IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 6, April 2025

Impact Factor: 7.67

Digital Forgery Detection

Prof. Rashmi Mahajan, Anirudh Jakkani, Norbert Parkhe

Shivajirao S. Jondhale College of Engineering Dombivali East, Mumbai, India

Abstract: *Digital forgery detection has emerged as a critical area of research in the age of ubiquitous digital media. With the proliferation of advanced editing tools and techniques, distinguishing authentic digital content from manipulated versions poses significant challenges. This study presents a comprehensive overview of state-of- the-art methodologies for detecting digital forgeries, encompassing image, video, and audio formats. We explore various detection techniques, including machine learning algorithms, statistical analysis, and digital watermarking, highlighting their effectiveness and limitations. Furthermore, we examine the role of deep learning in enhancing detection accuracy, alongside the integration of forensic tools that aid in real-time analysis. Our findings underscore the necessity for adaptive, robust detection systems that evolve with emerging technologies, ensuring the integrity of digital content in diverse applications. The paper concludes with recommendations for future research directions, emphasizing interdisciplinary collaboration and the development of standards for digital content verification.*

Keywords: Digital Forensics, Forgery Detection, Image Manipulation, Video Tampering, Audio Forgery, Machine Learning, Deep Learning, Statistical Analysis, Digital Watermarking, Content Authenticity, Media Integrity, Forensic Tools, Real-time Detection, Adaptive Systems, Digital Content Verification

I. INTRODUCTION

In an increasingly digital world, the authenticity of visual and auditory content has become a pressing concern. Digital forgery—where images, videos, or audio files are manipulated or altered—poses significant risks across various domains, including journalism, law enforcement, and social media. The ease with which digital content can be edited and shared has led to a surge in misinformation, identity theft, and other malicious activities, necessitating effective detection methods to identify and combat these forgeries.Digital forgery detection involves a range of techniques aimed at verifying the authenticity of digital media. These techniques can be broadly categorized into two main approaches: active and passive detection. Active detection methods rely on embedding information during the creation of digital content, such as digital watermarks, while passive methods analyze existing content for signs of manipulation without prior knowledge of its original state.Recent advancements in machine learning and computer vision have significantly enhanced the capabilities of forgery detection systems. By leveraging deep learning algorithms, researchers can now develop models that automatically identify subtle artifacts of manipulation that might be imperceptible to the human eye. These innovations have led to more robust and accurate detection systems, capable of adapting to the evolving landscape of digital forgery techniques.Despite these advancements, challenges remain. The continuous development of sophisticated editing tools and techniques complicates the detection process, requiring ongoing research and innovation. Moreover, the balance between detection efficacy and computational efficiency is crucial, particularly in real-time applications.As the implications of digital forgeries extend beyond individual cases to broader societal concerns, the importance of effective detection methods cannot be overstated. This introduction sets the stage for a deeper exploration of the methodologies, challenges, and future directions in the field of digital forgery detection, emphasizing the need for collaborative efforts to safeguard the integrity of digital media. [1]

II. LITERATURE REVIEW

Face Warping Artifacts [15] proposed a technique that identifies manipulation by examining discrepancies between generated facial areas and their neighboring regions using a specialized Convolutional Neural Network (CNN). This study highlights two main types of facial artifacts. The underlying concept relies on the fact that most deepfake

DOI: 10.48175/IJARSCT-25333

generation algorithms are limited to producing low-resolution images, which are later adjusted to align with the target faces in the source video. However, their approach does not incorporate temporal analysis across video frames.

Eye Blinking Detection [16] introduced a method to identify deepfakes based on eye blinking patterns, treating them as a critical factor in distinguishing authentic videos from manipulated ones. They employed a Long-term Recurrent Convolutional Network (LRCN) to analyze eye blinking in cropped video frames over time. However, with advancements in deepfake technology, eye blinking alone is no longer a reliable indicator. Other facial attributes—such as unnatural tooth visibility, distorted wrinkles, or incorrect eyebrow positioning—should also be included in detection strategies.

