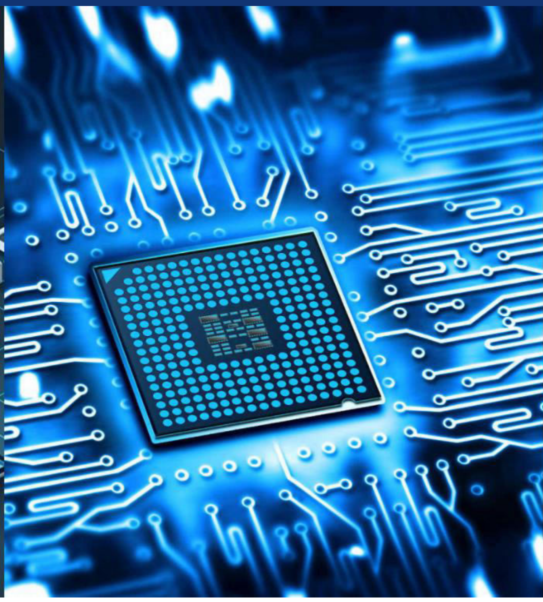
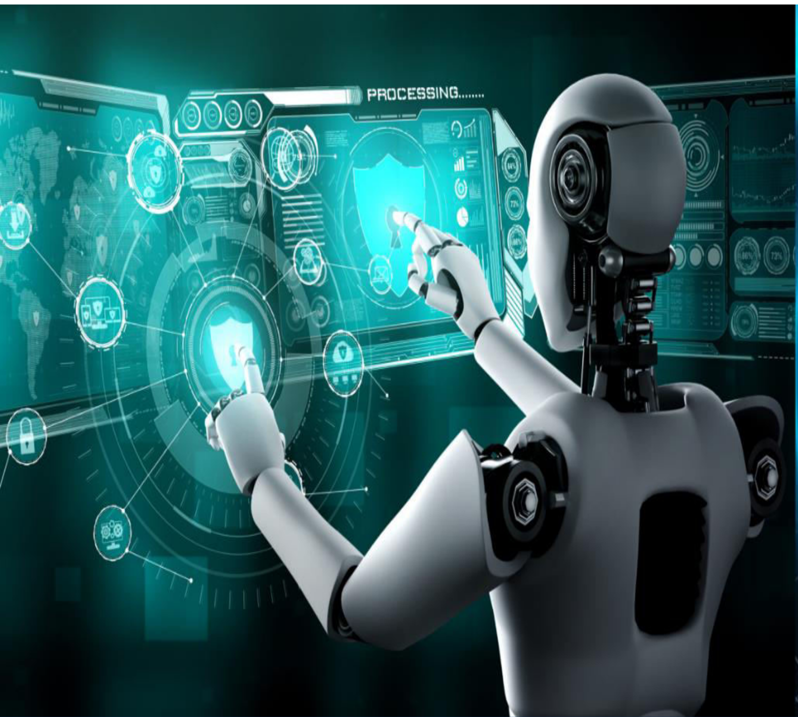
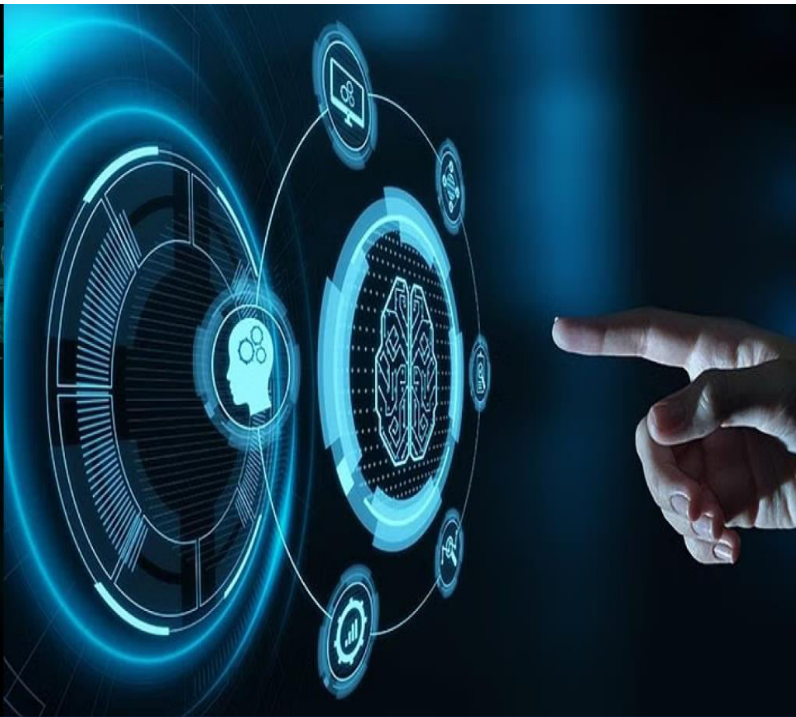
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927

