IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

HYBRID RESNET CNN-LSTM FOR DEEPFAKE VIDEO DETECTION

[1]Dr. Suresh M B, [2]Dinesh N, [3]Abhishek T N
[1]Professor, [2]B.E Student, [3]B.E Student
[1]Department Of Information Science and Technology,
[1]East West Institute of Technology, Bangalore, India

Abstract:

The proliferation of deepfake technology poses significant challenges to the integrity of digital content, necessitating robust detection mechanisms. In this study, we propose a hybrid approach that integrates ResNet, a convolutional neural network (CNN) architecture known for its feature extraction capabilities, with Long Short-Term Memory (LSTM) networks, specialized in capturing temporal dependencies. Our method aims to enhance the accuracy and effectiveness of deepfake detection by combining spatial and temporal information within a unified framework. We curate a diverse dataset containing authentic and deepfake videos, preprocess the data, and train the hybrid model using deep learning techniques. Evaluation on a separate test dataset demonstrates the superior performance of our approach, achieving high accuracy and precision in distinguishing between authentic and deepfake videos. Comparative analysis with baseline methods further validates the effectiveness of the proposed approach. Additionally, ethical considerations are carefully addressed throughout the research process, ensuring responsible development and deployment of the deepfake detection system. Through this study, we contribute to the advancement of techniques for combating deceptive visual media and preserving trust in digital content.

I. INTRODUCTION

Deepfake technology, an emerging form of synthetic media, has raised significant concerns regarding its potential misuse in spreading misinformation and manipulating visual content. These highly realistic but falsified videos can convincingly depict individuals saying or doing things they never did. Consequently, there is a pressing need for robust methods to detect and mitigate the proliferation of deepfake content. In this context, the proposed research introduces a novel approach that integrates the strengths of ResNet, a powerful convolutional neural network (CNN) architecture known for its feature extraction capabilities, with the temporal modeling capabilities of Long Short-Term Memory (LSTM) networks. By combining these techniques within a hybrid framework, our aim is to enhance the accuracy and effectiveness of deepfake detection. This study outlines the methodology, including dataset preprocessing, model training, and evaluation metrics, while also considering ethical implications associated with the development and deployment of deepfake detection systems. Through this research, we aim to contribute to the advancement of techniques for combating deceptive visual media and safeguarding the integrity of digital content. The rise of deepfake technology, driven by advancements in artificial intelligence and machine learning, has ushered in a new era of digital manipulation. Deepfake videos, which seamlessly superimpose one person's face onto another's body, have become increasingly sophisticated, blurring the lines between reality and fiction. With the potential to deceive viewers and manipulate public perception, deepfakes pose serious challenges to various sectors, including journalism, politics, and entertainment. In addition to technical considerations, this research also addresses ethical concerns surrounding the development and deployment of deepfake detection systems. Ethical considerations include issues related to privacy, consent, and potential misuse of the technology. By taking a holistic approach that considers both technical and ethical dimensions, we seek to develop responsible solutions that mitigate the harmful effects of deepfake technology while upholding ethical standards and individual rights. Through our research, we aim to contribute to the ongoing efforts to combat deceptive visual media and promote trust and authenticity in digital content.

II. BACKGROUND WORK :

2.1 RESNET:

ResNet (Residual Network) is utilized as a key component for feature extraction from the video frames. ResNet is a deep convolutional neural network architecture that is known for its ability to effectively capture and represent hierarchical features from images. Its unique architecture includes residual connections, or "skip connections," which allow the network to bypass certain layers, thus mitigating the vanishing gradient problem and enabling the training of very deep networks. Specifically in this project, pre-trained ResNet models are employed to extract spatial features from individual frames of the video sequences. The ResNet architectures are fine-tuned as needed to adapt them to the deepfake detection task and enhance their feature extraction capabilities. By leveraging the powerful feature representation learned by ResNet, the model can effectively capture intricate details and patterns present in the video frames, aiding in the discrimination between authentic and deepfake content.