Capsule Networks for Forgery Detection [17] implemented capsule networks to detect fake or altered visuals across various conditions, such as replay attacks and computer-generated video content. During training, random noise was introduced, which may reduce the model's effectiveness on real-world data despite promising results on their internal dataset. Our approach aims to avoid this issue by training on clean, real-time datasets.

Recurrent Neural Network (RNN) for Deepfake Detection [18] used a sequential processing method for video frames by integrating RNNs with an ImageNet pre-trained model. Their study used the HOHO [19] dataset, which contains only 600 videos with limited diversity. Due to the dataset's small size and lack of variability, the model's performance may be inadequate for real-time applications. In contrast, our model will be trained on a more extensive and diverse real-time dataset.

Biological Signal-Based Synthetic Portrait Detection [20] focuses on extracting biological signals from facial areas in both original and manipulated portrait videos. By applying spatial and temporal transformations, the method captures signal patterns as feature vectors and PPG (photoplethysmography) maps. These features are then used to train a probabilistic SVM and CNN, and the average authenticity score determines whether a video is genuine or fake. **FakeCatcher**, built on this principle, achieves high detection accuracy regardless of the video's origin, resolution, or quality. However, it lacks a proper discriminator, making it challenging to define a differentiable loss function that aligns with the biological signal extraction process.

III. METHODOLOGY

The methodology for digital forgery detection involves a systematic approach to identify manipulated content. This process can be broken down into several key phases:

Data Collection

The first step involves gathering a diverse dataset that includes both authentic and forged digital content. This dataset should encompass various types of media (images, videos, audio) and a range of manipulation techniques to ensure comprehensive training and testing.

Preprocessing

Once the data is collected, preprocessing techniques are applied to enhance the quality of the input. This may include:

Normalization: Standardizing image sizes and pixel values to create a uniform input format.

Noise Reduction: Removing unwanted artifacts that could interfere with feature extraction.

Enhancement: Applying filters to improve clarity and detail in images or videos.

Feature Extraction

In this phase, relevant features that may indicate forgery are extracted from the preprocessed data. Common techniques include:

Statistical Analysis: Evaluating pixel intensity distributions, histograms, and color statistics.

Spatial Features: Analyzing the spatial relationships among pixels to identify inconsistencies

Frequency Domain Analysis: Utilizing methods like Discrete Fourier Transform (DFT) to capture frequency patterns that may reveal manipulations.

Classification

Extracted features are then classified using various algorithms. This can involve:

Machine Learning Approaches: Techniques such as Support Vector Machines (SVM) or Random Forests that require feature engineering and training on labeled data.

Deep Learning Models: Convolutional Neural Networks (CNNs) that automatically learn hierarchical features from raw data, typically providing higher accuracy.

Decision Making

Based on the classification results, a decision is made regarding the authenticity of the content. The system will label the content as:

Authentic: The content is verified as genuine.

Forged: The content is identified as manipulated.

Suspicious: The system indicates potential forgery, requiring further investigation.

Post-Processing

To enhance the accuracy of detection, post-processing techniques are applied. This may include:

Ensemble Learning: Combining predictions from multiple classifiers to improve overall reliability.

Threshold Adjustments: Fine-tuning decision thresholds to balance between false positives and false negatives.

Evaluation and Validation

The system's performance is assessed using various metrics:

Accuracy: The overall rate of correct classifications.

Precision and Recall: Measures of the system's ability to identify forged content accurately.

F1 Score: The harmonic mean of precision and recall, providing a single metric for performance evaluation.

User Interface and Reporting

Finally, the methodology includes developing a user-friendly interface for stakeholders to upload content and receive results. The system should present findings clearly, including labels, confidence scores, and any relevant details about the detected manipulations.

