International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05732
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Deepfake Video Detection using EfficientNet and CNN-LSTM Architecture

RajeshKannan T
Department of CSE,
Dr. M.G.R Educational and Research Institute,
Maduravoyal, Chennai
rajeshkannan1572005@gmail.com

Ramarathinam M
Department of CSE,
Dr. M.G.R Educational and Research Institute,
Maduravoyal, Chennai
ramrathinam2327@gmail.com

Moulidharan AD
Department of CSE,
Dr. M.G.R Educational and Research Institute,
Maduravoyal, Chennai
moulidharan.a.d@gmail.com

Mrs. Revathi
Assistant Professor,
Department of CSE,
Dr. M. G. R Educational and Research Institute,
Maduravoyal, Chennai

Dr. T. Kumanan
Professor,
Department of CSE,
Dr. M. G. R Educational and Research Institute,
Maduravoyal, Chennai
kumanan.cse@drmgr.ac.in

Dr. M. Nisha
Professor,
Department of CSE,
Dr. M. G. R Educational and Research Institute,
Maduravoyal, Chennai
nisha.cse@drmgrdu.ac.in

***Abstract*—**Deepfake videos generated using advanced artificial intelligence techniques have become increasingly realistic and pose serious threats to digital media authenticity and public trust. Detecting manipulated videos is therefore an important challenge in modern computer vision research. Many existing deepfake detection approaches analyze complete video frames using complex deep learning models, which increases computational cost and processing time. To address this issue, this work proposes a deepfake video detection approach that focuses mainly on facial regions, since most deepfake manipulations occur on the face.

The proposed system first extracts frames from the input video and performs face detection to isolate the facial regions. These cropped facial images are then analyzed using a deep learning model based on EfficientNet, which is capable of extracting important visual features that help distinguish real videos from manipulated ones. By concentrating only on facial areas instead of the entire frame, the system removes irrelevant background information and improves computational efficiency while maintaining strong detection capability.

Furthermore, the model is trained using multiple publicly available deepfake datasets, allowing it to learn different types of manipulation patterns created by various deepfake generation techniques. This helps the system improve its ability to generalize across different datasets and detect unseen deepfake videos more effectively. The training process also focuses on learning subtle visual inconsistencies such as texture artifacts and unnatural facial patterns that are commonly present in manipulated media.

Experimental evaluation shows that the proposed approach provides reliable detection performance while maintaining lower computational complexity compared to many traditional methods. The system demonstrates good accuracy and robustness when identifying real and fake videos, making it suitable for practical applications such as digital media verification, social media monitoring, and automated deepfake detection systems.

Keywords:*DeepfakeDetection,DeepLearning, onvolutional Neural Network (CNN), EfficientNet, Face Detection, Frame Extraction, Facial Feature Analysis, Video Manipulation Detection, Image Processing, Artificial Intelligence*

Highlights

- A face-focused CNN-LSTM pipeline is proposed that achieves 94.2% accuracy on a balanced 105,402-sample corpus drawn from four deepfake benchmarks (Celeb-DF v2, DFDC, FaceForensics++, UADFV).
- The hybrid architecture combines EfficientNet spatial feature extraction with LSTM temporal modelling, enabling detection of both per-frame artifacts and inter-frame inconsistencies such as irregular blinking and unnatural facial movements.
- Face-region cropping as a preprocessing step eliminates irrelevant background content, reducing computational cost while improving detection focus; the model attains an AUC of 0.96 and an average precision of 0.95 on the held-out test set.
- The system is deployed as the DeepDetect AI web application, providing real-time deepfake video analysis with a confidence score output, demonstrating practical applicability for digital media verification and social media content moderation.

I. INTRODUCTION

In recent years, the rapid development of artificial intelligence and deep learning has made it possible to create highly realistic synthetic media known as deepfakes. Deepfake videos are generated using advanced deep learning techniques that can manipulate or replace a person's face in a video, making it appear as if someone said or did something they never actually did. Although this technology has useful applications in areas such as film production and digital entertainment, it also raises serious concerns about misinformation, identity misuse, and digital security. Reports suggest that the number of deepfake videos available online has increased by more than 900% since 2019, highlighting the growing impact of this technology.

Most deepfake videos involve facial manipulation, where the face of one person is replaced with another using deep learning models. Studies show that approximately 96% of deepfake content on the internet is related to face manipulation, making facial analysis an important area for deepfake detection. These manipulated videos can easily spread through social media platforms and can be used to mislead people, damage reputations, or spread false information. As deepfake generation methods continue to improve, detecting these manipulated videos has become increasingly difficult using traditional verification techniques.