2.2 LSTM

Long Short-Term Memory (LSTM) networks are employed for capturing temporal dependencies within the video sequences. LSTM is a type of recurrent neural network (RNN) architecture that is well-suited for modeling sequential data due to its ability to retain long-term dependencies and handle vanishing gradient issues. LSTM networks consist of recurrent units with a more complex structure than traditional RNNs. They incorporate three gates – input gate, forget gate, and output gate – which regulate the flow of information within the network, allowing it to selectively update and forget information over time. Specifically, in this project, LSTM networks are utilized to analyze the temporal dynamics present in the sequence of extracted features from the video frames. By capturing temporal dependencies, LSTM helps the model understand the context and temporal relationships between frames, which is crucial for distinguishing between authentic and deepfake videos. The LSTM component is integrated into the hybrid architecture alongside the spatial features extracted by ResNet. Through this fusion of spatial and temporal information, the model gains a comprehensive understanding of both the spatial and temporal aspects of the video sequences, enhancing its ability to detect subtle manipulations characteristic of deepfake content. LSTM plays a crucial role in capturing temporal information and contextual cues within the video sequences, contributing to the effectiveness of the hybrid ResNet CNN-LSTM approach for deepfake detection in the project.

III. LITERATURE SURVEY

The literature survey for the project on deepfake detection using a hybrid ResNet CNN-LSTM approach delves into various research domains. Firstly, it explores existing techniques for deepfake detection, ranging from traditional methods to more recent advancements in deep learning-based approaches. This includes examining studies that evaluate the efficacy of different features and classifiers in distinguishing between authentic and manipulated videos. Additionally, the survey delves into the realm of convolutional neural networks (CNNs), particularly ResNet models, focusing on their applications in computer vision tasks and feature extraction from visual data. Furthermore, it reviews the literature on recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks, highlighting their capabilities in modeling sequential data and capturing temporal dependencies in video sequences. Moreover, the survey explores hybrid architectures that integrate CNNs and RNNs for spatiotemporal data analysis, emphasizing how these architectures leverage both spatial and temporal information for improved performance. Ethical considerations surrounding deepfake technology, benchmark datasets for evaluation, and state-of-the-art approaches in deepfake detection are also examined to provide a comprehensive understanding of the research landscape. Through this literature survey, insights are gained to inform the development of the hybrid ResNet CNN-LSTM approach for deepfake detection, contributing to advancements in combating deceptive visual media.

