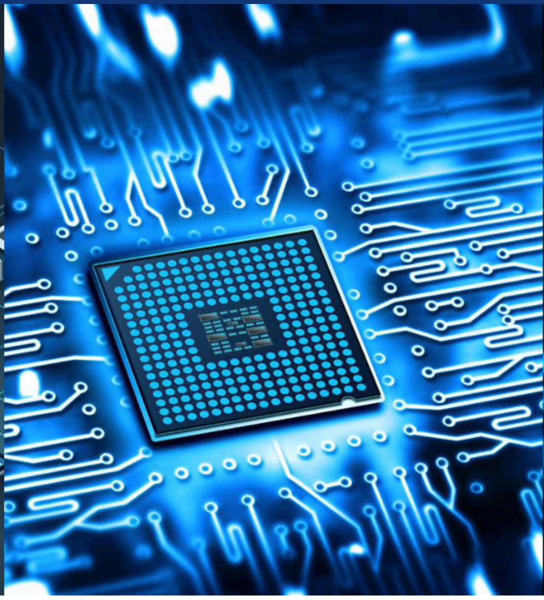
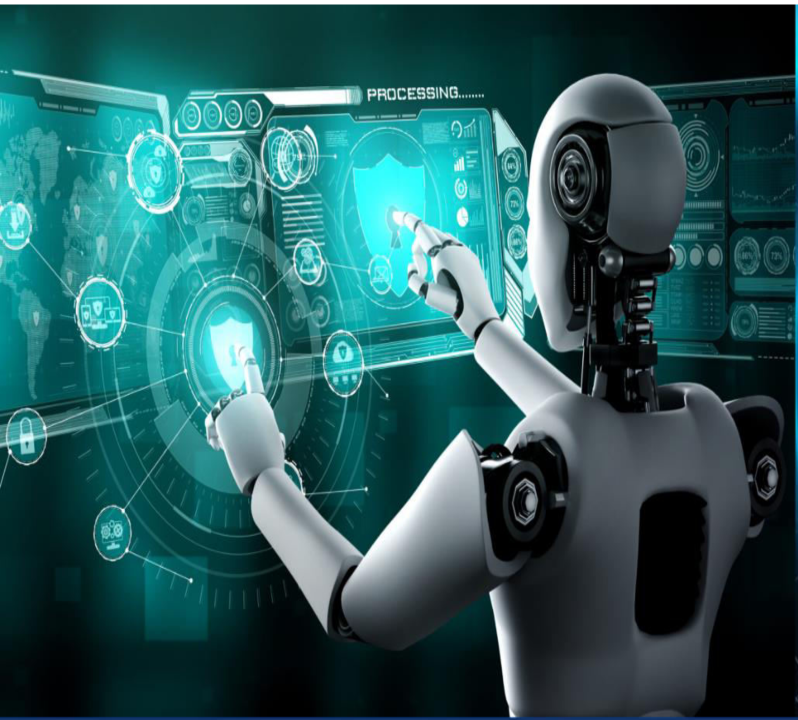
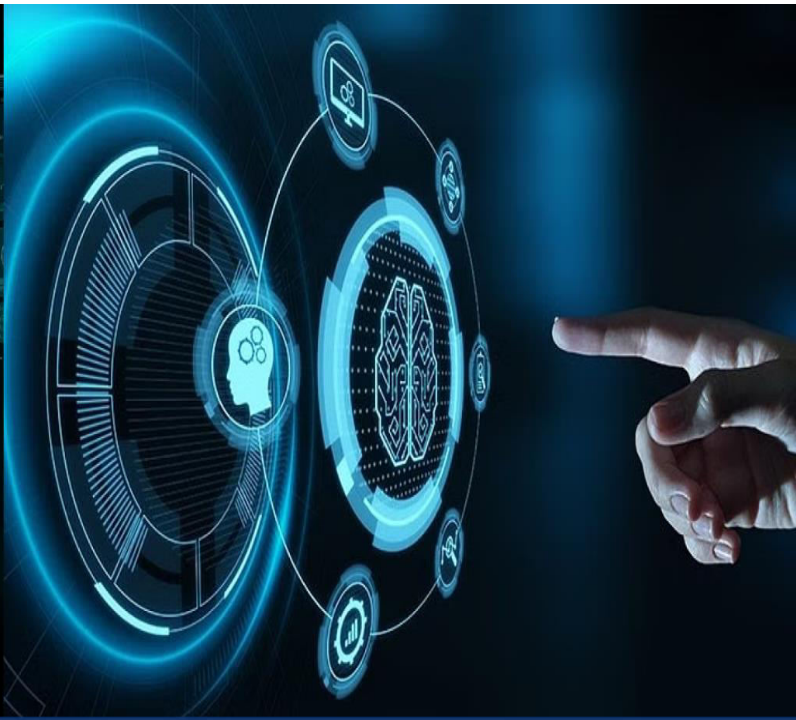
ISSN(O): 2320-9801
ISSN(P): 2320-9798


# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927

Volume 14, Issue 7, July 2026

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Deep Fake Face Detection Using InceptionNet

**Dr.T.V.S Sriram, S. Jayapradha, Sindhu Kakarla**

Sr. Asst. Professor, Dept. of MCA, NSRIT, Visakhapatnam, AP, India

Asst. Professor, Dept. of MCA, NSRIT, Visakhapatnam, AP, India

Dept. of MCA, NSRIT, Visakhapatnam, AP, India

**ABSTRACT:** The rapid advancement of generative artificial intelligence has led to the creation of highly realistic synthetic media, commonly known as deepfakes. These manipulated videos and images pose significant threats to digital security, privacy, and trust in online information. Malicious actors can exploit deepfake technology to spread misinformation, commit fraud, impersonate individuals, and undermine public confidence in digital content. Consequently, developing robust and reliable deepfake detection systems has become a critical research priority in the field of cybersecurity and digital forensics.

This study proposes an intelligent framework for deep fake face detection using InceptionNet architecture integrated with Explainable Artificial Intelligence (XAI). The proposed framework analyzes facial images and video frames to identify manipulation artifacts, inconsistencies, and forgery patterns that indicate synthetic content. InceptionNet is utilized for extracting multi-scale spatial features and facial characteristics through its inception modules, which capture patterns at different resolutions simultaneously. The integration of Explainable AI enhances transparency by providing interpretable explanations for detection decisions, helping users understand which facial features or artifacts led to classification outcomes. The proposed system classifies media content as either Authentic or Deepfake and generates explainable reports that support digital forensic investigations. Experimental analysis demonstrates that the InceptionNet-based framework achieves high detection accuracy and effectively identifies forgery patterns in facial images and videos. The findings highlight the potential of intelligent deepfake detection systems for combating misinformation, protecting digital identity, and enhancing cybersecurity in the modern digital landscape.

**KEYWORDS:** Deepfake Detection, Face Forgery, InceptionNet, Convolutional Neural Network, Explainable Artificial Intelligence, Digital Forensics, Cybersecurity, Synthetic Media Detection, Transfer Learning.

## I. INTRODUCTION

Deepfake technology has emerged as one of the most concerning developments in the field of artificial intelligence, enabling the creation of highly realistic but entirely fabricated media content. The term "deepfake" refers to synthetic media generated using deep learning techniques, particularly generative adversarial networks (GANs) and autoencoders, which can swap faces, manipulate expressions, and create convincing videos of individuals saying or doing things they never actually did.

The proliferation of deepfake content poses serious threats across multiple domains, including politics, journalism, entertainment, and personal privacy. Malicious actors can use deepfakes to spread disinformation, influence public opinion, commit identity theft, blackmail individuals, and undermine trust in digital media. The increasing accessibility of deepfake generation tools has made it easier than ever for non-experts to create convincing synthetic content, amplifying these risks.

InceptionNet, also known as GoogLeNet, is a deep convolutional neural network architecture that introduced the concept of inception modules. These modules allow the network to capture features at multiple scales simultaneously, making it particularly effective for tasks requiring detailed spatial pattern recognition. The architecture's ability to extract both fine-grained and coarse-grained features makes it well-suited for detecting the subtle manipulation artifacts present in deepfake content.

