

Enhanced Deepfake Detection Using Deep Learning

Tambe Vaishnavi Ramhari¹, Dr. Hemantkumar B. Jadhav, Prof. Jagruti R. Mahajan⁴, Prof. Pragati B. Chandane, Dr. Pradeep M. Patil⁵

¹PG Scholar, Adsul's Technical Campus, Ahilyanagar

²Professor, Adsul's Technical Campus, Ahilyanagar

³Professor, Adsul's Technical Campus, Ahilyanagar

⁴Assistant Professor, Adsul's Technical Campus, Ahilyanagar

⁵Principal, Adsul's Technical Campus, Ahilyanagar

Abstract: This study explores the growing challenge of deepfake videos and evaluates the effectiveness of various deep learning algorithms in detecting them. Models such as InceptionResNetV2, VGG19 CNN, Xception, InceptionV3, EfficientNetB1, DenseNet121, Hybrid architectures, LSTM, ResNeXt-LSTM, and MRIGAN

were tested to identify the most reliable approaches. The results highlight that Xception achieved the highest accuracy at 99.32%, while InceptionResNetV2 and DenseNet121 also demonstrated strong performance with accuracies above 99%. In contrast, models like VGG19 and LSTM produced comparatively lower results, indicating the need for further optimization. Overall, the findings emphasize the steady progress in detection techniques, which is crucial for countering the harmful effects of deepfakes, including misinformation and violations of privacy. The findings illustrate the rapid progress in detection techniques, reflecting the growing sophistication of deepfake generation methods. higher resilience and accuracy.

Ultimately, the work contributes to the development of reliable tools that can help mitigate the societal risks posed by deepfakes and strengthen trust in digital media environments.

Keywords: Deep Learning, Fake Detection, InceptionResnetV2, VGG19, CNN, and Xception

I. INTRODUCTION

Deepfakes, created through advanced machine learning techniques, have emerged as a powerful and increasingly popular way to generate highly realistic fake videos and images, often by swapping one person's face with another. While the technology demonstrates impressive capabilities, it has also raised serious concerns about misuse, as malicious actors may exploit deepfakes to spread misinformation, manipulate opinions, or violate privacy. This growing threat has prompted researchers and technologists to develop deep learning-based detection systems aimed at safeguarding digital media integrity.

II. LITERATURE REVIEW

In this review, we examine a range of deepfake detection strategies, with particular emphasis on deep learning approaches and the challenges of building resilient detection systems. The creation and identification of deepfakes using advanced neural networks has become a major focus in recent research. Khalil and Maged [1] highlight how deep learning can uncover subtle, often imperceptible changes in both images and videos, such as unnatural facial expressions, irregular movements, or inconsistent lighting. Their work stresses the importance of designing models that remain effective across diverse scenarios.

Zhao et al. [2] introduce a self consistency method that leverages the natural continuity of facial structures in authentic videos. By analyzing temporal coherence rather than just local spatial artifacts, their anomaly detection approach outperforms conventional techniques.

III. DATASET DESCRIPTION

- Kaggle deepfake dataset
- Data augmentation strategies.
- Preprocessing techniques used

IV. PROPOSED METHODOLOGY

1. Dataset Collection

The system is designed to counter the challenges posed by deepfake technology by employing a combination of advanced deep learning models. Architectures such as

InceptionResNetV2, VGG19, CNN, Xception, InceptionV3, EfficientNetB1, DenseNet121, hybrid approaches, LSTM, ResNext LSTM, and MRI GAN are integrated to analyze multimedia content for signs of manipulation. Among these, Xception demonstrates the highest accuracy. By combining these algorithms, the system aims to maximize detection accuracy across both images and videos. The architecture is modular, allowing future expansion with additional models or enhancements. A key focus is on throughput control, ensuring that malicious content can be flagged and addressed quickly to reduce the spread of fake information and mitigate its social consequences.

