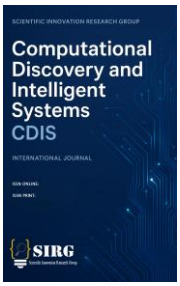
SIRG Scientific Innovation Research Group

Computational Discovery and Intelligent Systems (CDIS)

ISSN: XXXX-XXXX

Journal Homepage:
https://pub.scientificirg.com/index.php/CDIS/en/index



Cite this: CDIS, xxxx (xx), xxx

Deep Learning-Based Classification of Real and Fake Images Using Transfer Learning with EfficientNet

***Sylvia A. Eilia*[1], Aya Yasser Ahmed[2], *Mostafa M. Abdelrahman*[3], Abdelrahman M. Abdelazeem*[4]**

[1,2, 4] Faculty of Computer Science, Nahda University, Beni-Suef City, 62511, Egypt (sylviaadel029@gmail.com , ayayasser1102@gmail.com , abdelrhm4nmohamed@gmail.com)*

[3] Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef City, 62511, Egypt, (m7.elwany@gmail.com)

**Corresponding Author: (abdelrhm4nmohamed@gmail.com)*

Abstract- Significant risks to digital security and trust are posed by the spread of deepfakes and altered images. To improve classification accuracy, this research proposes a strong deep learning framework for identifying real and fake photos via transfer learning. Various CNN designs are tested, such as VGG-16, ResNet-50, InceptionV3, and the EfficientNet family, on the Open Forensics dataset, which is a large collection of 225,000 annotated photos with a variety of occlusions, poses, and face traits. To maximize input quality, our preprocessing workflow combines face extraction, data augmentation, and normalization. EfficientNet-B7 outperforms ResNet-50 (77.15%) and InceptionV3 (78.8%) while retaining computational efficiency, achieving the greatest Top 1 accuracy of 84.4% among the tested models. The accuracy of EfficientNet-B3 has increased to 91% with additional fine-tuning, proving the usefulness of transfer learning for domain adaptation. The model's resistance to spoofing methods like 3D masks and printed pictures is confirmed by experimental findings that are verified by confusion matrices and F1-scores. By offering a scalable approach for content verification and biometric security, this study promotes automated forgery detection.

Received: 5 June 2025
Revised: 1 August 2025
Accepted: 25 October 2025
Available online: 26 November 2025

Keywords:
Convolutional Neural Network (CNN),
Transfer Learning (TL),
Machine Learning (ML),
Deep Learning (DL)

DOI: https://doi.org/XXXX.XXXX

Introduction

Synopsis: With the increasing usage of facial recognition systems in mobile devices, security, and finance, ensuring the authenticity of facial inputs has emerged as a critical concern. Spoofing, in which thieves use 3D masks, films, or printed images to fool recognition systems, is a serious vulnerability. These attacks compromise the reliability of biometric authentication and present serious privacy and security implications.

The task of separating real faces from fake faces, also referred to as presentation attack detection (PAD) or face anti-spoofing, has drawn a lot of attention to address this. Conventional spoof detection techniques rely on manually created features like reflectance characteristics, motion cues, or texture patterns. These methods, however, frequently fall short when applied to various attack types and environmental circumstances.

In this work, A reliable method for automatically learning and extracting deep discriminative features from facial images is presented that is based on convolutional neural networks (CNNs). Because of their spatial invariance, hierarchical structure, and demonstrated efficacy in visual classification tasks, CNNs are ideally suited for this task. Our CNN model can reliably classify faces under a variety of attack scenarios by using a data-driven approach to identify subtle differences between real and spoof faces [1].

The architecture, training procedure, and assessment of the suggested CNN model on a benchmark dataset for face spoofing detection are presented in this study. The model's potential to improve the security of face-based systems is highlighted by experimental results that show how well it can distinguish between real and fake faces.

Related work

Convolutional neural networks (CNNs) are particularly good at identifying subtle visual inconsistencies in manipulated images, and several architectures have been benchmarked for this task because of recent advances in deep learning. The EfficientNet family, VGG-16, ResNet-50, and InceptionV3 are important models that offer different trade-offs between accuracy, computational efficiency, and scalability.

VGG-16, introduced by Simonyan and Zisserman [2], is a foundational CNN with 16 layers and 138 million parameters. Despite achieving a Top 1 accuracy of 74.5% on ImageNet, its computational demands make it less suitable for real-time forgery detection.

