International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 3, pp 148-154, March-2025

https://doi.org/10.47001/IRJIET/2025.903018

Video Deepfake Detection Using EfficientNet and LSTM

[1]Satwika.M, [2]Neha.K, [3]Pranavya.A, [4]Rishika.K, [5]Siva Sankar Namani

[1,2,3,4]Department of CSE (AI & ML), G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India

[5]Assistant Professor, Department of CSE (AI & ML), G. Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India

***Abstract* - As the prevalence of advanced face manipulation technologies continues to grow, the detection of deepfakes has emerged as a pivotal area of research, presenting ethical and security challenges. This study introduces an enhanced deepfake detection framework that integrates the EfficientNetB0 model, cutting-edge convolutional neural network (CNN) architecture, with an added long short-term memory (LSTM) layer. The CNN component, EfficientNetB0, excels at extracting spatial features from individual frames, capturing fine-grained details and artifacts indicative of manipulation. The LSTM layer processes sequential dependencies across frames, leveraging temporal inconsistencies that are often present in deepfake videos but difficult to detect from individual images alone. By combining spatial and temporal feature learning, this hybrid CNN-LSTM approach enhances the system's ability to distinguish real from manipulated media with greater accuracy. Leveraging video frame extraction and comprehensive data augmentation techniques, the system preprocesses inputs to improve generalization on limited training data. The EfficientNetB0 model, pre-trained on the ImageNet dataset, serves as the backbone for feature extraction, utilizing its highly efficient architecture, which includes depth wise separable convolutions. An evaluation on the Celeb-DF dataset demonstrates that the proposed system maintains high accuracy and robustness in detecting deepfake content, while preserving computational efficiency, making it well-suited for real-world applications. The experimental results validate the effectiveness of this extended approach, emphasizing its potential to significantly contribute to the mitigation of the detrimental impacts of deepfakes.**

Keywords: Convolutional neural networks, CNN, EfficientNet, Deepfake detection, Long short-term memory, LSTM, Temporal feature learning.

I. INTRODUCTION

Deepfakes refer to artificially generated media, including images, videos, and audio, created using sophisticated artificial intelligence (AI) techniques, primarily deep learning. These synthetic forms of media leverage neural networks to convincingly alter or fabricate visual and auditory content. The rise of deepfake technology has raised significant concerns regarding authenticity, privacy, and societal trust, as it enables the seamless manipulation of reality. The term "deepfake" is derived from "deep learning" and "fake," emphasizing the AI-driven approach underlying these manipulations.

Modern deepfake methodologies can be categorized based on their ability to modify facial expressions and features:

Face2Face: This real-time facial reenactment method employs 3D facial modeling and texture mapping to superimpose an actor's expressions onto another person's face. By transferring facial movements while retaining the target individual's identity, Face2Face enables highly realistic video manipulations, making it particularly effective for real-time applications.

Face Swap: As the name implies, Face Swap technology seamlessly replaces one person's facial features with another's. Using facial landmark detection, segmentation, and deep neural networks, it ensures that the swapped face maintains proper alignment, texture, and lighting. This technique is widely utilized in entertainment but has also been exploited for malicious purposes.

Neural Textures: This technique enhances the photorealism of deepfake videos by generating AI-based texture maps. These maps refine intricate details such as skin patterns, lighting, and dynamic facial expressions, making the manipulated content nearly indistinguishable from authentic footage.

Deepfake generation heavily relies on Generative Adversarial Networks (GANs), a class of neural networks consisting of two competing models: a generator and a discriminator. The generator is responsible for producing synthetic content, such as realistic human faces or voices, while the discriminator evaluates its authenticity, distinguishing real data from AI-generated fakes. Through

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 3, pp 148-154, March-2025

https://doi.org/10.47001/IRJIET/2025.903018

continuous adversarial training, the generator refines its outputs, progressively making them more realistic. Advanced variations, such as StyleGAN and CycleGAN, specialize in high-resolution image synthesis and domain-specific transformations, further advancing the quality and realism of deepfake media. The iterative nature of GANs, combined with large-scale datasets, has established them as the dominant architecture for deepfake creation.