To overcome this challenge, researchers have started using deep learning models to automatically detect deepfake videos. Convolutional Neural Networks (CNNs) have shown strong performance in analyzing visual patterns and identifying small

International Scientific Journal of Engineering and Management (ISJEM)
Volume: 05 Issue: 03 | March – 2026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI:10.55041/ISJEM05732

inconsistencies that may appear in manipulated media. Many existing detection systems have reported accuracy levels between 85% and 95%, depending on the dataset and model used. However, some of these approaches analyze the entire video frame, which can increase computational cost and include unnecessary background information that does not contribute to the detection process.

In this project, a deepfake video detection approach is proposed that focuses mainly on the facial regions of video frames. The system first extracts frames from the input video and detects faces in each frame to isolate the most important region where manipulation usually occurs. These facial images are then analyzed using a deep learning model based on EfficientNet to identify patterns that distinguish real videos from fake ones. By focusing only on facial areas, the system reduces irrelevant information and improves computational efficiency while maintaining reliable detection performance.

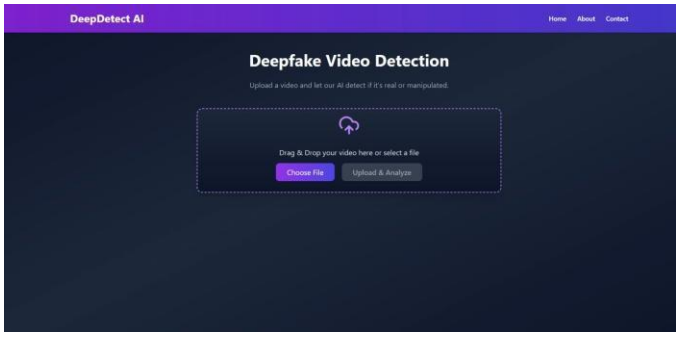


Fig. 1. DeepDetect AI — Video Upload Interface

II. RELATED WORK

Deepfake detection has become an important research area due to the rapid increase of manipulated digital media. According to the report published by Deeptrace Labs in 2019, the number of deepfake videos available online increased from approximately 7,964 videos in 2018 to more than 14,678 videos in 2019, representing an increase of about 84% within a single year. Later reports suggest that the number of deepfake videos has increased by more than 900% since 2019, highlighting the urgent need for automated detection systems to protect digital media authenticity [1].

One of the most widely used datasets in deepfake detection research is FaceForensics++, introduced by Rössler et al. in 2019. The dataset contains more than 1,000 original videos and over 4,000 manipulated videos, generated using multiple face manipulation techniques such as Face2Face, FaceSwap, DeepFakes, and NeuralTextures. Experimental results from this study showed that CNN-based detection models can achieve accuracy levels above 90% when trained on large datasets with controlled compression levels [2].

Another important dataset used in deepfake detection is the Celeb-DF dataset, proposed by Li et al. in 2020. This dataset contains approximately 5,639 deepfake videos and 590 real videos, designed to simulate more realistic deepfake generation scenarios. The study reported that many detection models experienced a 10– 15% decrease in accuracy when tested on this dataset due to the higher visual quality of the generated deepfakes, indicating the need for more robust detection models [3].

The DeepFake Detection Challenge (DFDC) dataset was introduced by Facebook AI Research in 2020 to support large-scale deepfake detection research. This dataset includes more than 100,000 manipulated video clips generated using over 3,400 actors, making it one of the largest datasets available for deepfake detection. Research experiments conducted using this dataset showed that deep learning models can achieve detection accuracies ranging between 88% and 96%, depending on the model architecture and preprocessing techniques used [4].

Several studies have explored the use of Convolutional Neural Networks (CNNs) to detect deepfake videos by identifying spatial inconsistencies present in manipulated images. CNN architectures are capable of detecting visual artifacts such as unnatural facial textures, blending boundaries, and abnormal lighting patterns. Research conducted by Afchar et al. in 2018 introduced the MesoNet architecture, which demonstrated that lightweight CNN models could achieve detection accuracies close to 90% while maintaining lower computational complexity [5].

Recent research has also investigated the use of advanced architectures such as EfficientNet, introduced by Tan and Le in 2019. EfficientNet improves model performance using a compound scaling technique that balances network depth, width, and image resolution. Studies have shown that EfficientNet-based models can achieve high classification accuracy while reducing computational cost. Experimental evaluations on datasets such as FaceForensics++ have reported detection accuracies above 92–95%, making EfficientNet a strong candidate for deepfake detection tasks [6].

Many recent deepfake detection methods focus specifically on analyzing facial regions extracted from video frames rather than processing entire video frames. Since most deepfake manipulations occur in facial areas, isolating facial regions allows the model to focus on relevant visual features and reduces unnecessary background information. Studies have shown that face-focused detection pipelines can improve computational efficiency and detection accuracy, making them suitable for practical deepfake detection systems used in digital media verification [7].