IV. PROPOSED SYSTEM

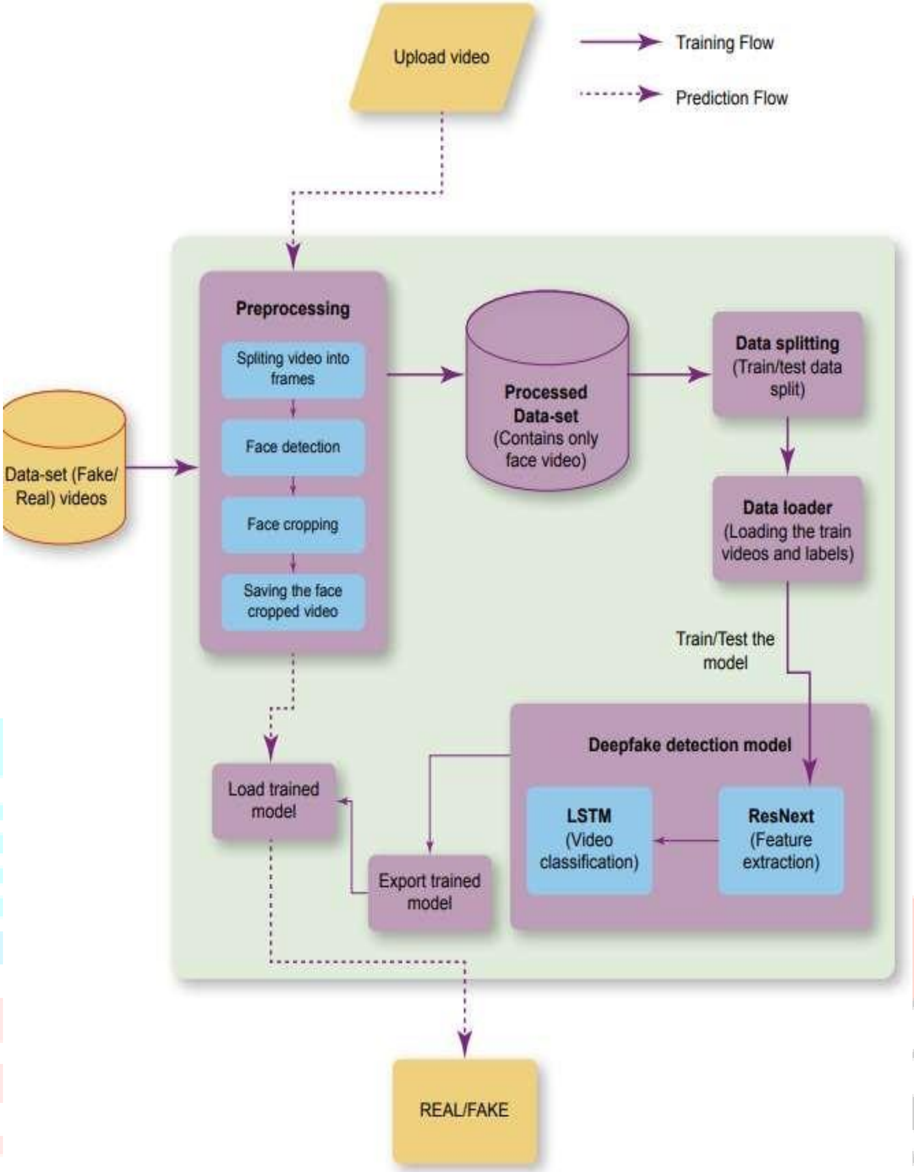

Upload video
Training Flow
Prediction Flow
Preprocessing
Spliting video into frames
Face detection
Face cropping
Saving the face cropped video
Data-set (Fake/ Real) videos
Processed Data-set (Contains only face video)
Data splitting (Train/test data split)
Data loader (Loading the train videos and labels)
Train/Test the model
Deepfake detection model
LSTM (Video classification)
ResNext (Feature extraction)
Load trained model
Export trained model
REAL/FAKE

Fig1: Architecture diagram of the Proposed System

The proposed system aims to develop a hybrid ResNet CNN-LSTM approach for deepfake video detection. This system integrates the strengths of ResNet, a powerful convolutional neural network (CNN) architecture known for its feature extraction capabilities, with Long Short-Term Memory (LSTM) networks, specialized in capturing temporal dependencies within sequential data. The system follows a multi-step process, beginning with dataset acquisition and preprocessing to ensure standardized input. Next, spatial features are extracted from individual frames of video sequences using pre-trained ResNet models. Concurrently, LSTM networks analyze temporal dependencies across the video frames. These spatial and temporal features are then fused within a hybrid architecture, enabling comprehensive analysis of both spatial and temporal aspects of the videos. The model is trained on the curated dataset using deep learning techniques, optimizing parameters to enhance performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are employed to assess the model's effectiveness in distinguishing between authentic and deepfake videos.