[9][12]

IV. SYSTEM ARCHITECTURE

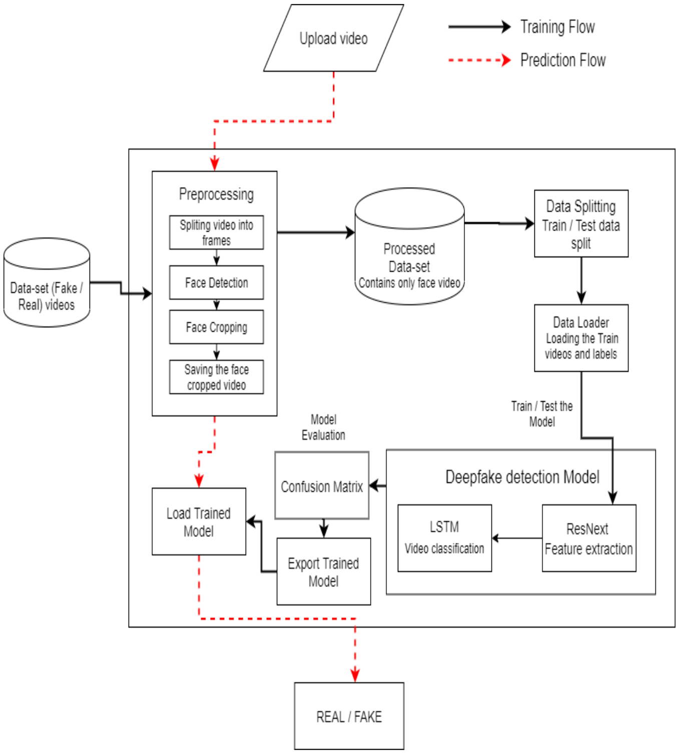


Fig 1: System Architecture

In our approach, we trained a deepfake detection model using PyTorch with a balanced dataset comprising an equal number of authentic and manipulated videos. This was done to eliminate any bias during model learning. The architecture of the model is illustrated in the accompanying diagram. During the development process, the initial dataset underwent preprocessing, resulting in a refined dataset that included only videos with cropped facial regions.

Generating Deepfake Videos

To accurately identify deepfake content, it is crucial to understand how such videos are produced. Most popular deepfake generation tools—utilizing technologies like GANs and autoencoders—require a source image and a target video. These tools extract individual frames from the video, detect the facial regions, and replace the face in each frame with the source face. Once the frame substitution is complete, the frames are reassembled using various pre-trained neural network models. These models also apply enhancements to improve visual quality and eliminate artifacts left behind by the face-swapping process. This results in highly realistic deepfakes that are difficult to distinguish from genuine videos using the naked eye.

We have applied this understanding to design our detection strategy. Despite their realism, deepfake generation methods often leave behind minor artifacts or inconsistencies—imperceptible to human observers. This research aims to identify and analyze these subtle anomalies and use them to differentiate between real and altered videos.

V. PROBLEM STATEMENT

The detection of digital forgeries has emerged as a major challenge in today's digital age, where the tampering of visual and audio content is becoming more widespread. The rapid development of digital manipulation tools has made it easier than ever to alter media, fueling the spread of misinformation. A key issue lies in the limited capability of current detection techniques, which often fall short in identifying advanced forgeries due to the ongoing innovation in manipulation methods.

Moreover, the variety of media types complicates the detection process, requiring adaptable solutions that can handle different formats. There is also an increasing demand for real-time detection—particularly across platforms like social media and news media—yet many existing methods are too resource-heavy for real-time deployment. The absence of standardized benchmarks for evaluating detection tools further hinders performance assessment and comparison. Lastly, the complexity of many detection systems makes them difficult for non-specialists to use. Addressing these limitations is essential for developing effective, user-friendly, and scalable solutions for digital forgery detection.

