International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

Leveraging EfficientNet for Deepfake Detection via Transfer Learning

Manpreet[1], Dr. Saurabh Sharma[2]

Research Scholar[1], Associate Professor[2]
Department of Computer Science and Applications, Sant Baba Bhag Singh University
Jalandhar, Punjab, India

How to Cite this Article:
Manpreet, (2026). Leveraging EfficientNet for Deepfake Detection via Transfer Learning. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.916

https://doi.org/10.55041/ijcope.v2i4.916

Abstract: — Deepfake technology has witness quick succession in previous years due to significant progress in deep learning and generative models, mainly Generative Adversarial Networks (GANs) and autoencoders. These technologies permit the foundation of extremely practical manufactured media, counting manipulated pictures and videos that are frequently hard to separate from genuine content. Whereas deepfakes offer valuable applications in areas such as entertainment, media generation,andvirtualsimulationsthey too present genuine dangers, countin g deception spread, personality robbery, political control, and cybercrime. As a result, the improvement of consistent and computerized deepfake detection systems has suit a critical area of research in digital forensics and cybersecurity. This research proposes a productive and scalable deepfake detection system based on the EfficientNet architecture built-in with transfer learning techniques. EfficientNet is selected due to its ability to attain high accuracy with less restriction through its complex scaling method, which balances connection profundity, width, and determination. The utilize of transfer learning allows the model to influence pre-trained weights from large-scale datasets, enabling earlier convergence, reduced computational cost, and enhanced performance even with imperfect labeled data.

Keywords: Deepfake Detection, EfficientNet, Transfer Learning, CNN, Image Classification, Digital Forensics

I. INTRODUCTION

Later progression in artificial intelligence (AI) and deep learning have driven to quick improvement of strategies able of creating profoundly realistic artificial media, commonly referred to as deepfakes. These deepfakes are unusually made or manipulated images, videos, or audio formed using difficult deep learning models such as Generative Adversarial Networks (GANs) and autoencoders. By learning compound plans from broad datasets, these models can replicate facial expressions, voice designs, and visual subtle elements with excellent accuracy, making it dynamically hard for peoples to recognize between real and fake content.

While deepfake development offers a few helpful applications in areas such as pleasure, filmmaking, virtual reality, and digital content creation, its mistreatment has raised actual concerns. Deepfakes have been broadly related with the spread of deception, identity theft, financial fraud, cybercrime, and political control. The ability to make persuading fake media poses an important threat to digital believes public safety, and information correctness. As a result, there is an upward require for dependable and automatic systems that can efficiently distinguish and moderate the risks connected with deepfake content.

Traditional detection methods, which depend on manual evaluation or handcrafted features, are no longer sufficient to handle the growing complexity and realism of modern deepfake era procedures. These approaches usually fail to identify subtle visual artifacts and irregularities nearby in high-quality manipulated media. Consequently, deep learning-based detection systems have risen as a more doing well solution, as they can mechanically learn discriminative features

directly from information and adjust to evolving manipulation techniques.

Among different approaches, transfer learning has gained important implication in deepfake detection tasks. Transfer learning permits models to use in sequence from pre-trained networks, typically ready on large-scale datasets, and adjust it to exacting tasks with limited data. This approach not only reduces preparation time but moreover improves model presentation and generalization ability, making it extremely suitable for deepfake detection scenarios.

EfficientNet, a state-of-the-art difficulty neural network design, has exposed exceptional execution in image organization assignments due to its inventive compound scaling method. Unlike traditional models that scale network capacity arbitrarily, EfficientNet logically balances network profundity, width, and input resolution, resulting in enhanced accuracy with less parameters and reduced computational cost. This makes it particularly sensible for applications needful both effectiveness and in height concert.

In this research, EfficientNet is mutual with transfer learning to generate a well-built and capable deepfake detection system. The proposed advance aims to precisely classify media substance as genuine or fake by capturing delicate handling artifact while keeping up computational efficiency. The study focuses on building an adaptable explanation that can be associated to real-world scenarios, contributing to the improvement of digital forensics and secures mixed media systems.

II. LITERATURE REVIEW

Deepfake detection has evolved radically over the past decade, driven by rapid advancements in artificial intelligence and generative modeling techniques. As deepfake generation methods have become more sophisticated, detection approaches have also progressed from easy handcrafted techniques to advanced deep learning-based models.

In the early stages, deepfake detection relied seriously on handcrafted features and traditional machine learning algorithms. Researchers focused on identifying observable inconsistencies such as irregular eye blinking patterns, abnormal facial expressions, and mismatched skin textures. These features were then used with classifiers like Support Vector Machines (SVM), K-Nearest Neighbors (KNN) and conclusion trees. Although these methods laid the foundation for deepfake detection, they lacked robustness and unsuccessful to generalize well to high- quality and complex deepfake content.

With the rapid improvement of Generative Adversarial Networks (GANs), deepfake generation techniques enhanced significantly, producing highly realistic outputs. This shift run to the acceptance of convolutional neural networks (CNNs) for detection tasks. CNN-based models such as MesoNet and XceptionNet demonstrated improved concert by automatically learning spatial and texture-based features from images. These models were able of identifying subtle artifacts introduced during the treatment process. However, they required large labeled datasets for effective training and often struggled to generalize crossways different datasets and unseen manipulation techniques.

