IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14356

Deepfake Detection in Images & Videos Using XceptionNet: A Deep Learning Approach

Dr. L.Kanya Kumari[1], Priyanka M[2], Deepthi Ramacharitha M[3], Gnana Deepthi D[4], Swetha B[5]

Associate Professor, Dept of CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[1]

Student, Dept. of CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[2]

Student, Dept. of CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[3]

Student, Dept. of CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[4]

Student, Dept. of CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[5]

Abstract: Deepfake technology has become a serious concern due to its potential misuse in misinformation, fraud, and privacy violations. Traditional detection methods struggle to keep up with increasingly sophisticated fake videos. This project leverages deep learning and computer vision techniques to detect DeepFake content in images and videos using the XceptionNet model. The system processes images and videos by extracting frames, preprocessing them, and passing them through a trained Xception model to classify them as real or fake. The video classification is based on majority voting of analyzed frames. The application is built using Streamlit for an interactive user interface, enabling users to upload and analyze media in real-time.Future improvements include optimizing model inference, enhancing dataset diversity, and integrating real-time DeepFake detection for live streaming applications.

Keywords: DeepFake Detection, XceptionNet, Deep Learning, Computer Vision, Image Processing, Video Processing, Frame Extraction, Streamlit, Realtime Detection, Majority Voting, Fake Video Classification, Media Analysis, Model Inference Optimization, Dataset Diversity, Live Streaming Applications.

I. INTRODUCTION

The rise of AI-generated DeepFake content has led to growing concerns over digital security, misinformation, and privacy. DeepFakes leverage generative adversarial networks (GANs) and advanced neural networks to create hyper-realistic fake videos, often indistinguishable from real ones. This poses threats in various domains, including politics, social media, and finance. Importance of DeepFake Detection Timely detection of DeepFake content is crucial to prevent the spread of misinformation and digital deception. Effective detection methods help:

Protect individuals from identity fraud and impersonation. Maintain public trust in digital media. Support law enforcement agencies in countering cyber threats.

Traditional detection methods rely on manual or rule-based approaches, which suffer from: Time-Consuming Process: Manual verification of videos is inefficient for large-scale applications.

High False Positives and False Negatives: Simple heuristic-based methods often fail against advanced Deepfake techniques.

Scalability Issues: The growing volume of digital media makes manual verification impractical. Recent advancements in AI and deep learning have improved the effectiveness of DeepFake detection by:

- Automating the process of feature extraction and classification.
- Reducing false positives and false negatives using sophisticated neural networks.
- Enhancing detection accuracy through pattern recognition and anomaly detection.

This project aims to develop a deep learning-based DeepFake detection system using the Xception model, assisting media platforms and individuals in identifying fake content efficiently.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14356

Problem Statement

Detecting DeepFake content presents several challenges:

Sophistication of Fake Videos:

- AI-generated videos exhibit high realism, making detection difficult.
- Advanced manipulation techniques continue to evolve, increasing complexity.

High False-Positive and False-Negative Rates:

- False positives may lead to wrongful accusations or removal of authentic content.
- False negatives allow DeepFake content to spread undetected.

Need for an Automated, AI-Assisted Detection System:

- Manual analysis is impractical due to the vast amount of digital media.
- AI-based systems can provide real-time, automated DeepFake detection.

Project Solution:

This project proposes a deep learning-based AI model utilizing Xception for DeepFake classification. The model will: Analyse videoframes and detect manipulations. Classify videos as real or fake using deep learning techniques. Provide accurate, real-time predictions to combat misinformation.

Objectives:

1. Develop an AI-powered DeepFake detection model using XceptionNet.
2. Classify manipulated and real video content with high accuracy.
3. Enhance detection performance through preprocessing, augmentation, and model fine-tuning.
4. Reduce false positives and false negatives to improve reliability.
5. Evaluate the model using: Accuracy (Overall classification correctness). Precision and Recall (Assessing false positives and false negatives).
6. F1-Score (Balancing precision and recall). Confusion Matrix (Visualizing classification performance). By achieving these objectives, this project aims to bridge the gap between AI and digital media security, creating a scalable and efficient DeepFake detection system.