Volume 14, Issue 8, August 2026

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408006

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Detection of Deepfake Video's Using Long-Distance Attention

Bhavani B E, Lakshmitha H Y

Dept. DoS in Computer Science, SBRR Mahajana First Grade College (Autonomous), PG Wing, Pooja Bhagavat Memorial Mahajana Education Centre, Mysore, Karnataka, India

ABSTRACT: Deepfake videos are videos that are created or manipulated using Artificial Intelligence (AI) and deep learning techniques. As these technologies continue to develop, it has become increasingly difficult to distinguish fake videos from real ones. The growing use of deepfake technology has led to serious concerns such as misinformation, identity theft, digital fraud, and cybersecurity threats. Therefore, developing an accurate and reliable deepfake detection system has become an important research area.

This research proposes a deepfake video detection system that combines YOLOv8 with a Long-Distance Attention model. The uploaded video is first divided into individual frames and pre-processed before being passed to the YOLOv8 model for accurate face detection and extraction. The detected facial regions are analyzed to identify important features such as facial texture, eye and lip movements, and manipulation artifacts. These features are then processed by the Long-Distance Attention model, which analyzes both spatial features and long-range temporal relationships across multiple frames. Finally, the extracted features are classified using a Fully Connected Layer and a Softmax classifier to determine whether the uploaded video is Real or Deepfake.

The proposed system improves the accuracy and reliability of deepfake detection by combining accurate face detection with temporal feature analysis. This approach helps enhance the authenticity, security, and trustworthiness of digital media.

KEYWORDS: Deepfake Video Detection, YOLOv8, Long-Distance Attention, Deep Learning, Face Detection, Video Forensics, Temporal Feature Analysis, Artificial Intelligence.

I. INTRODUCTION

Deepfake videos are videos that have been altered or created using **Artificial Intelligence (AI)** and deep learning techniques. As AI technology continues to improve, these videos are becoming more realistic, making it difficult to tell the difference between fake and real videos. The increasing use of deepfake technology has created serious problems such as the spread of false information, identity theft, cybercrime, and threats to digital media security. Because of these challenges, developing an accurate and reliable deepfake detection system has become an important area of research.

In this research, a deepfake video detection system is developed by combining **YOLOv8** with a **Long-Distance Attention** model. The **YOLOv8** model is used to detect and extract facial regions from video frames, while the **Long-Distance Attention** model analyzes both facial features and the relationships between multiple frames to identify signs of manipulation that may not be visible in a single frame. The extracted features are then passed through a **Fully Connected Layer** and classified using a **Softmax classifier** to determine whether the uploaded video is **Real** or **Deepfake**.

The proposed system aims to improve the accuracy and reliability of deepfake detection by combining precise face detection with effective temporal feature analysis. This approach helps provide a practical solution for improving the authenticity, security, and trustworthiness of digital media.

The main objectives of this research are:

- To develop an efficient and reliable system for detecting deepfake videos.
- To detect and accurately locate facial regions from video frames using the YOLOv8 model.
- To extract and analyze facial features for identifying signs of manipulation.
- To capture both spatial and temporal relationships between video frames using the Long-Distance Attention model.
- To classify videos as Real or Deepfake using a Fully Connected Layer and Softmax classifier.
- To improve the overall accuracy and robustness of deepfake video detection for enhancing digital media security.

II. RELATED WORK

[1] "Deepfake Video Detection Using CNN Features"

Yadav and Sharma proposed a CNN-based approach for detecting deepfake videos by extracting facial features such as texture irregularities and visual artifacts from individual video frames. Their model achieved an accuracy of **92.8%** on the **Face Forensics++** dataset. Although the method produced good detection results, it analyzed each frame separately and was unable to capture the temporal relationships between consecutive video frames.

[2] "Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers"

Thing compared the performance of Convolutional Neural Networks (CNNs) and Transformer-based models for deepfake detection. The study found that Transformer models were more effective at learning global features and achieved better generalization than CNN-based methods. However, the improved performance came with higher computational cost and increased resource requirements.

