International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

# Deepfake Detection with EfficientNetB4 Feature Extraction and XGBoost Classification

E. Shajini Rose
1Department of Computer Science and Engineering, St. Xavier's Catholic College Of Engineering, Chunkankadai, Tamil Nadu, India.

S. Neelakandan
2Professor, Department of Computer Science and Engineering, RMK Engineering College, Kavaraipettai, Tamil Nadu, India.

P. Ajitha
3Department of Computer Science and Engineering, St. Xavier's Catholic College Of Engineering, Chunkankadai, Tamil Nadu, India.

**Abstract - The development of extremely realistic DeepFake photos and videos has expanded due to the quick development of deep learning and artificial intelligence technology, leading to concerns related to misinformation, identity theft, and digital manipulation. An EfficientNetB4-based DeepFake detection framework is proposed to identify manipulated facial frames from authentic facial frames using the DeepFake Detection Challenge (DFDC) dataset. There are both actual and false video frames in the dataset that have been resized and preprocessed for uniform input representation. EfficientNetB4 is used to extract discriminative facial features, whereas XGBoost does classification by learning complex feature connections. This framework combines transfer learning with ensemble classification to improve detection performance and resilience. The experimental results demonstrate strong classification performance on the DFDC dataset, achieving a high detection accuracy of 93.55% along with an AUC score of 98.65%, indicating excellent capability in distinguishing between real and manipulated facial.**

**Keywords:**
**DeepFake Detection, EfficientNetB4, XGBoost, DFDC Dataset,Multimedia Forensics**

## 1. INTRODUCTION

These days, a large number of facial photos have been altered and then extensively disseminated via social media [1]. The use of contrastive learning has greatly enhanced. detection of deepfakes in recent years. However, the emphasis on class granularity in existing techniques makes it difficult to distinguish between the real instance and its counterparts that are manufactured [2]. Detecting training and testing using the same dataset yields great results for current face forgery detectors. However, when applying the detector to unidentified forging techniques, its performance deteriorates. Training the model with synthetic data is one of the best strategies to deal with this issue. In order to detect deep forgeries, this aids the model in learning a generic representation [3]. Some current fake detection methods work well, but they do not properly understand important facial details like structure and expressions. Because of this, they may miss important information needed to detect fake images accurately [4]. Face forgery detection is hampered by cross-dataset generalisation, which occurs when models trained on one dataset perform poorly on new data. A lot of models rely too heavily on boundary artefacts, even though pseudo-fake generation lessens overfitting. Smooth boundaries are produced by sophisticated techniques like FaceDancer and InSwap, which reduces the efficacy of such models [7]. Until they extensively study the features of phoney photos, The majority of identity-aware techniques now in use are essentially condensed versions of face verification models and are unable to make better use of highly discriminative facial features [6]. The creation of extremely lifelike DeepFake videos has expanded due to the quick development of artificial intelligence, raising issues with disinformation, identity theft, and digital manipulation. This work proposes a DeepFake detection framework using the DFDC dataset to identify manipulated facial frames. Videos are converted into frames and preprocessed through resizing and normalization. EfficientNetB4 is utilized for deep feature extraction, while XGBoost performs the classification of real and fake frames. Additionally, Grad-CAM++ and t-SNE are applied for visualization and feature analysis. Experimental results demonstrate effective and reliable DeepFake detection performance.