To overcome these limitations, researchers explored substitute approaches that focus on more intrinsic and generalizable features. Frequency-domain analysis emerged as a capable technique, where models analyze anomalies in the frequency spectrum of images, as deepfake generation methods often leave detectable traces in the frequency domain. Additionally, physiological signal-based methods were introduced, which detect inconsistencies in biological patterns such as heart rate, blood flow, and subtle facial color changes.

These approaches improved strength but often required compound preprocessing and additional computational resources.

More recently, transformer-based architectures have gained attention in deepfake detection. Vision Transformers (ViTs) and their variants are capable of capturing long range dependencies and global contextual information within images and videos.

These models have shown hopeful results in detecting complex manipulations; however, they are computationally costly and require significant training data and resources.

Transfer learning has become a key method in modern deepfake detection systems due to the limited availability of high-quality labeled datasets. By leveraging pre trained models such as ResNet, DenseNet and MobileNet researchers can reuse learned feature representations and adapt them to the deepfake detection task. This approach significantly reduces training time, improves concert and enhances generalization across different datasets.

Among the recent advancements, EfficientNet has emerged as a highly effective design for deepfake detection. It introduces a complex scaling method that systematically balances network depth, width, and input resolution, leading to enhanced performance with fewer parameters.

EfficientNet models offer strong feature extraction capabilities, enabling the detection of fine-grained artifacts and subtle inconsistencies in manipulated media. Furthermore, its computational competence makes it suitable for real time applications and deployment in resource- constrained environments.

Recent studies conducted between 2023 and 2025 indicate that EfficientNet-based models consistently perform well across various benchmark datasets, including FaceForensics++ and the DeepFake

Detection Challenge (DFDC). These models demonstrate high accuracy, robustness to compression and noise, and strong generalization capabilities. As a result, EfficientNet combined with transfer learning has become a preferred approach for developing scalable and efficient deepfake detection systems.

III. RESEARCH METHODOLOGY

This study follows an experimental research methodology to develop a proficient deepfake detection system using the EfficientNet model and transfer learning. Initially, publicly available datasets such as FaceForensics++ and the DeepFake Detection Challenge (DFDC) are used, which contain both real and manipulated media samples. The collected video data is preprocessed by extracting frames and detecting facial regions using standard face detection techniques. These extracted faces are then resized and normalized to maintain consistency in input data.

A. Research Design

This study adopts an experimental approach using supervised learning. The model is trained on labeled datasets containing real and fake media samples.

B. Dataset Selection

The model uses the techniques like:

a) FaceForensics++
b) DeepFake Detection Challenge (DFDC)

C. Data Preprocessing

Preprocessing plays a crucial role in improving model performance:

a) Frame Extraction: Videos are converted into frames

b) Face Detection: Faces are extracted using MTCNN or Dlib

c) Normalization: Images resized to 224×224

d) Data Augmentation: Rotation, flipping, zooming, brightness adjustment

D. Model Architecture

The proposed system uses EfficientNet (B0–B4 variants) with transfer learning:

a) Pre-trained on ImageNet
b) Initial layers are frozen
c) Final layers are fine-tuned

Custom Layers Added:

d) Global Average Pooling
e) Dense Layer (ReLU)
f) Dropout Layer
g) Output Layer (Sigmoid/SoftMax)

E. Model Training

Training configuration includes:

a) Optimizer: Adam
b) Loss Function: Binary Cross-Entropy
c) Batch Size: 16–64
d) Learning Rate Scheduling

Fine-tuning is performed by unfreezing top layers to improve feature learning.

F. Model Evaluation

Performance is evaluated using:

a) Accuracy
b) Precision
c) Recall
d) F1-score
e) Confusion Matrix

The model is also tested for robustness against:

f) Low-resolution images
g) Compressed videos
h) Noisy data

G. Implementation & Deployment

The proposed system is implemented as a web- based application to provide an accessible and user- friendly interface for deepfake detection. The web application is developed using frameworks such as Flask or FastAPI, allowing users to upload images or videos and receive real-time predictions.

Although the current implementation focuses on a web-based platform, the system architecture is designed to be flexible and can be extended to other platforms in the future, if required.

Pipeline:

a) Upload media
b) Extract frames
c) Detect faces
d) Predict real/fake
e) Display result

IV. RESULTS

The proposed EfficientNet-based deepfake detection model demonstrates strong performance in identifying manipulated media content. The model effectively distinguishes between real and fake samples by learning complex visual patterns and subtle inconsistencies present in deepfake images and videos. The use of transfer learning further enhances the model's capability by utilizing pre-trained knowledge, resulting in improved accuracy and reduced training time.