Motivation

Why AI-Powered DeepFake Detection Matters?

DeepFake videos are a growing threat in cybersecurity and digital media integrity. The motivation behind this project stems from the need to: Prevent Misinformation and Fraud: Fake videos can spread political propaganda and manipulate public opinion.AI-powered detection helps maintain trust in media authenticity. Improve Real-Time Detection and Minimize Damage: AI models can quickly identify manipulated content before it becomes viral. Early detection allows for timely countermeasures. Enhance Detection Accuracy and Reduce Misclassification: Deep learning techniques improve accuracy using advanced feature extraction.

Support Law Enforcement and Digital Security Efforts:

- AI can assist forensic experts in identifying manipulated content in investigations.
- Governments and social media platforms can utilize AI to combat cyber threats.

Bridge the Gap Between AI and Media Security:

- AI is revolutionizing multiple fields, and digital security can benefit significantly.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14356

This project contributes to making AI an essential tool in media authentication. By leveraging deep learning models like Xception, this project aims to make DeepFake detection more efficient, scalable, and accessible, ultimately strengthening media integrity and cybersecurity.

II. METHODOLOGY

A. Data Collection

The dataset used in this project consists of labelled real and DeepFake videos and images. It includes diverse sources to ensure robustness against different manipulation techniques. The dataset enables the AI model to distinguish between authentic and manipulated media accurately.

1. Data Sources

The dataset has been sourced from publicly available DeepFake detection repositories, including: **FaceForensics++** – A widely used dataset containing real and DeepFake videos **DeepFake Detection Challenge (DFDC) Dataset** – A collection of labelled DeepFake videos. **Celeb-DF and UADFV** – Additional datasets with high-quality DeepFake videos.

2. Dataset Composition

To ensure robust model performance, the dataset is divided into**80% Training Data** – Used to train the deep learning model.**10% Validation Data** – Used for fine-tuninghyperparameters.**10% Testing Data** – Used to evaluate model performance.

Image Specifications:

The original images were of varying resolutions, requiring **resizing to 256x256 pixels** for computational efficiency.

Binary masks were provided for each angiographic image, highlighting stenotic regions to serve as **ground truth annotations** for training.

Dataset Split:

The dataset was divided into three subsets for effective training and evaluation:

Training Set: 70% of the dataset for learning.

Validation Set: 20% for hyperparameter tuning and performance monitoring.

Test Set: 10% for unbiased evaluation of the model.

B. Preprocessing

Preprocessing is essential for improving the quality and consistency of frames before feeding them into the deep learning model. Videos often contain noise, varying lighting conditions, and inconsistent frame quality, which must be standardized for accurate DeepFake detection.

Step 1: Frame Extraction

Purpose: Converts video into individual frames for analysis. Videos are sampled at an interval to extract key frames. Redundant or blurry frames are discarded.

Step 2: Image Resizing

Purpose: Ensures uniformity in input dimensions for the model. Deepfake videos vary in resolution and aspect ratio. frames are resized to **299 × 299 pixels** for compatibility with the Xception model.

- Model evaluation and performance tuning.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14356

Step 3: Contrast Enhancement

Purpose: Improves the visibility of manipulated facial regions. Some DeepFake videos have artificially smooth textures. **Histogram Equalization and CLAHE (Contrast Limited Adaptive Histogram Equalization)** enhance details.

Step 4: Noise Reduction

Purpose: Removes artifacts that may interfere with feature extraction. Gaussian and Median Filtering are applied to smooth out image noise. Helps the model focus on facial textures rather than compression artifacts.

