WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Cutting-Edge Real-Time System for the Detection of AI-Generated and Manipulated Video Content.

Neha Pawar, Sanika Arekar, Sanika Gaikwad, Prajakta Teli, Ms. Amrapali Babar

Department of Computer Science and Engineering, Annasaheb Dange College of Engineering and Technology, Ashta, Maharashtra
abcse@adcet.in, kp90040@gmail.com, sanikaarekar01@gmail.com, gaikwadsanika45@gmail.com, teliprajakta8@gmail.com

ABSTRACT—

This study provides a social media network inspired by Instagram that incorporates an integrated AI-based system for video tampering detection. For deepfake signatures, the system analyses submitted video information using MesoNet, a compact convolutional neural network. To guarantee the authenticity of the content, detected altered videos are immediately marked with a visible watermark before being shared. To create a safe, scalable multipage application, the project makes use of Node.js, MongoDB, Python, and MoviePy. Results reveal 85–90

Index Terms—Deepfake Detection, Social Media, Video Water- marking, MesoNet, Content Authentication, Web Security

I. INTRODUCTION

This paper suggests a solution by integrating real-time deepfake detection directly into the media upload process of a web application, with the goal of enhancing the authenticity of shared content and protecting users from maliciously altered videos. The rapid advancement of AI technologies, particularly deepfake generation, has raised significant concerns about the authenticity of media shared across digital platforms. Deepfakes, especially in video form, are increasingly being used to spread misinformation, disrupt trust, and manipulate public opinion. Social media platforms, which are central to global communication, have become a key vector for the propagation of such deceptive content..

II. OBJECTIVE

• To instantly identify deepfake video uploads. • To add a noticeable FAKE tag to watermarked videos that have been altered. • To use MongoDB to safely store and show videos.

• To offer a safe and easy-to-use media sharing platform • To stop false information by warning consumers of modified content.

III. LITERATURE REVIEW

Andreas et al [1] this paper examines the realism of state-of the-art image manipulations, and how difficult it is to detect them, either automatically or by humans. After the collecting data it is manipulated, then the image is detected whether it is fake or real using CNNs convolutional neural networks. Yuezun Li et al .

[2] The need to develop and evaluate Deep Fake detection algorithms calls for large-scale datasets. However, current Deep Fake datasets suffer from low visual quality and do not resemble Deep Fake videos circulated on the Internet. The use of DNNs has made the process to create convincing fake videos increasingly easier and faster. In this work, they present a new large-scale and challenging Deep Fake video dataset, Celeb- DF3, for the development and evaluation of Deep Fake detection algorithms

Brian et al [3] The DFDC is the largest currently and pub- licly available face swap video dataset. The dataset contains over 100,000 clips from 3,426+ paid actors. The dataset is created using several Deep fakes and GAN-based and non- learning techniques.

Ricard et al[4] By analysing a low-resolution video se- quence of Face Forensics++ dataset, our method detects ma- nipulated videos with 90

Ruben et al [5] This survey offers a comprehensive overview of methods to detect and manipulate face images, including Deep Fake techniques. Specifically, four categories of face manipulation are examined: i) the full face; ii) switching iden- tities; iii) modifying characteristics; iv) switching

expressions. Nicol'o et al [6] Take up the challenge of detecting face alteration in video sequences that use contemporary facial manipulation methods. Using more than 10,000 videos, the CNN approach is used to recognize false videos.

Bojia et al [7] In this research, we offer a new dataset called Wild Manipulated, which comprises of 7,314 face sequences derived from 707 deep fake videos acquired entirely from the internet, to better enhance detection against real-world deep fakes. Two Attention-based Manipulated Detection Networks (ADDNets) were presented by the researcher.

Kaede et al [8] In order to identify deep fakes, we introduce in this paper new synthetic training data dubbed self- blended images (SBIs). To replicate forging artifacts, SBIs are created by merging source and target photos that have been marginally altered from one authentic image.

IV. Proposed Solution

The system is designed with three core components: fron- tend, backend, and an AI detection engine.

The frontend provides a user-friendly interface where users can upload video content. It also manages user sessions and displays results (REAL or FAKE) after analysis.

The backend handles video processing tasks. It checks the file type, sends video files to the Python-based AI engine, and manages the storage of results. If a video is detected as fake, it is watermarked using MoviePy, and both the watermarked video and metadata (such as prediction score and authenticity label) are stored in a secure database.

The AI engine, powered by MesoNet, analyzes extracted video frames to detect signs of manipulation. If the prediction score exceeds a set threshold, the video is classified as FAKE; otherwise, it's marked REAL.

The final output is a verified or watermarked video that is stored securely and can be accessed by users through a protected content feed.