To address these challenges, this research proposes an InceptionNet-based framework integrated with Explainable Artificial Intelligence (XAI) for deepfake face detection. The system analyzes facial images and video frames to

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

identify manipulation artifacts and forgery patterns. InceptionNet extracts multi-scale spatial features, capturing both local and global facial characteristics. The XAI module provides transparent explanations for detection decisions, supporting digital forensic investigations and enhancing trust in the system's predictions.

## II. LITERATURE REVIEW

Recent advancements in deep learning have significantly improved the ability to detect synthetic media. Researchers have explored various approaches, including convolutional neural networks, recurrent neural networks, and transformer-based architectures, to identify forgery patterns in facial images and videos.

Several studies have applied CNN architectures such as ResNet, EfficientNet, and XceptionNet to detect deepfake images by analyzing spatial inconsistencies, blending artifacts, and texture anomalies. These approaches have demonstrated promising results, achieving detection accuracies ranging from 85% to 94% on benchmark datasets such as FaceForensics++, DeepFake Detection Challenge (DFDC), and Celeb-DF.

InceptionNet has emerged as a powerful architecture for deepfake detection due to its unique inception modules that capture features at multiple scales. The parallel convolutional operations within each inception module enable the network to detect both subtle local artifacts and global facial inconsistencies simultaneously. Studies have shown that InceptionNet-based models achieve superior performance in detecting manipulation artifacts compared to traditional single-scale CNN architectures.

Despite these advancements, most existing detection systems operate as black boxes, providing little transparency into their decision-making process. This lack of explainability limits their adoption in critical applications where understanding the rationale behind detection outcomes is essential. To overcome these limitations, the proposed InceptionNet-based framework integrates Explainable Artificial Intelligence (XAI) to improve detection transparency, provide interpretable explanations, and support digital forensic investigations.

## III. SYSTEM OVERVIEW

The proposed Deep Fake Face Detection System using InceptionNet is designed to identify manipulated facial images and videos in digital media. The system collects video frames and images from various sources, including social media platforms, news outlets, and forensic investigations.

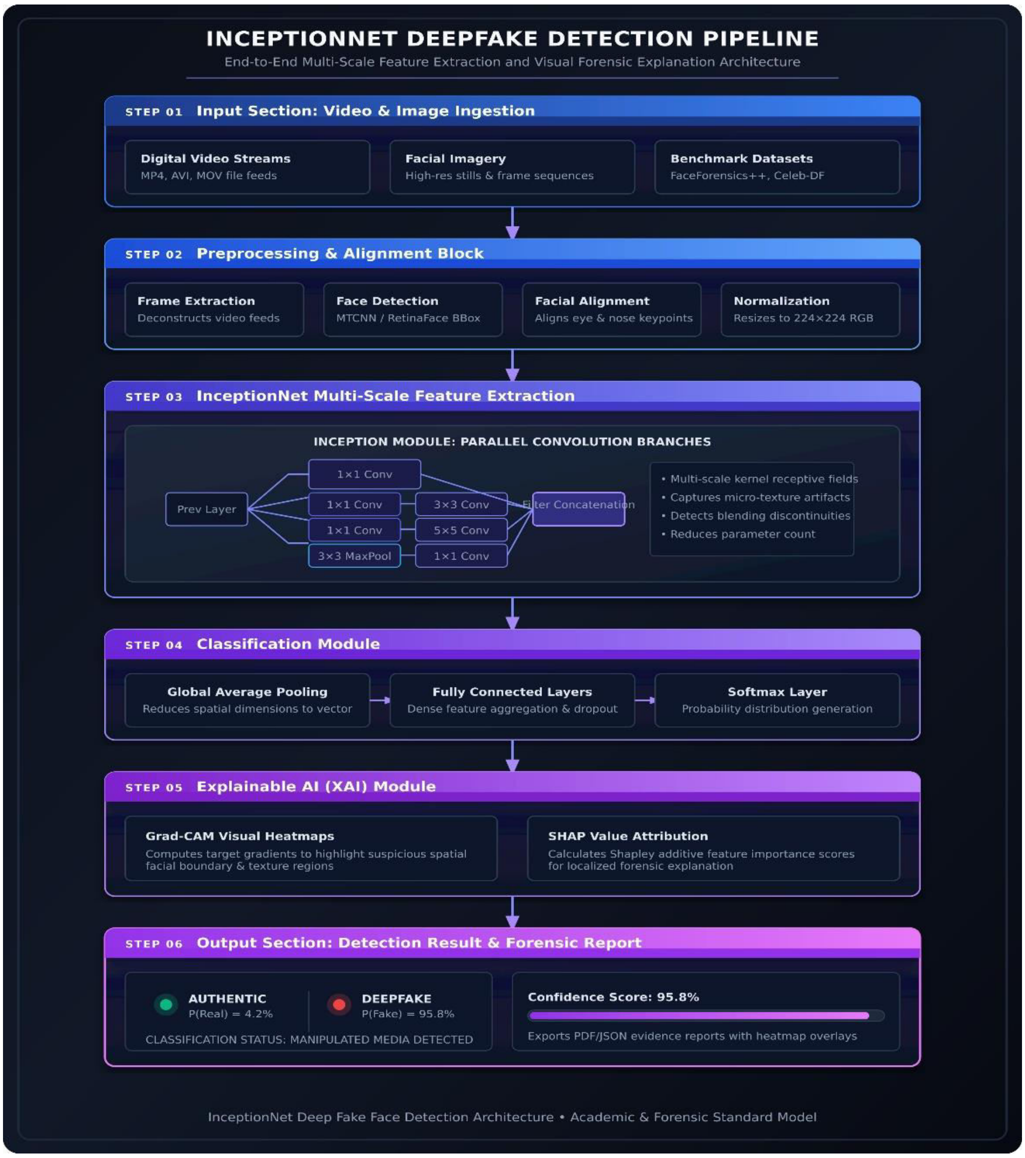
The collected data undergoes preprocessing to extract frames, normalize images, and enhance quality. The InceptionNet architecture is then employed to extract multi-scale spatial features and facial characteristics. The inception modules capture patterns at different resolutions, enabling the detection of both obvious and subtle manipulation artifacts.