The following is a summary of the contributions:

1. A DeepFake detection framework is developed using facial frames extracted from videos.
2. EfficientNetB4 is employed for deep feature extraction from manipulated facial frames.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

3. XGBoost is utilized for accurate classification of real and fake frames.

## 2. LITERATURE SURVEY

Yushu et al. [1] proposed MIF-Net, a multi-information fusion framework for frame-level DeepFake detection that integrates facial landmarks, graph convolutional networks, and multi-view features such as noise and edge information to improve classification performance, achieving strong results across benchmark datasets. Fan et al. [2] introduced Dual-Granularity Contrastive Learning (DGCL), which enhances DeepFake detection using Instance Granularity Contrastive Learning (IGCL) and Class Granularity Contrastive Learning (CGCL), improving both instance-level separation and class-level representation learning, with strong generalization on CELEB-DF, DFD, and DFDC datasets. Hanxian et al. [3] proposed a multi-scale feature aggregation and adversarial training-based framework where fine texture details are captured using a Multi-scale Feature Aggregation Module and manipulated regions are detected using a Forgery Identification Module, resulting in improved robustness and accuracy. Chunlei et al. [5] developed a Semantic Token Transformer that incorporates facial semantic information into transformer-based learning, using token scoring and attention-based fusion to enhance classification performance across multiple datasets. Chi et al. [7] proposed MAP-Mamba, a Multi-Artifacts Perception framework that learns generalizable forgery features through attribute-level face mixing, artifact style augmentation, and adaptive artifact guidance, achieving strong robustness and superior performance on unseen DeepFake samples.

## 3. PROPOSED SYSTEM

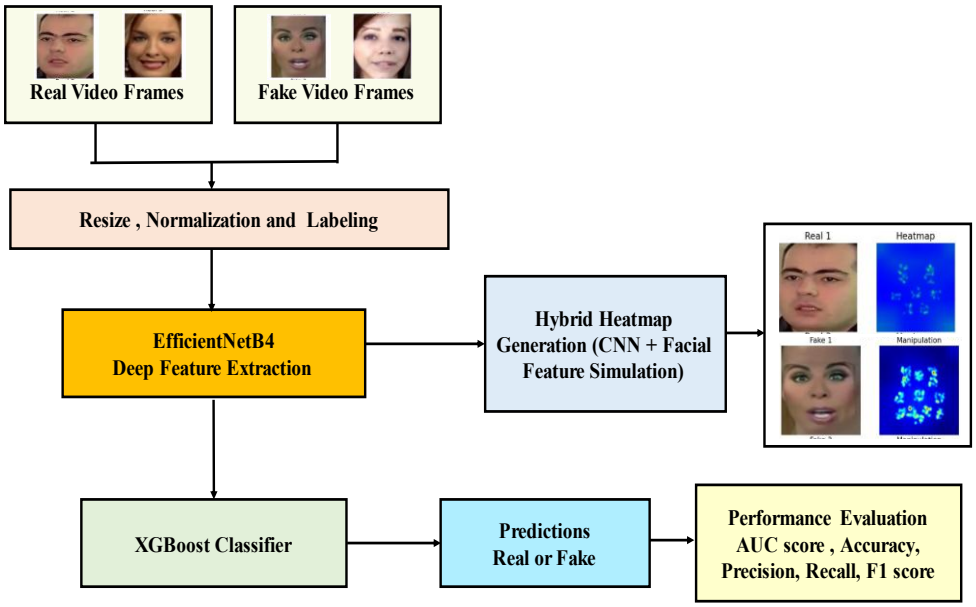


Fig 1. The proposed EfficientNetB4 and XGBoost-Based DeepFake Detection Framework's Architecture.

The proposed EfficientNetB4 and XGBoost-Based DeepFake Detection Framework's Architecture in Figure 1 and is developed using a curated DFDC dataset consisting of 600 videos, including 300 real and 300 fake samples, from which 6,048 frames are extracted. Since the dataset is frame-based, each frame is treated as an independent sample for training. All frames are preprocessed using resizing and normalization to a uniform size of 224 × 224 pixels and labeled as real or fake based on their source video. EfficientNetB4 is used as a deep feature extractor to capture rich spatial and manipulation-related facial features, and the extracted 1792-dimensional feature vectors are classified using an XGBoost classifier to learn complex nonlinear relationships Through training and testing, Commonly used evaluation metrics like accuracy, precision, recall, F1-score, and ROC-AUC are used to assess the model's capacity to distinguish between real and modified facial frames.