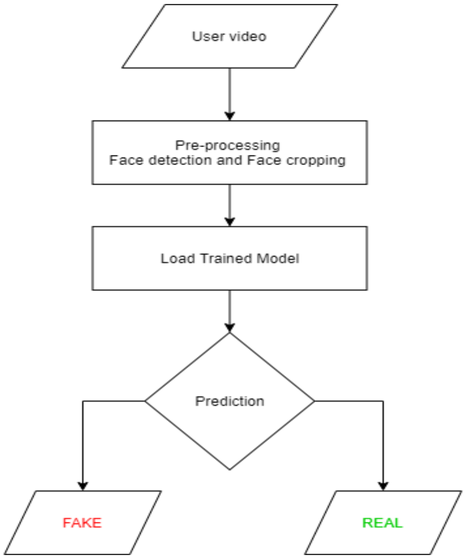
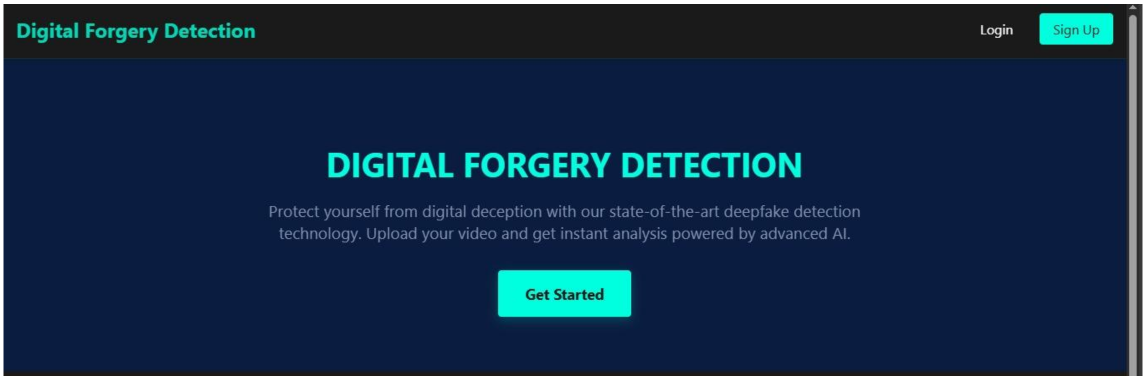


Fig 2: Testing workflow

DOI: 10.48175/IJARSCT-25333

VI. RESULTS



Protect yourself from digital deception with our state-of-the-art deepfake detection technology. Upload your video and get instant analysis powered by advanced AI.