Initially considered a technological breakthrough, deepfakes have now become a growing societal concern due to their potential for misuse. Their unethical applications have led to serious challenges spanning ethical, economic, and security domains. One of the most pressing issues is the spread of disinformation, where fabricated videos of political figures making false announcements or statements have triggered confusion and eroded public trust in legitimate news sources. Privacy violations are another critical concern, as deepfake technology has been exploited to create non-consensual explicit content, leading to reputational damage, psychological distress, and legal repercussions. Economically, deepfakes pose security risks, including financial fraud through AI-driven impersonation of executives, facilitating unauthorized transactions, and voice cloning in spear-phishing attacks, thereby compromising cybersecurity. On a broader scale, deepfakes pose national and global security threats, with the potential to fabricate videos of world leaders declaring military action or resignations, thereby inciting geopolitical instability and diplomatic crises.

II. LITERATURE SURVEY

Recent advancements in deepfake detection have employed sophisticated deep learning models to analyze forged media by identifying manipulation traces and extracting distinctive features.

Li et al. (2018) introduced an anomaly detection technique that leveraged the absence of human eye blinking in GAN-generated videos as a distinguishing factor for deepfake identification. Sabir et al. proposed a spatio-temporal R-CNN framework, which effectively captured inconsistencies in manipulated videos by extracting both temporal and spatial features. However, this approach proved computationally intensive and was constrained to video-based detection. Guera and Delp further refined this methodology by integrating Convolutional Neural Networks (CNNs) for feature extraction with Recurrent Neural Networks (RNNs) to model temporal dependencies, improving video-based detection at the cost of increased computational overhead.

Wang et al. (2020) focused on static image detection by identifying artifacts commonly found in CNN-generated images. Utilizing a ResNet-50 architecture pre-trained on ImageNet, their approach demonstrated high detection accuracy but struggled with generalizing across varied image datasets. Yang et al. developed a method that analyzed head pose inconsistencies in deepfake videos using Support Vector Machines (SVMs). While this approach successfully detected unnatural movements, its effectiveness was hampered by limited dataset diversity and errors in pose estimation. Chen and Yang proposed a CNN-based manipulated face detector designed to identify tampered faces under varying conditions such as lighting and pose, yet it faced difficulties in achieving robustness across different datasets.

Durall et al. introduced a novel detection approach by shifting from spatial domain analysis to frequency domain analysis. By employing the power spectrum as a forensic feature, they achieved significant success in detecting deepfakes, particularly in low-resolution media. However, their method was less effective under low-light conditions. These methodologies, evaluated on well-known datasets like FaceForensics++, CelebA, and ForenSynths, underscore the rapid progress in deepfake detection while highlighting persistent challenges in scalability, robustness, and computational efficiency.

Recent studies have expanded deepfake detection methodologies with innovative techniques and diverse datasets. Enhancing Authenticity Verification with Transfer Learning and Ensemble Techniques in Facial Feature-Based Deepfake Detection (Nadeem Qazi & Iftikhar Ahmed, 2024) employed Transfer Learning with DenseNet201, ResNet152V2, and Xception, integrated with a Deep Artificial Neural Network (DANN). Their model achieved 98% accuracy on the FaceForensics++ dataset containing 1000 videos but suffered from high computational complexity. Enhancing Deepfake Detection: An Ensemble Deep Learning Approach for Efficient Attribute Manipulation Identification (Sudharsana P P et al., 2024) utilized InceptionV3 and EfficientNetB0, demonstrating 93% accuracy on the DFDC dataset (800 videos). Despite its efficiency, the model faced computational complexity challenges. Generalizable Deepfake Detection with Phase-Based Motion Analysis (Ekta Prashnani et al., 2024) introduced the PhaseForensics method, achieving 92.4% AUC across DFDC, CelebDFv2, and FaceForensics++. The method relied heavily on detecting motion in facial regions, limiting its applicability in scenarios lacking significant motion. An Improved Dense CNN Architecture for Deepfake Image Detection (Yogesh Patel et al., 2023) developed a Deep-CNN model that attained 97.2% accuracy on CelebA and FFHQ datasets (5000 images each). However, it was restricted to image-based detection. Exposing Deepfake Face Forgeries With Guided Residuals (Zhiqing Guo et al., 2023) proposed an Adaptive Fusion-based Guided Residuals Network (AdapGRnet), achieving 96.52% accuracy on the