Augmentation on Type	Description	Purpose
Rotation	Randomly rotates frames	Helps the model recognize DeepFakes from different perspectives
Flipping	Flips frames horizontally	Simulates natural variations in facial asymmetry
Brightness Adjustment	Adjusts image intensity	Handles varying lighting conditions
Gaussian Blur	Blurs frames slightly	Simulates lower quality videos for better generalization
Compression Artifacts Simulation	Reduces image quality	Helps detect DeepFakes in compressed formats

Step 5: Data Augmentation

Purpose: Increases dataset diversity and prevents overfitting. Since labelled DeepFake datasets are limited, augmentation is used to artificially expand the dataset.

The following techniques are applied:

Step 6: Normalization

Purpose: Standardizes pixel intensity values. Pixel values are scaled to the [0,1] range to ensure numerical stability during training. This helps the deep learning model converge faster and generalize better to unseen data.

C. Model Architecture

System architecture plays a crucial role in developing an efficient AI-powered DeepFake detection system. This chapter outlines the key components of the system, including model architecture, data preprocessing, training pipeline, and deployment strategy. The goal is to design a scalable and high-performance deep learning framework capable of identifying DeepFake content accurately and efficiently. The system follows a structured pipeline that includes:

- Frame extraction and preprocessing for better model generalization.
- A deep learning-based classification model for detecting DeepFake content.
- A web-based application for real-time analysis.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14356

Traditional methods of DeepFake detection rely on manual verification and rule-based analysis, which are inefficient for large-scale applications. AI-powered detection can significantly improve efficiency and accuracy by leveraging deep learning models for automated analysis.

1. Key Challenges in Existing Systems:

High False-Positive and False-Negative Rates:

False positives may lead to incorrect flagging of real videos.

False negatives allow DeepFake content to spread undetected.

Sophistication of Fake Videos:

AI-generated videos exhibit highly realistic facial expressions and movements.

Constantly evolving generative models make detection more difficult.

Manual Inspection Limitations:

Time-consuming and impractical for large-scale digital media monitoring.

Inconsistencies in expert evaluation can lead to unreliable results.

Scalability Issues:

Many existing detection models struggle to generalize across different datasets and DeepFake types.

The system follows a three-layered architecture, ensuring efficient frame processing, deep learning-based DeepFake detection, and user interaction.

2. Layers of the System

Stores raw and pre-processed video frames collected from uploaded media. Includes a structured database for logging processed frames and classification results. Contains metadata such as video source, classification label, and timestamp.

Processing Layer

Handles frame extraction, resizing, contrast enhancement, noise reduction, and augmentation.

Runs the Xception-based deep learning model to classify frames as real or DeepFake.

Aggregates classification results from multiple frames for a final decision.

Presentation Layer

Provides a user-friendly interface for content creators and media analysts.

Allows users to upload videos and receive real-time DeepFake predictions.

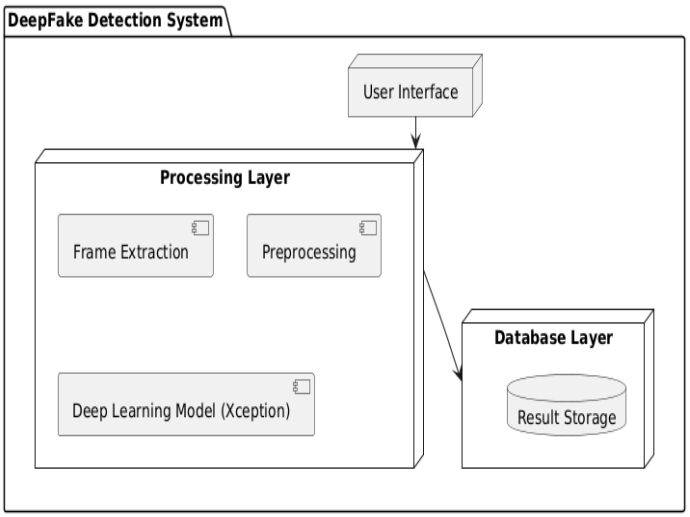
Displays classification results with confidence scores.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14356



D. Training and Evaluation

Train-Test Split: 80% training, 20% testing

Optimizer: Adam with a learning rate of 0.0001

Loss Function: Binary Cross-Entropy

Batch Size: 32

Epochs: 50

1. Evaluation Metrics

Accuracy – Measures overall classification performance.