4.1 DATASET

There are several datasets that have been used in deepfake video detection research. Here are some of the most commonly used datasets: Face Forensics ++: This is one of the largest and most widely used deepfake video datasets. It contains over 1,000 real videos and over 1,000 deepfake videos generated using several different methods, including Deep Fake, Face2Face, and Neural Textures. Celeb-DF: This is another popular dataset for deepfake video detection. It contains over 890 real videos and over 5,639 deepfake videos generated using the Deep Fake method. Deep Fake Detection Challenge (DFDC) dataset: This is a dataset created by Facebook for a competition aimed at developing better deepfake detection methods [11]. It contains over 100,000 videos, including both real and deepfake videos generated using several different methods. DeeperForensics-1.0 This is a relatively new dataset that contains over 5,000 videos, including real videos and deepfake videos generated using several different methods. Self-created dataset: This is a dataset created by our own in order improve training and prediction accuracy and to pre detect fake video in real time scenarios, and to make the system work better. These datasets typically contain labeled videos, where each video is labeled as either real or fake. They are often used for training and evaluating deepfake video detection models. However, there are also some challenges associated with using these datasets, such as the potential for bias in the labeling process and the lack of diversity in the types of deepfake videos included [12]. Researchers must be careful to consider these limitations when using these datasets for deepfake video detection.

V. EXPERIMENT

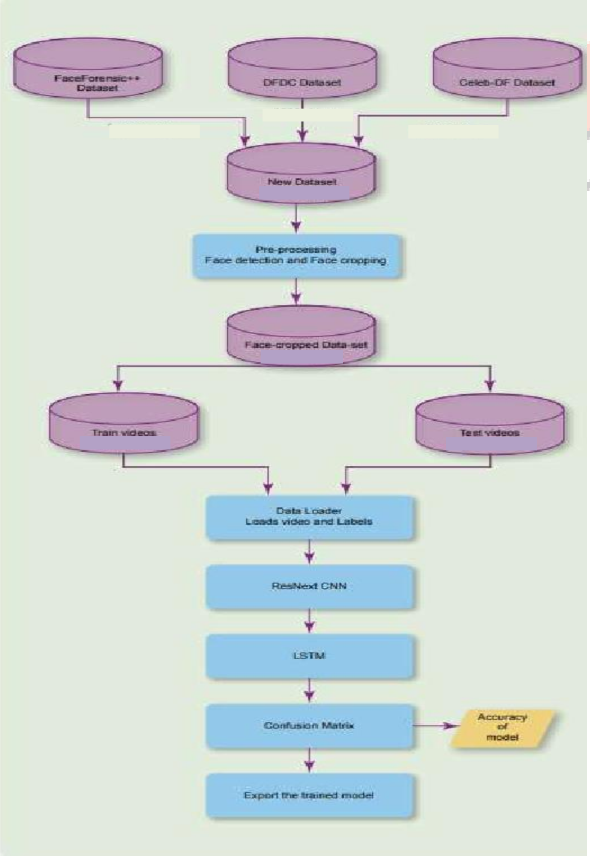
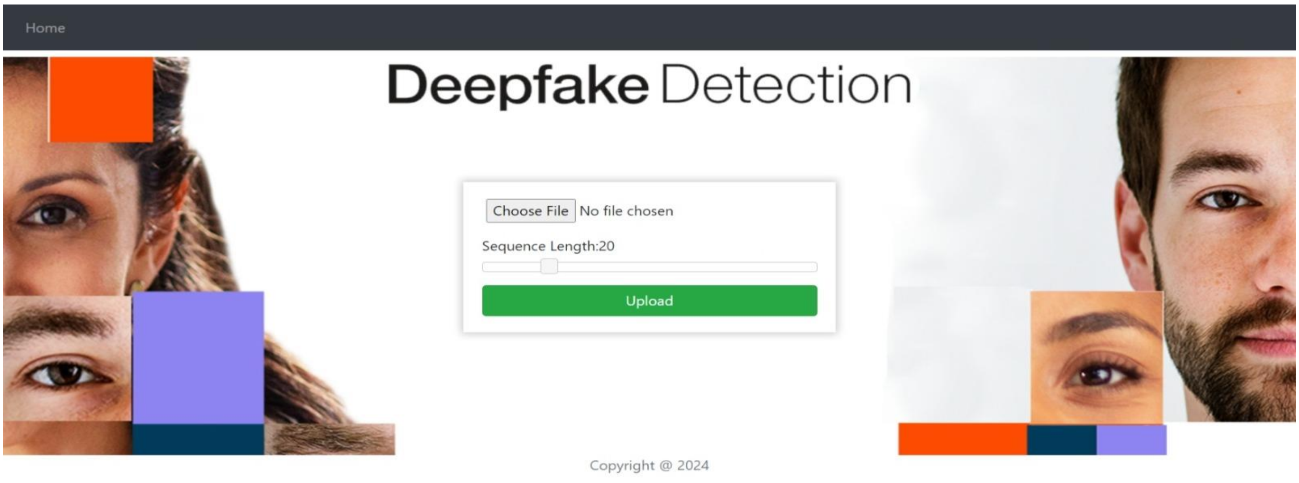