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 3, pp 148-154, March-2025

https://doi.org/10.47001/IRJIET/2025.903018

Hybrid Fake Face (HFF) dataset. This approach struggled with compressed, low-quality video frames. Generalization of Forgery Detection With Meta Deepfake Detection Model (Van-Nhan Tran et al., 2023) introduced a Meta Deepfake Detection model (MDD), yielding an 86.1% accuracy on DFDC, CelebDFv2, and FaceForensics++. However, its low accuracy limited its generalization capabilities. Unsupervised Learning-Based Framework for Deepfake Video Detection (Li Zhang et al., 2023) applied an unsupervised learning system on a custom dataset of 200 real and 100 deepfake videos, achieving 93.1% accuracy, albeit with high computational complexity. An Efficient Deep Video Model for Deepfake Detection (Ruipeng Sun et al., 2022) leveraged Sequential-Parallel Networks (SPNet), reaching 91.13% accuracy on DFDC, though it remained computationally expensive. Deepfake Video Detection Based on Convolutional Neural Networks (Sarah Riyadh Adnan et al., 2022) developed a CNN model using ResNet-50, obtaining 98% accuracy on DFDC, but was limited by a small dataset. Deepfake Video Detection Using Spatiotemporal Convolutional Network and Photo Response Non-Uniformity (Sio Jurnalis Pipin et al., 2022) employed a Spatiotemporal Convolutional Network (SCN) combined with PRNU analysis, achieving 97.89% accuracy across FaceForensics, CelebDF, and DFDC. The model's high computational demand posed practical challenges. Detecting Manipulated Facial Videos: A Time Series Solution (Zhewei Zhang et al., 2021) utilized a Bi-LSTM time series model, achieving 87.9% accuracy on FaceForensics++ but suffered from low robustness to facial variations. Detection of Fake and Fraudulent Faces via Neural Memory Networks (Tharindu Fernando et al., 2021) introduced a Hierarchical Attention Memory Network (HAMN) that yielded 75% accuracy on FaceForensics++, but dataset quality issues impacted its generalization. Eff-YNet: A Dual Task Network for DeepFake Detection and Segmentation (Eric Tjon et al., 2021) combined EfficientNet and U-Net in the Eff-YNet model, achieving 80% accuracy on the Deepfake Detection Challenge dataset, but was computationally intensive.

Despite these advancements, several challenges persist. Most models perform well on specific datasets like FaceForensics++ or Celeb-A but struggle with unseen data. Deep learning-based models, such as CNNs and RNNs, demand significant processing power, hindering real-time application feasibility. Many detection algorithms falter under diverse conditions, affecting robustness. Deepfake techniques are evolving, introducing more realistic manipulations that are difficult to detect. The continuous advancement of deepfake generation techniques necessitates constant updates in detection methodologies.

III. PROPOSED SYSTEM

In this work, we propose a novel system for deepfake detection using a Convolutional Neural Network (CNN) in combination with a Recurrent Neural Network (RNN), specifically leveraging the EfficientNetB0 model for feature extraction and an LSTM network for classification on the Celeb-DF dataset. The system utilizes EfficientNetB0, a state-of-the-art pre-trained model, as the backbone for extracting spatial features from video frames, followed by an LSTM network to capture temporal dependencies and classify real vs. deepfake videos. The primary goal is to classify the authenticity of video frames while optimizing both efficiency and accuracy in detecting manipulated content.

3.1 System Overview

The proposed deepfake detection system is structured in two main phases:

1. Data Preprocessing and Augmentation

Video Frame Extraction: The first step involves extracting individual frames from video files. Since deepfake videos often show subtle inconsistencies in each frame, extracting frames is crucial for isolating these artifacts and analyzing them.