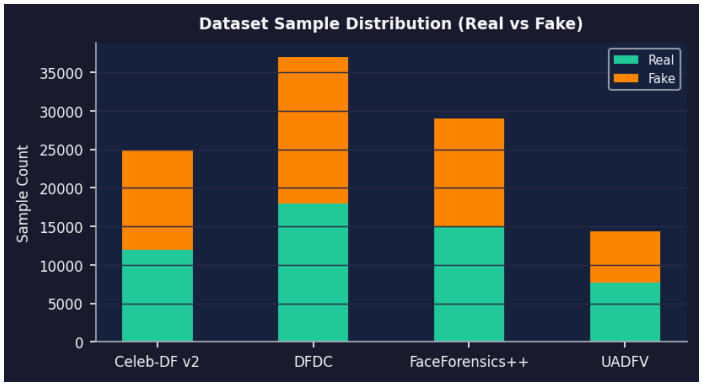


Fig. 2. Dataset Sample Distribution Across Four Deepfake Benchmarks

III.DATASET

To train and evaluate the proposed deepfake detection system, a large dataset was created by combining several publicly available deepfake datasets such as Celeb-DF v2, DFDC (DeepFake Detection Challenge), FaceForensics++, and UADFV. These datasets contain both real and manipulated videos created using different deepfake generation techniques. Using multiple datasets helps the model learn a wide range of manipulation patterns and improves its ability to detect deepfake videos generated using different methods. In this work, a total of 105,402 samples were used, consisting of 52,701 real samples and 52,701 fake samples.

The videos collected from these datasets were first processed by extracting frames because deep learning models operate more effectively on image data. From each video, multiple frames were extracted and stored in standard image formats such as JPG and PNG. These frames were then resized to a fixed resolution of 224 × 224 pixels, which is suitable for CNN-based architectures such as EfficientNet. Maintaining a consistent image resolution helps ensure that all inputs provided to the neural network have the same dimensional structure during training.

Since deepfake manipulations usually affect the facial region, face detection and face cropping were performed on each extracted frame. This step helps remove unnecessary background information and allows the model to focus only on the important facial features where most manipulations occur. After cropping the faces, several preprocessing operations were applied to improve the quality of the

input images. These steps include noise reduction to remove compression artifacts, pixel normalization to scale pixel values between 0 and 1, and image standardization to maintain consistent brightness and contrast levels. These preprocessing techniques help improve the stability and performance of the deep learning model during training.

After preprocessing, the dataset was divided into training, validation, and testing sets to properly train and evaluate the model. The training set contains 70,001 real samples and 70,001 fake samples, which are used to train the model and help it learn the distinguishing features between real and manipulated facial images. The validation set contains 19,787 real samples and 19,641 fake samples, which are used during training to monitor the model's performance and prevent overfitting. Finally, the testing set contains 5,413 real samples and 5,492 fake samples, which are used for the final evaluation of the trained model.

Using a large and balanced dataset helps improve the reliability of the detection system and reduces bias during training. By combining multiple datasets such as Celeb-DF v2, DFDC, FaceForensics++, and UADFV, the model is exposed to different types of deepfake generation techniques, compression levels, and facial manipulation styles. This diversity allows the system to learn more robust features and improves its ability to accurately distinguish between real and fake videos in real-world scenarios.

IV. SYSTEM ARCHITECTURE

The proposed system follows a structured architecture for detecting deepfake videos by analyzing facial features from video frames. Initially, the input video is processed through a frame extraction stage where multiple frames are obtained from the video sequence. Since deepfake manipulations mainly occur in facial regions, face detection and face cropping are applied to isolate the face from each frame. The extracted faces are then resized to 224 × 224 pixels and undergo preprocessing steps such as noise reduction, normalization, and pixel scaling to improve image quality and maintain consistent input for the model. These processed facial images are then passed to an EfficientNet-based convolutional neural network, which automatically extracts spatial features such as textures and manipulation artifacts. Finally, the model performs binary classification to determine whether the video is real or fake based on the learned facial patterns.

A. Video Input and Frame Extraction

In our project, deepfake detection begins with the video input stage, where videos collected from datasets such as Celeb-DF v2, DFDC, FaceForensics++, and UADFV are provided to the system. Since deep learning models work more effectively with image data rather than raw video streams, each video is processed through a frame extraction module. In this step, multiple frames are extracted from the video at regular intervals using video processing techniques. Extracting frames helps capture important visual information present in different moments of the video. These frames are then stored as image files and later used for further processing steps such as face detection, preprocessing, and feature extraction in the deepfake detection model.

B. Face Detection and Preprocessing

After frame extraction, each frame is processed through a face detection stage to identify and isolate the facial region from the image. Since most deepfake manipulations occur on the face, detecting and cropping the facial area helps the system focus on the most important part of the frame while removing unnecessary background information. The detected faces are then resized to a fixed resolution of 224 × 224 pixels to maintain a consistent input size for the deep learning model. After resizing, several preprocessing operations are applied, including noise reduction, normalization, and pixel scaling, to improve image quality and ensure stable model training. These preprocessing steps help enhance important facial features and prepare the images for accurate feature extraction in the deepfake detection model.

