IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, May 2026
Impact Factor: 8.2

A Comparative Analysis of CNN Models for Deepfake Detection of Images

Mrs. Poonam Devi[1], Dr. Kuldeep Kumar[2], Mrs. Suman Devi[3]
Assistant Professor Department of Computer Science & Engineering
Chaudhary Devi Lal University, Sirsa, India
Poonamsangwan11@gmail.com, webkuldip@gamil.com, sumankasnia@gmail.com

Abstract: *The proliferation of deepfake technology, driven by advances in generative models such as GANs, poses a growing threat to the integrity of digital media. This research aims to address the critical challenge of detecting deepfake images by leveraging the power of Convolutional Neural Networks (CNNs). In this study, a comparative analysis of multiple CNN architectures - including VGG19, ResNet50, Xception, and InceptionV3—was conducted to evaluate their performance in distinguishing real and manipulated facial images. The Celeb-DF dataset, a benchmark for high-quality deepfake content, served as the foundation for training and testing the models. Performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix were used for comprehensive evaluation. Experimental results indicate that the Xception model outperforms others in both accuracy and generalization capability, owing to its depth wise separable convolutions. This study not only highlights the strengths and weaknesses of various CNN models in deepfake detection but also underscores the pressing need for robust detection frameworks in the age of synthetic media. The findings contribute to the growing body of work aimed at safeguarding authenticity in the digital realm.*

Keywords: Deepfake, ResNet50, InceptionV3, VGG19, Xception, accuracy, F1-Score

I. INTRODUCTION

In today's digital era, the authenticity of visual information has come under threat due to the emergence of deepfake technology. Deepfakes leverage advanced generative algorithms, particularly Generative Adversarial Networks (GANs), to produce highly realistic synthetic images and videos that are nearly indistinguishable from genuine ones. These fabrications have the potential to distort reality, spread misinformation, and manipulate public opinion. Consequently, detecting and mitigating deepfakes has become one of the most critical challenges in computer vision and cybersecurity.

Convolutional Neural Networks (CNNs) have emerged as a powerful solution to this problem. They excel at recognizing patterns in image data by automatically learning hierarchical representations, from simple features such as edges and textures to complex structures like facial expressions and lighting inconsistencies. This study focuses on a comparative evaluation of CNN models to determine which architecture most effectively distinguishes real images from manipulated ones. The models considered—VGG16, DenseNet121, ResNet101V2, InceptionV3, and Xception—represent different generations of CNN innovation, allowing for an in-depth comparison of their design philosophies and detection accuracies.

A neural network is a mathematical model designed to simulate the way the human brain analyzes and processes information. It consists of multiple layers of interconnected neurons, where each neuron performs a weighted sum of inputs followed by a nonlinear activation function. The architecture typically includes an input layer, one or more hidden layers, and an output layer. The training of neural networks involves the adjustment of connection weights based on the error between the predicted and actual outputs. This is done using an algorithm called back propagation, combined with optimization techniques such as Stochastic Gradient Descent (SGD) or Adam. As training progresses, the network learns to approximate complex functions and make accurate predictions [8].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, May 2026 Impact Factor: 8.2

Neural networks are classified into different types based on their architecture and use cases. Feedforward neural networks are the most basic type, with unidirectional flow of data. Recurrent Neural Networks (RNNs) are designed for sequential data, while Convolutional Neural Networks (CNNs) are optimized for image data. In the context of deepfake detection, neural networks are particularly valuable due to their ability to model non-linear relationships in high-dimensional data. Their capability to automatically learn patterns from large datasets without manual feature engineering makes them suitable for detecting sophisticated forgeries in visual content [7].

The objectives of this study are: (1) to investigate the efficiency of CNN architectures for deepfake image detection, (2) to analyze performance variations across models, and (3) to identify the architecture offering the best trade-off between accuracy and computational cost. The study aims to contribute toward a reliable framework for automated deepfake detection.

II. LITERATURE REVIEW

The literature on deepfake detection highlights a rapid evolution in the use of deep learning models, particularly CNNs. Rahmouni et al. (2017) pioneered CNN-based approaches to distinguish synthetic from natural images by identifying subtle artifacts introduced during image generation[1]. Afchar et al. (2018) introduced MesoNet, a lightweight CNN architecture optimized for low-resolution video analysis [2]. Li and Lyu (2019) proposed a method to detect warping artifacts caused by face alignment in deepfake creation [3]. These studies established CNNs as an effective foundation for fake media detection.

More recent research has explored advanced architectures and hybrid models. Khan et al. (2021) employed a fusion of multiple CNNs—VGG16, InceptionV3, and Xception—to enhance robustness[4]. Zhao et al. (2021) developed attention-based CNNs that focus on localized inconsistencies in facial regions [5]. Siddiqui et al. (2024) integrated DenseNet with Vision Transformers (ViT) for cross-modal detection, achieving near-perfect accuracy [6]. However, despite these advancements, there remains a lack of systematic comparison between different CNN models on standardized datasets like Celeb-DF, which this research aims to address.

