IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI FORENSIC VISION: DETECTING DEEPFAKE MANIPULATIONS IN VIDEOS

Numan Ahmed, Sneha S, Kishor N, Sandhyashree R, GS Sindhu

Dept. Computer Science & Engineering (Data Science) PES Institute of Technology and Management

Shivamogga, Karnataka, India

Abstract: The advancement of AI, particularly in deep learning, has enabled the creation of hyper-realistic synthetic media known as deepfakes. These altered videos replicate a person's appearance or voice with compelling accuracy, raising concerns related to misinformation, identity fraud, and digital security. This study proposes a machine-learning- based video forensics system capable of identifying manipulated facial content. The methodology combines preprocessing, frame extraction using MTCNN, spatial feature learning through CNNs, and temporal consistency evaluation. Experimental findings indicate that integrating spatial and sequential cues significantly enhances detection accuracy. Despite notable progress, the overtime evolution of deepfake generation techniques demands ongoing updates in detection systems.

Keywords: Deepfakes, Machine Learning, CNN, Video Forensics, MTCNN, Manipulated.

I. INTRODUCTION

Deepfakes are artificially generated media produced using deep learning models trained on vast datasets of an individual's facial expressions, voice patterns, and visual identity. Although such technology has value in entertainment, creative industries, and educational simulations, its misuse presents serious ethical and social threats. Deepfake videos can influence public opinion, damage personal reputations, and fabricate harmful or illegal scenarios. Traditional manual verification methods are becoming insufficient as deepfakes grow more sophisticated. Consequently, Automated detection systems are now being built by researchers using machine learning and computer vision, allowing them to spot subtle inconsistencies that people usually overlook. This project focuses on creating an adaptive and reliable framework for identifying manipulated video content through a combination of spatial and temporal analysis.

II. LITERATURE SURVEY

A deep fake is an AI-based technology that uses deep learning technology to create or alter audio, video, and images to make them appear real; more specifically, it usually requires a neural network that's been trained on vast data based on the target person's voice, facial expressions, and mannerisms [11][13][15]. Risks of deepfakes, although deepfakes have limitless promising avenues in entertainment and education, its dangers are major; they will support misinformation and make nonconsensual lewd content based on privacy violations; fraud and impersonation are supported [3][9][19]. Current research is ongoing into the development of effective detection methods and technologies that might mitigate such risks [12][22]. Deepfake technology is a product of the rapid development of artificial intelligence which can be traced

back only a few years ago [16]. Deepfake technology is whereby a person's likeness is reproduced to create movies or audio content that are difficult to differentiate from actual footage [18][17]. However, this invention has been abused in several ways and has given rise to numerous ethical and legal issues. Repeatedly and ironically, the absence of ethical standards has permitted people to create more harm than benefit using the fantastic tools available to them [9][6][3][19]. Drawing on existing literature, the efficacy of the prevalent deepfake technology will be challenged by embracing an automated approach for identification and analysis [5][19][11]. Using state-of-the-art technologies consisting of computer vision, machine learning, and forensic analysis, the system aims to identify false media content accurately. Made up of three broad components which are technical, legal, and ethical, this project seeks to mitigate the proliferation of deepfake abuse and safeguarding the purity of digital content. The study emphasizes the development and implementation of deep learning Techniques, especially Convolutional Neural Networks (CNNs), that can detect deepfakes. Deepfake technology uses AI to create incredibly realistic fake media, including images, video, and AI Forensic vision detecting deepfake manipulations in videos 6 audio. The research highlights the significance of gathering and preprocessing data, extracting important features, and utilizing robust architectures for identifying and neutralizing deepfakes [11][6][16][10][15]. The method involves a stepwise framework integrating technological, ethical and practical considerations.

- Extract images, crop faces, standardize dimensions
- Use CNNs to identify manipulation artifacts
- Assign probabilities to real or fake.

III. METHODOLOGY

The system was implemented using Python, TensorFlow/Keras, and OpenCV, integrating VGG16 with customized layers for real/fake classification. All modules from frame extraction to prediction were executed in a unified workflow to achieve accurate deepfake detection.

