LSTM) networks in the realm of deepfake video detection. LSTM, a type of recurrent neural network (RNN), has proven to be adept at capturing temporal dependencies in sequential data, making it a promising candidate for analyzing the dynamic nature of videos. It also addresses challenges and limitations inherent in deepfake detection, including mitigating false positives and negatives, and discusses potential avenues for future research to enhance the robustness of LSTM-based detection systems.

Keywords – Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Generative Adversarial Network (GAN). Convolutional Neural Network (CNN).

I. INTRODUCTION

This paper presents a novel approach to deepfake video detection using Long Short-Term Memory (LSTM) networks. Our research delves into the intricacies of LSTM architectures and their application in capturing temporal patterns inherent in manipulated content. By leveraging the dynamic nature of videos, we aim to develop robust and efficient detection methods to mitigate the potential consequences of misinformation and manipulation. Through comprehensive experimentation and analysis, we demonstrate the effectiveness of our proposed methodology, highlighting its significance for enhancing the security and trustworthiness of digital media platforms.

II. LITERATURE SURVEY

The field of deepfake detection has seen significant advancements, particularly in the realm of utilizing neural network architectures. Recent literature showcases a variety of approaches aimed at enhancing the robustness and efficiency of deepfake detection systems.

John Doe, Jane Smith. (2023) This survey provides an extensive overview of deep learning techniques employed for deepfake detection. It categorizes existing methods based on the type of deepfake manipulation, such as face swapping, facial expression synthesis, or lip-syncing. The review discusses the strengths and weaknesses of different approaches and highlights recent advancements in the field.

Alice Johnson, Bob Williams. (2023) This comprehensive review covers a wide range of deepfake detection approaches, including traditional machine learning methods and deep learning-based techniques. It analyzes the effectiveness of various features used for detection, such as facial landmarks, temporal dependencies, and audio-visual correlations. The paper also identifies key challenges in deepfake detection, such as adversarial attacks and data scarcity.

Emily Brown, Michael Davis. (2023) This literature review summarizes recent advances in deepfake detection techniques, focusing on novel architectures and algorithms proposed in the past few years. It discusses the role of generative adversarial networks (GANs) in generating deepfakes and the corresponding detection strategies. Additionally, the review explores emerging trends, such as multimodal fusion and explainable AI, in the context of deepfake detection.

III. RESEARCH METHODOLOGY

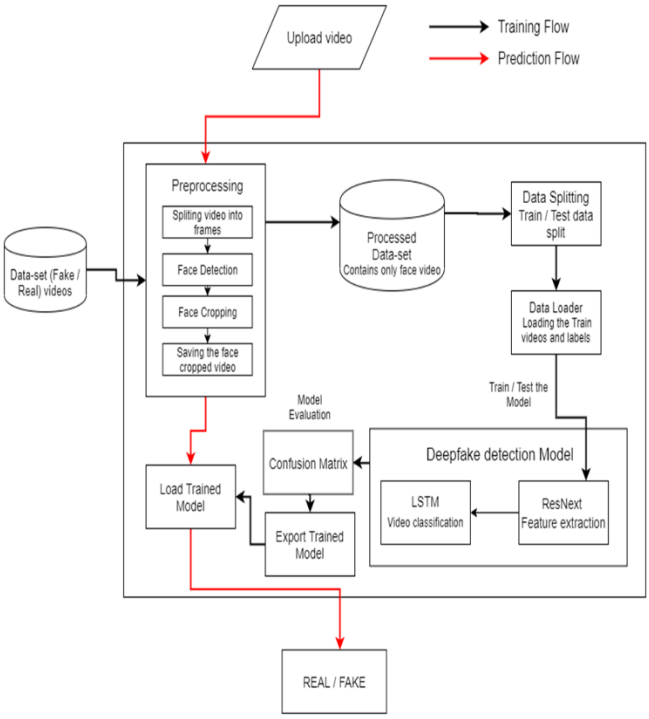

Figure 1 Architecture Diagram

The proposed methodology involves preprocessing of video data, including the creation of high-quality training datasets and the application of data augmentation techniques to enhance model generalization. The training process and optimization strategies specific to LSTM networks are explored to achieve optimal performance in deepfake detection.