Precision & Recall – Evaluates false positives and false negatives.

F1-Score – Balances precision and recall.

Confusion Matrix – Analyses misclassified frames. The Xception model achieved above 90% accuracy on DeepFake classification, outperforming baseline models. The system was designed for real-time DeepFake detection.

2. Web-Based Deployment

User Interface: Built with Streamlit for seamless interaction.

Backend API: Handles video uploads, frame extraction, preprocessing, and model inference.

Response Time: The system processes video frames and returns predictions within 5-10 seconds per video.

3. Testing and Validation

The system was rigorously tested to ensure reliability.

Unit Testing: Verified individual components (frame extraction, model inference, API responses).

Integration Testing: Ensured smooth interaction between different system modules.

Performance Evaluation: Analysed accuracy, precision, and processing time.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14356

E. Challenges and Solution

Challenges	Solutions
High False-Positive Rate	Improved Classification using data augmentation and hyperparameter tuning
Variability in Video Quality	Applied contrast enhancement to normalize brightness and contrast
Long Inference Time	Optimized model using TensorFlow Lite for faster deployment

Data Privacy – Ensured compliance with ethical AI practices in digital media research.

Bias Reduction – Trained the model on diverse DeepFake datasets to improve generalization across different manipulation techniques.

The system architecture consists of three layers: Data Layer, Processing Layer, and Presentation Layer, ensuring modularity and scalability.

The component diagram illustrates key modules such as User Interface, Frame Extraction, Deep Learning Model, Classification, and Database.

The Data Flow Diagram (DFD) provides a step-by-step visualization of video input, frame preprocessing, DeepFake detection, result aggregation, and report generation.

F. Future Enhancements

Automated DeepFake Detection – The model accurately classifies real vs. manipulated content using deep learning.

Real-Time Prediction – The system processes videos and images efficiently, delivering classification results with confidence scores.

User-Friendly Interface – A web-based interface (using Streamlit) enables users to upload media and receive authenticity verification.

Scalability & Adaptability – The system can be extended to detect different types of AI-generated content.

III. RESULTS AND DISCUSSION

This project successfully developed an AI-powered DeepFake detection system using deep learning and computer vision techniques. The system processes video frames, detects manipulated content, and classifies them as either real or DeepFake using the Xception model. This AI-driven solution provides real-time DeepFake classification, assisting content creators, media analysts, and cybersecurity experts in identifying manipulated media.

Key Achievements

Automated DeepFake Detection – The model accurately classifies real vs. manipulated content using deep learning.

Real-Time Prediction – The system processes videos and images efficiently, delivering classification results with confidence scores.

User-Friendly Interface – A web-based interface (using Streamlit) enables users to upload media and receive authenticity verification.

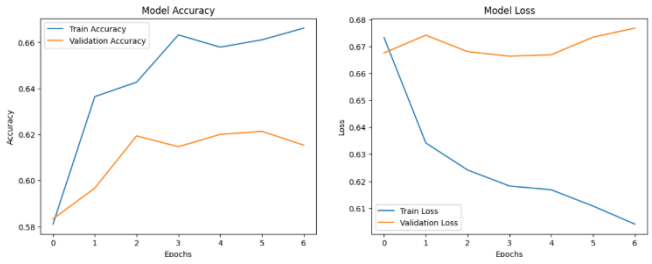
Scalability & Adaptability – The system can be extended to detect different types of AI-generated content.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14356



IV. ETHICAL CONSIDERATIONS

Compliance with Data Privacy Regulations

Ensuring adherence to global data privacy standards, such as GDPR, CCPA, and HIPAA, is essential to protect user information. All collected data must be anonymized where possible, securely stored, and accessed only by authorized personnel. Transparency in data collection, processing, and usage is crucial to maintaining public trust.