Fig 3. Home page

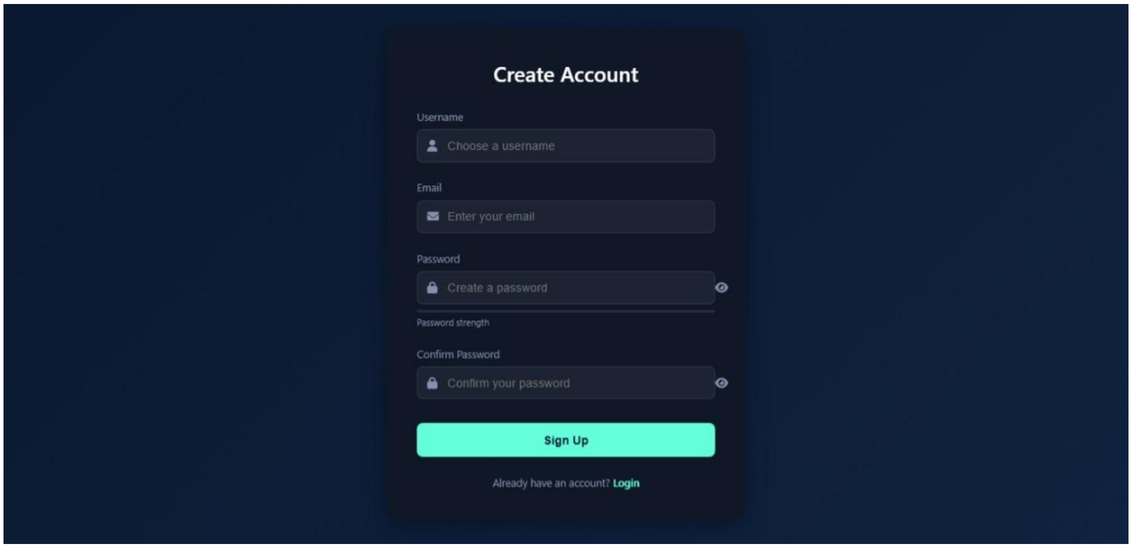


Fig 4. Sign up page

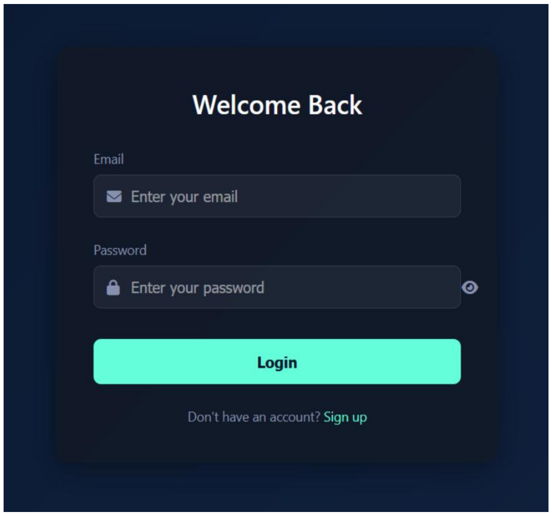
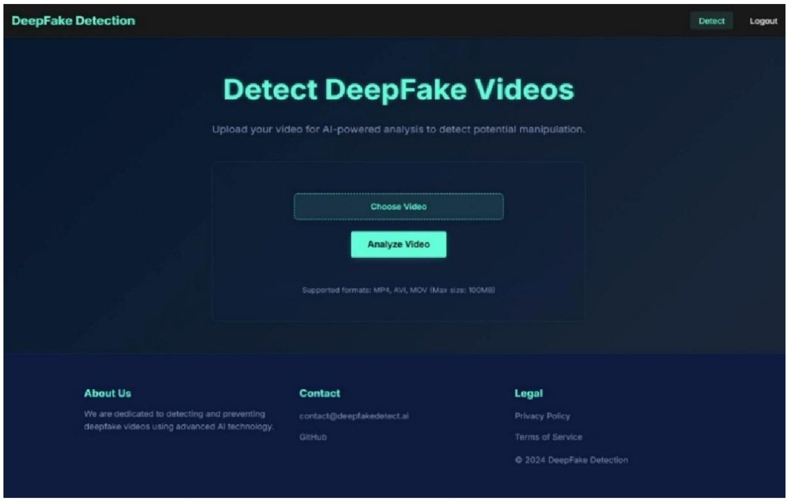


Fig 5. Login page



DeepFake Detection

Detect

Logout

Detect DeepFake Videos

Upload your video for AI-powered analysis to detect potential manipulation.

Choose Video

Analyze Video

Supported formats: MP4, AVI, MOV (Max size: 100MB)

About Us

We are dedicated to detecting and preventing deepfake videos using advanced AI technology.