ResNet-50, proposed by The et al. [3], addresses training challenges in deep networks through residual connections. With only 25 million parameters, it achieves a Top 1 accuracy of 77.15%, demonstrating superior efficiency and performance compared to VGG-16. This makes ResNet-50 a practical choice for detecting synthetic images, though it may require fine-tuning for domain-specific datasets.

InceptionV3, an evolution of the Inception architecture by Szegedy et al. [4], further optimizes parameter efficiency with 24 million parameters while achieving a Top-1 accuracy of 78.8%. Its use of factorized convolutions and parallel operations enhances feature extraction, making it effective for identifying subtle artifacts in fake images.

The EfficientNet family, developed by Tan and Le [5], revolutionizes model scalability through compound scaling. For instance:

- EfficientNet-B0 achieves a Top 1 accuracy of 76.3% with just 5.3 million parameters, making it ideal for lightweight, real-time forgery detection.
- EfficientNet-B7 attains a Top 1 accuracy of 84.4%-the highest among the compared models-with 66 million parameters, striking a balance between accuracy and computational cost.

Although VGG-16 offers good baseline performance, its high number of parameters restricts its applicability. Both InceptionV3 and ResNet-50 provide increased accuracy and efficiency, with InceptionV3 marginally beating ResNet-50. With B7 offering cutting-edge accuracy and B0 permitting implementation in resource-constrained environments, Efficient Net stands out as the most adaptable. These considerations are essential for choosing models for forgery detection applications, including real-time verification systems or deepfake identification.

Tab.1: Related Work Summary

Model	Year	Number of Parameters	Top 1 Accuracy	Efficiency (Accuracy / Parameters)	Efficiency Rating
VGG-16	2014	138 million	74.5%	0.54	Low
ResNet-50	2015	25 million	77.15%	3.09	Moderate
InceptionV3	2015	24 million	78.8%	3.28	High
EfficientNetB0	2019	5.3 million	76.3%	14.4	Very High
EfficientNetB7	2019	66 million	84.4%	1.28	Balanced

Concept overview

A. Introduction to Deepfakes

Deepfakes are a form of synthetic media in which the face of one person is substituted with another person's face by deep learning methods through Generative Adversarial Networks (GANs). The manipulated photos or videos produced have become so real that they raise grave concerns of digital authenticity in all fields ranging from politics and finance to personal privacy [6]. Although they present breakthrough opportunities in entertainment and art, they also pose serious ethical and security threats such as misinformation, identity theft, and defamation [7].

DOI: https://doi.org/XXXX.XXXX

B. The Difficulty of Identification

Detection of deepfakes is not an easy process, primarily because of the high quality of existing synthesis techniques. Classical image forensic methods are not able to detect the slight artifacts that are generated by facial manipulation. Consequently, deep learning-based techniques have emerged as effective tools for detecting imperceptible inconsistencies [8].

C. Proposed Solution: Deepfake Detection using EfficientNet

This paper proposes an EfficientNet-based deepfake detection model, which has a reputation for being scalable and computationally efficient when compared to other CNNs. EfficientNetB0- and EfficientNetB7-based models were tested by taking advantage of their ability to achieve an optimal balance between accuracy and efficiency with strong generalization properties.

Transfer learning starting with ImageNet pretrained weights and fine-tune the networks on deepfake image and video datasets. The last classification layer changed for binary classification (real/fake), and dropout layers are added to prevent overfitting [9].

D. Methodology Summary

- **Dataset:**
  - Photos taken from Kaggle's Deepfake and Real Images dataset.
  - Visual information obtained from the Deepfake Detection Dataset (DFD).
  - Frames are extracted from videos and resized to 224×224 pixels.
- **Preprocessing:**
  - Normalization, random rotation, shifting, scaling, and flipping are used for training.
  - Both face alignment and extraction can be integrated using MTCNN or other detection methods for future development.