1.1 Image Processing:

Preprocessing is a critical step in computer vision tasks. Two major techniques are applied:

1.2 Image resizing:

Ensures consistency across inputs and prevents computational overload.

1.3 Image data augmentation: Introduces variations such as rotations, zooms.

2. Data Visualization

Videos are segmented into frames to capture changes. Visualization methods are then used to represent frame data in statistical and graphical formats. Interactive visualization tools allow researchers to closely examine specific frames, annotate anomalies, and identify patterns that may indicate manipulation. This step enhances interpretability and supports the validation of detection results.

3. Algorithms

Inception-ResNetV2:

Inception-ResNetV2 is a deep convolutional neural network that combines the concepts of Inception modules and residual connections. By integrating these two architectures, it aims to enhance both accuracy and computational efficiency. The residual connections help in training deeper networks by reducing issues like vanishing gradients, while the Inception modules improve feature extraction at multiple scales.

VGG19:

VGG19 is a 19-layer deep convolutional neural network developed by the Visual Geometry Group at the University of Oxford, introduced by K. Simonyan and A. Zisserman. It is widely used for image classification tasks due to its simple and uniform architecture, which relies on stacked convolutional layers. Despite its simplicity, VGG19 has proven to be highly effective in extracting visual features and delivering strong performance in image recognition tasks.

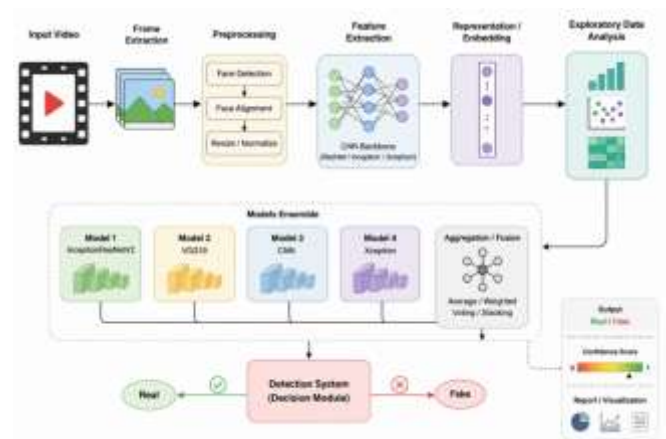
CNN (Convolutional Neural Network):

A Convolutional Neural Network (CNN) is a type of deep learning model commonly used for image recognition and classification tasks. It consists of multiple layers, including convolutional layers and pooling layers, followed by fully connected layers. These layers work together to automatically learn and extract important features from images, which are then used for high-level representation and final classification.

Xception:

Xception is an advanced deep learning architecture inspired by the Inception model, but it replaces standard convolutions with depthwise separable convolutions. This design efficiently captures the relationship between spatial and channel-wise information in feature maps. As a result, it improves model performance while also reducing computational cost, making it both accurate and efficient for image-related tasks.

V. SYSTEM ARCHITECTURE



The proposed deepfake detection framework follows a structured multi-stage pipeline designed to identify manipulated video content with high reliability. Initially, video datasets are collected and converted into individual frames to enable detailed frame-wise analysis. Each extracted frame undergoes preprocessing operations such as resizing, normalization, noise reduction, and face region extraction to improve image quality and ensure consistency across the dataset.

After preprocessing, Exploratory Data Analysis (EDA) and visualization techniques are applied to study frame distributions, pixel intensity variations, facial patterns, and inconsistencies between authentic and manipulated images. Data augmentation and image data generators are further utilized to increase dataset diversity and improve the generalization capability of the models during training.

The processed frames are then passed through multiple deep learning architectures, including InceptionResNetV2, VGG19, Convolutional Neural Networks, and Xception. Each model independently learns discriminative facial features, texture inconsistencies, compression artifacts, and spatial abnormalities commonly present in deepfake media. Combining multiple architectures helps improve robustness and reduces the possibility of biased predictions from a single model.