V. Related Work

Several studies have explored deepfake detection techniques using deep learning algorithms. Afchar et al. (2018) in their pa- per "MesoNet: a Compact Facial Video Forgery Detection Net- work" introduced MesoNet, a lightweight deep learning model designed to identify manipulated facial features in videos. This model is particularly effective in detecting deepfakes with high accuracy. Similarly, Zhao et al. (2018) in "Deepfake Detection Through Visual Artifacts" examined the use of CNNs for detecting inconsistencies in facial movements, lighting, and textures, which are often present in deepfake videos.

To secure media and ensure its authenticity, Wang et al. (2019) proposed using blockchain technology in their paper "Blockchain for Media Authentication: A Survey". The au- thors discussed the potential of blockchain to create immutable records of video uploads, allowing for the verification of video authenticity through timestamping and metadata storage. This method ensures that users can trace the origin and history of the media content, providing an additional layer of security and trustworthiness.

Watermarking has also been explored as a method to mark manipulated videos. In their paper "Dynamic Watermarking for Video Integrity", Yao et al. (2020) proposed dynamic wa- termarking techniques where visible or invisible watermarks are applied to video content once deepfake detection is com- pleted. MoviePy, a Python-based tool, has been widely used in these applications to overlay watermarks on videos, clearly indicating whether the content is genuine or manipulated.

Some studies have examined integrating user feedback to improve deepfake detection. McMahan et al. (2017) in their paper "Federated Learning: Collaborative Machine Learning Without Centralized Training Data" discussed the use of federated learning to update detection models without com- promising user privacy. Additionally, Panda et al. (2020) in their paper "Real-Time Deepfake Detection in Media Upload Pipelines" proposed embedding real-time deepfake detection directly into media upload pipelines, offering immediate feed- back to users during the video upload process.

In recent work, Korshunov and Marcel (2020) in their paper "Deepfake Video Detection using Deep Learning and Audio- Visual Synchronization" explored multi-modal approaches combining audio, video, and metadata analysis for deepfake detection. They found that combining these data sources could significantly improve the accuracy and reliability of detecting manipulated media.

VI. System Architecture and Modules

The cutting-edge real-time system for the detection of AI- generated and manipulated video content comprises three main components: the Database, Admin, and AI Detection modules.

A. Admin Module

This module manages users, monitors system activities, and safeguards the integrity of the platform. It oversees flagged content and controls the detection process to ensure proper functionality and compliance.

B. User Module

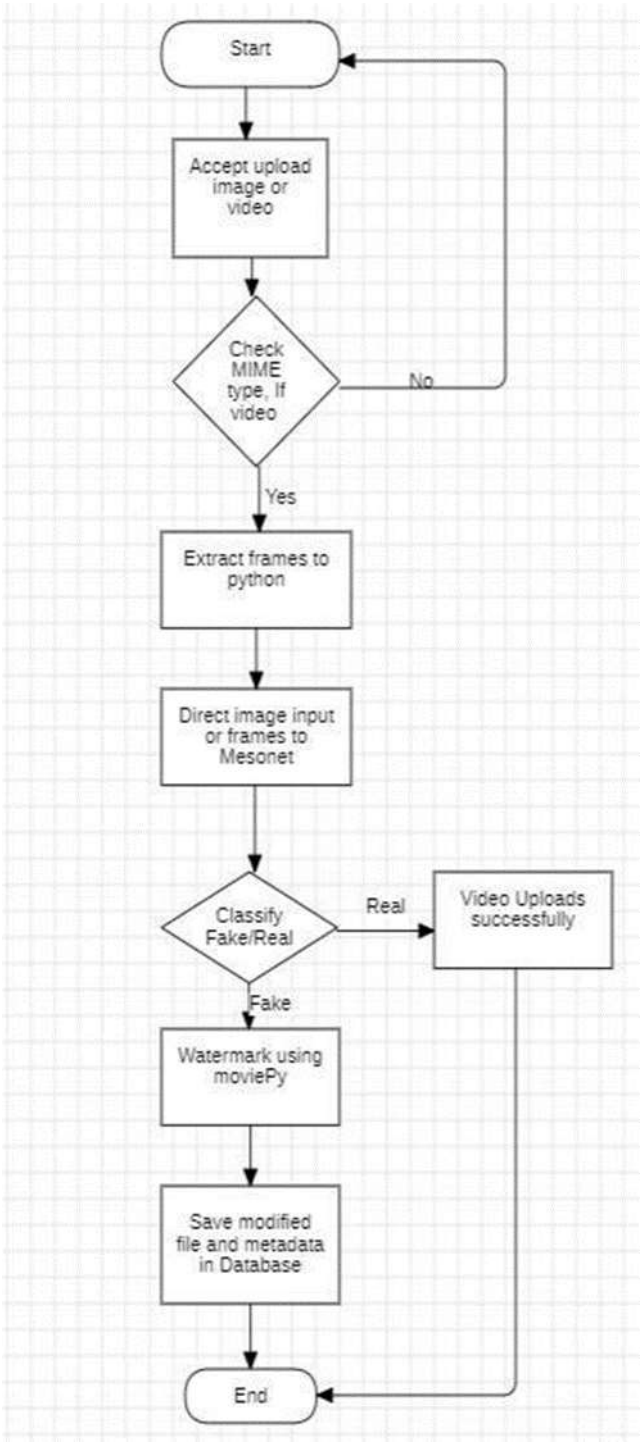
Users can register, log in, and upload videos for verification. The system provides real-time feedback along with detailed detection results, ensuring a seamless user experience.