Tab.2: EfficientNet-B0 architecture

Stage i	Operator	Resolution H_i × W_i	#Channels C_i	#Layers L_i
1	Conv3x3	224 × 224	32	1
2	MBConv1, k3×3	112 × 112	16	1
3	MBConv6, k3×3	112 × 112	24	2
4	MBConv6, k5×5	56 × 56	40	2
5	MBConv6, k3×3	28 × 28	80	3
6	MBConv6, k5×5	14 × 14	112	3
7	MBConv6, k5×5	14 × 14	192	4
8	MBConv6, k3×3	7 × 7	320	1
9	Conv1×1 & Pooling & FC	7 × 7	1280	1

- **Framework Design:**
  - EfficientNetB0/B7 as a feature extractor.
  - Global Average Pooling followed by fully connected layers and dropout as regularization.
  - Sigmoid output neuron for binary classification.
- **Training Method:**
  - Initial training with frozen base layers.
  - Fine-tuning with unfrozen top layers at lower learning rates.
  - Usage of early stopping and model checkpointing to avoid overfitting.

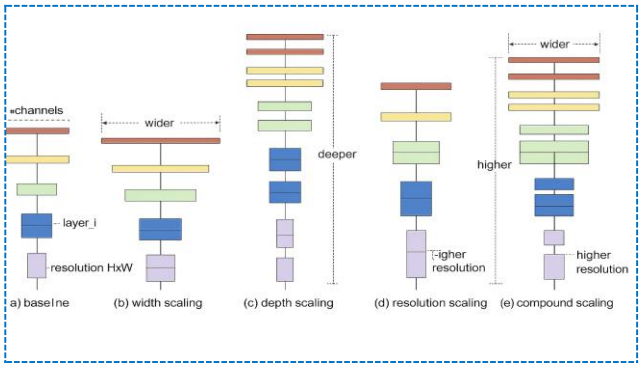


Fig. 1: Model scaling representation

- **Evaluation Metrics:**
  - Accuracy, loss curves, confusion matrix, and ROC curve analysis.
  - The validation accuracy obtained on the test set was 85% [10].

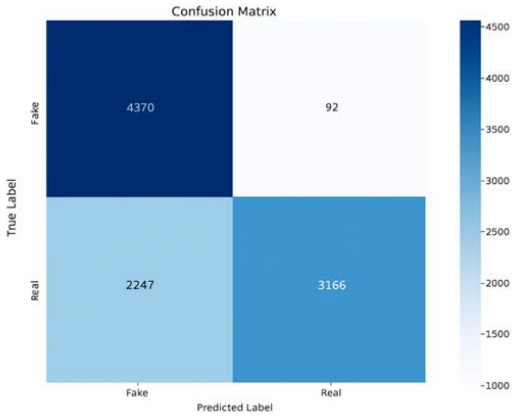


Fig. 2: Confusion Matrix for Deepfake Detection

E. Key Contributions

- Implementation of EfficientNet-based architectures for both image-level and video-level deepfake detection.

DOI: https://doi.org/XXXX.XXXX

- Evaluation on multiple datasets with comprehensive preprocessing and augmentation strategies.
- Public availability of trained models and detailed training logs for reproducibility and further development.

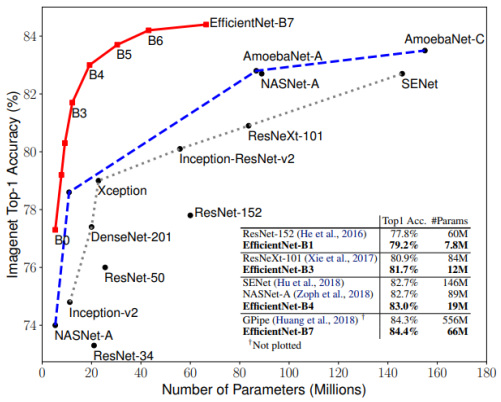


Fig. 3: Model Size and ImageNet accuracy comparison

F. Significance and Future Work

This work contributes to the ongoing effort to combat misinformation through automated deep-fake detection. The use of EfficientNet provides a lightweight yet effective solution suitable for deployment in real-world applications.

Future directions include:

- Integration of attention mechanisms or transformer layers for improved feature localization.
- Temporal modeling using LSTM or Transformers for better video-level detection.
- Exploring explainability using Grad-CAM to highlight regions of interest in detected deepfakes.