[3] "A Timely Survey on Vision Transformer for Deepfake Detection"

Wang et al. presented a detailed survey of Vision Transformer (ViT)-based approaches for deepfake detection. The survey explained that Transformer architectures are capable of learning long-range spatial dependencies, allowing them to identify complex manipulation patterns more effectively than traditional CNN-based methods. The study also highlighted the growing importance of Transformer models in improving deepfake detection accuracy.

[4] "Self-supervised Transformer for Deepfake Detection"

Zhao et al. introduced a self-supervised Transformer framework that learns meaningful feature representations without relying on a large amount of labelled training data. Their approach improved the overall detection performance and demonstrated better generalization across different deepfake datasets, making it more suitable for identifying unseen manipulations

.

[5] Summary of Identified Research Gaps

The literature review indicates that existing deepfake detection methods have shown promising performance, but several challenges still remain. Most CNN-based methods focus mainly on spatial information extracted from individual frames and do not effectively capture long-range temporal relationships within videos. Although Transformer-based approaches improve temporal analysis, they generally require high computational power and large training datasets. In addition, some existing methods do not provide accurate face localization, which can reduce the overall detection performance. Many models also struggle to identify newly generated and highly realistic deepfake videos. To address these limitations, the proposed system combines **YOLOv8** for accurate face detection with a **Long-Distance Attention** model for analyzing long-range temporal relationships. This combination improves both the accuracy and reliability of deepfake video detection.

III. MATERIALS AND METHODS

[1] Deepfake Video Dataset (Kaggle)

The proposed deepfake detection system was evaluated using the **Deep Fake Detection (DFD) Entire Original Dataset** available on the Kaggle platform. The dataset consists of **3,431 videos**, including **363 original (real) videos** and **3,068 manipulated (deepfake) videos**. It contains videos with different facial expressions, head poses, lighting conditions, and background scenes, making it suitable for testing the system under realistic conditions. This diverse dataset provides a reliable benchmark for evaluating the performance and effectiveness of the proposed deepfake detection model.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408006

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[2] Input Module
The deepfake detection process starts when the user uploads a video to the system. The uploaded video serves as the input, and its visual content is prepared for analysis to determine whether it is Real or Deepfake.

[3] Frame Extraction Module
The uploaded video is converted into a sequence of individual frames. Extracting frames allows the system to examine facial movements and visual changes over time. These frames are then used as input for the next processing stage.

[4] Preprocessing Module
The extracted frames are pre-processed to improve their quality and maintain a consistent input format. Each frame is resized to **640 × 640 pixels**, normalized using **Min-Max normalization**, and denoised to remove unwanted noise. After preprocessing, the frames are prepared for accurate face detection and feature extraction.

[5] Face Detection Using YOLOv8
The pre-processed frames are passed to the **pre-trained YOLOv8** model developed by **Ultralytics** for accurate face detection and localization. The model is fine-tuned using the deepfake dataset with an **image size of 640 × 640 pixels**, **batch size of 16**, and **640 training epochs** to improve detection performance. YOLOv8 accurately detects and crops facial regions while removing unnecessary background information. The extracted face regions contain important spatial features such as **facial texture, eye movements, lip details, skin appearance, and facial expressions**, which are forwarded to the Long-Distance Attention model for further analysis.

[6] Long-Distance Attention Model
The cropped face sequences obtained from the **pre-trained YOLOv8** model are processed using the **Long-Distance Attention** model. It analyzes multiple video frames together and learns long-range temporal relationships by focusing on important facial changes across the video. The model extracts both **spatial** and **temporal** features and identifies manipulation artifacts such as **unnatural facial movements, inconsistent expressions, irregular blinking, and lip-sync mismatches**. By combining these features, the model improves the overall accuracy and reliability of deepfake video detection.

[7] Softmax Classification Module
The spatial and temporal features extracted by the Long-Distance Attention model are passed to a Fully Connected Neural Network, followed by a Softmax classifier. The Softmax classifier calculates the probability of each class and predicts whether the uploaded video is Real or Deepfake. The class with the highest probability is selected as the final prediction, providing an accurate and reliable classification of the input video.

[8] Output Module
The final prediction is displayed after the analysis is completed. Based on the classification result, the system identifies the uploaded video as either **Real**, indicating that no manipulation is detected, or **Deepfake**, indicating that the video contains AI-generated or manipulated facial content.