C. AI Detection Module

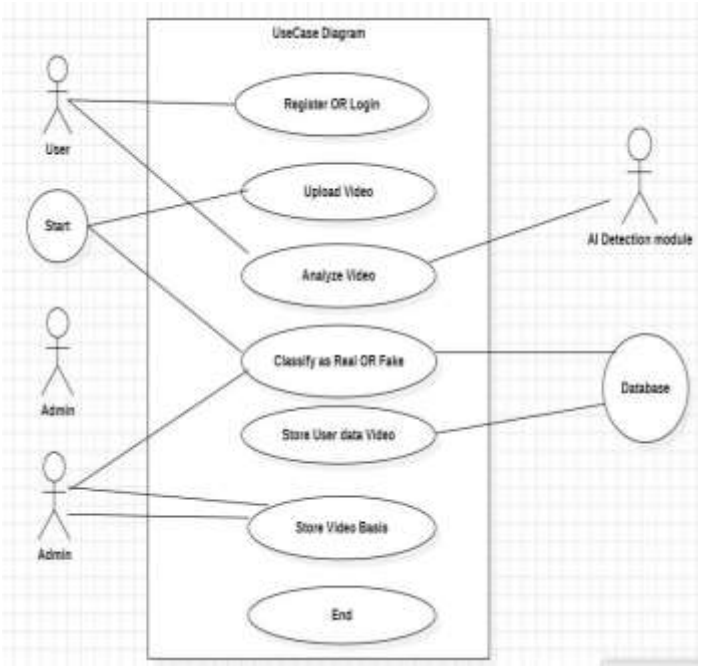
Powered by Python, this backend module uses the MesoNet deep learning model to efficiently and accurately classify uploaded video content as real or AI-generated (deepfake).

D. Database Module

Utilizing MongoDB, this module ensures secure storage and efficient retrieval of user credentials, uploaded video information, detection logs, and file locations.



Use case



VII. TECHNOLOGY IMPLEMENTATION

HTML: All users can easily navigate and engage with the platform thanks to the user interface's responsive and intuitive design.

CSS: CSS ensures a unified user experience throughout the system by producing an aesthetically pleasing and consistent design.

JavaScript: JavaScript provides a smooth user experience by managing client-side interactivity, such as dynamic updates, form validation, and video previews.

Real-time video processing and safe connection with the database and AI models are made possible by the combination of Node.js and Express.js, which handle backend logic and API routing.

MongoDB: MongoDB provides scalable, effective data stor- age and retrieval by storing user data, video information, and detection results.

Python TensorFlow (MesoNet): The MesoNet model is used to analyse and categorise video authenticity with high accuracy in the deepfake detection process, which is powered by Python and TensorFlow.

MoviePy: For safe tracking and traceability, also it is used to watermark modified videos.

VIII. ALGORITHM

1. Prompt user to upload media file (image or video)
2. Check MIME type of the uploaded file If MIME type == ”video” then

Send file to Python backend Else

Display message: ”Only video files are supported”

3. In Python backend:

a. Extract frames from video using OpenCV

b. Initialize total*score* = 0, *framecount* = 0

FOR each frame in video:

i. Feed frame into MesoNet model

ii. Get prediction*score*

iii. total*score*+ = *predictionscore*

iv. frame*count*+ = 1

4. avg*score* = *totalscore/framecount*
5. If avg*score* > *thresholdTHEN*

Label video as Fake

Apply watermark ”FAKE” using MoviePy Else

Label video as Real

6. Save modified video to server storage
7. Store metadata in database:

- filename
- prediction*score*
- label (FAKE/REAL)
- timestamp

8. Display result to user

IX. Future Scope

In order to maintain the authenticity of digital content,

the suggested solution integrates cutting-edge technology and presents a strong foundation for real-time identification of AI- generated and modified video content. To guarantee un- changeable media integrity verification across platforms and promote transparency and traceability in content distribution, future advancements might use blockchain logging techniques (Kumar Das, 2022). By using federated learning, the system could allow for model updates that preserve privacy based on localised data, improving detection accuracy while protecting user privacy (Li et al., 2021). A comprehensive multi-modal deepfake detection suite that can recognise phoney text and speech would be established by expanding to include audio and text forgery detection (Zhang et al., 2020). Furthermore, by improving model performance over time through user- reported errors, a user feedback loop could enable ongoing learning and adaptation, fostering dependability and commu- nity confidence in the detection process (Chen Liu, 2023).

X. Conclusion

In conclusion, this study shows that integrating real-time deepfake detection into a web platform for media sharing is feasible. Users are instantly notified about the legitimacy of videos when AI analysis is included into the upload process, which helps to cut down on false information.

XI. REFERANCES

[1] Abbas,F., Taeihagh,A.Unmaskingdeepfakes:Asystem atic reviewof deepfakedetectionandgeneration techniques usingartificial intelligence.ExpertSyst.Appl.124260(2024).

https://doi.org/10.1016/j.eswa.2024.124260

[2] Xia,R., Liu,D., Li, J.,Yuan, L.,Wang,N., Gao,X. Mmnet:multi-collaborationandmulti-supervisionnetworkfor sequentialdeepfakedetection. IEEETrans. Inf.Foren.Secur. (2024). https://www.science.org/doi/10.1126/sciadv.ads7159

[3] Rathoure,N.,Pateriya,R.K.,Bharot,N.,Verma,P.Com bat- ingdeepfakes:Acomprehensivemultilayerdeepfakevideo detec- tionframework.Multimed.ToolsAppl.1–18(2024).

[4] Kingra,S.,Aggarwal,N. Kaur,N.Emergenceofdeep fakes andvideo tamperingdetection approaches:Asurvey. Multimed. ToolsAppl. 82(7), 10165–10209 (2023). Article GoogleScholar https://dl.acm.org/doi/abs/10.1007/s11042-022-13100-x

[5] Khan, A. A. et al. Digital forensics for the socio cyber world (DF-SCW): Anovel framework for deepfake multimediainvestigationonsocialmediaplatforms.Egypt.Inf.J.27,100502(2024).GoogleScholar https://www.sciencedirect.com/science/article/pii/S1110866524000653

[6] Khan,A.A. et al. IMG-forensics:Multimedia-enabled information hiding investigation using convolutional neu- ral network. IETImageProc.16(11),2854–2862(2022).Article GoogleScholar

https://doi.org/10.1049/ipr2.12272

[7] Ahmed,S.R., Sonuc¸,E.Evaluatingtheeffectivenessof rationale-augmentedconvolutional neural networks fordeep fakedetection.SoftComput.1–12. (2023). https://www.researchgate.net/publication/374506947RETRACTEDARaugmentedconvolutionalneuralnetworksf ordeepfakedetection

[8] Qadir, A.,Mahum, R., El-Meligy,M. A., Ragab, A. E., AlSalman, A., Awais,M. An efficient deepfake video detec-tionusingrobustdeeplearning.Heliyon10(5). (2024). https://www.sciencedirect.com/science/article/pii/S2405844024017882

[9] Pang,G.,Zhang,B.,Teng,Z.,Qi,Z. Fan, J.MRE-Net: Multi-rate excitation network for deepfake video detection. IEEETrans.CircuitsSyst.VideoTechnol. 33(8), 3663–3676 (2023).ArticleGoogleScholar https://ieeexplore.ieee.org/document/10025759

[10] Kumar,M. Sharma,H.K.AGAN-basedmodel of deepfakedetection insocialmedia. Proc.Comput. Sci. 218, 2153–2162(2023).ArticleGoogleScholar https://www.researchgate.net/publication/367603800AGAN−BasedM odelofDeepfakeDetectioninSocialM edia

[11].Mcuba,M.,Singh,A., Ikuesan,R.A. Venter,H.The ef- fectofdeeplearningmethodsondeepfakeaudiodetection for digi- tal investigation. Proc. Comput. Sci. 219, 211–219 (2023).Ar- ticleGoogleScholar https://www.sciencedirect.com/science/article/pii/S1877050923002910

[12] Patel,Y.,Tanwar,S.,Gupta,R.,Bhattacharya,P.,David son, I.E.,Nyameko,R., Vimal,V.Deepfakegenerationand detection:Casestudyandchallenges. IEEEAccess. (2023). https://ieeexplore.ieee.org/document/10057390