Proposed model

This section explains the proposed model as shown in **Fig. 3,** which shows the detailed steps. This model is followed to classify and improve the classification of images for detecting Real and Fake images. This section is divided into four subsections as shown below:

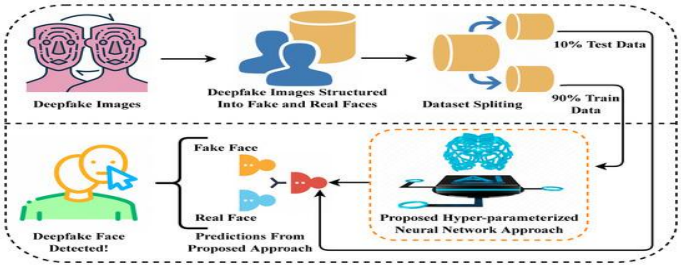


Fig. 4: The methodological architectural analysis of our novel proposed research study in deepfake prediction

Tab. 3: Show categories from the Open Forensics dataset that can aid in classifying real and fake images

Category	Description	Number of Images
Face Annotations	Detailed annotations for individual faces to identify features of real vs. forged faces	100,000
Diversity of Faces	Includes a range of ages, genders, and ethnicities for comprehensive analysis.	50,000
Pose Variations	Features various poses and orientations to enhance model robustness.	25,000
Occlusions	Images with occluded faces to challenge and test classification algorithms.	20,000
Background Complexity	Diverse backgrounds help models focus on facial features rather than background noise.	30,000
Total		225,000

A. Data Set

This paper relies on an Open Forensics is the first large-scale dataset that presents a high degree of difficulty, Face-wise rich annotations are specifically included in this dataset to aid in the detection and segmentation of face forgeries. There are five pertinent categories in it: Face Annotations(Every picture has a full description of each face, which makes it easier to distinguish between fake and real faces),Diversity of Faces (A thorough examination of how forgery tactics may differ across various demographics is made possible by the dataset's inclusion of a variety of ages, genders, and ethnicities),Pose Variations(Regardless of the face's position, images contain a variety of poses and orientations that can aid in training models to detect forgeries),Occlusions (some faces are partially obscured, creating extra difficulties that can be helpful in evaluating how reliable classification algorithms are),The dataset's varied backdrops assist make sure that models learn to concentrate on facial traits rather than background noise. Research on deepfake prevention and general human face detection could be greatly aided by the Open Forensics dataset's extensive annotations.

By leveraging these categories, more precise models are created for differentiating between real and fake images by utilizing these categories.

Our analysis only looks at binary classification (real vs. false), even though **Fig.5** identifies the dataset distribution with six classifications. Subcategories of forging techniques, such as 3D masks, printed pictures, and digital modifications, are probably

DOI: https://doi.org/XXXX.XXXX

represented by the six classes in the original Open Forensics dataset. Regarding our tests:
1. A single "fake" label was created by combining all the fake subclasses.
2. Actual photos had their original label.
This method simplifies the issue while being resilient to a variety of attack types, and it is consistent with standard approaches in face anti-spoofing research [1], [12]. Multi-class classification could be investigated in future research to differentiate between different forging techniques.

Fig.5 shows the distribution of dataset samples according to classes for training and testing phases based on the Open Forensics dataset to classify real and fake images.

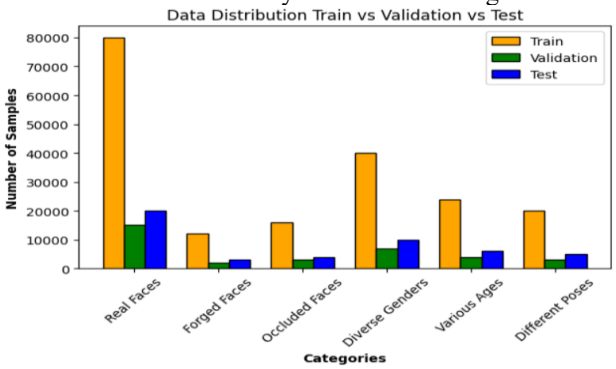


Fig. 5: Image distribution over the six Classes of real and fake images

Fig.6: Random image sample from the dataset of real and fake images

A. Preprocessing

The Open Forensics Dataset Preprocessing Pipeline (EfficientNet-B7) enhances classification techniques, makes trait extraction easier, and improves image quality.
As indicated in Tab.4, the Paper suggests an ideal preprocessing pipeline that incorporates normalization and data augmentation (e.g., Horizontal Flip, Width/Height Shift, Small Rotation, Shearing, Zoom, Brightness/Contrast, Occlusion (cutout, etc.).