IV. SYSTEM ARCHITECTURE

The proposed system architecture illustrates the overall workflow of the deepfake video detection process. It begins with an input video, which is converted into individual frames for analysis. The extracted frames are then pre-processed and passed through the YOLOv8 model for accurate face detection. The detected facial regions are further analyzed using the Long-Distance Attention model to capture both spatial and temporal features. Finally, the extracted features are classified using a Softmax classifier to determine whether the uploaded video is Real or Deepfake.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408006
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

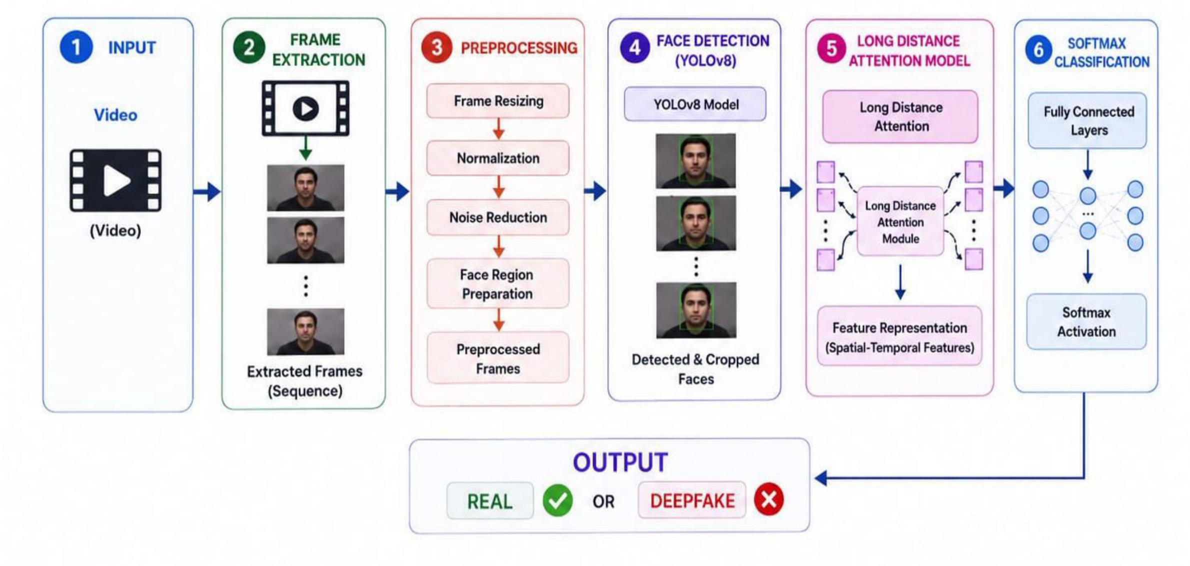


Figure 1: System Architecture Diagram for Deepfake Video Detection System

The proposed deepfake video detection system follows a structured process to analyze the uploaded video. The process starts with the video input, which is then divided into individual frames and pre-processed to improve their quality for further analysis. These pre-processed frames are passed to the YOLOv8 model for accurate face detection, while the Long-Distance Attention model extracts both spatial and temporal features from the detected facial regions. Finally, the extracted features are classified using a Softmax classifier, which determines whether the uploaded video is **Real** or **Deepfake**.

V. PSEUDO CODE

Input:
Upload a video
output
video divided into multiple frame
analyze frame
final result prediction
confidence score
recent history

1) Frame extraction
 # Import libraries and define constants
IMPORT OpenCV
DEFINE FRAMES_PER_VIDEO = 10
 # Load video dataset
LOAD input video
 # Extract frames
OPEN video file
CALCULATE total frames
COMPUTE frame interval = total frames / FRAMES_PER_VIDEO
FOR i = 1 TO FRAMES_PER_VIDEO
    READ frame at interval × i
    SAVE extracted frame
END FOR
CLOSE video
END

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2) Pre processing

```
# Import libraries and define constants
IMPORT OpenCV, NumPy
DEFINE IMAGE_SIZE = 640 × 640
# Load extracted frames
LOAD extracted frames
# Preprocess frames
FOR each extracted frame
   RESIZE frame to IMAGE_SIZE
   CONVERT frame to RGB
   NORMALIZE pixel values
   REMOVE noise (optional)
   SAVE preprocessed frame
END FOR
END
```