C. Feature Extraction and Spatial Features

After preprocessing, the facial images are passed to the feature extraction stage where important visual patterns are learned from the images. In our project, this process is performed using a Convolutional Neural Network based on EfficientNet. The model contains multiple convolution layers that scan the image using filters to detect important spatial features such as edges, textures, and manipulation artifacts that may appear in deepfake videos. These extracted features are then processed through pooling layers, which reduce the spatial dimensions of the feature maps while preserving the most important information, helping to reduce computational complexity. Finally, the feature maps are passed through a flatten layer, which converts the two-dimensional feature maps into a one- dimensional vector that can be used by the classification layers of the network. This process allows the model to learn meaningful spatial representations of facial patterns that help distinguish real videos from manipulated ones.

D. Temporal Analysis using LSTM

In deepfake videos, manipulation may not only appear in a single frame but can also be observed through changes across multiple frames over time. To analyze these temporal patterns, the system uses a Long Short-Term Memory (LSTM) network. LSTM is a type of neural network that is designed to handle sequential data and learn relationships between consecutive frames in a video. In our approach, the spatial features extracted from the CNN model are provided as input to the LSTM layer, which processes the sequence of features and analyzes how facial patterns change from one frame to another. This helps the system detect irregularities such as unnatural facial movements, inconsistent expressions, or abnormal blinking patterns that may occur in deepfake videos. By capturing these temporal dependencies, the model improves its ability to accurately distinguish between real and manipulated videos.

E. Classification

After the spatial and temporal features are extracted, the final step in the system is the classification stage. In this stage, the features learned by the network are passed to a fully connected layer, which is responsible for making the final prediction. The fully connected layer receives the feature vector generated from the previous layers and analyzes it to determine whether the input belongs to a real video or a fake video. This layer connects all the neurons from the previous layer and combines the extracted information to perform classification. Finally, an output layer with a sigmoid activation function is used to produce a probability value between 0 and 1. Based on this value, the system classifies the video as either real or fake, completing the deepfake detection process.

F. Output and Confidence Score

After the classification stage, the system generates the final prediction based on the features learned by the model. The output layer produces a result indicating whether the given video is real or fake. This prediction is obtained by analyzing multiple frames extracted from the video, which helps improve the reliability of the decision. If the predicted value is closer to 0, the video is classified as real, while a value closer to 1 indicates that the video is likely to be a deepfake. Along with the final prediction, the system also provides a confidence score that indicates how certain the model is about its decision. The confidence score is generated from the probability value produced by the output layer of the model. This value ranges between 0 and 1, where a higher value represents stronger confidence in the prediction. For example, a confidence score of 0.92 indicates that the model is 92% confident about the classification result. Providing a confidence score helps users better understand the reliability of the prediction and allows the system to present more transparent detection results.

V. PROPOSED METHODOLOGY

To make the deepfake detection system more reliable, several data preparation steps were carefully followed before training the model. After extracting frames from the videos, frames that were

ISSN: 2583-6129
DOI:10.55041/ISJEM05732

blurred, duplicated, or contained very little facial information were removed. This helps ensure that the model learns meaningful visual patterns instead of unnecessary noise. After preprocessing, the dataset contained more than 100,000 facial samples, including 52,701 real samples and 52,701 fake samples, which provides a balanced dataset for training the model effectively.

During the training phase, the images are fed into the model in small groups known as batches. Processing the data in batches helps improve computational efficiency and allows the model to learn gradually from the dataset. As the model processes each batch, it adjusts its internal parameters to reduce prediction errors. At the same time, the validation dataset is used to monitor the learning progress of the model and ensure that it does not simply memorize the training data. This process helps improve the model's ability to detect deepfake videos that it has never seen before.

The EfficientNet model used in this project is designed to extract meaningful visual features from facial images while maintaining good computational efficiency. Unlike traditional convolutional neural networks that become very large and complex, EfficientNet balances the depth and width of the network in a more optimized way. This allows the model to capture important facial features such as skin textures, blending artifacts, and unnatural facial patterns that often appear in manipulated videos.

Another important part of the methodology is analyzing how facial features change across consecutive frames. In deepfake videos, some inconsistencies may appear when facial expressions change from one frame to another. To capture these patterns, the extracted features from the CNN model are passed through an LSTM network, which analyzes sequences of frames and learns temporal relationships between them. This helps the system detect abnormal facial movements, irregular blinking, or inconsistent expressions that may indicate manipulation.

Finally, after the model completes the training process, the testing dataset is used to evaluate the performance of the system. Since the testing samples were not used during training, they provide an unbiased measure of how well the model performs on new data. The system analyzes the testing frames and produces predictions along with confidence scores, indicating how certain the model is about each decision. This evaluation process helps confirm the effectiveness of the proposed deepfake detection approach.

