STM JOURNALS

Journal of
Image Processing & Pattern Recognition Progress

ISSN: 2394-1995
Volume 12, Issue 2, 2025
May–August
DOI (Journal): 10.37591/JoIPPRP

https://journals.stmjournals.com/joipprp

Review JoIPPRP

Enhancing Facial Recognition: Assessing CNNs for Detecting Image Manipulation

Karshit Bhaskar[1,*], Anurag Aeron[2], Hemang Shukla[1], Kanika Yadav[1], Iesha[1]

Abstract

Deepfake technology, powered by highly advanced deep learning models, has raised significant concerns regarding media manipulation, identity theft, and the spread of online disinformation. Due to the increasing sophistication of deepfake content, traditional forensic methods often fail to detect such artificially generated images with high accuracy. Consequently, deep learning-based approaches have become essential in combating this challenge. This study compares six prominent deep learning architectures: VGG16, ResNet50, MobileNetV2, InceptionV3, EfficientNetB0, and DenseNet121, to analyze the effectiveness of Convolutional Neural Networks (CNNs) in detecting deepfake images. The models were trained on a dataset containing both real and manipulated facial images, incorporating various levels of complexity. Among the tested architectures, ResNet50 emerged as the top-performing model, achieving an impressive accuracy rate of 96.58%. The findings of this research highlight the critical role of deep learning in media verification and cybersecurity. By providing an in-depth analysis of CNN-based deepfake detection methods, this study lays the foundation for future advancements in safeguarding digital content against deceptive media attacks.

Keywords: Efficient Net, CNN, VGG19, Keras, accuracy

INTRODUCTION

Deepfake technology, powered by advanced deep learning models, poses significant threats such as media manipulation, identity theft, and online disinformation. Traditional forensic tools often fail to detect highly realistic deepfakes, making deep learning-based approaches essential. This study evaluates the effectiveness of six Convolutional Neural Network (CNN) architectures: VGG16, ResNet50, MobileNetV2, InceptionV3, EfficientNetB0, and DenseNet121, in identifying deepfake images. The models were trained and tested on a dataset containing both real and manipulated facial images with varying levels of complexity.

***Author for Correspondence**
Karshit Bhaskar
E-mail: karshit.bhaskar.cse.2021@miet.ac.in

[1]Student, Department of Computer Science and Engineering, Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India
[2]Professor, Department of Computer Science and Engineering, Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India

Received Date: March 13, 2025
Accepted Date: April 12, 2025
Published Date: April 28, 2025

Citation: Karshit Bhaskar, Anurag Aeron, Hemang Shukla, Kanika Yadav, Iesha. Enhancing Facial Recognition: Assessing CNNs for Detecting Image Manipulation. Journal of Image Processing & Pattern Recognition Progress. 2025; 12(2): 27–36p.

Among the architectures examined, ResNet50 demonstrated the highest performance, achieving an accuracy of 96.58%. The results highlight the importance of deep learning in combating deepfake threats, as CNN-based models can effectively distinguish between genuine and altered images. This study lays a foundation for future advancements in media verification, cybersecurity, and digital forensics, ensuring stronger safeguards against malicious deepfake usage.

As deepfake technology continues to evolve, the development of robust detection techniques becomes increasingly critical. By comparing multiple

CNN architectures, this research provides insights into the strengths and weaknesses of different models, guiding further improvements in deepfake detection systems. The findings emphasize the necessity of integrating deep learning solutions into security frameworks to counteract digital deception and protect online integrity.

LITERATURE REVIEW

Classical forgeries are identified through methods such as: Noise Print CNN was built by Cozzolino *et al*. for trace fingerprint detection for forgeries [1]; Face manipulation detection through Two-stream CNN was proposed by Zhou *et al*. [2]. Li *et al*. further state that DeepFake videos are not realistic from an eye blink point of view because training images as a source would generally not include images of individuals with their eyes closed [3].

MesoNet utilized CNNs for the purpose of discriminating genuine face images from synthesized face images using DeepFake pipeline [4]. Combining RNN with CNN, recurrent neural network-based DeepFake video detection further prolonged the MesoNet technique into the time axis [5]. Not an exception is it, i.e., it requires huge volumes of training samples, real and synthetic images, leading to significant time and resource consumption.

With deep component characteristics of CNN, a more sophisticated architecture CNN, was introduced [6]. The second, long VGG CNN was used for classification using the partially extracted features from the initial VGG (Visual Geometry Group) CNN. Wen *et al*. drew out various feature categories to achieve the objective of incursion face detection [7]. The four feature types are cascaded into one feature vector to train two individual SVM classifiers to detect the fake face.