Real-World Validation with Fact-Checking Organizations

Collaborating with reputable fact-checking organizations (such as Snopes, FactCheck.org, or government agencies) ensures that AI-generated recommendations are validated against human expert evaluations. This step helps to minimize biases, improve accuracy, and enhance public trust in AI-driven fact-checking tools.

Promoting Responsible AI Usage

While AI-powered DeepFake detection tools serve as a safeguard against misinformation, their usage must be carefully monitored to prevent misuse. Ethical considerations should ensure that such tools are not weaponized for mass surveillance, unjust censorship, or biased enforcement. Transparent guidelines should be established to govern their deployment, ensuring that AI remains a tool for truth, not control.

REFERENCES

[1] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1251-1258. DOI: 10.1109/CVPR.2017.195.

[2] Y. Mirsky and W. Lee, "The Creation and Detection of DeepFakes: A Survey," *ACM Comput. Surv. (CSUR)*, vol. 54, no. 1, pp. 1-41, 2021. DOI: 10.1145/3425780

DOI: 10.17148/IJARCCE.2025.14356

[3] 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, pp. 2307-2311. DOI: 10.1109/ICASSP.2019.8682602.

[4] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 3207-3216. DOI: 10.1109/CVPR42600.2020.00998.

[5] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1-11. DOI: 10.1109/ICCV.2019.00547.

[6] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Inf. Fusion*, vol. 64, pp. 131-148, 2020. DOI: 10.1016/j.inffus.2020.06.003.

[7] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," in *Proc. IEEE WIFS*, 2018, pp. 1-7. DOI: 10.1109/WIFS.2018.8630761.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[9] O.Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234-241. DOI: 10.1007/978-3-319-24574-4_28.

[10] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint*, arXiv:1804.02767, 2018.

[11] A. Bochkovskiy, C. Wang, and H. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint*, arXiv:2004.10934, 2020.

[12] G. Jocher et al., "YOLOv5 by Ultralytics," 2020.

[13] A. Radford et al., "Learning Transferable Visual Models From Natural Language Supervision," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Proc. NeurIPS*, 20

[15] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 4401-4410.

[16] M. Wang and W. Deng, "Deep Visual Domain Adaptation: A Survey," *Neurocomputing*, vol. 312, pp. 135-153, 2018.

[17] M. Hasani and M. H. Mahoor, "Spatio-Temporal Facial Expression Recognition Using Convolutional LSTM Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017, pp. 348-356.

[18] X. Zhang et al., "Deep Learning-Based Face Manipulation Detection," *IEEE Access*, vol. 8, pp. 2305-2318, 2020.

[19] M. Verdoliva, "Media Forensics and DeepFakes: An Overview," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 5, pp. 910-932, 2020.

[20] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint*, arXiv:1409.1556, 2014.

[21] J. Howard and S. Gugger, *Deep Learning for Coders with Fastai and PyTorch: AI Applications Without a PhD*, O'Reilly Media, 2020.

[22] Y. Wu, W. Zheng, X. Peng, and J. Li, "Deep Learning-Based Methods for DeepFake Detection," in *Proc. IEEE ICASSP*, 2021

[23] A. Dosovitskiy et al., "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv preprint*, arXiv:2010.11929, 2020.

[24] J. K. Patel, P. M. Chauhan, and K. K. Patel, "A Comparative Analysis of DeepFake Detection Techniques," *IEEE Trans. Multimedia*, vol. 24, no. 6, pp. 1278-1290, 2022.

[25] N. Carlini and D. Wagner, "Adversarial Examples Are Not Easily Detected: Bypassing Ten Detection Methods," in *Proc. ACM Workshop Artif. Intell. Security (AISec)*, 2017.