Contact

contact@deepfakedetect.ai

GitHub

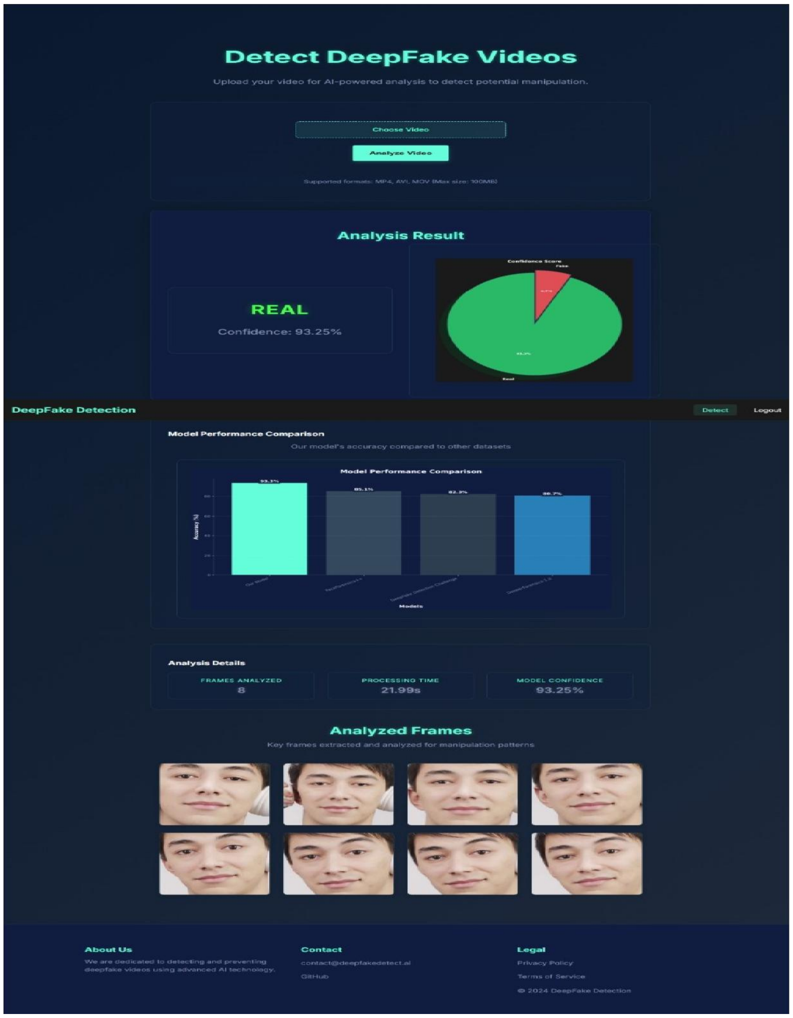
Legal

Privacy Policy

Terms of Service

© 2024 DeepFake Detection

Fig 6. Upload page



Detect DeepFake Videos

Upload your video for AI-powered analysis to detect potential manipulation.

Choose Video

Analyze Video

Supported formats: MP4, AVI, MOV (Max size: 100MB)

Analysis Result

REAL

Confidence: 93.25%

DeepFake Detection

Detect

Logout

Model Performance Comparison

Our model's accuracy compared to other datasets

Analysis Details

FRAMES ANALYZED	PROCESSING TIME	MODEL CONFIDENCE
8	21.99s	93.25%

Analyzed Frames

Key frames extracted and analyzed for manipulation patterns

About Us

We are dedicated to detecting and preventing deepfake videos using advanced AI technology.