Fig2: Flow Diagram

In the experimentation phase of the project on deepfake detection using a hybrid ResNet CNN-LSTM approach, rigorous steps are undertaken to train and evaluate the model. Initially, the curated dataset is split into training, validation, and test sets, ensuring adequate representation of authentic and deepfake videos across each subset. The hybrid model is then trained on the training data, leveraging deep learning techniques to optimize its parameters and minimize loss. Throughout training, hyperparameters are fine-tuned via techniques like grid search to enhance model performance. Validation on the validation set occurs iteratively to monitor the model's progress and prevent overfitting. Following training, the model's performance is evaluated on the test dataset, with key metrics such as accuracy, precision, recall, and F1-score calculated to quantify its effectiveness in identifying deepfake content. Comparative analysis against baseline methods and state-of-the-art approaches provides insights into the superiority of the proposed model. Ethical considerations, including privacy and consent, are consistently addressed throughout the experimentation phase. The comprehensive documentation and reporting of experimental findings ensure transparency and reproducibility, advancing the understanding and development of deepfake detection system.

VI. RESULTS AND DISCUSSION

The results and discussion section of the project on deepfake detection using a hybrid ResNet CNN-LSTM approach encapsulates the culmination of extensive experimentation and analysis. Evaluating the model's performance on the test dataset reveals its efficacy in accurately discerning between authentic and deepfake videos, as evidenced by high accuracy, precision, recall, and F1-score metrics. Comparative analysis with baseline methods demonstrates the superiority of the hybrid model, showcasing its ability to outperform traditional approaches in detecting deceptive visual media. However, the discussion also delves into the impact of dataset size and quality on model performance, acknowledging the challenges posed by variations in dataset characteristics. Ethical considerations surrounding the responsible development and deployment of deepfake detection systems are thoroughly addressed, underscoring the importance of mitigating potential risks and safeguarding individual privacy. Despite the achievements of the proposed approach, limitations such as dataset biases and model constraints are acknowledged, paving the way for future research directions. Real-world applications of the deepfake detection system are explored, highlighting its potential to combat misinformation and uphold the integrity of digital content. In conclusion, the results and discussion section provides a comprehensive analysis of the hybrid ResNet CNN-LSTM approach, emphasizing its significance in addressing the evolving threat of deceptive visual media in today's digital landscape.



Deepfake Detection

No file chosen

Sequence Length:20

Copyright @ 2024

Fig 1: Home Page of Deepfake model application.