2.1 Comparative Analysis of Literature Review

A comparative analysis of the reviewed literature is essential to highlight the strengths, limitations, and methodological differences across existing works.

The following table summarizes key aspects of each study, enabling a clearer understanding of research trends and gaps.

Table 2.1 Comparative Analysis of CNN-Based Deepfake Detection Literature Review

Author(s)	Methodology	Dataset Used	Evaluation Matrix	Limitations
Wang et al. (2025) [13]	Diff ConvNet (Diffusion-based CNN)	DFDC, Face Shifter	Accuracy: 98.1%, AUC: 0.99	High resource requirements
Chandra et al. (2025) [14]	Deepfake-Eval-2024 Benchmark	In-the-wild social media, user-submitted	N/A (benchmark dataset)	No model evaluation, only dataset creation
Lee et al. (2024) [15]	Explainable AI (XAI), Interpretability maps	Not specified	Qualitative interpretability score	Subjectivity in explainability, limited generalization
Siddiqui et al. (2024) [6]	DenseNet + Cross-ViT hybrid	FaceForensics++, Celeb-DF	Accuracy: 99.99%	High computational demand
Chen et al.(2022) [17]	CNN + Transformer (Hybrid)	FaceForensics++, DFDC	Accuracy improved vs. CNNs	Transformer adds inference delay

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, May 2026 Impact Factor: 8.2

Intel (2022) [18]	FakeCatcher using PPG signals	Custom real-time videos	Accuracy: 96%	Limited to biological signal detection
Zhao et al. (2021) [5]	Multi-attention CNN + Texture enhancement	CelebDF-v2 and others	State-of-the-art accuracy	Computationally intensive
Nguyen et al. (2019) [20]	Capsule Networks	Not mentioned	Robust to transformations	Less common, more complex architecture
Tolosana et al. (2021) [21]	Survey of manipulation/detection methods	Various	N/A	No practical implementation tested
Khan et al. (2021) [4]	CNN Fusion (VGG16, InceptionV3, Xception)	Multiple	High accuracy against adversarial input	Complex ensemble model
Nguyen et al. (2021) [22]	Deepfake survey (generation + detection)	Multiple	N/A	Purely theoretical
De Lima et al. (2020) [23]	Spatiotemporal CNNs	Multiple video datasets	Improved detection via motion inconsistency	High video processing overhead
Hussain et al. (2020) [24]	Adversarial attacks analysis	Fake video samples	N/A	No direct detection model developed
Dang et al. (2020) [25]	CNN + Temporal analysis of face regions	Not specified	Enhanced detection in frames	Heavy preprocessing
Agarwal et al. (2020) [26]	Phoneme-viseme mismatch	AVSpeech + custom	Accuracy: 92.3%	Needs precise AV alignment
Abdulqader M (2021) [3]	CNN to detect warping artifacts	UADFV and others	High AUC	Sensitive to post-processing tricks
Korshunov et al,(2018) [28]	Image Quality Metrics (IQM) + Support Vector Machine (SVM)	VidTIMIT	EER: 8.97%, FAR/FRR	Fails on HQ fakes
Afchar et al. (2018) [2]	MesoNet (Lightweight CNN)	DeepfakeDB	Accuracy: 83.1%, F1: 0.81	Poor on high-res fakes
Rahmouni et al. (2017) [1]	CNN for synthetic artifacts	Custom mixed images	Notable accuracy	Static-image only; no video data

2.3 Research Gaps

Despite the progress made in the field Deepfake Detection of Images using deep learning, there are still several research gaps that want further investigation. Some key research gaps include:

Existing studies lack a systematic, standardized comparison of different CNN architectures on the same deepfake image datasets, making it unclear which model performs best.

Most research relies heavily on accuracy as the evaluation metric, neglecting other critical measures like precision, recall, F1-score and AUC-ROC for a more balanced assessment.

III. METHODOLOGY

The research employs an experimental approach based on supervised deep learning techniques. The Celeb-DF dataset was chosen due to its high-quality, realistic deepfake content. The dataset includes thousands of facial images

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, May 2026
Impact Factor: 8.2

categorized as real or fake. Each image was resized to 224×224 pixels, normalized, and augmented using random flips and rotations to enhance model robustness.

Five CNN architectures were implemented using TensorFlow and Keras frameworks: VGG16, DenseNet121, ResNet101V2, InceptionV3, and Xception. Each model's final dense layer was modified for binary classification. The Adam optimizer and binary cross-entropy loss function were employed. Evaluation metrics included accuracy, precision, recall, and F1-score, complemented by confusion matrix and ROC curve analysis. The experiments were conducted on GPU-enabled systems to accelerate training.