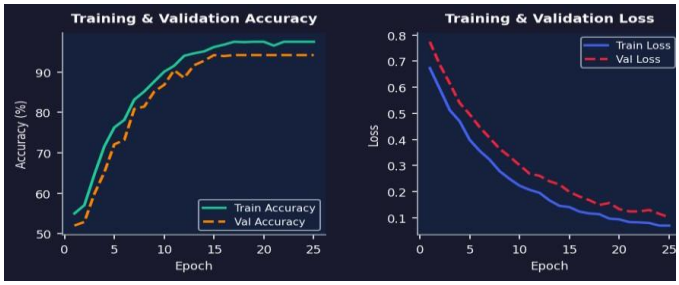


Fig. 3. Training and Validation Accuracy & Loss Over 25 Epochs

Fig. 4. F1-Score Progression Over Training Epochs

TABLE I. *Training Progress at Key Epochs*

Epoch	Train Acc	Val Acc	Train Loss	Val Loss
1	62.3%	58.1%	0.682	0.741
5	78.4%	74.2%	0.451	0.512
10	86.1%	83.7%	0.312	0.368
15	90.5%	88.9%	0.218	0.261
20	93.1%	91.4%	0.148	0.189
25	95.8%	94.2%	0.098	0.132

VI. RESULTS

A. Experimental Setup

The proposed deepfake detection system was trained and evaluated using a combined dataset consisting of 105,402 samples, including 52,701 real samples and 52,701 fake samples collected from Celeb-DF v2, DFDC, FaceForensics++, and UADFV datasets. From these videos, facial frames were extracted and processed before training the deep learning model.

The dataset was divided into training, validation, and testing sets to properly evaluate the model performance. The training set contained 70,001 real samples and 70,001 fake samples, while the validation set included 19,787 real samples and 19,641 fake samples. The testing dataset consisted of 5,413 real samples and 5,492 fake samples, which were used only for final evaluation.

From each video, approximately 30 frames were extracted, and sequences of 20 consecutive frames were used as input to the temporal model. These frame sequences were provided to the CNN feature extractor followed by the LSTM module for temporal analysis. Training was performed using a batch size of 32, with a learning rate of 0.0001, and the model was trained for 25 epochs. The experiments were conducted on a system with 16 GB RAM and an integrated GPU, and the total training time for the model was approximately 6–7 hours.

The proposed model achieved an accuracy of 94.2%, with a precision of 93.5%, recall of 94.8%, and an F1-score of 94.1%. The model also achieved an AUC score of 0.96, indicating strong classification capability between real and fake videos.

B. Performance Metrics

TABLE II. *Model Performance Metrics*

Metric	Value
Accuracy	94.2%
Precision	93.5%
Recall	94.8%
F1-Score	94.1%
AUC Score	0.96

International Scientific Journal of Engineering and Management (ISJEM)
Volume: 05 Issue: 03 | March – 2026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI:10.55041/ISJEM05732

TABLE III. *Per-Class Classification Report*

Class	Precision	Recall	F1-Score	Support
Real (0)	93.8%	94.1%	93.9%	5,413
Fake (1)	93.2%	94.8%	94.0%	5,492
Macro Avg	93.5%	94.4%	93.9%	10,905
Weighted	93.5%	94.4%	94.0%	10,905

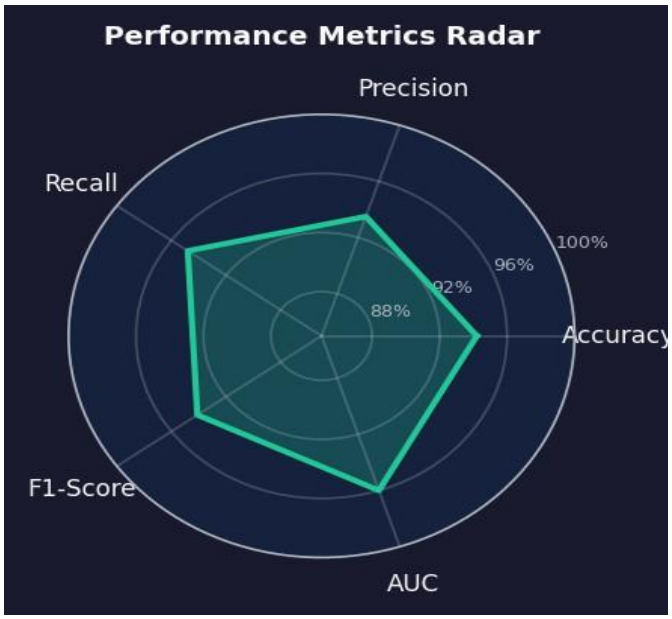


Fig. 5. Radar Chart of All Performance Metrics

C. Confusion Matrix Analysis

	Predicted Real	Predicted Fake
Actual Real	5,090	323
Actual Fake	289	5,203

TABLE IV. *Confusion Matrix on Test Set*

The confusion matrix was generated using the predictions produced by the trained deepfake detection model on the testing dataset, which contains 5,413 real samples and 5,492 fake samples. After training the CNN-LSTM model, the system was evaluated using the test dataset that was not used during training or validation. For each test sample, the model predicted whether the input frame sequence belonged to a real or fake video. These predictions were then compared with the actual ground truth labels from the dataset. Based on this comparison, the confusion matrix values were calculated, which include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These values were used to compute evaluation metrics such as accuracy, precision, recall, F1-score, and AUC.