3) YOLOv8

```
# Import libraries and define constants
IMPORT Ultralytics YOLOv8, OpenCV, NumPy
DEFINE MODEL = "yolov8n.pt"
DEFINE IMAGE_SIZE = 640
DEFINE EPOCHS = 640
# Load dataset
LOAD training, validation, and testing images
LOAD corresponding annotation files
# Load and train YOLOv8 model
model = YOLO(MODEL)
model.train(
   data = "data.yaml",
   epochs = EPOCHS,
   imgsz = IMAGE_SIZE
)
# Perform object detection
FOR each test image
   results = model.predict(image)
   DISPLAY bounding boxes, class labels, and confidence scores
# Evaluate model
CALCULATE Precision, Recall, and mAP
SAVE trained model and detection results
END
```

4) Long-Distance Attention

```
# Import libraries and define constants
IMPORT TensorFlow, NumPy, OpenCV
DEFINE ATTENTION_THRESHOLD = 0.5
# Load face feature
LOAD face feature sequence
# Apply Long Distance Attention
FOR each feature vector
   COMPUTE attention scores
   GENERATE attention-weighted features
END FOR
# Detect face regions
FOR each detected face
   IF confidence > THRESHOLD
```

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
        CROP face
    END IF
END FOR
```

 # Predict result

```
CLASSIFY as Real or Deepfake
SAVE prediction results
```

END

5) SOFTMAX

 # Import libraries and define constants

```
IMPORT TensorFlow, NumPy
DEFINE INPUT_DIM = 512
DEFINE NUM_CLASSES = 3
```

 # Load features

```
LOAD feature vectors
LOAD class labels
```

 # Build classifier

```
INITIALIZE Softmax classifier
```

 # Train model

```
FOR each epoch
    COMPUTE Softmax probabilities
    CALCULATE Cross-Entropy Loss
    UPDATE model weights
END FOR
```

 # Predict and evaluate