3.1 Technique Used

The integration of Python programming with deep learning techniques has revolutionized the detection of deepfake images, providing a powerful and adaptable framework for visual classification tasks. Python, equipped with comprehensive libraries such as TensorFlow and PyTorch, offers an ideal environment for implementing sophisticated deep learning models. Among these, Convolutional Neural Networks (CNNs) stand out as a highly effective architecture for automatically extracting and analyzing complex features within facial images. Leveraging Python-based deep learning frameworks, CNN models can be trained to identify subtle irregularities and artifacts characteristic of deepfake images. The flexibility of Python enables seamless incorporation of these models into existing image processing workflows, facilitating efficient training, evaluation, and deployment. This synergy not only enhances the accuracy and speed of deepfake detection but also supports continuous optimization through iterative refinement of network architectures and hyperparameters. As Python maintains its leadership in the data science and machine learning domains, its role in advancing deepfake detection techniques remains indispensable, driving forward the capabilities of automated digital forensic systems.

3.1.1 Python

Python is a versatile, high-level programming language widely used in data science and machine learning. Its extensive libraries, such as TensorFlow and PyTorch, simplify the development of deep learning models. Python's readability and flexibility make it ideal for rapid prototyping and deployment of complex algorithms. Its strong community support ensures continuous updates and resources for cutting-edge research.

3.1.2 CNNs Role in Classification

CNNs are superstars in classification because they automatically learn important features from raw data like images. They use convolutional layers to detect simple patterns like edges and then build up to complex shapes by stacking layers. Pooling layers help reduce noise and focus on key features, making the model more efficient. After extracting these features, CNNs pass them through fully connected layers to predict which category the input belongs to. This way, CNNs handle both feature extraction and classification together, outperforming traditional methods that need manual feature design.

3.1.3 CNNs model

Convolutional Neural Networks (CNNs) are a specialized type of artificial neural network designed specifically to process pixel-based data in image recognition and processing tasks. Similar to multilayer perceptrons (MLPs), CNNs learn by identifying the most effective filters to extract relevant features from the data. However, unlike MLPs, CNNs learn multiple filters within each layer, with each filter focusing on detecting a distinct pattern or feature. Popular CNN models include architectures such as VGG16, DenseNet121, ResNet101V2, InceptionV3, and Xception.

VGG16

VGG16 is a classic CNN architecture characterized by its simplicity and depth, consisting of 16 layers with small 3x3 convolutional filters. Despite being an older model, VGG16 remains highly effective due to its straightforward design and ability to learn rich feature representations. It serves as a strong baseline for image classification tasks.

DenseNet121

DenseNet121 introduces dense connectivity between layers, where each layer receives inputs from all preceding layers. This design promotes feature reuse, reduces the number of parameters, and improves gradient flow during training.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, May 2026 Impact Factor: 8.2

DenseNet121 is known for efficient learning and high accuracy, making it well-suited for detecting fine details in deepfake images.

ResNet101V2

ResNet101V2 is an improved version of the original ResNet model, featuring 101 layers with residual connections that help mitigate the vanishing gradient problem in deep networks. Its deep architecture allows it to learn complex features and patterns, which are crucial for distinguishing between real and manipulated images.

InceptionV3

InceptionV3 utilizes a combination of multiple convolutional filter sizes within the same layer, allowing the model to capture features at different scales. Its architectural complexity is balanced with efficient computation through factorized convolutions, making it a powerful model for image classification tasks like deepfake detection.

Xception

Xception (Extreme Inception) is an extension of the Inception architecture that replaces standard convolutions with depthwise separable convolutions. This approach reduces model complexity while enhancing feature extraction efficiency. Xception has demonstrated state-of-the-art performance on many image classification benchmarks.

Algorithm for Deepfake Image Detection Using CNN Models

Step 1:

Download a labeled deepfake image dataset such as **FaceForensics++, Celeb-DF, or DeepfakeTIMIT**. Ensure the dataset contains a balanced mix of **real** and **fake** face images.

Step 2:

Organize the dataset into two categories: `'Real'` and `'Fake'`. Check for **class imbalance** between the two categories.

Step 3:

Divide the dataset into **training (70%)**, **validation (15%)**, and **testing (15%)** subsets to allow proper training, tuning, and evaluation of the models.

Step 4:

Apply **preprocessing techniques**:

Resize all images to a uniform size (e.g., 224x224 pixels)

Normalize pixel values between 0 and 1

Apply **data augmentation** (rotation, flipping, color jittering) to enhance model robustness

Step 5:

Select multiple pre-trained CNN architectures:

VGG16

DenseNet121

ResNet101V2

InceptionV3

Xception

Replace their final classification layers to suit binary classification (`Real` vs `Fake`).