1. Face extraction (face cropping)
Each face in the picture can be cropped using the bounding box information. To preserve more face context, you might choose to add padding around the clipped area.
2. Resize the picture
Each cropped face should be resized to a fixed resolution of 60 0 x 600 pixels to comply with EfficientNet-B7's input size requirements.
3. Normalization Techniques
To provide consistent input for the model, normalize the intensi ty values of the pixels. Two common techniques:
Option 1: Min-Max Normalization Equation.

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Normalization transforms pixel values into a [0, 1] range using the original value X , the dataset's minimum X_{min} (often 0), and X_{max} (typically 255).

When utilizing custom training, compatibility with pretrained models, such as EfficientNet-B7.
Option 2: ImageNet-Based Normalization (for pretrained EfficientNet-B7)

$$X_{normalized} = \frac{X - \mu}{\sigma} \qquad (2)$$

Tab.4: shows ImageNet Channel-wise Mean and Standard Deviation (RGB)

Channel	Mean (μ\mu)	Std Dev (σ\sigma)
Red	0.485	0.229
Green	0.456	0.224
Blue	0.406	0.225

4. Data Augmentation (to improve generalization and robustness)
Data augmentation works to increase data artificially to improve the image quality and model performance. Data augmentation includes many techniques, where any number of them can be applied, such as Horizontal Flip, Width/Height Shift, Small Rotation, Shearing, Zoom, Brightness/Contrast adjustments, Occlusion methods (e.g., Cutout), etc. These techniques were performed on the original dataset. This improves the resilience and generalization of the classifying real and fake images [13].

DOI: https://doi.org/XXXX.XXXX

Tab.5: Summary of Preprocessing Steps

Step	Transformation Type	Purpose
Face Extraction	Cropping with optional padding	Isolates faces from images while preserving contextual information around the face.
Resize	Resize to 600 x 600 pixels	Ensures input images meet EfficientNet-B7 size requirements for consistent processing.
Normalization	Min-Max Scaling or ImageNet Norm	Standardizes pixel values for stable training and compatibility with pretrained models.
Rotation	Small random rotations (±2°)	Improves the model's ability to recognize faces under slight rotations.
Width Shift	Horizontal translation (±10%)	Prevents overfitting to fixed horizontal face positions.
Height Shift	Vertical translation (±10%)	Increases robustness to vertical variations in face placement.
Shear	Slant distortion (10%)	Helps the model learn to detect tilted faces.
Zoom	Random zooming (5%)	Enables scale invariance for face recognition.
Horizontal Flip	Flip images left/right	Useful for symmetrical face features and generalization.
Occlusion	Cutout or similar occlusion techniques	Improves robustness by simulating partial face occlusion.

Network Architecture

Architecture of Models: To distinguish between real and fake facial images, a Convolutional Neural Network (CNN) was created. CNNs work well because they don't require human feature extraction; instead, they automatically extract both low-level and high-level features from raw images, as shown in **Fig.7**. The Layer of Input 128 × 128 × 3 (height, width, RGB channels) is the new size for the input images. To increase training speed and stability, pixel values are normalized to the [0, 1] range. Layers of Convolution Hierarchical features are extracted using multiple convolutional layers: 32 3x3 filters make up the first layer. 64 filters in the second layer, 128 filters in the third layer. The ReLU activation function is used in each layer to introduce non-linearity and facilitate the learning of intricate patterns. Normalization of Batches. In order to improve training speed and model convergence, Batch normalization is used to normalize activations and decrease internal covariate shift. Layers of Dropout. By randomly deactivating some neurons during training, dropout is used between dense layers with rates ranging from 0.3 to 0.5 to lessen overfitting. Output and Dense Layers. The feature maps are flattened, and the information is then sent to 128 or 256 neurons in a dense layer with ReLU activation. One neuron and sigmoid activation in an output layer. A probability is provided by **the sigmoid output: ≥ 0.5 = real < 0.5 = fake**