```
PREDICT class labels
CALCULATE Accuracy
SAVE trained model
```

END

VI. RESULT AND DISCUSSION

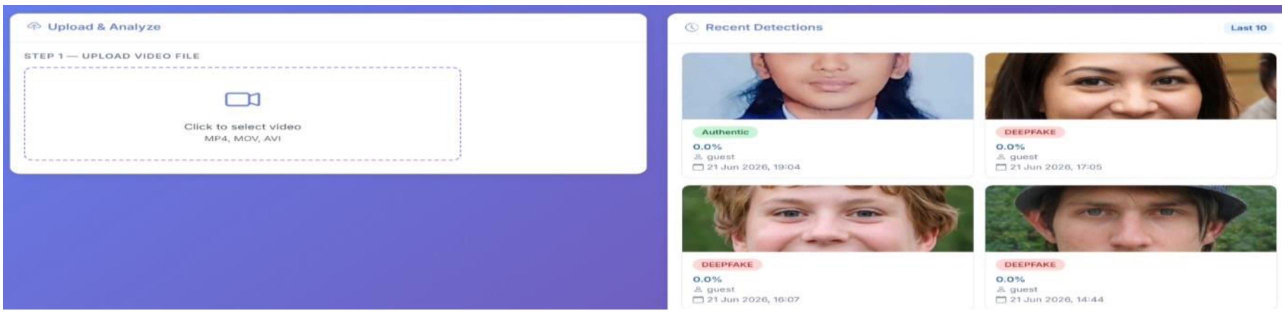


Figure 2: Deepfake Detection System – Video Upload and Recent Detection Interface

The figure shows the main interface of the proposed deepfake detection system. It allows users to upload videos in supported formats for analysis and also displays the details of recently analyzed videos. After the analysis is completed, the system presents the predicted result, confidence score, and other detection information. This simple and user-friendly interface enables users to upload videos easily and review previous detection results whenever required.

© 2026 IJIRCCE | Volume 14, Issue 8, August 2026 DOI: 10.15680/IJIRCCE.2026.1408006

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

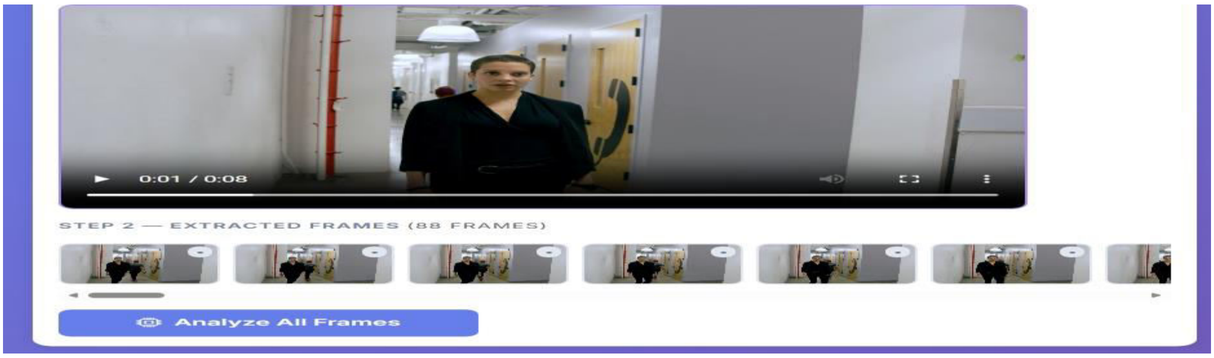


Figure 3: Video Playback and Extracted Frame Visualization for Deepfake Detection Analysis

The figure illustrates the video analysis stage of the proposed deepfake detection system. Once a video is uploaded, it is displayed in the playback window, and the system extracts individual frames for further processing. These frames are analyzed to detect facial features and identify temporal inconsistencies that may indicate deepfake manipulation. In this example, a total of **88 frames** are extracted during preprocessing and processed for face detection, feature extraction, and final classification. This frame-by-frame analysis helps the system accurately determine whether the uploaded video is **Real** or **Deepfake**.

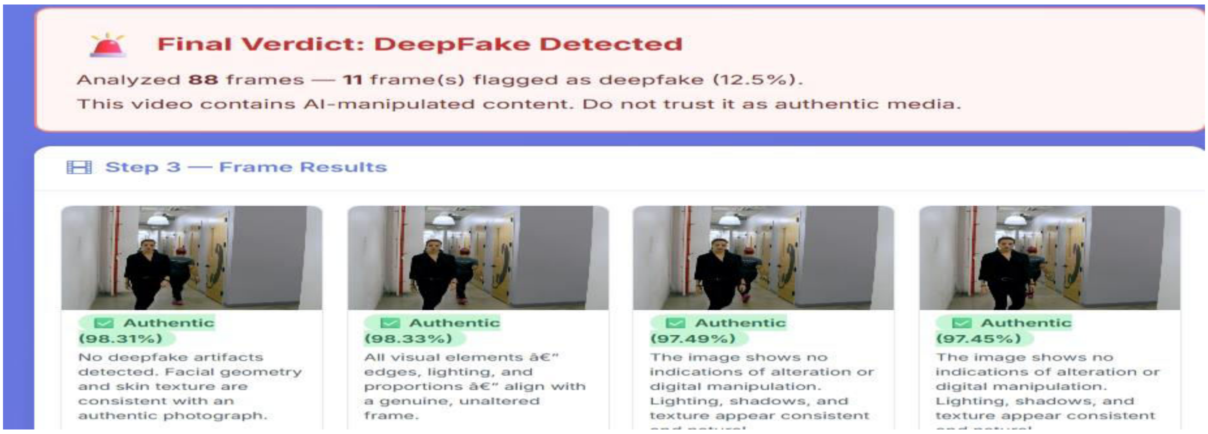


Figure 4: Final Deepfake Detection Result and Frame-wise Classification

The figure presents the final output generated by the proposed deepfake detection system after analyzing all extracted video frames. In this example, the system processed **88 frames**, and **11 frames (12.5%)** were identified as containing possible deepfake characteristics. Based on the overall frame analysis, the final result was displayed as **"Deepfake Detected."** A warning message is also shown to inform the user that the uploaded video may contain AI-generated manipulation. The frame results section displays the classification and confidence score for each frame, allowing users to understand how the individual frame predictions contribute to the final decision.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Figure 5 Frame-wise Deepfake Detection Results with Confidence Scores

The figure shows the classification results for each extracted frame along with the corresponding confidence scores. Every frame is analyzed individually and classified as either Authentic or Deepfake. The system identifies manipulated frames by examining facial textures, lighting variations, facial expressions, and other visual inconsistencies. Displaying the prediction for every frame improves the transparency of the detection process and helps increase the reliability of the final video-level classification.

VII. CONCLUSION AND FUTURE WORK

[1] Conclusion

This research presents a deepfake video detection system that combines YOLOv8 and a Long-Distance Attention model to identify manipulated videos. The pre-trained YOLOv8 model accurately detects and extracts facial regions, while the Long-Distance Attention model analyzes both spatial and temporal features to identify subtle manipulation artifacts. The extracted features are classified using a Fully Connected Layer and a Softmax classifier to determine whether the uploaded video is Real or Deepfake. The proposed system improves the accuracy and reliability of deepfake detection and can support applications such as digital forensics, cybersecurity, and social media content verification.

[2] Future Work

In the future, the system can be improved by training it on larger and more diverse datasets to detect advanced deepfake techniques more effectively. Additional improvements include developing a real-time detection system, optimizing the model for mobile and edge devices, and integrating it with web applications, social media platforms, and digital forensic tools for faster and more reliable deepfake detection.