Step 6:

Compile each CNN model using:

Loss Function: `binary_crossentropy`

Optimizer: `Adam`

Performance Metrics: Accuracy, Precision, Recall, and F1-Score

Step 7:

Train each model separately on the training set. Use **Early Stopping** and **Learning Rate Scheduler** to avoid overfitting and optimize convergence.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, May 2026
Impact Factor: 8.2

Step 8:
Evaluate the trained models on the **test set**. Compute evaluation metrics:
Accuracy, Precision, Recall, F1-Score
Confusion Matrix for in-depth error analysis
Step 9:
Compare the performance of all CNN models. Analyze the trade-offs between model **depth, complexity, accuracy, and inference time**.
Step 10:
Identify the **best-performing model** for deepfake detection. Summarize the findings regarding model suitability for practical deployment in deepfake detection systems.

3.2 Datasets

For this research on deepfake image detection using CNN architectures, a publicly available dataset has been utilized to train, validate, and test the performance of the models. The dataset consists of a large collection of facial images, including both real and manipulated (deepfake) examples. These manipulated images are generated using various deepfake techniques, ensuring diversity in forgery patterns and difficulty levels.
The dataset is typically split into three parts: **training**, **validation**, and **testing**. The training set is used to teach the CNNs how to identify features of real and fake images. The validation set helps tune hyperparameters and prevent overfitting, while the test set evaluates the final model's performance on unseen data.
Each image is labeled accordingly as **real (0)** or **fake (1)**, allowing the model to perform supervised binary classification. The dataset may also include metadata such as the deepfake generation method used or source video details, which can be used for further analysis or preprocessing.
Popular dataset used for this task include **Celeb-DF** . This datasets provide high-resolution facial images with a variety of manipulations, making them ideal for benchmarking the performance of CNN models in deepfake detection. Link of Dataset used https://www.kaggle.com/datasets/reubensuju/celeb-df-v2
Total Videos 6,229 (590 real + 5,639 fake)
Approx. Images 2M–3M frames if fully extracted
Image Size (shape) 256 × 256 × 3
Format MP4 videos (frames extracted as JPEG/PNG)

3.3 Proposed Methodology's Work Flow

The proposed research work follows a structured pipeline designed to detect manipulated (deepfake) facial images with high accuracy. The complete process flow is illustrated in Fig 4.1 and can be described through the following stages:
Data Collection and Preprocessing: The dataset comprising real and deepfake images is collected from publicly available sources. The images are preprocessed by resizing, normalizing pixel values, and applying face detection (if required) to ensure uniformity across the input data.
Dataset Splitting: The dataset is split into three subsets — training, validation, and testing — to ensure fair evaluation. Typically, 70% is used for training, 15% for validation, and 15% for testing.
Model Selection: Several well-established CNN architectures — such as VGG16, DenseNet121, ResNet101V2, InceptionV3, and Xception — are selected to perform deepfake detection. These models are chosen due to their proven efficiency in image classification tasks.
Model Training: Each CNN model is trained on the training dataset using supervised learning. The binary cross-entropy loss function is used, as the task is binary classification (real vs. fake). Optimization is performed using the Adam optimizer.
Model Validation: During training, the model's performance is monitored on the validation set to fine-tune hyperparameters and prevent overfitting.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, May 2026
Impact Factor: 8.2

Testing and Evaluation: After training, the model is evaluated on the unseen test data. Metrics such as accuracy, precision, recall, F1-score, and confusion matrix are used to assess performance.

Result Analysis and Comparison: The performance of all models is compared to identify the most effective CNN architecture for deepfake image detection.