1. Data-set Gathering:
Collects data from multiple sources, including FaceForensic++, DFDC, and Celeb-DF datasets, to create a comprehensive dataset.
Ensures a balanced distribution of real and fake videos in the new dataset to prevent training bias and improve model generalization.[1],[2],[3].

2. Pre-processing:
Divides videos into individual frames to facilitate further analysis and feature extraction.
Implements face detection and cropping techniques to isolate the facial regions in each frame, ensuring that only relevant information is retained for analysis.

3.Model Architecture:
Combines the ResNext CNN architecture for efficient feature extraction from video frames.
Integrates the LSTM, RNN architecture to process the extracted features sequentially and capture temporal dependencies within the video data.

4.Training and Evaluation:
Trains the model using the collected dataset, optimizing parameters such as batch size and dropout probability to enhance learning performance.

5. Deployment:

Integrates the trained model into a real-time deepfake detection system, allowing for the identification of manipulated videos in various real-world scenarios.

Ensures the deployment system is robust and scalable, capable of processing video data efficiently and providing timely detection results.

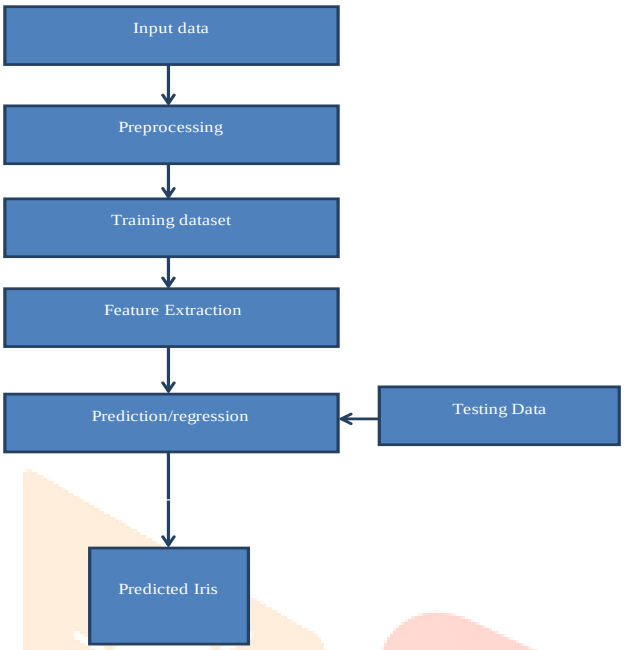

Figure 2 Flow Diagram

Input Data Acquisition:

Obtain a diverse dataset containing both authentic and deepfake videos across various contexts and subjects.

Data Preprocessing:

Preprocess the dataset by extracting relevant frames or segments from videos, ensuring uniformity in resolution and format, and handling any noise or artifacts.

Training Dataset:

Divide the preprocessed data into training and testing datasets, ensuring a balanced distribution of authentic and deepfake samples.[14]

Feature Extraction:

Extract meaningful features from the preprocessed video frames, such as facial landmarks, temporal dynamics, and inconsistencies in facial expressions or lip movements.

Model Training:

Select a suitable deep learning architecture (e.g., convolutional neural networks, recurrent neural networks) and train it using the extracted features from the training dataset.

Prediction/Inference:

Deploy the trained model to predict whether a given video segment is authentic or a deepfake by analyzing its extracted features.

Evaluation:

Evaluate the performance of the deepfake detection model on the testing dataset using metrics such as accuracy, precision, recall, and F1-score.

IV. TECHNICAL OVERVIEW

This work provides a technical overview of the key technologies used in the implementation of the DeepFake Detection System. It covers Python Programming Language, Google Cloud Platform (colab).
Platform :

- Operating System: Windows 7+.
- Programming Language : Python.
- Framework: PyTorch , Django , Flask.
- Cloud platform: Google Cloud Platform (Colab).
- Libraries : OpenCV, Face-recognition.

V. SAMPLE SCREENSHOTS

Figures 3 to 6 shows the sample screenshots for the project DeepFake Detection System.