Generative adversarial networks (GANs) allow very realistic human images to be generated in the static image space [8, 9]. Moreover, realistic videos of any person speaking anything that the artist desires can be generated in the video space [10].

Our own effort comes closest to that of Agarwal *et al.* [11]. To identify unusual and persistent head motion and facial expression patterns, the authors of the study monitored hours of footage of some people in this instance, a panel of world leaders and presidential candidates. In particular, Ekman and Friesen extracted each individual's frame-by-frame facial movements (parameterized in 18 action units and two axes rotation of the 3-D head) from each 10-sec clip of a subject [12]. All of the 20 features generated a 190-D feature vector characterizing an individual's temporal mannerisms based on their correlation.

METHODOLOGY

About the Dataset

High-quality, expertly altered facial photos produced using cutting-edge image editing techniques make up the dataset. Several faces have been combined into these photos, and either the mouth, nose, eyes, or the complete face have been altered. Real and false are the two primary classes into which the dataset is divided. Training deep learning models to differentiate between real and fake faces is made easier with the help of this dataset [13]. The classifier gains the ability to recognize patterns that are frequently present in images produced by Generative Adversarial Networks (GANs). However, since more complex forgeries produced by skilled human manipulation follow an entirely separate generation process, these patterns might not always be successful in identifying them (Figure 1).

Necessary Modules

TensorFlow

Our experiment takes advantage of TensorFlow, a mature deep learning environment, to learn and test the models that differentiate between real images and deepfake images. It provides efficient techniques of convolutional neural network (CNN) deployment, handling gigantic picture datasets, and model improvement. We used VGG16, ResNet50, MobileNetV2, InceptionV3, EfficientNetB0, and DenseNet121 to learn

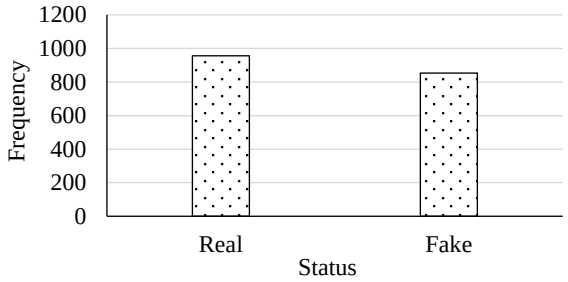

Figure 1. Faces data distribution.

differentiating between fake and real faces using TensorFlow. The seamless model deployment offered by TensorFlow's GPU capabilities and TensorFlow-Keras API optimized computational speed and precision. It was the most suitable for deepfake detection and real face recognition in our research as it could seamlessly process complex image data.

Keras

Development of neural networks is eased using Keras, a high-level deep learning library. Keras plays an important role in our project in model training and testing for actual face recognition and Deep Fake detection. It has an easy-to-use API for designing and optimizing deep learning models such as VGG16, ResNet50, MobileNetV2, InceptionV3, EfficientNetB0, and DenseNet121. It allows effective model training using pre-trained weights by utilizing transfer learning. Additionally, it provides transparent TensorFlow integration and the ability to use GPU acceleration for accelerated computation. Keras is the best fit for our DeepFake detection system because it is flexible.

Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) form the core of real face detection and deepfake detection within this work. CNNs are more apt to deal with image data through convolutional, pooling, and fully connected layers in order to identify spatial features. The model uses a series of architectures like MobileNetV2, EfficientNetB0, VGG16, DenseNet121, ResNet50, and InceptionV3 for the real and manipulated face differentiation. Its classification accuracy is enhanced through hierarchical feature extraction, allowing it to detect minor abnormalities in deepfakes like texture problems and abnormal face variations. The success of CNNs relies on the fact that the accuracy of theirs naturally varies.

Keras Layers Used

Conv2D

One of the most critical components of the Convolutional Neural Networks (CNNs) implemented in our project for deepfake detection and real face identification is the Conv2D layer. To identify spatial information from input images like edges, textures, and patterns, it executes 2D convolution operations. They are the features required for classification. Significant facial features are identified at different levels by each Conv2D layer by a series of learnable filters (kernels) traversing the image. Our deep learning models VGG16, ResNet50, MobileNetV2, InceptionV3, EfficientNetB0, and DenseNet121 are able to classify real and fake faces by layering multiple Conv2D layers, with very high accuracy for the classification operation.

Flatten

Keras' Flatten layer is central to conversion from convolutional layers to fully connected layers in our deepfake detector model. As the convolutional layers possess feature maps, which are of multi-dimensional type, Flatten layer converts the latter into one-dimensional vector that dense layers can accept. In the process, it facilitates the use of models such as MobileNetV2, EfficientNetB0, VGG16, DenseNet121, ResNet50, and InceptionV3 for feature extraction processing. Support for spatial hierarchies in addition to facilitating classification, Flatten layer allows unbridled feature propagation to allow the model to detect real and deepfake images with ease.