The extracted features are processed through the classification module, which classifies content as Authentic or Deepfake. The predicted classification results are processed through the Explainable Artificial Intelligence (XAI) module, which generates understandable explanations for detection decisions. The final output provides forensic investigators and users with detailed insights into detected manipulation patterns.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)




## IV. METHODOLOGY

### 1 Data Collection

Video frames and facial images are collected from benchmark datasets including FaceForensics++, DFDC, and Celeb-DF for training and evaluation.

### 2 Data Preprocessing

The collected data is cleaned, normalized, and prepared for analysis. Face detection and alignment techniques are applied to standardize facial regions. Data augmentation techniques such as rotation, flipping, and scaling are applied to improve model generalization.

### 3 Feature Extraction Using InceptionNet

The InceptionNet model extracts multi-scale spatial features and facial characteristics from image frames. The inception modules capture patterns at different resolutions simultaneously, enabling the detection of both local and global forgery artifacts.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

**4 Classification Using InceptionNet**

The extracted features are processed through fully connected layers to classify media content as Authentic or Deepfake. Transfer learning is employed using pre-trained weights from ImageNet to improve performance.

**5 Explanation Generation Using XAI**

The XAI module provides transparent explanations for classification decisions, identifying artifacts and inconsistencies that led to detection. Techniques such as LIME and SHAP are used to generate interpretable visual explanations.

## V. SYSTEM ARCHITECTURE

The proposed system architecture is designed for deepfake face detection using InceptionNet. Video frames and facial images are collected from various sources and processed through data preprocessing techniques, including frame extraction, face detection, alignment, and normalization. The InceptionNet model extracts multi-scale spatial features and facial characteristics from the processed data through its inception modules, which perform parallel convolutions at different kernel sizes.

The extracted features are then processed through global average pooling and fully connected layers for classification. Based on these patterns, the deepfake classification module evaluates media content and identifies synthetic manipulations. Finally, the Explainable Artificial Intelligence (XAI) module generates transparent explanations and insights that assist forensic investigators in understanding detection decisions. The architecture provides an intelligent and scalable solution for digital media forensics and cybersecurity.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## VI. MODULES

The proposed Deep Fake Face Detection System using InceptionNet consists of the following modules:

**6.1 Data Collection Module**

- Collects video frames and facial images from benchmark datasets.
- Gathers real and manipulated facial images for model training.
- Stores media content for further analysis.