Data Augmentation: Given the limited availability of labeled deepfake data, data augmentation techniques are employed to artificially expand the training dataset. This includes rotation, shifting, shearing, zooming, and flipping, which helps the model generalize better and prevents overfitting.

2. Feature Extraction and Classification using EfficientNetB0 and LSTM

EfficientNetB0 Model: The EfficientNetB0 model, pre-trained on the ImageNet dataset, serves as the backbone of the proposed system. EfficientNetB0 is known for its efficiency in terms of both model size and performance, providing an excellent trade-off between accuracy and computational resources.

Fine-tuning: We use EfficientNetB0 without its top (classification) layers. The feature extraction part of the model is frozen to retain the pre-trained weights, while the top layers are customized for deepfake detection.

Global Average Pooling (GAP): After extracting features using EfficientNetB0, a Global Average Pooling (GAP) layer is applied to convert the 2D feature maps into a 1D feature vector. GAP helps reduce dimensionality while retaining essential spatial information.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 3, pp 148-154, March-2025

https://doi.org/10.47001/IRJIET/2025.903018

LSTM Network: The extracted feature vectors from multiple frames are sequentially fed into an LSTM network to capture temporal dependencies. The LSTM processes these sequential features and learns patterns across frames.

Fully Connected (Dense) Layer: The LSTM output is passed through a dense layer with a sigmoid activation function, producing a probability score that classifies the video as either real or deepfake.

3.2 Training and Evaluation

Loss Function and Optimizer: The model is trained using binary crossentropy as the loss function, suitable for a binary classification task. The Adam optimizer is employed with a learning rate of 0.0001 to fine-tune the model effectively.

Model Evaluation: The model's performance is evaluated using accuracy as the primary metric to assess its capability in distinguishing real from fake content.

3.3 Detailed System Architecture

Input Layer: The system takes input video frames resized to 128x128x3 dimensions (height, width, and color channels).

3.4 EfficientNetB0 Backbone

Stem: Initial convolution followed by batch normalization and Swish activation.

MBConv Blocks with Squeeze-and-Excitation (SE): These blocks involve depthwise separable convolutions and utilize the Squeeze-and-Excitation technique to adaptively recalibrate channel-wise feature responses.

Global Average Pooling (GAP): Converts the final feature maps into a 1D vector of features that represent the image content.

3.5 LSTM Network

Sequential feature vectors are processed to capture temporal dependencies.

The LSTM network learns patterns across frames for robust classification.

3.6 Fully Connected (Dense) Layer

After feature extraction and sequence modeling, the model uses a dense layer with sigmoid activation for binary classification (real vs deepfake).

The final output is a single scalar representing the probability of a video being real or deepfake.

3.7 Advantages of the Proposed System:

Efficiency: By using EfficientNetB0 as the backbone, the model benefits from a highly optimized architecture, achieving superior accuracy with fewer parameters compared to other deep learning models.

Temporal Feature Learning: The integration of LSTM enhances deepfake detection by analyzing sequential frame patterns rather than treating frames independently.

Data Augmentation: The inclusion of data augmentation techniques ensures better generalization by increasing the diversity of the training data.

Pre-trained Weights: Using pre-trained weights from ImageNet helps the system leverage learned features that are useful in detecting even subtle discrepancies in deepfake content.

Scalability: The system is scalable to handle large datasets, and EfficientNetB0's lightweight nature makes it suitable for real-time applications.

3.8 Experimental Setup:

Dataset: The Celeb-DF dataset is used, which contains deepfake videos of celebrities. This dataset includes both real and fake video frames, allowing for binary classification.

3.9 Evaluation Metric:

Accuracy: The percentage of correctly classified videos.This system aims to provide an efficient and accurate deepfake detection method by combining spatial feature extraction using EfficientNetB0 with temporal modeling using an LSTM network.