Feature extraction layers identify subtle manipulations such as unnatural eye movements, blending artifacts, irregular facial boundaries, and inconsistencies in lighting or expressions. The outputs generated by the different models are integrated using ensemble learning techniques to

achieve improved detection performance and stability.

Finally, the system classifies the input content as either real or fake.

VI. RESULTS AND PERFORMANCE EVALUATION

Accuracy:

Accuracy represents the overall effectiveness of the model in correctly identifying both real and fake media samples. It indicates how well the system performs in distinguishing manipulated content from authentic content across the complete dataset. Higher accuracy values reflect better classification capability and improved reliability of the detection framework.

Accuracy Formula

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- **TP** = True Positive
- **TN** = True Negative
- **FP** = False Positive
- **FN** = False Negative

Recall:

Recall measures the ability of the model to successfully identify all relevant manipulated samples within the dataset. It evaluates how effectively the system detects actual deepfake instances without missing them. A high recall

value indicates that the model can recognize most fake media samples accurately.

$$Recall = \frac{TP}{TP + FN}$$

Precision:

Precision evaluates the proportion of correctly predicted fake samples among all samples identified as fake by the model. It reflects the reliability of positive predictions and helps determine whether the system minimizes false detections. Higher precision indicates fewer incorrect classifications of genuine media as fake.

VII. CONCLUSION

This study demonstrates the growing effectiveness of deep learning techniques in detecting deepfake content through advanced architectures such as Xception and InceptionResNetV2. These models achieved strong performance in terms of accuracy, precision, and recall when differentiating authentic media from manipulated content. Although these approaches require substantial computational resources, they provide a balanced combination of reliability and performance for complex deepfake detection tasks. Simpler convolutional neural network (CNN) models can serve as a foundational approach, but further optimization is necessary before they can be effectively deployed in real-world environments.

The findings suggest that advanced detection models are especially valuable in sensitive domains such as digital journalism, cybersecurity, identity protection, and social media content verification, where manipulated media can create serious social and security concerns. The research also emphasizes the broader capability of deep learning systems to address challenges related to misinformation, fabricated visual content, and image-based defamation.

Future work should focus on enhancing the robustness and adaptability of detection frameworks to keep pace with increasingly sophisticated deepfake generation methods. Further investigation into modern techniques, including reinforcement learning and explainable artificial intelligence, may improve both detection efficiency and model transparency. Collaboration among academic researchers, industry professionals, and policymakers will also be essential for establishing ethical standards and effective protocols for managing deepfake-related risks. In addition, future studies should prioritize real-time detection mechanisms and interpretable model designs that improve user confidence and practical usability in real-world applications.

VIII. REFERENCES

- [1] International Conference on Advancements in Smart, Secure and Intelligent Computing, "Enhanced Deep Fake Detection Using Deep Learning" 2025.
- [2] H. A. Khalil and S. A. Maged, "Deepfakes creation and detection using deep learning," in 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). IEEE, 2021, pp. 1–4.

[3] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning selfconsistency for deepfake detection," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 15 023–15 033.

[3] Z. Chen and H. Yang, "Attentive semantic exploring for manipulated face detection," in ICASSP 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 1985–1989.

[4] Z. Guo, L. Hu, M. Xia, and G. Yang, "Blind detection of glow-based facial forgery," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 76877710, 2021.

[5] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "Wilddeepfake: A challenging real-world dataset for deepfake detection," in Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 2382– 2390.

[6] S. Fernandes, S. Raj, R. Ewetz, J. S. Pannu, S. K. Jha, E. Ortiz, I. Vintila, and M. Salter, "Detecting deepfake videos using attributionbased confidence metric," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 308–309.

[7] U. A. Ciftci, I. Demir, and L. Yin, "Fakecatcher: Detection of synthetic portrait videos using biological signals," *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[8] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "Cnn-generated images are surprisingly easy to spot... for now," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 8695–8704.