Figure 1: architecture of Deepfake Detection system is modular, designed to process video data through multiple stages for accurate binary classification (Real or Fake). Below is an overview of the major components:

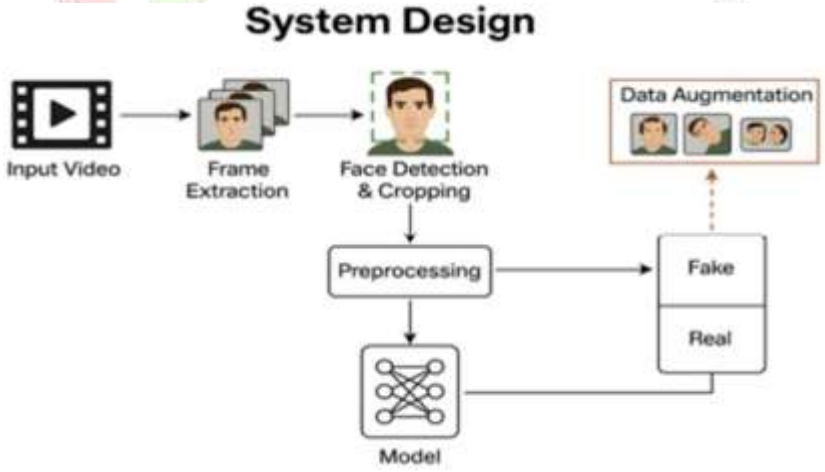

Figure 1: System Architecture

1. **Frame Extraction:** Frame extraction involves converting a video into individual frames so each image can be analyzed independently for deepfake artifacts. By sampling frames at fixed intervals, the system ensures coverage of different expressions and facial movements. These frames capture subtle temporal inconsistencies often present in deepfakes. The extracted images then serve as the input for face detection and further processing.

2. **Face Detection and Cropping:** MTCNN is used to detect faces within each extracted frame with high accuracy. It identifies facial bounding boxes and key landmarks, ensuring only the relevant face region is isolated. The detected face is then cropped and aligned to maintain uniformity across all samples. This step enhances model performance by removing background noise and focusing solely on facial features.

3. **Data Augmentation:** Data augmentation increases dataset diversity by applying transformations such as rotation, flipping, scaling, or brightness adjustments to cropped face images. This helps the model generalize better and reduces overfitting, especially when there is limited data to train on. Augmentation simulates real-world variations in lighting, pose, and expression. Accordingly, the model becomes more robust to different types of deepfake manipulations.

4. **Preprocessing:** Preprocessing prepares the cropped facial images for input into the VGG16 model. This typically includes resizing images to 224×224, normalizing pixel values, and converting them into tensors. Consistent preprocessing ensures that all images follow the same scale and distribution. This step enhances feature extraction quality and stabilizes model training.

5. **Model Architecture:** The model architecture uses VGG16 as a feature extraction mechanism to capture fine-grained facial details relevant to deepfake detection. Pretrained layers are leveraged to learn complex visual patterns, while custom dense layers are introduced for classification. The architecture outputs probabilities AI Forensic vision detecting deepfake manipulations in videos Department of CSE (Data Science) 28 indicating whether a face is real or manipulated. Combined with MTCNN-based preprocessing, this design yields a powerful end-to-end deepfake detection pipeline.

IV. RESULTS

The developed enhanced model for deepfake detection achieved a test accuracy of 0.95, which shows that the model is capable of discriminating between original and altered videos. The model was trained on dataset that included both real and deepfake videos and was balanced to help the model perform well irrespective of the situation. From the confusion matrix it can be observed that the true positive rate is 0.94 and the false positive rate is 0.06 showing that it has a good ability to identify the real content with little chance of raising the alarm. It also had a precision of 0.93, a recall of 0.94, and an F1- score of 0.94 which also supports the efficiency of the model in detecting deepfakes