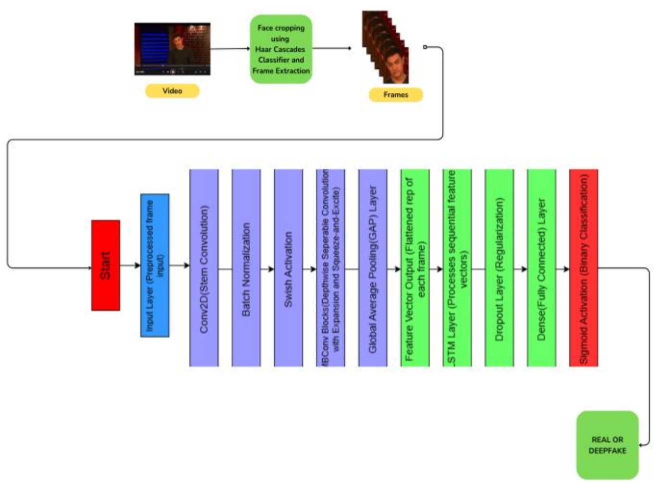


Figure 3.1: Architecture of the Proposed System

Figure shows the architecture of the proposed system.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 3, pp 148-154, March-2025

https://doi.org/10.47001/IRJIET/2025.903018

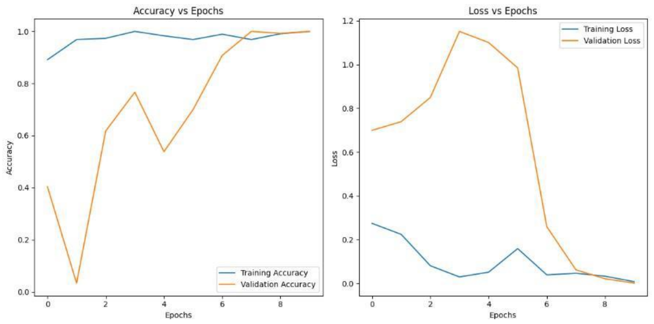
IV. RESULTS

The proposed deepfake detection system, leveraging EfficientNetB0 for feature extraction and an LSTM network for classification, was trained and evaluated on the Celeb-DF dataset. The model was assessed using accuracy as the primary evaluation metric.

Training Accuracy: The model achieved an accuracy of 98% on the training set, demonstrating effective learning of spatial and temporal features from real and deepfake videos.

Testing Accuracy: On the test set, the model attained an accuracy of 95%, indicating strong generalization performance in distinguishing manipulated content from authentic videos.

The results highlight the effectiveness of EfficientNetB0 in capturing spatial details and LSTM's ability to model temporal dependencies in video sequences. The high accuracy suggests that the model successfully detects subtle inconsistencies in deepfake videos while maintaining computational efficiency.



(a) (b)

Figure 4.1: (a) Accuracy vs Epochs graph, (b) Loss vs Epochs graph

V. CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this work, we proposed a deepfake detection system that combines EfficientNetB0 for feature extraction and an LSTM network for classification to analyze video frames and detect manipulated content. By leveraging a pre-trained EfficientNetB0 model, the system effectively captures spatial features, while the LSTM component models temporal dependencies across frames, improving classification accuracy. The experimental results demonstrate that the model achieves high accuracy (98% on training and 95% on testing) on the Celeb-DF dataset, highlighting its capability to distinguish between real and deepfake videos.

Despite these promising results, challenges remain in ensuring the model's robustness against diverse deepfake generation techniques. The system's performance may vary across datasets, emphasizing the need for broader generalization.

5.2 Future Work

To further enhance the effectiveness and reliability of the deepfake detection system, we plan to explore the following directions:

Improving Generalization: Enhancing the model's ability to detect deepfakes across different datasets and unseen manipulation techniques by incorporating additional augmentation strategies and domain adaptation methods.

Expanding the Dataset: Training the model on larger and more diverse datasets, including various deepfake benchmarks (such as DFDC, FaceForensics++, and WildDeepfake) to improve robustness across different video sources and deepfake generation methods.

Optimizing Model Efficiency: Exploring lightweight architectures and model compression techniques to improve inference speed while maintaining high accuracy, making the system more suitable for real-time applications.