REFERENCES

[1] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, **"The DeepFake Detection Challenge (DFDC) Dataset,"** arXiv preprint arXiv:2006.07397, 2020.
[2] Vrizlynn L. L. Thing, **"Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers,"** arXiv preprint arXiv:2304.03698, 2023.
[3] Zhikan Wang, Zhongyao Cheng, Jiajie Xiong, Xun Xu, Tianrui Li, Bharadwaj Veeravalli, and Xulei Yang, **"A Timely Survey on Vision Transformer for Deepfake Detection,"** arXiv preprint arXiv:2405.08463, 2024.
[4] Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Weiming Zhang, and Nenghai Yu, **"Self-supervised Transformer for Deepfake Detection,"** arXiv preprint arXiv:2203.01265, 2022.
[5] Pranav Sathe, Varad Sabane, Chaitanya Undale, Atharva Uttarkar, Vaibhav Chavhan, and Nilesh P. Sable, **"Deepfake Image Detection Using YOLOv8,"** Proceedings of the IEEE Pune Section International Conference (PuneCon), 2024.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[6] Glenn Jocher, Ayush Chaurasia, and Jing Qiu, **YOLOv8 Documentation and Ultralytics Framework**, Ultralytics, 2023.
[7] Alexey Dosovitskiy et al., **"An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale,"** International Conference on Learning Representations (ICLR), 2021.
[8] FaceForensics++, **"A Large-Scale Video Dataset for Forgery Detection and Benchmarking,"** 2019.
[9] DeepFake Detection Challenge (DFDC), Kaggle Dataset, 2020.
[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, **Deep Learning**, MIT Press, 2016.
[11] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner, **"FaceForensics++: Learning to Detect Manipulated Facial Images,"** Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
[12] David Güera and Edward J. Delp, **"Deepfake Video Detection Using Recurrent Neural Networks,"** 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2018.
[13] Yuezun Li and Siwei Lyu, **"Exposing DeepFake Videos by Detecting Face Warping Artifacts,"** Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.
[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, et al., **"An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale,"** International Conference on Learning Representations (ICLR), 2021.
[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, **"Attention Is All You Need,"** Advances in Neural Information Processing Systems (NeurIPS), 2017.
[16] YoloV8 Pretrained model-Pretrained Weights (GitHub): https://github.com/ultralytics/assets/releases
[17] Y. Li, M.-C. Chang, and S. Lyu, **"In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking,"** in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), 2018.
[18] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, **"Multi-task Learning for Detecting and Segmenting Manipulated Facial Images and Videos,"** in Proc. IEEE Int. Conf. Biometrics: Theory, Applications and Systems (BTAS), 2019.
[19] T. Karras, S. Laine, and T. Aila, **"A Style-Based Generator Architecture for Generative Adversarial Networks,"** in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2019**.**
[20] H. H. Nguyen, J. Yamagishi, and I. Echizen, **"Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos,"** in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2019.
[21] H. H. Nguyen, J. Yamagishi, and I. Echizen, **"Deep Learning for Deepfakes Creation and Detection: A Survey,"** Computer Vision and Image Understanding, vol. 223, Art. no. 103525, 2022.
[22] Y. Zhou and S.-N. Lim, **"Joint Audio-Visual Deepfake Detection,"** in Proc. IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
[23] L. Verdoliva, **"Media Forensics and DeepFakes: An Overview,"** IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, Aug. 2020.
[24] B. Bayar and M. C. Stamm, **"A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer,"** in Proc. ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec), 2016.
[25] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," Information Fusion, vol. 64, pp. 131–148, Dec. 2020.