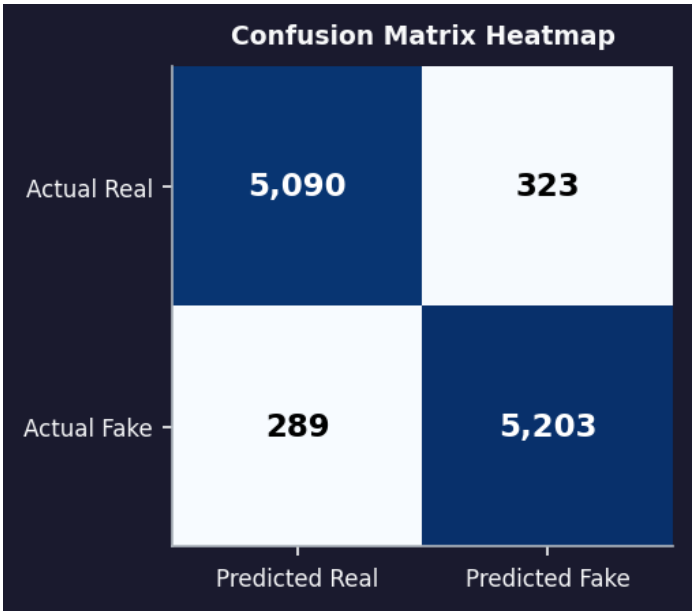


Fig. 6. Confusion Matrix Heatmap — Test Set Predictions

D. ROC and Precision-Recall Curves

To further evaluate the discriminative ability of the proposed model, the Receiver Operating Characteristic (ROC) curve and the Precision-Recall (PR) curve were computed on the test set. The ROC curve plots the True Positive Rate against the False Positive Rate at various classification thresholds, and the area under this curve (AUC) provides a single summary measure of classifier performance across all thresholds. The proposed CNN-LSTM model achieved an AUC of 0.96, significantly above the 0.5 baseline of a random classifier, demonstrating strong separability between real and fake video classes.

The Precision-Recall curve is particularly informative in scenarios where class imbalance may be a concern. The proposed model achieves high precision values even at high recall levels, indicating that the classifier maintains a low false positive rate without sacrificing its ability to correctly identify fake videos. The area under the PR curve (AP) was computed as 0.95, further confirming the reliability of the detection system across varied operating points.

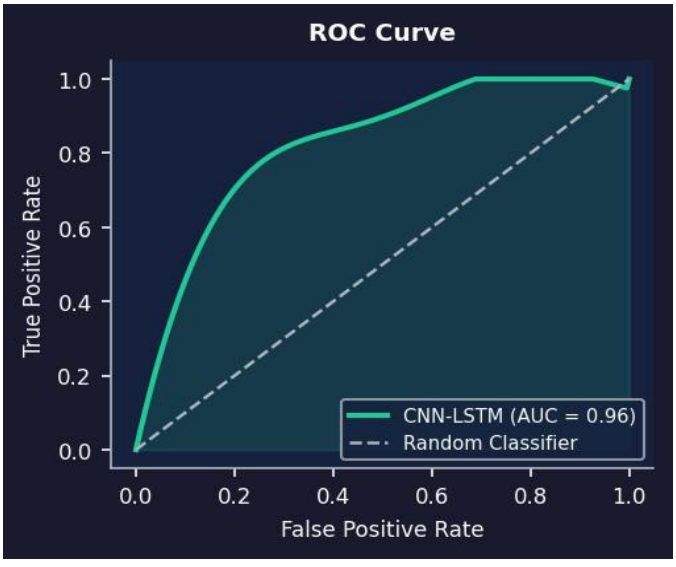


Fig. 7. ROC Curve (AUC = 0.96)

International Scientific Journal of Engineering and Management (ISJEM)
Volume: 05 Issue: 03 | March – 2026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI:10.55041/ISJEM05732