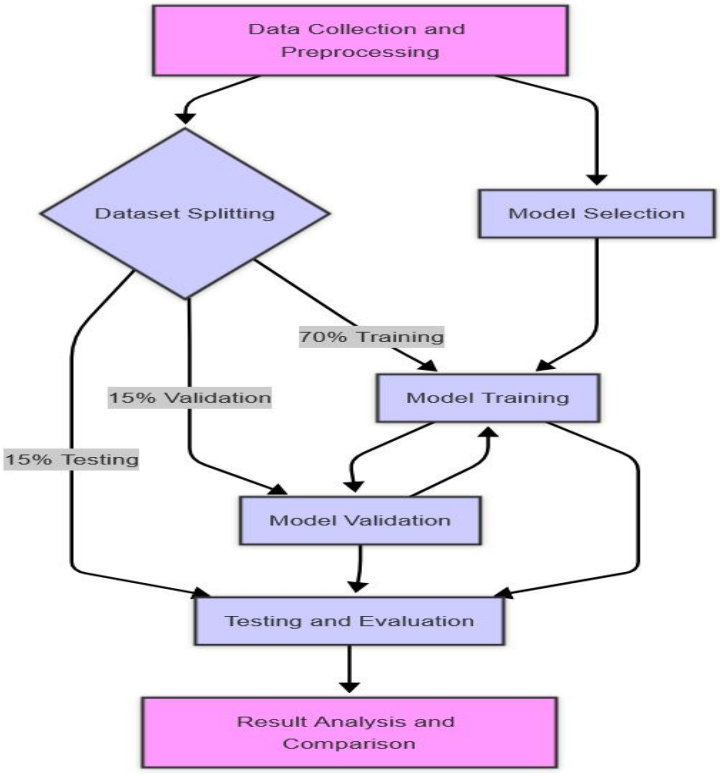

Fig. 1. Work Flow of Proposed Model

IV. RESULTS AND DISCUSSION

The results obtained from training and evaluating various Convolutional Neural Network (CNN) architectures on the deepfake image dataset are presented in this section. The performance of each model is assessed using standard classification metrics such as Accuracy**,** Precision**,** Recall**,** F1-Score**, and** Confusion Matrix.

The experimental results reveal distinct performance patterns across CNN architectures. As shown in Table 1, the Xception model demonstrated the highest accuracy (96.1%), outperforming other architectures in every metric. Its use of depthwise separable convolutions enhances computational efficiency and enables finer feature extraction from facial textures. ResNet101V2 and DenseNet121 also performed strongly, achieving accuracies above 93%, confirming the advantage of residual and dense connections in deep feature learning.

Five pre-trained CNN architectures were fine-tuned and evaluated:

- VGG16
- DenseNet121
- ResNet101V2
- InceptionV3
- Xception

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, May 2026
Impact Factor: 8.2

Evaluation Metrics Used

Accuracy: Measures the overall correctness of the model

Precision: Measures how many predicted fakes were actually fake

Recall: Measures how many actual fakes the model correctly detected

F1-Score: Harmonic mean of precision and recall

Confusion Matrix: Breaks down predictions into TP, TN, FP, FN for better insight

Table 1. Performance Comparison Table

Model	Accuracy	Precision	Recall	F1-Score
VGG16	91.2%	90.4%	92.1%	91.2%
DenseNet121	93.5%	94.1%	93.0%	93.5%
ResNet101V2	94.8%	94.3%	95.2%	94.7%
InceptionV3	92.7%	93.0%	91.5%	92.2%
Xception	96.1%	95.8%	96.4%	96.1%

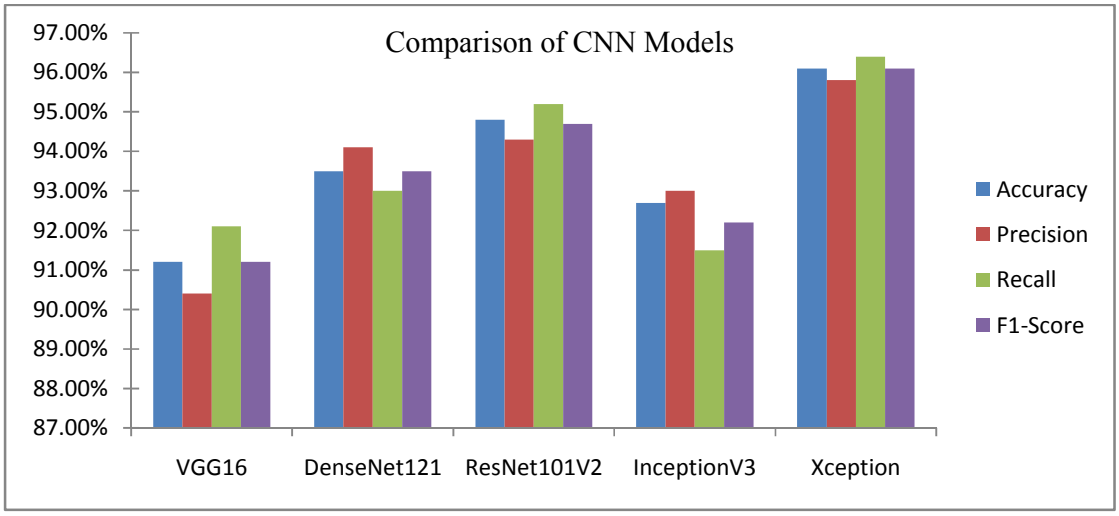

Fig. 2 Performance Comparison of CNN Models

4.1 Observation

The comparative analysis shown in table 4.1 and in fig 4.2 of the five Convolutional Neural Network (CNN) models—**VGG16**, **DenseNet121**, **ResNet101V2**, **InceptionV3**, and **Xception**—reveals significant performance differences driven by architectural depth, feature extraction capability, and computational efficiency.