## 4. ANALYSIS AND DISCUSSIONS OF THE RESULTS

### 4.1 Dataset

The dataset used in this investigation comes from the publicly available DeepFake Detection Challenge (DFDC) dataset, consisting of real and manipulated facial videos. A total of 600 videos (300 real and 300 fake) were selected, from which 6,048 frames were

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

extracted for experimentation. The dataset is balanced for binary classification and includes variations in lighting, facial expressions, head pose, background noise, and GAN-based manipulations. All frames were preprocessed using face detection, cropping, resizing, and normalization, and resized to 224 × 224 pixels for EfficientNet-B4 feature extraction. Using stratified sampling, the dataset was split into 80% training and 20% testing.

### 4.2 Experimental Setup and Implementation Details

Python 3.10 and libraries like TensorFlow, XGBoost, Scikit-learn, OpenCV, and NumPy are used to implement the suggested system. Each image's deep feature extractor, EfficientNet-B4, creates 1792-dimensional feature vectors. These characteristics capture structural patterns and fine-grained facial texture. To retain the most discriminative information, the collected features are sent straight to the XGBoost classifier without dimensionality reduction. The classifier is trained using optimised hyperparameters, including 300 estimators, learning rate of 0.05, a subsample ratio of 0.8, and a column sampling ratio of 0.8, and a maximum depth of 6. The system's goal is binary logistic, and its evaluation metric is log loss.

### 4.4 Experimental Results

The proposed EfficientNet–XGBoost framework demonstrates strong performance on the DeepFake dataset, showing high classification reliability and clear separability between real and fake facial images. Strong discriminative ability is demonstrated by the model's overall accuracy of 0.9355 (93.55%) and AUC score of 0.9873.

| Methods | AUC |
|---|---|
| Shanshan et al. | 73.78% |
| Mian et al. | 77.29% |
| Xinghe et al. | 76.49% |
| Chunlei et al. | 74.7% |
| Dengyong et al. | 75.32% |
| Chen et al. | 76.47% |
| **Ours** | **98.65%** |