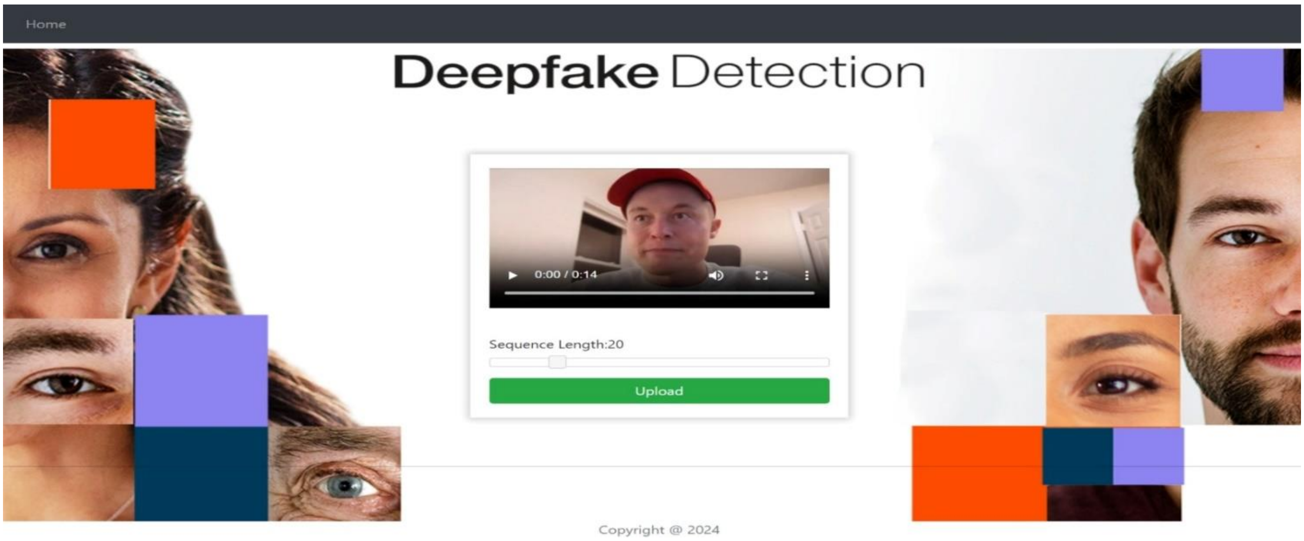


Fig 2: Home page of Deepfake model depicting the process of uploading a video.