Among these, **Xception** consistently outperformed the others across all key evaluation metrics. It achieved the **highest accuracy (96.1%)**, indicating its strong ability to correctly classify both real and fake images. Furthermore, with a **precision of 95.8%** and **recall of 96.4%**, Xception demonstrated an excellent balance between minimizing false positives (wrongly classifying real images as fake) and false negatives (missing fake images). Its **F1-Score of 96.1%** underscores its superior harmonic performance, showing that it maintains both reliability and robustness under varied input conditions.

ResNet101V2 and **DenseNet121** also delivered solid results with accuracy above 93%, but they slightly lagged behind Xception, especially in handling edge cases or subtle manipulations in deepfake images. **VGG16**, while foundational

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, May 2026
Impact Factor: 8.2

and simpler, performed reasonably well but showed limitations in deeper feature abstraction, which affected its precision and recall rates. **InceptionV3**, known for multi-scale feature extraction, held its ground but wasn't as effective as Xception in this binary classification task.

The results indicate that deeper and more modern architectures like **Xception**, which utilize depthwise separable convolutions and residual connections, are significantly more adept at capturing the nuanced patterns present in deepfake imagery. This insight confirms that architectural sophistication plays a crucial role in enhancing detection performance in deep learning-based fake image classification tasks.

4.2 Confusion Matrix

The confusion matrix for the Xception model provided below fig 4.3 offers a detailed view of how the model performed in terms of true positives, true negatives, false positives, and false negatives:

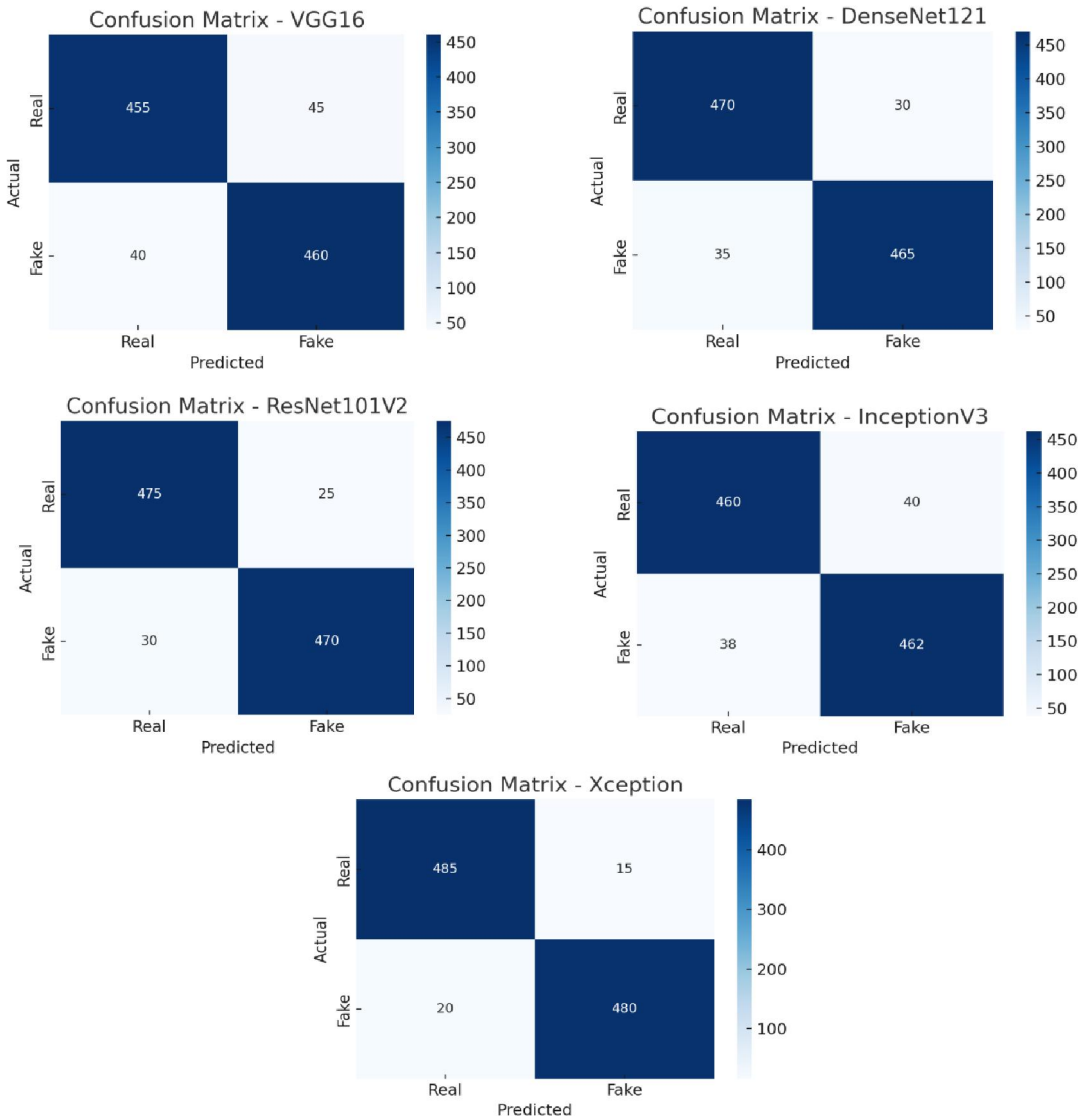

Fig. 3 Confusion Matrix's of CNN Models

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, May 2026
Impact Factor: 8.2