Contact

contact@deepfakedetect.ai

GitHub

Legal

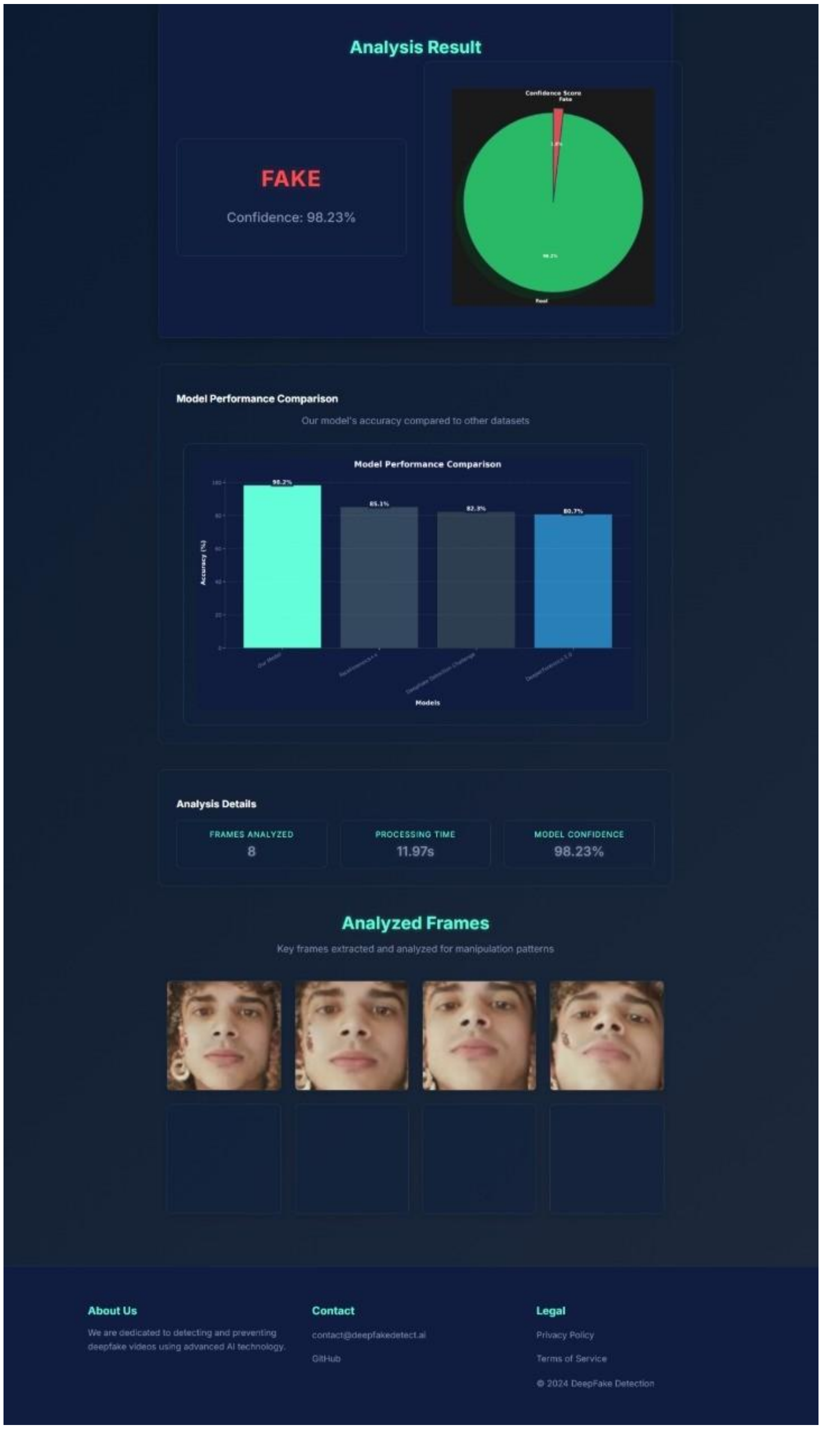
Privacy Policy

Terms of Service

© 2024 DeepFake Detection

Fig 7. Real Image result

DOI: 10.48175/IJARSCT-25333



Analysis Result

FAKE

Confidence: 98.23%

Model Performance Comparison

Our model's accuracy compared to other datasets

Analysis Details

FRAMES ANALYZED	PROCESSING TIME	MODEL CONFIDENCE
8	11.97s	98.23%

Analyzed Frames

Key frames extracted and analyzed for manipulation patterns

About Us

We are dedicated to detecting and preventing deepfake videos using advanced AI technology.

Contact

contact@deepfakedetect.ai

GitHub

Legal

Privacy Policy

Terms of Service

© 2024 DeepFake Detection

Fig 8. Fake Image Result

VII. CONCLUSION AND FUTURE SCOPE

Conclusion

We propose a neural-network framework that classifies videos as either genuine or deepfake and outputs a confidence score for each decision. By processing just one second of footage sampled at 10 fps, our approach delivers strong predictive performance. We use a ResNeXt CNN—pretrained to extract rich, frame-level features—and feed those into an LSTM to capture temporal dynamics and pinpoint differences between consecutive frames (t versus t–1). Moreover, our system is adaptable, handling input sequences of various lengths (10, 20, 40, 60, 80, or 100 frames) without loss of accuracy.

Future Scope

There is always a scope for enhancements in any developed system, especially when the project build using latest trending technology and has a good scope in future.

Web based platform can be upscaled to a browser plugin for ease of access to the user.

Currently only Face Deep Fakes are being detected by the algorithm, but the algorithm can be enhanced in detecting full body deep fakes.

REFERENCES

[1] Key resources include Andreas Rössler et al.'s FaceForensics++ study on detecting manipulated facial imagery (arXiv:1901.08971)

[2]; the Deepfake Detection Challenge dataset released in March 2020 [2]; Yuezun Li and colleagues' extensive Celeb DF deepfake forensics dataset (arXiv:1909.12962)

[3]The Fortune report detailing the viral Mark Zuckerberg deepfake ahead of the congressional AI hearing

[4]And an online compilation of ten particularly striking deepfake examples that both alarmed and amused internet audiences Accessed on 26 March, 2020

[5]TensorFlow: https://www.tensorflow.org/ (Accessed on 26 March, 2020)

[6]Keras: https://keras.io/ (Accessed on 26 March, 2020)

[7]PyTorch : https://pytorch.org/ (Accessed on 26 March, 2020) G. Antipov, M. Baccouche, and J.-L. Dugelay. Antipov, Baccouche, and Dugelay (February 2017) explored the problem of simulating facial aging by employing conditional generative adversarial networks (arXiv:1702.01983). In a separate effort, Thies et al. unveiled Face2Face at CVPR 2016, a system that captures live RGB video of a subject's face and instantly reenacts those expressions on another person's visage in real time (pp. 2387–2395)

[8] Las Vegas, NV.