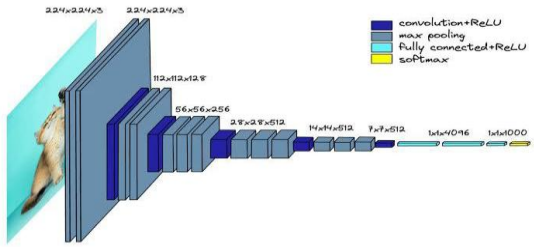


Fig. 7: Architectural diagram of the CNN Architecture

A. Feature Extractor: Transfer Learning

Using EfficientNet-B7 for Transfer Learning with Feature Extraction. A popular deep learning technique is transfer learning, which involves using a model that has been trained on a sizable dataset for a related but distinct task. **Fig. 8** employed a pre-trained network to extract valuable features from input images rather than starting from scratch when training a deep model. This method works particularly well when there is a lack of data. Feature extraction is a popular tactic where, To extract deep features from input images, the convolutional layers of a pre-trained model employed EfficientNet-B7. To complete a particular classification task, like differentiating between real and fake faces, a new classified, typically fully connected layers is added on top.

Trained Efficient Net [14] is used as a feature detector in this study. Specifically, EfficientNetB3, a convolutional neural network (CNN) model from the EfficientNet family, is employed due to its balance between accuracy and computational efficiency. The EfficientNetB7 model has been trained on the ImageNet-1K dataset, which consists of approximately 1.2 million images [12].

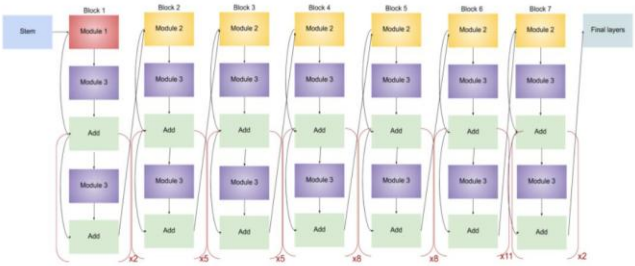


Fig. 8: Architecture for EfficientNet-B7 [12]

Images with dimensions of 600 × 600 are used to train the EfficientNetB7 model. As a result, rather than being reduced in size, the photographs are resized to 600 × 600. EfficientNetB7 is specifically made to handle larger images, which improves its performance for challenging classification jobs; therefore, this scaling is crucial.

An input layer that can handle 600 x 600 photos is the first part of the design. After that, a pre-processing stage uses Lambda

DOI: https://doi.org/XXXX.XXXX

(preprocess input) to scale the images according to the model. By utilizing the rich characteristics discovered from a sizable data set, this pre-trained model functions as a reliable feature extractor.

The architecture incorporates a Global-Average-Pooling layer after the input and preprocessing, which successfully shrinks the feature maps' spatial dimensions. This stage is essential for preserving the model's effectiveness while guaranteeing the preservation of significant elements. Effective classification is made possible by a dense layer with a SoftMax activation function that performs the final classification and outputs the probabilities of each category.

The model's performance is improved by EfficientNetB7's use of pre-trained weights, which enables it to extract high-level features while modifying the final layers to meet the demands of the task. This model is especially well-suited for applications needing high accuracy and precision in image classification since it excels at handling high-dimensional data well while retaining computing efficacy [13].

Experimental results

In contrast to other models, the EfficientNetB7 model, as shown in the accuracy chart Fig.4, exhibits remarkable efficiency in achieving high performance with a relatively lower number of parameters. In our experiments, the EfficientNetB7 model was trained on the same data distributions and sample sizes as the previous models, using consistent hyperparameters like the Stochastic Gradient Descent optimizer with a learning rate of 0.01and momentum of 0.9. This consistency ensured a fair comparison between the models.

After fine-tuning, EfficientNetB7 attained an accuracy of about 84.4%, despite the dataset's complexity and the noise that had previously hampered performance. Compared to the EfficientNetB3 model, which had already shown a notable gain to 81.7% accuracy through transfer learning, this represented a 4% improvement.

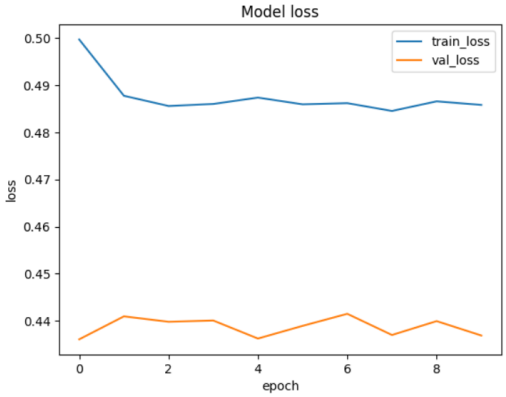


Fig. 9: Shows EfficientNetB7 model loss

Fig.9 shows the EfficientNetB7 model promising results in minimizing loss during training over the specified number of epochs; the loss stabilizes around 0.44. that after approximately 10 epochs, which is essential for achieving high accuracy in real vs. fake image classification. Monitoring both training and validation losses will be key in ensuring the model continues to generalize well with new data.