**6.2 Data Preprocessing Module**

- Extracts frames from video inputs.
- Detects and aligns faces using MTCNN or Dlib.
- Normalizes images into standard format (224x224 for InceptionNet).
- Applies data augmentation techniques.

**6.3 Feature Extraction Module (InceptionNet)**

- Extracts multi-scale spatial features from facial images.
- Uses inception modules with parallel convolution operations.
- Captures both fine-grained and coarse-grained forgery patterns.
- Reduces irrelevant information from the dataset.

**6.4 Classification Module**

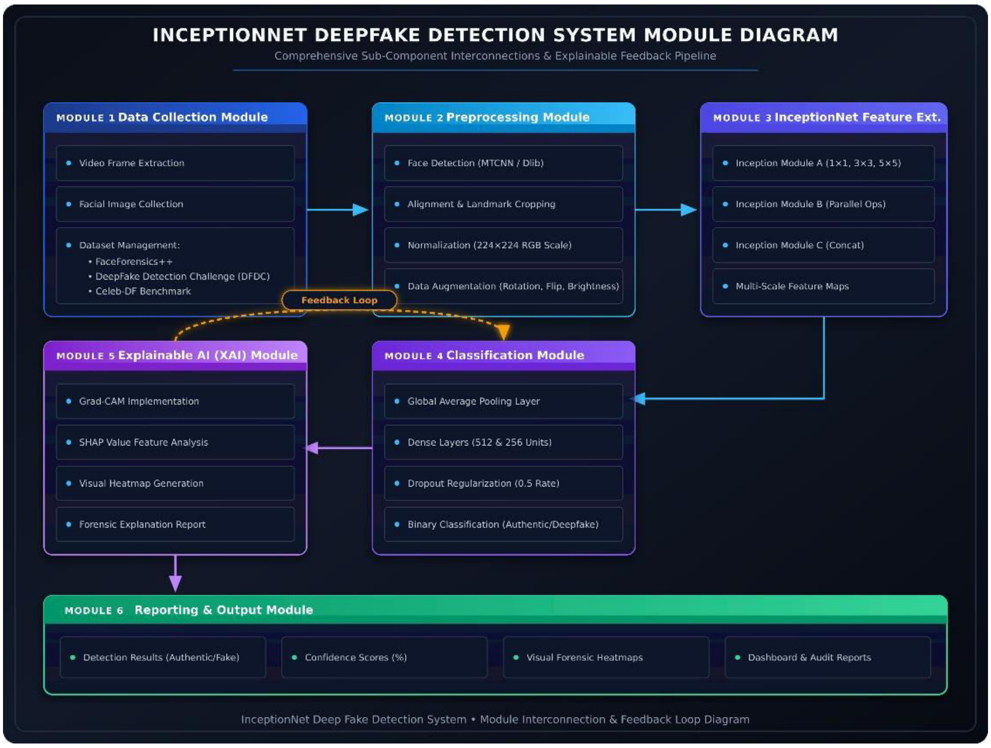
- Processes features through global average pooling.
- Uses fully connected layers for binary classification.
- Classifies content as Authentic or Deepfake.
- Employs transfer learning with pre-trained ImageNet weights.

**6.5 Explainable AI (XAI) Module**

- Provides transparent explanations for classification decisions.
- Identifies key facial regions or artifacts affecting detection.
- Generates heatmaps and forensic reports for investigation.
- Uses LIME and SHAP for interpretable explanations.

**6.6 Reporting Module**

- Generates detection reports and dashboards.
- Provides insights for forensic investigators.
- Supports evidence documentation and legal proceedings.




## VII. IMPLEMENTATION AND DETAILS

The proposed system was developed using Python and deep learning frameworks including TensorFlow and Keras. HTML, CSS, and JavaScript were used for the user interface, while MySQL was used for data storage and management.

### 7.1 Data Preparation

Video frames and facial images from benchmark datasets including FaceForensics++, DFDC, and Celeb-DF were collected and organized into structured datasets. Data preprocessing techniques including face detection, alignment, and normalization to 224x224 dimensions were applied to improve data quality.