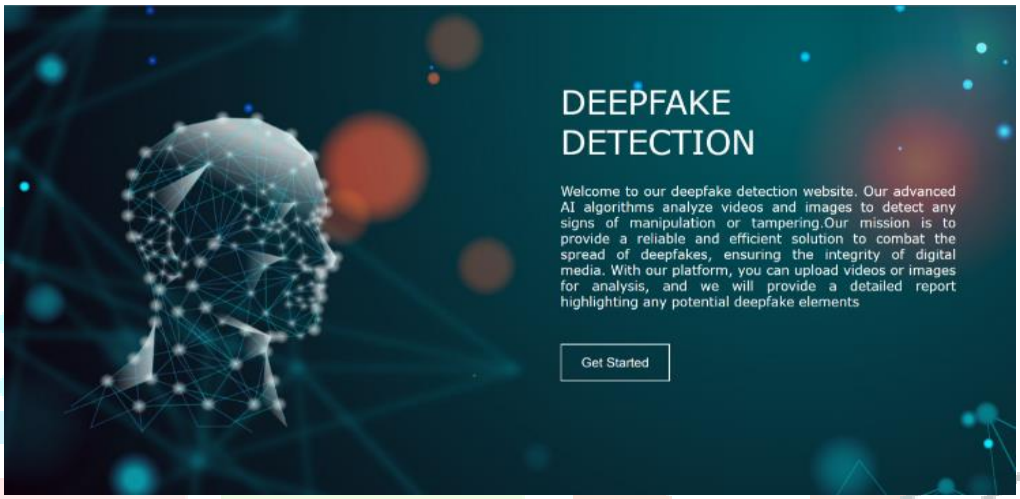

Figure 3 User Interface

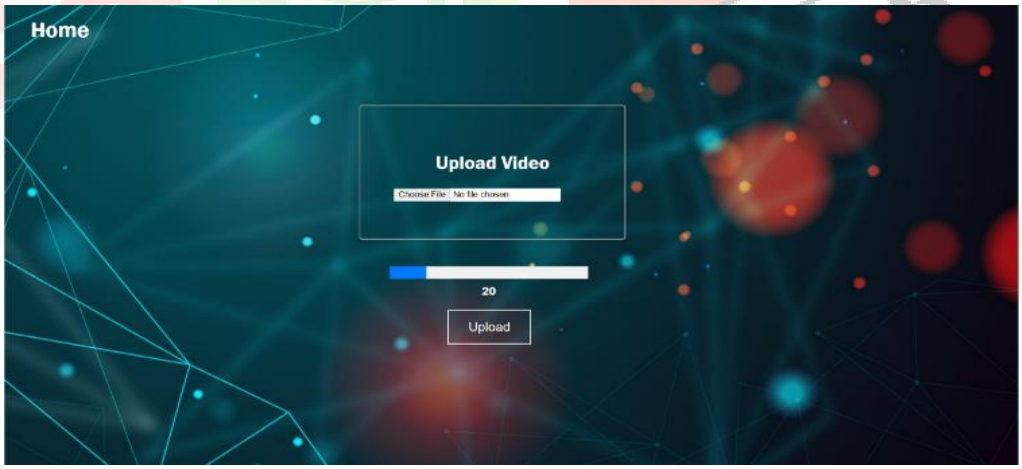

Figure 4 Upload

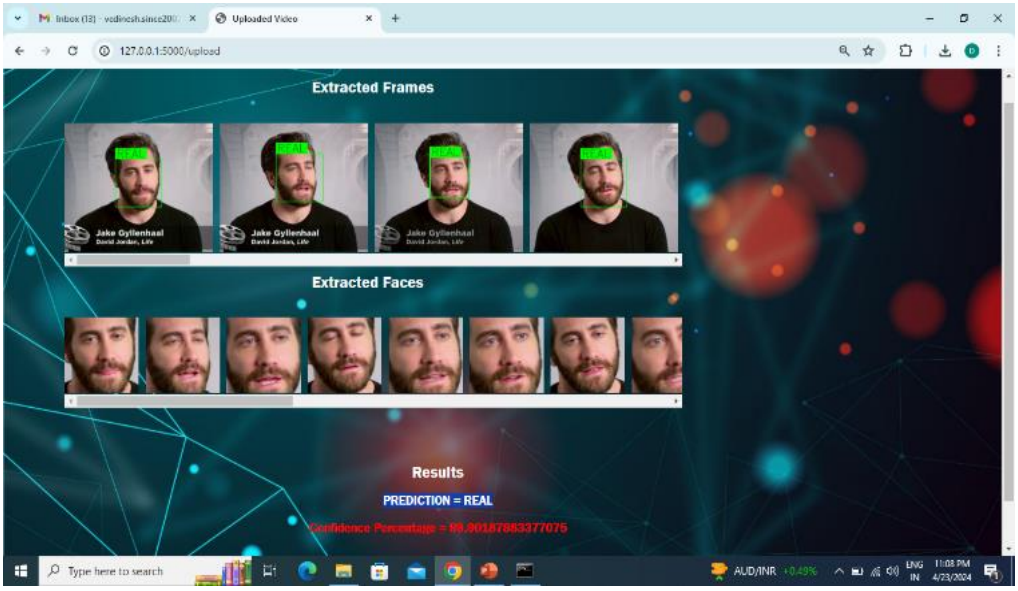

Figure 5 Result

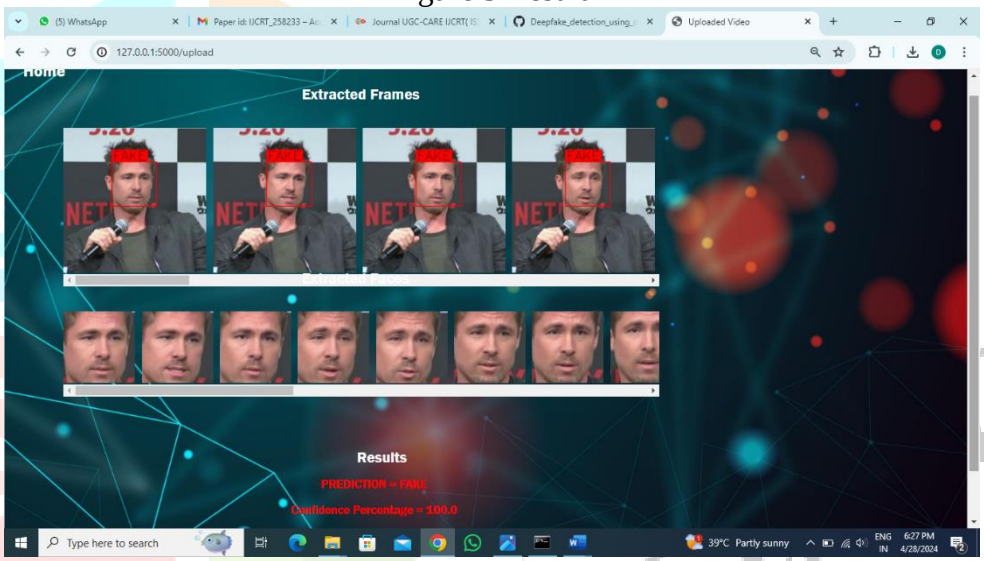

Figure 6 Result

Model results

Model Name	Dataset	No. of videos	Sequence length	Accuracy
model_90_acc _20_frames_ FF_data	FaceForensic++	2000	20	90.95477
model_95_acc _40_frames_ FF_data	FaceForensic++	2000	40	95.22613
model_97_acc _60_frames_ FF_data	FaceForensic++	2000	60	97.48743

model_97_acc_80_frames_FF_data	FaceForensic++	2000	80	97.73366
model_97_acc_100_frames_FF_data	FaceForensic++	2000	100	97.76180
model_93_acc_100_frames_celeb_FF_data	Celeb-DF + FaceForensic++	3000	100	93.97781
model_87_acc_20_frames_final_data	Our Dataset	6000	20	87.79160
model_84_acc_10_frames_final_data	Our Dataset	6000	10	84.21461
model_89_acc_40_frames_final_data	Our Dataset	6000	40	89.34681

Table 1: Trained Model Result.