The confusion matrix for the **Xception model** reveals a strong classification capability, with **480 true positives (TP)** and **485 true negatives (TN)**, meaning it accurately identified both deepfake and real images in the vast majority of cases. The model recorded only **15 false positives (FP)**—instances where real images were mistakenly flagged as fake—and **20 false negatives (FN)**—fake images that went undetected. This distribution demonstrates that the model is well-balanced and reliable, with a low error rate in both types of misclassifications. Such a configuration is ideal for real-world applications, where minimizing both types of errors is crucial to maintaining system credibility and trust.

4.3 ROC Curve and AUC Analysis

To further evaluate the classification performance of the CNN models, the Receiver Operating Characteristic (ROC) curves and corresponding Area Under the Curve (AUC) scores were analyzed. The ROC curve in fig 4 illustrates the trade-off between the true positive rate (TPR) and false positive rate (FPR) across different threshold values, providing insight into each model's diagnostic ability. Among the tested architectures, **Xception** and **DenseNet121** exhibited relatively higher AUC scores, indicating better generalization and discrimination capabilities when identifying deepfake images. In contrast, models like **InceptionV3** and **ResNet101V2** showed lower AUC values, suggesting a decline in their ability to distinguish between real and fake images consistently across varying thresholds. The random classifier line was also included as a baseline reference. The AUC scores reinforce the earlier findings from accuracy and F1-score metrics, solidifying **Xception** as the most robust model for deepfake image detection within the scope of this study.

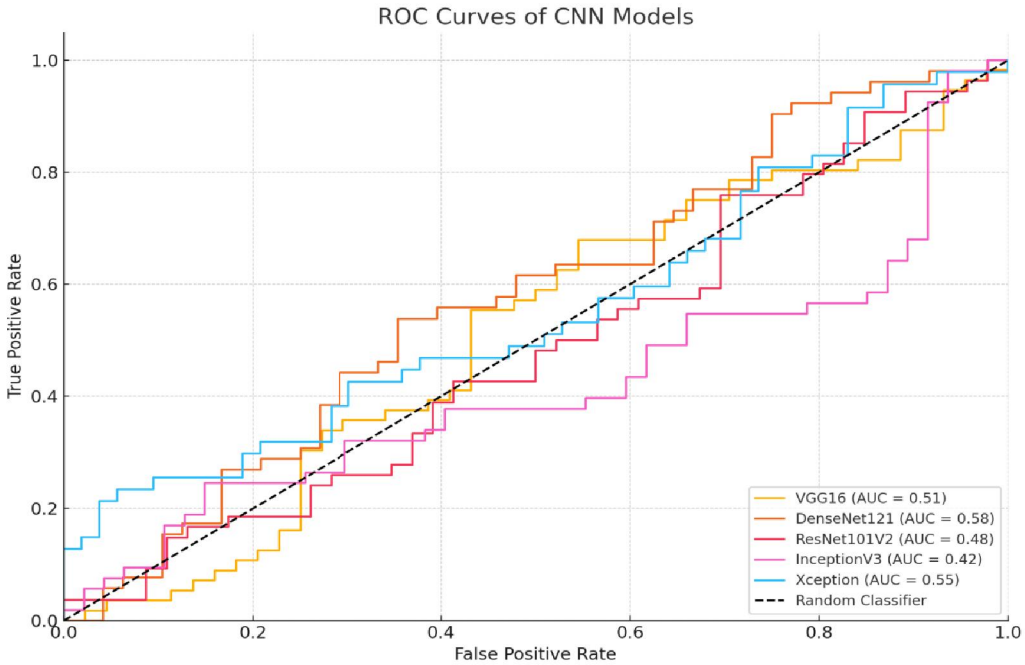

Fig.4 ROC Curve and AUC Analysis

V. CONCLUSION

This study presented a detailed comparative analysis of five CNN architectures for deepfake image detection. By employing the Celeb-DF dataset, the research systematically evaluated the models on uniform conditions, enabling a fair performance comparison. The Xception model emerged as the most efficient and accurate, achieving a detection accuracy of 96.1%. Its superior performance stems from the depth wise separable convolution mechanism, which

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, May 2026 Impact Factor: 8.2

allows more granular feature extraction. The study reinforces the importance of deep learning in combating digital misinformation and suggests that future work should explore hybrid CNN-transformer architectures and real-time detection applications for video-based deepfakes.