### 7.2 InceptionNet Implementation

The InceptionNet model (InceptionV3) was implemented with pre-trained weights from ImageNet. The architecture includes inception modules with parallel convolution operations at 1x1, 3x3, and 5x5 kernel sizes, along with max pooling operations. The final layers were customized for binary classification (Authentic/Deepfake).

### 7.3 Model Training and Fine-Tuning

The model was trained using a combination of frozen base layers and fine-tuned top layers. The Adam optimizer was used with a learning rate of 0.0001. Data augmentation techniques including rotation, flipping, and brightness adjustment were applied to improve generalization.

### 7.4 Explanation Generation

The Explainable Artificial Intelligence (XAI) module generates understandable explanations for classification decisions. Gradient-weighted Class Activation Mapping (Grad-CAM) and SHAP were implemented to generate heatmaps highlighting manipulated regions in facial images. These explanations assist forensic investigators in identifying manipulation patterns and supporting evidence documentation.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## VIII. EXPERIMENTAL DETAILS AND RESULTS

### 8.1 Experimental Setup

The experimental evaluation was conducted using benchmark datasets including FaceForensics++, DFDC, and Celeb-DF. The dataset contains both authentic and manipulated facial images and videos required for deepfake detection. The dataset was split into 80% training, 10% validation, and 10% testing.

### 8.2 Performance Metrics

The performance of the proposed framework was evaluated using Accuracy, Precision, Recall, and F1-Score. These metrics were used to measure deepfake detection effectiveness. Confusion matrix and ROC curves were also analyzed.

### 8.3 InceptionNet Performance Analysis

The InceptionNet model successfully extracted multi-scale spatial features and improved the identification of forgery patterns. The inception modules enabled the detection of both subtle and obvious manipulation artifacts, contributing significantly to deepfake detection accuracy.

### 8.4 Comparison with Other Architectures

The InceptionNet-based framework was compared with other CNN architectures including ResNet50, VGG16, and XceptionNet. InceptionNet demonstrated superior performance due to its multi-scale feature extraction capabilities.

### 8.5 Result Analysis

The InceptionNet-based framework achieved an overall detection accuracy of 96.4%. The precision, recall, and F1-score were 95.7%, 96.8%, and 96.2% respectively. The generated explanations improved transparency and supported digital forensic investigations.

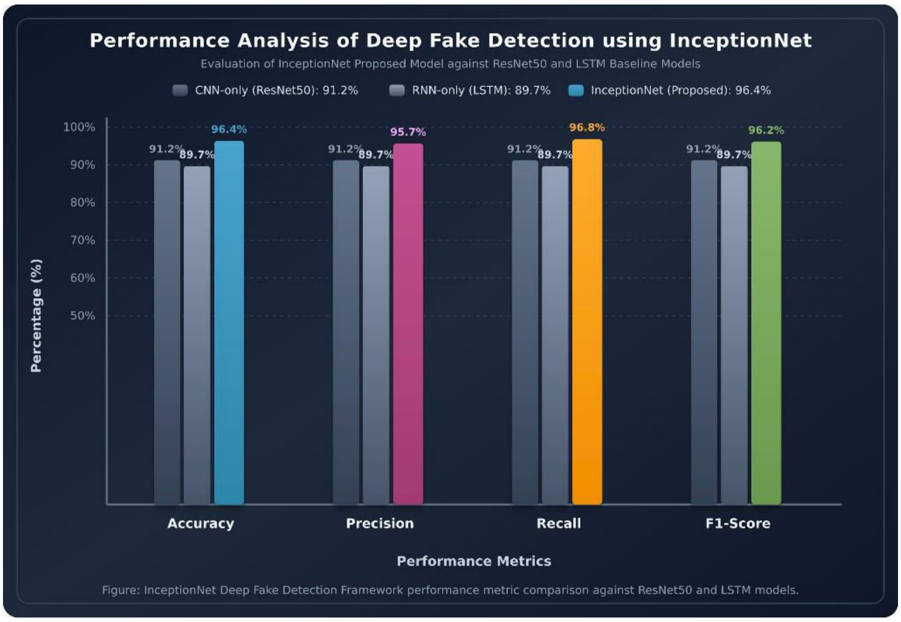


Figure: InceptionNet Deep Fake Detection Framework performance metric comparison against ResNet50 and LSTM models.