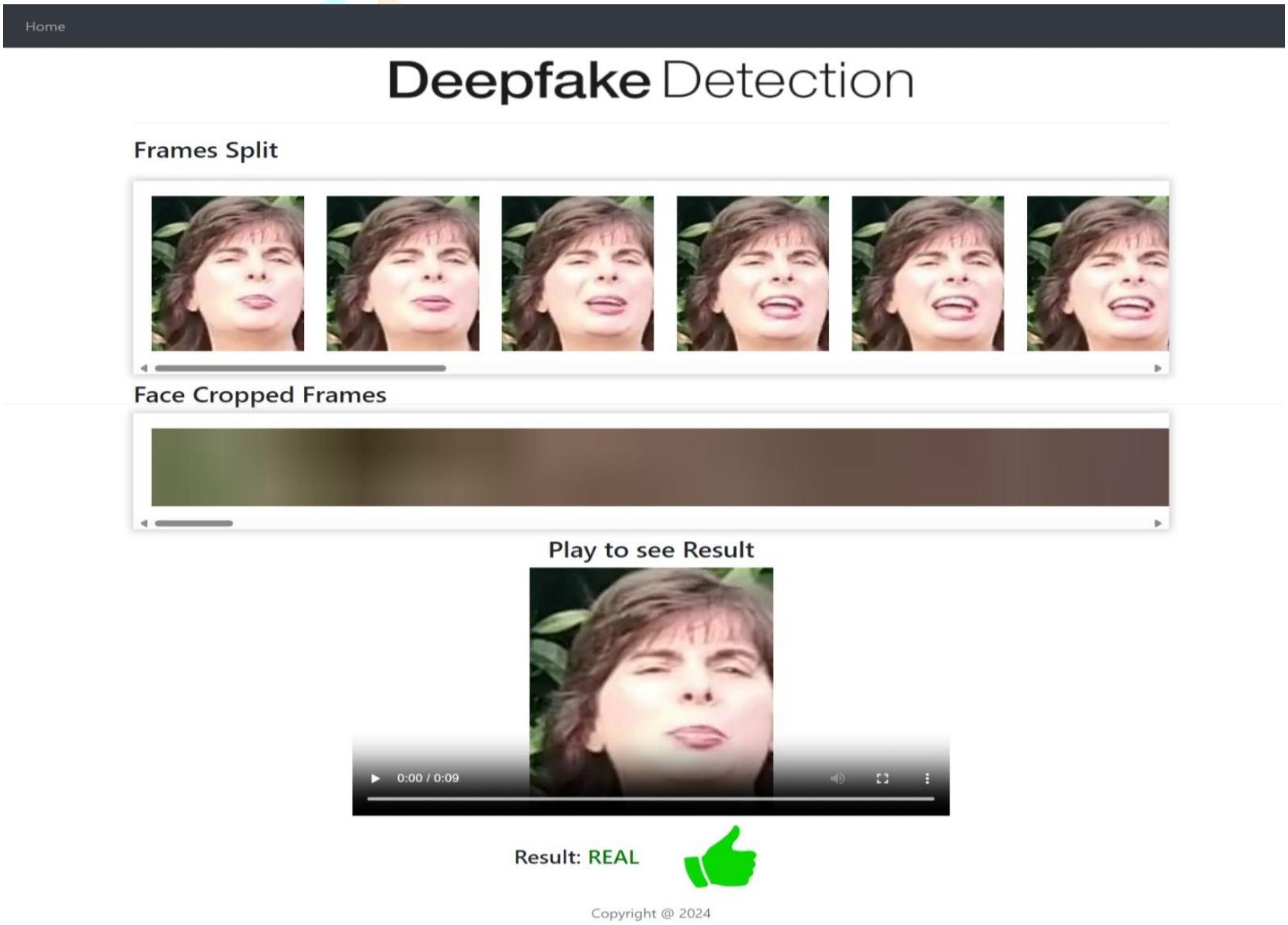


Fig 3: Result depicting the model showing that the uploaded video is real.