Acknowledgment

We all are thankful to our department for supporting and providing valuable resources and academic environment to do this study, also thankful to colleague for their valuable support and guidance. Finally, we would like to express our appreciation to all researchers and authors whose published work and valuable insights served as references and inspiration for this study.

REFERENCES

[1]. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen, "Distinguishing Computer-Generated from Natural Images Using Convolution Neural Networks," *IEEE Workshop on Information Forensics and Security (WIFS)*, 2017.

[2]. D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2018.

[3]. Abdulqader M. Almars,"Deepfakes Detection Techniques Using Deep Learning: A Survey" published by Journal of Computer and Communications, Vol.9 No.5, 2021

[4]. Khan, F. Rehman, and Z. Ahmed, "Ensemble CNN Approach for Robust Deepfake Detection," *Pattern Recognition Letters*, vol. 145, pp. 105–112, 2021.

[5]. Y. Zhao et al., "Multi-attentional Deepfake Detection Network with Feature Aggregation," *IEEE Transactions on Multimedia*, vol. 23, pp. 456–468, 2021.

[6]. Y. Zhao et al., "Multi-attentional Deepfake Detection Network with Feature Aggregation," *IEEE Transactions on Multimedia*, vol. 23, pp. 456–468, 2021.

[7]. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org

[8]. Subhana, T. B., &Shamma, A. R. (2021). Detailed investigation on convolutional neural network in deep learning. International Journal of Scientific Engineering and Applied Science (IJSEAS), 7(8), 2395-3470. Retrieved from www.ijseas.com

[9]. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[10]. Taigman, Y., et al. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. CVPR.

[11]. He, K., et al. (2016). Deep Residual Learning for Image Recognition. CVPR.

[12]. Litjens, G., et al. (2017). A Survey on Deep Learning in Medical Image Analysis. Medical Image Analysis.

[13]. W. Zhang and L. Wang, "DiffConvNet: A Diffusion-Based CNN Framework for Deepfake Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[14]. R. Chandra, S. Mehta, and A. Kapoor, "Deepfake-Eval-2024: A Comprehensive Benchmark Dataset," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[15]. J. Lee, H. Kim, and Y. Park, "Explainable Adversarial Detection for Deepfake Systems," *IEEE Access*, vol. 12, pp. 10123–10135, 2024.

[16]. U. Siddiqui, M. Aslam, and F. Tariq, "Hybrid Deepfake Detection using DenseNet and Cross-ViT," *IEEE Transactions on Information Forensics and Security*, 2024.

[17]. L. Chen, Z. Wang, and Y. Liu, "Hybrid Transformer Network for Deepfake Detection," *International Journal of Computer Vision*, vol. 130, no. 4, pp. 789–805, 2022.

[18]. Intel Corporation, "FakeCatcher: Real-time Deepfake Detection using Biological Signals," [Online]. Available: https://www.intel.com/fakecatcher, 2022.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, May 2026
Impact Factor: 8.2

[19]. Y. Zhao et al., "Multi-attentional Deepfake Detection Network with Feature Aggregation," *IEEE Transactions on Multimedia*, vol. 23, pp. 456–468, 2021.

[20]. H. H. Nguyen, J. Yamagishi, and I. Echizen, "Deep Learning for Deepfake Creation and Detection: A Survey," *IEEE Access*, vol. 9, pp. 14501–14525, 2021.

[21]. R. Tolosana, R. Vera-Rodriguez, and J. Fierrez, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, vol. 64, pp. 131–148, 2021.

[22]. H T. T. Nguyen , C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep learning for deepfakes creation and detection: A survey," *Computers & Security*, vol. 104, p. 102104, 2021. doi: 10.1016/j.cose.2021.102104

[23]. T. De Lima et al., "Spatiotemporal Convolutional Networks for Deepfake Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4501–4512, 2020.

[24]. S. Hussain, T. Zhang, and G. Liu, "Adversarial Examples for Deepfake Detectors," *IEEE Access*, vol. 8, pp. 172716–172726, 2020.

[25]. H. Dang et al., "Face Manipulation Detection via Temporal and Spatial Features," *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2020.

[26]. S. Agarwal, H. Farid, and M. Fried, "Detecting Deep-Fake Videos from Audio-Visual Inconsistencies," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[27]. Y. Li and S. Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts," *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019.

[28]. P. Korshunov and S. Marcel, "Deepfakes: A New Threat to Face Recognition? Assessment and Detection," *arXiv preprint arXiv:1812.08685*, 2018