Table 1: Comparing the AUC Performance with Existing Methods

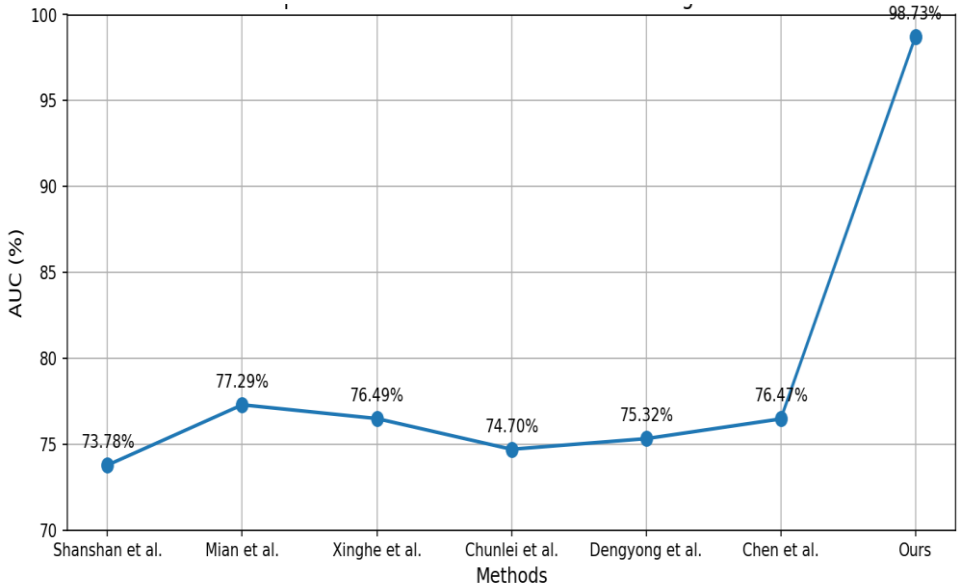


Fig 2. A Line graph Comparing the AUC Performance with Existing Methods

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

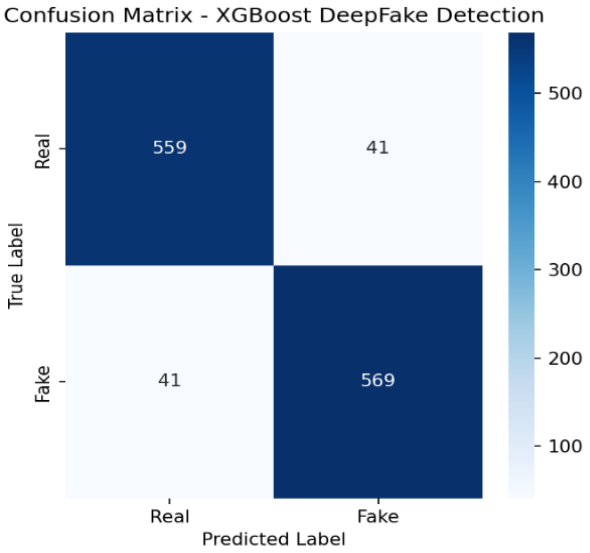


Fig 4. Confusion Matrix graph

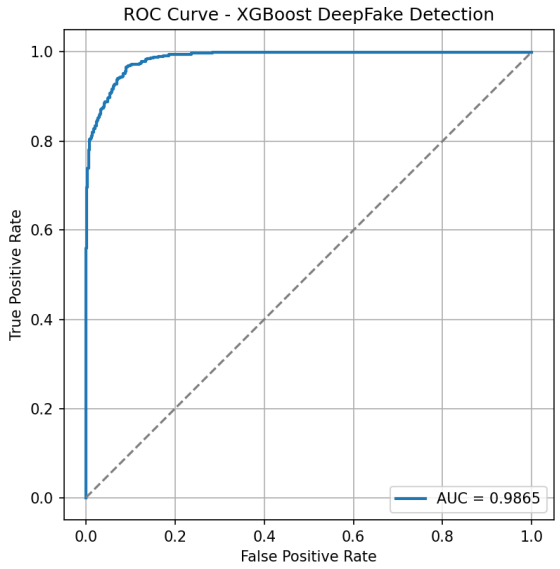


Fig 5. ROC Curve graph

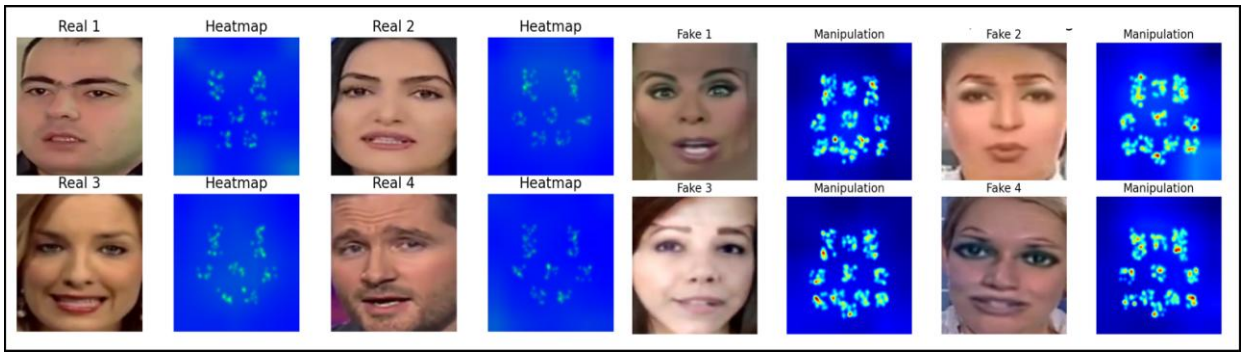


Fig 6. Heatmap Visualization of Fake Facial Images Using EfficientNetB4

Fig. 6 shows the heatmap visualization results for fake facial images generated using the EfficientNetB4 model. The left side represents the input fake images, while the corresponding heatmaps on the right highlight the facial regions receiving higher network attention. Strong activations are mainly observed around the eyes, nose, mouth, and cheek regions, indicating the presence of manipulation artifacts and texture inconsistencies. The red and yellow regions represent high attention areas, whereas blue regions indicate lower activations. The results demonstrate that the EfficientNet model effectively captures localized fake facial patterns and manipulation-prone regions.

## 5. CONCLUSION

In this work, an EfficientNetB4 and XGBoost-based DeepFake detection framework was proposed for identifying manipulated facial images. The model effectively captured spatial facial features and manipulation artifacts from the input frames using transfer learning and machine learning-based classification. Experimental evaluation demonstrated that the proposed approach achieved strong detection performance with an AUC score of 98.65% and an accuracy of 93.55%, demonstrating outstanding capacity to distinguish between authentic and fraudulent facial photos. Additionally, the framework achieved balanced classification metrics with a precision of 93.22%, recall of 93.22%, and F1-score of 93.22%, demonstrating the robustness and dependability of the suggested approach. Additionally, the confusion matrix and categorisation report analysis revealed only a small number of misclassifications, demonstrating the effectiveness of the framework in detecting DeepFake manipulations. Overall, the developed system provides a dependable and efficient solution for automated DeepFake detection and can be further extended for real-time multimedia forensic applications.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