Enhancing Temporal Analysis: Investigating alternative recurrent architectures, such as Transformer-based models, to further improve the system's capability in learning long-term dependencies in video sequences.

Adversarial Training: Strengthening model robustness by incorporating adversarial examples in training, making the system more resilient against evolving deepfake generation methods.

By addressing these aspects, we aim to develop a more generalized and scalable deepfake detection system, capable of detecting manipulated content more effectively across various real-world scenarios.

REFERENCES

[1] Agarwal S, Farid H, Fried O, and Agrawala M (2020) Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA.

[2] Amerini I, Galteri L, Caldelli R, and Del Bimbo A (2019) Deepfake Video Detection through Optical Flow Based CNN. In: IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South).

[3] Ding X, Raziei Z, Larson E.C. et al. (2020).

ISSN (online): 2581-3048
Volume 9, Issue 3, pp 148-154, March-2025
https://doi.org/10.47001/IRJIET/2025.903018

[4] Dolhansky B, Howes R, Pflaum B, Baram N, Ferrer C (2019) The Deepfake Detection Challenge (DFDC) Preview Dataset. arXiv:1910:08854.

[5] Fei J, Xia Z, Yu P, Xiao F (2020) Exposing AI-generated videos with motion magnification. Multimedia Tools Appl. 80(20), 30789–30802.

[6] Fernando T., Fookes C., Denman S. and Sridharan S. (2021) Detection of Fake and Fraudulent Faces via Neural Memory Networks. In: IEEE Transactions on Information Forensics and Security, vol. 16, pp. 1973-1988, doi: 10.1109/TIFS.2020.3047768.

[7] Hashmi M.F., Ashish B.K.K., Keskar A.G., Bokde N.D., Yoon J.H., Geem Z.W. (2020) An Exploratory Analysis on Visual Counterfeits Using Conv-LSTM Hybrid Architecture. IEEE Access 8, 101293–101308.

[8] Hsu C.-C., Zhuang Y-Xiu., Lee C.-Y. (2020) Deep fake image detection based on pairwise learning. Applied Sciences 10(1), 370.

[9] Huy H. Nguyen, Fuming Fang, Junichi Yamagishi, Isao Echizen (2019) Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos. In: IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS).

[10] Komal Chugh, Parul Gupta, Abhinav Dhall, and Ramanathan Subramanian (2018) Not made for each other – Audio-Visual Dissonance-based Deepfake Detection and Localization. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05.

[11] Li, Haodong; Li, Bin; Tan, Shunquan; Huang, Jiwu: Identification of deep network generated images using disparities in color components. Signal Process. (2020).

[12] Li, Y.; Chang, M.; Lyu, S. (2018) In Ictu Oculi Exposing AI Created Fake Videos by Detecting Eye Blinking. In: IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, Hong Kong.

[13] Li Y, Lyu S (2018) Exposing Deep Fake Videos By Detecting Face Warping Artifacts. In: IEEE Conference Computer. Vision Pattern Recognition.

[14] L. Zhang, T. Qiao, M. Xu, N. Zheng and S. Xie (2023) Unsupervised Learning-Based Framework for Deepfake Video Detection. In: IEEE Transactions on Multimedia, vol. 25, pp. 4785-4799, doi: 10.1109/TMM.2022.3182509.

[15] Matthews T.F., Cootes J.A., Bangham S.C., Harvey R. (2002) Extraction of visual features for lip reading. IEEE Trans. Pattern Anal. Mach. Intell. 24(2), 198–213.

[16] Matern F, Riess C, Stamminger M (2019) Exploiting Visual Artifacts to Expose Deep fakes and Face Manipulations. In: IEEE Winter Applications of Computer Vision Workshops (WACVW), Waikoloa Village, HI, USA.

[17] Minh Dang L., Hassan S.I., Im S., Moon H. (2019) Face image manipulation detection based on a convolutional neural network. Expert Syst. Appl. 129, 156–168.