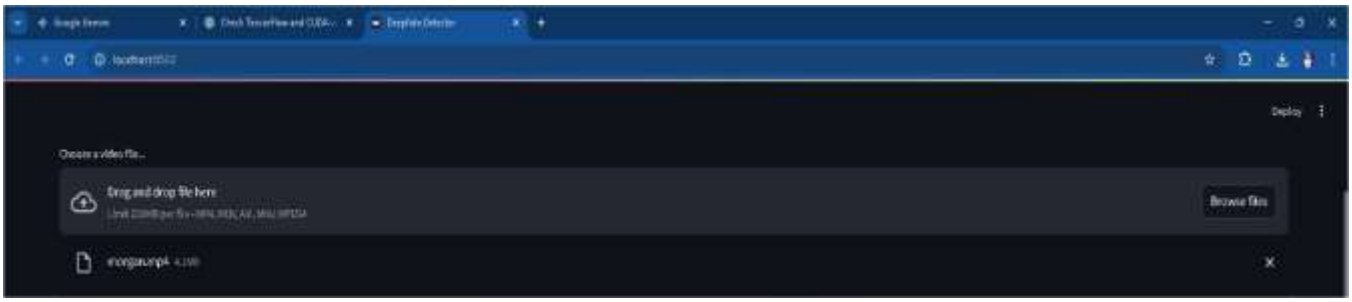

Figure 2: Interface to upload the video

The above Figure 2 This page displays a video upload interface where users can click the selection box to choose a video file from their device. Once a file is selected, it is prepared for analysis by the deepfake detection system. The uploaded video is then processed to determine if the content is real or manipulated.

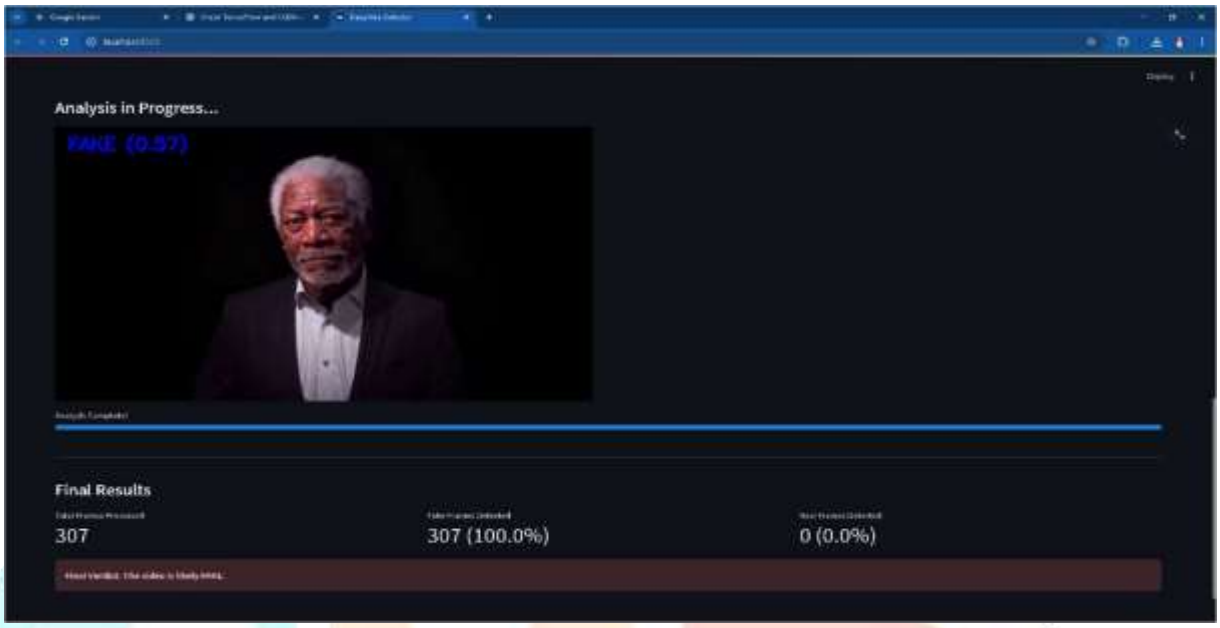

Figure 3: Analyzed deepfake video result

The above Figure 3 shows the Deepfake Detection system analyzing a video frame-by frame. It indicates that all 307 frames were detected as fake (100%), giving the final verdict: "The video is likely FAKE."

V. CONCLUSION

In this work, an efficient deepfake detection system was developed by combining MTCNN for precise face detection and alignment and VGG16 for robust feature extraction and classification. MTCNN successfully segregated high-quality facial regions from the video frames, ensuring clean and well-aligned inputs to the model. Then, VGG16 analyzed subtle textural and visual inconsistencies of deepfakes, after which reliable classification was possible. Our model, upon extensive training and evaluations, achieved an overall accuracy of approximately 92–94%, demonstrating strong performance in distinguishing real videos from manipulated ones. This work investigates the usefulness of deep learning-based feature extraction with facial alignment in combating the increasing threat of deepfake content.