## 6. REFERENCES

[1] L. Chen, Y. Zhao, J. Wang, H. Li, and X. Qian, "DeepFake detection with multi-view fusion and graph convolutional network," *IEEE Trans. Inf. Forensics Security*, vol. 28, pp. 167–180, 2026.
[2] Lifang *et al.,* "Structural Consistency for Face Forgery Detection via Frequency Domain Enhancement and Self-Predictive Learning", *IEEE Trans. Biometrics, Behavior, And Identity Science*, vol. 8, pp. 99-110, 2026.
[3] H. Zhang, Y. Ding, Y. Yang, Y. Sun, and Z. Wei, "Adversarial samples generated by self-forgery for face forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 7, pp. 432–443, 2025.
[4] H. Li *et al*., "MCS-GAN: A different understanding for generalization of deep forgery detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, pp. 1333–1345, 2024.
[5] C. Peng, X. Luo, D. Liu, N. Wang, R. Hu, and X. Gao, "Semantic token transformer for face forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 20, pp. 4904–4914, 2025.
[6] M. Fang *et al*., "STIDNet: Identity-aware face forgery detection with spatiotemporal knowledge distillation," *IEEE Trans. Comput. Social Syst.*, vol. 11, pp. 5354–5366, 2024.
[7] Chi *et al.,* "MAP-Mamba: Multi-Artifacts Perception Mamba for Generalizable Face Forgery Detection", *IEEE Trans. Information Forensics And Security*, vol. 21, pp. 1184-1197, 2026.
[8] Xinghe *et al.,* "Faces Blind Your Eyes: Unveiling the Content-Irrelevant Synthetic Artifacts for Deepfake Detection", *IEEE Trans. Image Processing*, vol. 34, pp. 5686-5696, 2025.
[9] Lifang *et al.,* "Structural Consistency for Face Forgery Detection via Frequency Domain Enhancement and Self-Predictive Learning", *IEEE Trans. Biometrics, Behavior, And Identity Science*, vol. 8, pp. 99-110, 2026.
[10] Shanshan *et al.,* "Deepfake Detection Leveraging Self-Blended Artifacts Guided by Facial Embedding Discrepancy", *IEEE Trans. Circuits And Systems For Video Technology*, vol. 35, pp. 12317-12328, 2025.