A. High Detection Accuracy:

The model achieves high accuracy in classifying real and fake media samples, indicating its strong capability in identifying deepfake content. This high accuracy is achieved due to EfficientNet's ability to extract meaningful and discriminative features from input data. The model performs consistently well across training and testing datasets, showing reliable classification performance.

a) Highlight author and affiliation lines of affiliation 1 and copy this selection.
b) Formatting: Insert one hard return immediately after the last character of the last affiliation line. Then paste down the copy of affiliation 1. Repeat as necessary for each additional affiliation.
c) Reassign number of columns: Place your cursor to the right of the last character of the last affiliation line of an even numbered affiliation Go to Column icon and select "2 Columns". If you have an odd number of affiliations, the final affiliation will be centered on the page; all previous will be in two columns.

B. Strong Generalization Ability :

The model demonstrates good generalization across different datasets and unseen samples. It is capable of detecting various types of deepfake manipulations, even those not explicitly present in the training data. This indicates that the model has learned generalized features rather than memorizing specific patterns, making it more suitable for real-world applications.

C. Computational Efficiency:

EfficientNet requires significantly fewer parameters compared to traditional convolutional neural networks, which reduces the computational cost of training and inference. This efficiency allows the model to be trained faster and deployed easily on systems with limited hardware resources, making it practical for real-time applications.

D. Improved Performance over Traditional Models:

When compared with conventional architectures such as ResNet and VGGNet, the proposed model shows improved performance in terms of both accuracy and efficiency. While traditional models often require more computational power, EfficientNet achieves better results with optimized resource usage, making it a more effective solution.

Table 1 Comparison of Traditional Deepfake Detection and EfficientNet-Based Approach

Features	Traditional Deepfake Detection	EfficientNet-Based Approach
Detection Method	Uses handcrafted features (eye blinking, texture)	Uses deep learning feature extraction
Accuracy	Moderate	High
Feature Learning	Manual feature engineering	Automatic feature learning
Computational Efficiency	Low to moderate	High (optimized architecture)
Model Complexity	Simple models	Optimized deep neural network
Training Time	Higher due to manual tuning	Reduced using transfer learning
Adaptability	Less adaptable to new techniques	Easily adaptable with fine-tuning

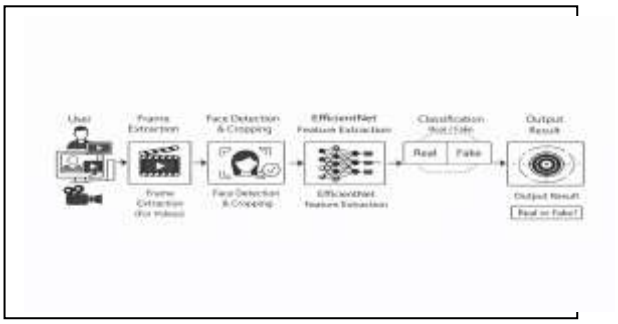

Figure.2 Result that is shown after the detection of the image

V. CONCLUSION

This research presents an efficient and reliable deepfake detection system based on the EfficientNet architecture combined with transfer learning. The proposed approach effectively addresses the growing challenges posed by sophisticated deepfake generation techniques. By leveraging pre-trained models, the system is able to learn complex visual patterns and identify subtle inconsistencies in manipulated media with high accuracy.

The experimental results demonstrate that EfficientNet provides a strong balance between performance and computational efficiency. Compared to traditional convolutional neural network architectures, the proposed model achieves improved accuracy while using fewer parameter and reduced processing time. This makes it a suitable choice for practical applications where both speed and accuracy are critical.

EfficientNet proves to be particularly effective in capturing fine-grained features such as texture inconsistencies, facial distortions, and pixel-level artifacts that are commonly present in deepfake content. The integration of transfer learning further enhances the model's capability by reducing training complexity and improving generalization across different datasets and manipulation techniques.

Overall, the proposed system contributes to the field of deepfake detection by offering a scalable, efficient, and accurate solution. Its implementation as a web-based application also demonstrates its usability in real-world scenarios, providing users with an accessible platform for detecting manipulated media.

REFERENCES

[1]. Li, Y., et al. "Celeb-DF: A large-scale dataset for deepfake forensics," CVPR, 2020.

[2]. Dolhansky, B., et al. "DeepFake Detection Challenge Dataset," 2020.

[3]. Tolosana, R., et al. "Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection," Information Fusion, 2020.

[4]. Dang, S., et al. "Deepfake Detection Survey," ACM Computing Surveys, 2021.

[5]. Dosovitskiy, A., et al. "An Image is Worth 16x16 Words: Vision Transformers," ICLR, 2021.

[6]. Tan, M., and Le, Q. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," ICML, 2019.

[7]. B. Dolhansky et al., "The DeepFake Detection Challenge (DFDC) dataset," *arXiv: 2006.07397*, 2020.

[8]. G. Guarnera, M. Barni and A. Del Bimbo, "A survey on deepfake detection: Data, methods and evaluation," *Information Fusion*, 2022.

[9]. M. Banerjee et al., "Deepfake detection using transfer learning," *IEEE ICCCNT*, 2021.

[10]. Dang et al., "Deep learning-based face manipulation detection: A survey," *ACM Computing Surveys*, 2021.

[11]. Khan, S., et al. "Transformer-Based Deepfake Detection: A Survey," IEEE Access, 2023.