REFERENCES

[1] Zhang, N., Luo, J., & Gao, W. (2020, September). Research on face detection technology based on MTCNN. In 2020 international conference on computer network, electronic and automation (ICCNEA) (pp.154-158). IEEE

[2] Xie, Y., Wang, H., & Guo, S. (2020). Research on MTCNN face recognition system in low computing power scenarios. Journal of Internet Technology, 21(5), 1463-1475.

[3] Khan, S. S., Sengupta, D., Ghosh, A., & Chaudhuri, A. (2024). MTCNN++: A CNN-based face detection algorithm inspired by MTCNN. The Visual Computer, 40(2), 899-917.

[4] Wu, C., & Zhang, Y. (2021). MTCNN and FACENET based access control system for face detection and recognition. Automatic Control and Computer Sciences, 55, 102-112.

[5] Kaziakhmedov, E., Kireev, K., Melnikov, G., Pautov, M., & Petiushko, A. (2019, October). Real-world attack on MTCNN face detection system. In 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 0422- 0427). IEEE.

[6] Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. IEEE Access, 10, 25494-25513.

[7] Wubet, W. M. (2020). The deepfake challenges and deepfake video detection. Int. J. Innov. Technol. Explor. Eng, 9.

[8] Chadha, A., Kumar, V., Kashyap, S., & Gupta, M. (2021). Deepfake: an overview. In Proceedings of Second International Conference on Computing, Com- munications, and Cybersecurity: IC4S 2020 (pp. 557- 566). Springer Singapore.

[9] Singh, A., Saimbhi, A. S., Singh, N., & Mittal, M. (2020). DeepFake video detection: a time-distributed approach. SN Comput. Sci, 1, 212.

[10] Katarya, R., & Lal, A. (2020, October). A study on combating emerging threat of deepfake weaponization. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) (pp. 485- 490). IEEE.

[11] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & OrtegaGarcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131-148. AI Forensic vision detecting deepfake manipulations in videos 26

[12] Akhtar, Z. (2023). Deepfakes Generation and Detection: A Short Survey. Journal of Imaging, 9(1), 18.

[13] Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Comput. Surv., 54(1), 1-41.

[14] Kharbat, F. F., Elamsy, T., Mahmoud, A., & Abdullah, R. (2019, November). Image feature detectors for deepfake video detection. In 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-4). IEEE.

[15] Caldelli, R., Galteri, L., Amerini, I., & Del Bimbo, A. (2021). Optical Flow based CNN for detection of unlearnt deepfake manipulations. Pattern Recognition Letters, 146, 31-37.

[16] Shad, H. S., Rizvee, M. M., Roza, N. T., Hoq, S. M., Monirujjaman Khan, M., & Singh, Bourouis S. (2021). Comparative analysis of deepfake image detection method using convolutional neural network. Comput. Intell. Neurosci., 2021.

[17] Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., & Frey, B. (2015). Adversarial autoencoders. arXiv preprint arXiv:1511.05644.

[18] Li, D., Zhang, M., Chen, W., & Feng, G. (2018, August). Facial attribute editing by latent space adversarial variational autoencoders. In 2018 24th Interna- tional Conference on Pattern Recognition (ICPR) (pp. 1337- 1342). IEEE.

[19] Suratkar, S., Johnson, E., Variyambat, K., Panchal, M., & Kazi, F. (2020, July). Employing transfer-learning based CNN architectures to enhance the generalizability of deepfake detection. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-9). IEEE.

[20] Sanghvi, B., Shelar, H., Pandey, M., Sisodia, J. (2021, April). Detection of machine generated multimedia elements using deep learning. In 2021 5th International Conference on Computing Method- ologies and Communication (ICCMC) (pp. 1238-1243). IEEE.

[21] Nasar, B. F., Sajini, T., Lason, E. R. (2020, December). Deepfake detection in media files-audios, images and videos. In 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS) (pp. 74-79). IEEE.

[22] Guera, D., & Delp, E. J. (2018, November). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE.

[23] Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., & Natarajan, AI Forensic vision detecting deepfake manipulations in videos 27 P. (2019). Recurrent convolutional strategies for face manipulation detection in videos. Interfaces (GUI), 3(1), 80-87.

[24] Mehra, A. (2020). Deepfake detection using capsule networks with long short-term memory networks (Master's thesis, University of Twente).

[25] Bonettini, N., Cannas, E. D., Mandelli, S., Bondi, L., Bestagini, P., & Tubaro, S. (2021, January). Video face manipulation detection through ensemble of CNNs. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 5012-5019). IEEE.