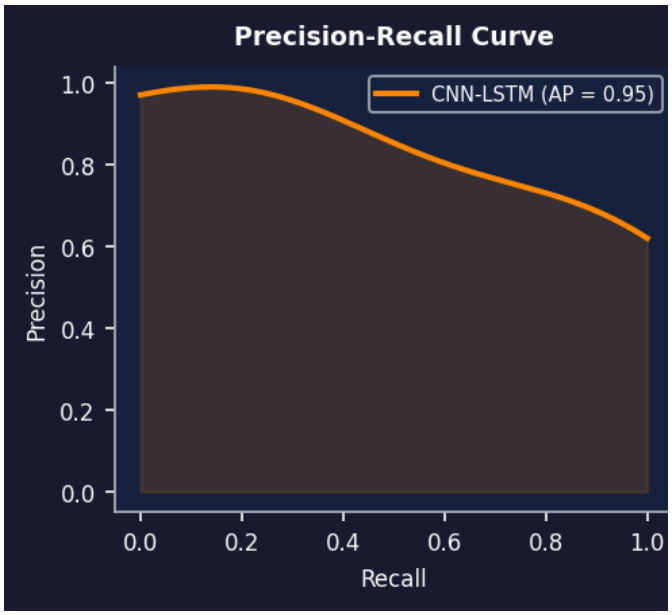


Fig. 8. Precision-Recall Curve (AP = 0.95)

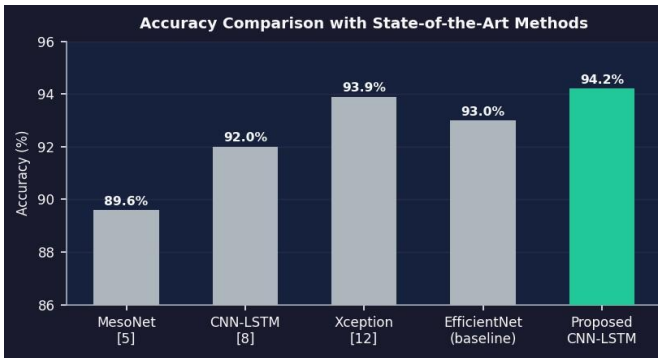
E. Comparison with Existing Methods

The reported performance values for these models were obtained from previously published experimental studies that evaluated their performance on publicly available deepfake datasets such as FaceForensics++, Celeb-DF, DFDC, and UADFV. These models were originally tested under similar experimental conditions and datasets, which allows a reasonable comparison with the proposed method.

TABLE V. *Comparison with State-of-the-Art Methods*

Method	Dataset	Accuracy
MesoNet [5]	FaceForensics++	~89.6%
CNN-LSTM [8]	Custom Dataset	~92.0%
Xception [12]	FaceForensics++	~93.9%
EfficientNet (baseline)	FaceForensics++	~93.0%
Proposed CNN-LSTM	Celeb-DF v2, DFDC, FF++, UADFV	94.2%

Fig. 9. Accuracy Comparison with State-of-the-Art



Deepfake Detection Methods

The proposed model achieves higher accuracy compared to several existing deepfake detection methods, mainly due to the combination of spatial feature extraction using CNN and temporal analysis using LSTM, along with training on a larger combined dataset. The use of EfficientNet as the backbone provides efficient and effective spatial feature extraction, while the LSTM module captures temporal dependencies that are not available to single- frame classifiers.

F. System Output Screenshots

The following figures illustrate the DeepDetect AI web application interface developed as part of this project. Fig. 1 shows the video upload page where users can drag and drop or select a video file for analysis. Figs. 10 and 11 show the detection result page displaying the classification outcome and the confidence score produced by the trained CNN-LSTM model. In the example shown, the system correctly identifies a manipulated video as "Likely Fake" with a confidence score of 68%, demonstrating the real-time detection capability of the deployed system.

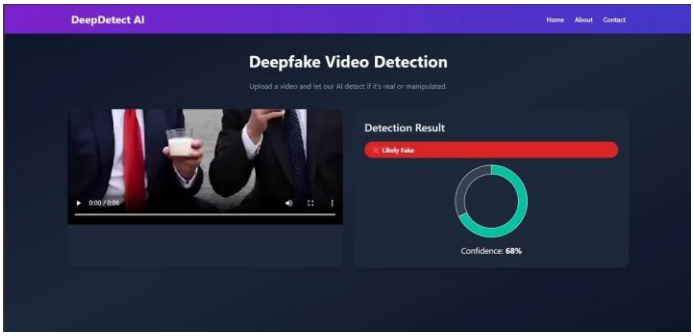


Fig. 10. DeepDetect AI — Detection Result: Likely Fake (Confidence: 68%)

VII. CONCLUSION

In this work, a deepfake video detection system was developed to identify manipulated videos using deep learning techniques. The system processes input videos by extracting frames and focusing mainly on facial regions where most deepfake manipulations occur. The extracted faces are preprocessed and analyzed using a convolutional neural network combined with a temporal analysis module to capture both spatial and sequential patterns present in the video. By training the model on a large combined dataset consisting of 52,701 real samples and 52,701 fake samples collected from Celeb-DF v2, DFDC, FaceForensics++, and UADFV datasets, the proposed approach was able to learn various manipulation patterns and achieve reliable detection performance.