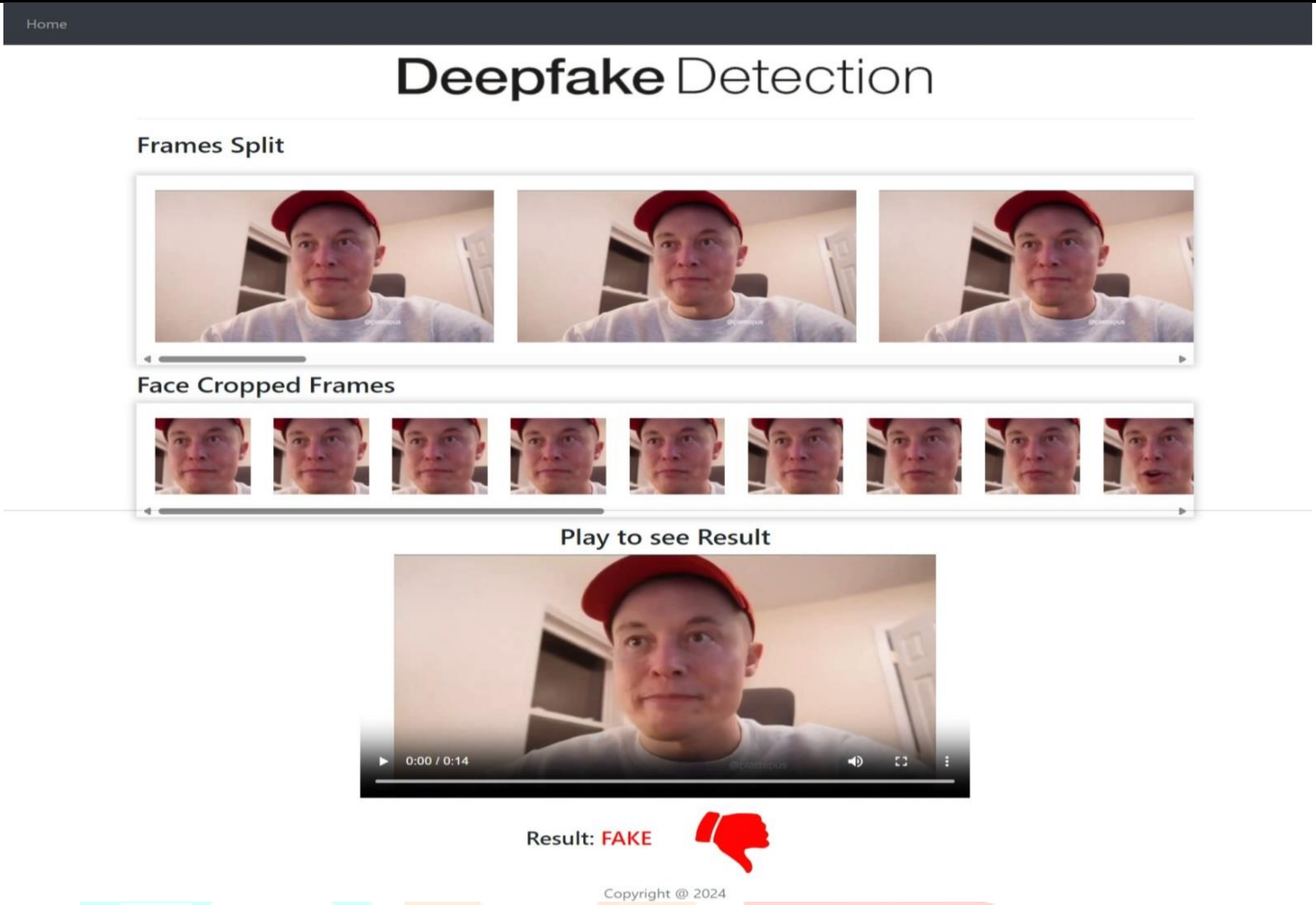


Fig 4: Result depicting the model showing that the uploaded video is fake.

REFERENCES

[1] Zhou, X., & Chellappa, R. (2020). Detecting Deepfake Videos From the Transitions. IEEE Transactions on Image Processing, 29, 9495-9508.

[2] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 1-11.

[3] Li, Y., Yang, X., Sun, P., Qi, H., Liu, M. Y., & Zhou, X. (2020). Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3207-3216.

[4] Nguyen, H. A., Yamagishi, J., Echizen, I., & Kankanhalli, M. (2019). Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 211-219.

[5] Li, Y., Chang, M. C., Lyu, S., & Guo, H. (2018). In ictu oculi: Exposing AI-Generated Fake Face Videos by Detecting Eye Blinking. arXiv preprint arXiv:1806.02877.

[6] Li, C., & Lyu, S. (2019). Exposing DeepFake Videos by Detecting Face Warping Artifacts. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 3230-3239.

[7] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). Mesonet: a Compact Facial Video Forgery Detection Network. Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), 1-7.

[8] Sabir, E., Cheng, H., & AbdAlmageed, W. (2020). Recurrent Convolutional Strategies for Face Manipulation Detection in Videos. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2497-2506.

[9] Agarwal, S., & Farid, H. (2020). Protecting World Leaders Against Deep Fakes. arXiv preprint arXiv:2004.02657.

[10] Li, Y., Yang, X., Sun, P., & Lyu, S. (2020). Face X-Ray for More General Face Forgery Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5346-5355.

[11] Guera, D., Saeedan, F., Bayat, A., & Alkinani, M. H. (2020). Deepfake Video Detection Using Convolutional Neural Networks and Bidirectional Long Short-Term Memory Networks. IEEE Access, 8, 222023-222032.

[12] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., & Morales, A. (2020). Deep Learning for Face Forgery Detection: Passive and Active Techniques. IEEE Access, 8, 202019-202036.

[13] Marra, F., Gragnaniello, D., Verdoliva, L., & Cozzolino, D. (2020). DeepFake Detection Based on Deep Learning Techniques: A Review. IEEE Access, 8, 134415-134460.

[14] Yang, X., Li, Y., & Lyu, S. (2019). Exposing GAN-Generated Faces Using Inconsistent Corneal Specular Highlights. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 10338-10347.

[15] Niu, Y., Li, X., Deng, W., & Chen, X. (2020). Learning a Model to Generate Face Images That Are Not Exist in the Training Set and Detect Deepfake Videos. IEEE Access, 8, 102158-102167.