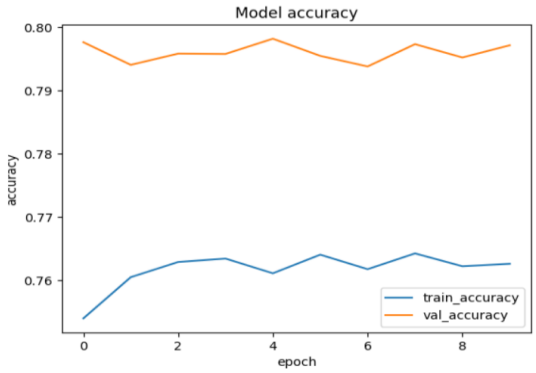


Fig. 10: shows the EfficientNetB7 model accuracy

Fig.10 shows EfficientNetB7 demonstrates strong performance in classifying real and fake images, with training accuracy stabilizing around 0.79 and validation accuracy reaching approximately 0.80. Continuous monitoring of both accuracies is essential to ensure that the model maintains its generalization capabilities as it is trained further.

Overall, the results highlight EfficientNetB7's capability to balance accuracy and efficiency, establishing it as a leading architecture in the realm of deep learning for image classification.

Conclusions

The exploration of pre-trained models for image classification reveals the remarkable advancements in the field of Computer Vision. Each model discussed, VGG-16, Inception, ResNet50, and Efficient Net, represents significant strides in achieving near-human-level accuracy in recognizing and categorizing images. Thus, pre-trained models have transformed the landscape of image classification, making state-of-the-art techniques accessible for a wide range of applications. By understanding and utilizing these models, practitioners can significantly enhance the efficiency and accuracy of their computer vision tasks, paving the way for further innovations and applications in the field. An overall accuracy of 84% is achieved, which is very encouraging.

References

[1] A. Liu, X. Li, J. Wan, Y. Liang, S. Escalera, H.J. Escalante, M. Madadi, Y. Jin, Z. Wu, X.J.I.B. Yu, Cross-ethnicity face anti-spoofing recognition challenge: A review, 10(1) (2021) 24-43.

DOI: https://doi.org/XXXX.XXXX

[2] K. Simonyan, A.J.a.p.a. Zisserman, Very deep convolutional networks for large-scale image recognition, (2014).
[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.
[4] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
[5] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, International conference on machine learning, PMLR, 2019, pp. 6105-6114.
[6] S.J.a.p.a. Bai, An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, (2018).
[7] Y. LeCun, L. Bottou, Y. Bengio, P.J.P.o.t.I. Haffner, Gradient-based learning applied to document recognition, 86(11) (2002) 2278-2324.
[8] Tan, M. and Le, Q., 2019, May. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
[9] Koritala, S.P., Chimata, M., Polavarapu, S.N., Vangapandu, B.S., Gogineni, T.K. and Manikandan, V.M., 2024, June. A Deepfake detection technique using Recurrent Neural Network and EfficientNet. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.
[10] Suratkar, S. and Kazi, F., 2023. Deep fake video detection using transfer learning approach. Arabian Journal for Science and Engineering, 48(8), pp.9727-9737.
[11] Le, T.N., Nguyen, H.H., Yamagishi, J. and Echizen, I., 2021. Openforensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10117-10127).
[12] Tan, M. and Le, Q., 2019, May. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.
[13] Kornblith, S., Shlens, J. and Le, Q.V., 2019. Do better imagenet models transfer better?. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2661-2671).
[14] Li, H., He, P., Wang, S., Rocha, A., Jiang, X. and Kot, A.C., 2018. Learning generalized deep feature representation for face anti-spoofing. IEEE Transactions on Information Forensics and Security, 13(10), pp.2639-2652.

DOI: https://doi.org/XXXX.XXXX