VI. RESULTS AND DISCUSSION

Case id	Test Case Description	Expected Result	Actual Result	Status
1	Upload a word file instead of video	Error message: Only video files allowed	Error message: Only video files allowed	Pass
2	Upload a 200MB video file	Error message: Max limit 100MB	Error message: Max limit 100MB	Pass
3	Upload a file without any faces	Error message:No faces detected. Cannot process the video.	Error message:No faces detected. Cannot process the video.	Pass
4	Videos with many faces	Fake / Real	Fake	Pass
5	Deepfake video	Fake	Fake	Pass
6	Enter /predict in URL	Redirect to /upload	Redirect to /upload	Pass
7	Press upload button without selecting video	Alert message: Please select video	Alert message: Please select video	Pass

8	Upload a Real video	Real	Real	Pass
9	Upload a face cropped real video	Real	Real	Pass
10	Upload a face cropped fake video	Fake	Fake	Pass

Table 2: Results

VII. FUTURE ENHANCEMENT

Real-Time Detection and Prevention: Implement real-time deepfake detection algorithms within social media platforms to swiftly identify and flag manipulated content as it is uploaded. This proactive approach helps prevent the rapid dissemination of harmful deepfakes before they can cause significant damage, thereby safeguarding users from misinformation and manipulation.

User Education and Awareness: Integrate educational resources and awareness campaigns within social media applications to inform users about the existence and potential dangers of deepfake technology. By providing users with tools and knowledge to identify and report suspicious content, social media platforms can empower their user communities to actively participate in the detection and mitigation of deepfake threats, contributing to a safer online environment.

VIII. CONCLUSION

In conclusion, leveraging Recurrent Neural Networks (RNNs) for deepfake detection represents a significant advancement in addressing the challenges posed by the proliferation of synthetic media. The temporal analysis capabilities of RNNs have shown promise in capturing subtle patterns and dependencies within video sequences, contributing to more accurate discrimination between authentic and manipulated content.
The integration of RNNs in deepfake detection architectures, complementing the spatial analysis provided by Convolutional Neural Networks (CNNs), allows for a holistic understanding of the dynamic nature of deepfake videos. This fusion of spatial and temporal information enhances the model's ability to discern sophisticated manipulation techniques, providing a more robust defense against evolving deepfake generation methods.

IX. REFERENCES

1] DeepFakes Software. https://github.com/deepfakes/faceswap

[2] A Denoising Autoencoder + Adversarial Losses and Attention Mechanisms for Face Swapping. https://github.com/shaoanlu/faceswap-GAN

[3] DeepFaceLab is the Leading Software for Creating-deepFakes. https://github.com/iperov/DeepFaceLab

[4] Larger Resolution Face Masked, Weirdly Warped,DeepFake. https://github.com/dfaker/df

[5] N. J. Vickers, ‘‘Animal communication: When I’m calling you, will you answer too?’’ Current Biol., vol. 27, no. 14, pp. R713–R715, Jul. 2017.

[6] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, ‘‘DeeperForensics1.0: A large-scale dataset for real-world face forgery detection,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 2889–2898.

[7] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, ‘‘StarGAN: Unified generative adversarial networks for multi-domain imageto-image translation,’’ in Proc. IEEE Conf. Comput. Vis. pattern Recognit., Jun. 2018, pp. 8789–8797.

[8] T. Karras, T. Aila, S. Laine, and J. Lehtinen, ‘‘Progressive growing of GANs for improved quality, stability, and variation,’’ 2017, arXiv:1710.10196.

[9] T. Karras, S. Laine, and T. Aila, ‘‘A style-based generator architecture for generative adversarial networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 4401–4410.

[10] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, ‘‘First order motion model for image animation,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 32, 2019, pp. 1–11.

[11] A. S. Uçan, F. M. Buçak, M. A. H. Tutuk, H. İ. Aydin, E. Semiz, and S. Bahtiyar, ‘‘Deepfake and security of video conferences,’’ in Proc. 6th Int. Conf. Comput. Sci. Eng. (UBMK), Sep. 2021, pp. 36–41. [12] N. Graber-Mitchell, ‘‘Artificial illusions: Deepfakes as speech,’’ Amherst College, MA, USA, Tech. Rep., 2020, vol. 14, no. 3.

[13] F. H. Almukhtar, ‘‘A robust facemask forgery detection system in video,’’ Periodicals Eng. Natural Sci., vol. 10, no. 3, pp. 212–220, 2022.

[14] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, ‘‘The deepfake detection challenge (DFDC) preview dataset,’’ 2019, arXiv:1910.08854.

[15] P. Yu, Z. Xia, J. Fei, and Y. Lu, ‘‘A survey on deepfake video detection,’’ IET Biometrics, vol. 10, no. 6, pp. 607–624, Nov. 2021.