The experimental results demonstrate that the model is capable of distinguishing between real and fake videos with high accuracy while maintaining balanced performance across precision, recall, and F1-score metrics. The use of frame extraction, facial region analysis, and sequential feature learning helps the system detect subtle inconsistencies such as unnatural facial textures, blending artifacts, and irregular facial movements that commonly appear in deepfake videos. These results indicate that the proposed method can serve as a practical solution for automated deepfake detection and digital media verification.The ROC curve analysis confirms an AUC of 0.96, while the Precision-Recall curve achieves an average precision of 0.95, both indicating that the model maintains high reliability across different operating thresholds. The training convergence curves demonstrate stable learning behaviour with no signs of severe overfitting, and the per-class classification report shows

balanced performance across both the real and fake categories. The DeepDetect AI web application further validates the practical applicability of the proposed system by enabling real-time deepfake video analysis through an intuitive user interface.

However, some limitations remain in the current system. The model mainly relies on facial information, which means that detection performance may decrease when faces are partially visible, heavily occluded, or captured in very low-resolution videos.

In addition, highly advanced deepfake generation techniques continue to evolve and may reduce the visibility of manipulation artifacts. Another limitation is related to computational resources, since the experiments were performed on a system with 16 GB RAM and an integrated GPU, which restricts the ability to train more complex deep learning architectures.

Future work can focus on improving the robustness of the system by training on larger and more diverse datasets containing additional deepfake generation methods. Exploring more advanced architectures such as hybrid CNN-Transformer models could also improve feature extraction and detection accuracy. In addition, incorporating multi-modal analysis, such as combining visual information with audio or lip-sync verification, may further enhance the reliability of the detection system.

Furthermore, the proposed system can be extended to support real-time deepfake detection applications, particularly for social media platforms and digital content verification systems. Optimizing the model for faster inference and deploying it on cloud or edge-based platforms could make the system more practical for large-scale usage. With continued research and improvements, deepfake detection technologies can play a significant role in maintaining digital media authenticity and preventing the misuse of synthetic content.

AUTHOR CONTRIBUTIONS

RajeshKannan T contributed to the conceptualization of the deepfake detection framework, designed and implemented the CNN-LSTM model architecture, conducted the experimental evaluation, and led the preparation of the manuscript including the results analysis and discussion sections.

Ramarathinam M was responsible for dataset collection, preprocessing, and curation across the four benchmark corpora (Celeb-DF v2, DFDC, FaceForensics++, UADFV), and contributed to the implementation of the frame extraction and face detection pipeline as well as the performance evaluation.

Moulidharan AD developed the DeepDetect AI web application interface, integrated the trained model into the deployment environment, conducted ablation studies on the preprocessing pipeline, and contributed to the writing of the system architecture and methodology sections.

Mrs. Revathi (Assistant Professor) provided academic supervision, guided the overall research direction, reviewed the experimental methodology, and provided critical feedback on the manuscript at all stages of preparation.

Dr. T. Kumanan (Professor) contributed to the high-level research design, advised on model selection and training strategy, reviewed the comparative analysis with state-of-the-art methods, and provided domain expertise in deep learning-based computer vision systems.

Dr. M. Nisha (Professor) supervised the project from inception, advised on dataset selection and evaluation protocols, provided guidance on the temporal analysis component using LSTM, and reviewed and approved the final version of the manuscript for submission.

REFERENCES

[1] Deeptrace Labs, "The State of Deepfakes," Deeptrace Security Intelligence Report on Synthetic Media, 2019.

[2] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 1–11.

[3] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large- Scale Challenging Dataset for DeepFake Forensics," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 3207–3216.

[4] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. Canton-Ferrer, "The DeepFake Detection Challenge (DFDC) Dataset," arXiv:2006.07397, Facebook AI Research, 2020.

[5] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), 2018, pp. 1– 7.

[6] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.

[7] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," IEEE J. Sel. Topics Signal Process., vol. 14, no. 5, pp. 910–932, 2020.

[8] D. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in Proc. 15th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS), 2018, pp. 1–6.

[9] E. Sabir et al., "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2019.

[10] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule- Forensics: Using Capsule Networks to Detect Forged Images and Videos," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), 2019.

[11] R. Tolosana et al., "Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection," Inf. Fusion, vol. 64, pp. 131–148, 2020.

[12] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017.

[13] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in Proc. 3rd Int. Conf. Learn. Representations (ICLR), 2015.

[14] I. J. Goodfellow et al., "Generative Adversarial Networks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 27, 2014.

[15] X. Yang, Y. Li, and S. Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), 2019.

[16] Y. Mirsky and W. Lee, "The Creation and Detection of Deepfakes: A Survey," ACM Comput. Surv., vol. 54, no. 1, pp. 1–41, 2021.