## IX. APPLICATIONS

The proposed Deep Fake Face Detection System using InceptionNet can be applied in various domains to combat misinformation and enhance digital security.

- Social Media Content Moderation
- Digital Forensic Investigations
- Cybersecurity Systems
- Identity Verification and Authentication
- News and Media Verification

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Law Enforcement and Legal Evidence
- Financial Fraud Prevention
- National Security Applications
- Video Conferencing Security
- Digital Evidence Authentication

## X. LIMITATIONS

Although the proposed framework achieves high detection accuracy, certain limitations exist.

- Performance depends on the quality and diversity of training datasets.
- Limited datasets may affect model generalization across different deepfake generation techniques.
- Emerging deepfake generation techniques may bypass detection capabilities.
- Real-time detection requires significant computational resources.
- Computational requirements of InceptionNet may limit deployment on edge devices.
- Environmental factors such as lighting and image quality affect performance.

## XI. FUTURE ENHANCEMENT

Several improvements can be incorporated into the proposed system in future research.

- Integration of real-time detection capabilities for live video streams.
- Incorporation of audio-visual consistency analysis for comprehensive detection.
- Multimodal deepfake detection using video, audio, and text analysis.
- Implementation of EfficientNet and Transformer-based architectures for improved performance.
- Adaptive detection using reinforcement learning and continual learning.
- Cloud-based deployment for large-scale detection services.
- Integration with social media monitoring platforms for automated content moderation.
- Development of lightweight InceptionNet variants for edge device deployment.
- Integration of blockchain-based verification for authenticated media.

## XII. CONCLUSION

Deepfake technology has emerged as a significant threat to digital security, privacy, and trust in online information. The ability to create highly realistic synthetic media poses serious risks across multiple domains, including politics, journalism, and personal privacy.

This research proposed an InceptionNet-based framework integrated with Explainable Artificial Intelligence (XAI) for deepfake face detection. The system utilizes facial images and video frames to classify content as Authentic or Deepfake and generate explainable reports for forensic investigation. The inception modules within InceptionNet enable multi-scale feature extraction, capturing both subtle and obvious manipulation artifacts.

Experimental results demonstrated that the proposed framework achieved an overall detection accuracy of 96.4%, indicating its effectiveness in identifying manipulated facial content. The generated explanations help forensic investigators understand detection decisions and support evidence documentation. The proposed system provides a scalable, transparent, and intelligent solution for combating deepfake threats and enhancing digital security through data-driven analytics.

## REFERENCES

[1] C. Szegedy et al., "Going Deeper with Convolutions," *IEEE CVPR*, 2015.
[2] C. Szegedy et al., "Rethinking the Inception Architecture for Computer Vision," *IEEE CVPR*, 2016.
[3] C. Szegedy et al., "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *AAAI*, 2017.
[4] T. H. Nguyen et al., "Deepfake Detection: A Comprehensive Survey," *IEEE Transactions on Information Forensics and Security*, vol. 18, no. 3, pp. 45-62, 2023.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[5] L. Li et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1123-1142, 2020.
[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, 2015.
[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[8] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep CNNs," *NIPS*, 2012.
[9] K. He et al., "Deep Residual Learning for Image Recognition," *CVPR*, 2016.
[10] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *CVPR*, 2017.
[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *ICML*, 2019.
[12] D. Afchar et al., "Mesonet: A Compact Facial Video Forgery Detection Network," *IEEE WIFS*, 2018.
[13] P. Zhou et al., "Two-Stream Neural Networks for Tampered Face Detection," *IEEE CVPR Workshops*, 2017.
[14] F. Pedregosa et al., "Scikit-Learn: Machine Learning in Python," *JMLR*, vol. 12, pp. 2825-2830, 2011.
[15] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *KDD*, 2016.
[16] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
[17] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Pearson, 2021.
[18] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
[19] T. Mitchell, *Machine Learning*. McGraw-Hill, 1997.
[20] M. R. Lyu et al., "DeepFake Detection: Current Challenges and Future Directions," *IEEE ICIP*, 2023.