Face app: https://www.faceapp.com/ (Accessed on 26 March, 2020)

Face Swap : https://faceswaponline.com/ (Accessed on 26 March, 2020)

[9]Deepfakes, Revenge Porn, And The Impact On Women : https://www.forbes.com/sites/chenxiwang/2019/11/01/deepfakes-revenge-porn-and- the-impact-on-women/

[10]The rise of the deepfake and the threat to democracy : https://www.theguardian.com/technology/ng-interactive/2019/jun/22/the-rise-of- the-deepfake-and-the-threat-to-democracy(Accessed on 26 March, 2020)

[11]Li, Chang, and Lyu (2018) developed a technique for uncovering AI□generated videos by analyzing eye□blinking patterns (arXiv:1806.02877v2). Nguyen, Yamagishi, and Echizen (2018) leveraged capsule networks to distinguish forged images and video manipulations (arXiv:1810.11215).

[12]Güera and Delp (2018) demonstrated deepfake detection via recurrent neural networks at the AVSS conference, showing how temporal sequence modeling can flag tampered footage. Earlier, Laptev et al. (2008) explored learning lifelike human actions from movie clips using large□scale video datasets (CVPR, pp. 1–8).

[13] More recently, Ciftci, Demir, and Yin (2019) harnessed subtle biological signals—encoded in photoplethysmography maps and spatial□temporal coherence—to spot synthetic portrait videos (arXiv:1901.02212v2).

[14] Kingma and Ba’s Adam optimizer (2014) remains a foundational method for efficiently training these and other deep networks.. arXiv:1412.6980, Dec. 2014.

[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/

[19]FakeApp (FakeApp 2020), DFaker (DFaker github 2019), faceswap-GAN (faceswapGAN github 2019),faceswap(faceswap github 2019), andDeepFaceLab(DeepFaceLab github 2020).

[20]Frank J, Eisen hofer T, Schönherr L, Fischer A, Kolossa D, Holz T (2020) Leveraging frequency analysis for deep fake image recognition.

[21]Fernando, Fookes, Denman, and Sridharan (2019) drew on human social cognition principles—embedding memory□network structures inspired by how people recall and recognize faces—to spot manipulated or counterfeit facial images at CVPR [28].

[22]In 2020, Carlini and Farid revealed that both white box and black box adversarial strategies can effectively fool state of the art deepfake image detectors [23]. More recent work has also shown that common up sampling (up convolution) layers in generative CNNs fail to faithfully reproduce the true spectral characteristics of real images, leading to detectable inconsistencies

[24].Fridrich, Soukal, and Lukas (2003) presented a method aimed at uncovering copy–move forgeries within digital photographs. Their approach was detailed in the proceedings of the Digital Forensic Research Workshop (DFRWS).

[25] In 2006, Bayram, Sencar, and Memon introduced a highly reliable and efficient strategy to identify copy–paste forgeries. Their technique was featured at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

[26] Popescu and Farid (2005) proposed a system for revealing digital image forgeries by locating repeated regions within an image. This technique was elaborated in a technical report (TR2004-515) from Dartmouth College's Computer Science department.

[27] Salloum, Ren, and Kuo (2018) developed a technique for identifying areas in spliced images by leveraging a multi-task fully convolutional neural network (MFCN), as documented in the Journal of Visual Communication and Image Representation.

[28] Cozzolino, Poggi, and Verdoliva (2015) introduced Splicebuster, a tool capable of detecting image splicing without prior knowledge of the original image. This work was presented at the IEEE Workshop on Information Forensics and Security (WIFS).

[29] Rössler et al. (2019) proposed FaceForensics++, a framework for detecting facial image tampering using deep learning methods. Their research was showcased at the IEEE Conference on Computer Vision (ICCV).

[30] Verdoliva (2020) provided a broad overview of advancements in media forensics, focusing specifically on deepfake detection techniques. This comprehensive study appeared in the IEEE Journal of Selected Topics in Signal Processing.