[18] Mingxing Tan, Quoc V. Le (2019) EfficientNet: Re-thinking Model Scaling for Convolutional Neural Networks. Cornell University.

[19] Montserrat D.M. et al. (2020) Deepfakes Detection with Automatic Face Weighting In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA.

[20] N. Qazi and I. Ahmed (2024) Enhancing Authenticity Verification with Transfer Learning and Ensemble Techniques in Facial Feature-Based Deepfake Detection. In: 2024 14th International Conference on Pattern Recognition Systems (ICPRS), London, United Kingdom, pp. 1-6, doi: 10.1109/ICPRS62101.2024.10677831.

[21] Prashnani E., Goebel M. and Manjunath B. S. (2025) Generalizable Deepfake Detection with Phase-Based Motion Analysis. In: IEEE Transactions on Image Processing, vol. 34, pp. 100-112, doi: 10.1109/TIP.2024.3441821.

[22] Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M. (2018) FaceForensics: A large-scale video dataset for forgery detection in human faces.

[23] Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Niessner M. (2019) FaceForensics++: learning to detect manipulated facial images. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV).

[24] Sabir E, Cheng J, Jaiswal A, Almageed W.A., Masi I, Natarajan P. (2019) Recurrent Convolutional Strategies for Face Manipulation Detection in Videos. IEEE Conf. Comput. Vision Pattern Recogn. 3, 80–87.

[25] S. J. Pipin, R. Purba and M. F. Pasha (2022) Deepfake Video Detection Using Spatiotemporal Convolutional Network and Photo Response Non Uniformity. In: 2022 IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM), Laguboti, North Sumatra, Indonesia, pp. 1-6, doi: 10.1109/ICOSNIKOM56551.2022.10034890.

[26] S. R. Adnan and H. A. Abdulbaqi (2022) Deepfake Video Detection Based on Convolutional Neural Networks. In: 2022 International Conference on Data Science and Intelligent Computing (ICDSIC), Karbala, Iraq, pp. 65-69, doi: 10.1109/ICDSIC56987.2022.10075830.

[27] Torfi A, Iranmanesh S.M., Nasrabadi N., Dawson J. (2017) 3D Convolutional Neural Networks for Cross

ISSN (online): 2581-3048
Volume 9, Issue 3, pp 148-154, March-2025
https://doi.org/10.47001/IRJIET/2025.903018

Audio-Visual Matching Recognition. IEEE Access 5, 22081–22091.

[28] T. Fernando, C. Fookes, S. Denman and S. Sridharan (2021) Detection of Fake and Fraudulent Faces via Neural Memory Networks. In: IEEE Transactions on Information Forensics and Security, vol. 16, pp. 1973-1988, doi: 10.1109/TIFS.2020.3047768.

[29] Tran V. -N., Kwon S. -G., Lee S. -H., Le H. -S. and Kwon K. -R. (2023) Generalization of Forgery Detection With Meta Deepfake Detection Model. In: IEEE Access, vol. 11, pp. 535-546, doi: 10.1109/ACCESS.2022.3232290.

[30] Y. Patel et al. (2023) An Improved Dense CNN Architecture for Deepfake Image Detection. In: IEEE Access, vol. 11, pp. 22081-22095, doi: 10.1109/ACCESS.2023.3251417.

[31] Z. Guo, G. Yang, J. Chen and X. Sun (2023) Exposing Deepfake Face Forgeries With Guided Residuals. In: IEEE Transactions on Multimedia, vol. 25, pp. 8458-8470, doi: 10.1109/TMM.2023.3237169.

[32] Z. Zhang, C. Mal, B. Ding and M. Gao (2020) Detecting Manipulated Facial Videos: A Time Series Solution. In: 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, pp. 2817-2823, doi: 10.1109/ICPR48806.2021.9412610.

Citation of this Article:

Satwika.M, Neha.K, Pranavya.A, Rishika.K, Siva Sankar Namani. (2025). Video Deepfake Detection Using EfficientNet and LSTM. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(3), 148-154. Article DOI https://doi.org/10.47001/IRJIET/2025.903018

****