Max Pooling

Our Max Pooling-based deep learning models for face detection and deepfake detection highly rely on Max Pooling itself. Max Pooling prevents overfitting and saves computational resources without losing informative contents by lowering the spatial dimension of the feature maps. Our Max Pooling compresses feature maps with the maximum content of each region of the inputs following convolution layers. Thus, most significant features like edges and texture that are useful in distinguishing deepfake and original images are preserved. Our VGG16, ResNet50, and InceptionV3 models generalize and optimize best with Max Pooling.

Average Pooling

Our face recognition and deepfake detection deep learning models heavily rely on the Average Pooling layer. It assists in the preservation of useful information through the reduction of feature maps' spatial dimension using down-sampling. Average Pooling also simplifies feature extraction using the average input frame values, unlike Max Pooling that uses the maximum value. It assists in the preservation of useful image patterns without compromising computing efficiency. Our VGG16 and InceptionV3 models presented in this study are highly enhanced by Average Pooling, which allows them to generalize and filter features more in real vs. fake facial images discrimination.

Dense

In our real-life face recognition and deepfake issue, the Dense layer does the classification. Dense layer is a fully connected layer where all of the neurons in the previous layer are connected with each of the neurons in the current layer. It learns the abstract representations and ultimately makes predictions. After the feature extraction by the convolutional layers, the Dense layer transforms the abstract representations into discriminative real and fake images. It uses activation functions like ReLU for feature mapping and Softmax or Sigmoid for classification. The Dense layer gives the model's generalization ability, which leads to the overall better accuracy.

Activation

On the inclusion of non-linearity, the activation layer plays a vital role in our deep learning models since it enables the network to recognize sophisticated patterns in facial images. Activation functions like Sigmoid, Softmax, and ReLU (Rectified Linear Unit) are implemented throughout our project at various stages. ReLU is usually applied to convolutional layers to gain efficient feature extraction by preventing the vanishing gradient problem. On the final layer of classification, Sigmoid facilitates the binary classification operations, while Softmax assigns a probability value to each class (real or fake). ResNet50 and InceptionV3 models are better optimized because of these activation layers, improving deepfake detection with higher accuracy.

Dropout

Our deep models utilize dropout, a regularization technique, to generalize and prevent overfitting. To compel the network to learn stronger features instead of relying on specific neurons, dropout layers in this study randomly disable some neurons during training. This improves the model's ability to distinguish between real and fake faces, even under challenging circumstances. In the VGG16, ResNet50, and EfficientNetB0 models, dropout is employed to reduce the model's sensitivity towards noise and variability in deepfakes. In detecting manipulated facial images on other datasets, the approach guarantees improved generalization.

MODELS USED

VGG16

Our work employs VGG16 as a deep learning model for real face recognition and deepfake detection. VGG16 is a 16-layer Convolutional Neural Network (CNN) with a good reputation for possessing an easy but effective architecture made up of max-pooling layers, fully connected layers, and a number of convolutional layers with very small 3×3 filters. VGG16 can recognize fine differences between real and imitated faces through extraction of deep hierarchical features from images.

With an accuracy of 0.6927 in testing, VGG16 was moderately efficient in distinguishing deepfake from real images. VGG16 is inferior to more complex models like ResNet50, though with good texture feature capture. Computational intensity and instability to manipulated image changes are VGG16's primary weaknesses. It remains relevant to our work, however, as it provides us with a baseline by which to compare the performance of more intricate and deeper networks in identifying deepfakes.

ResNet50

ResNet50 (Residual Network with 50 layers) is one of the deep learning networks used in our project for deepfake detection and real face identification. Residual connections of this high-performance CNN architecture are renowned for their contribution to solving the vanishing gradient problem of deep networks. ResNet50 facilitates deeper learning of features without compromising performance through identity shortcuts. ResNet50 performed the best with a value of 0.9657 in our test. ResNet50 was the strongest to distinguish between deepfake and actual faces. It is able to learn subtle features such as texture anomaly and facial defects of abnormality that exist in deepfake images because its deep architecture enables it to do so. The model is very accurate for classification since it can handle fine-grained patterns and spatial hierarchies well. ResNet50 is more accurate and reliable compared to other models, and its strong feature extraction ability is essential to improving the robustness of our deepfake detection system.

EfficientNetB0

One of the deep learning models implemented in our project for identifying deepfakes and authentic faces is EfficientNetB0. This CNN attempts a balance between computational expense and model size through compound scaling. It is powerful and efficient. This allows it to have low processing costs but remain incredibly accurate. With isolation of finer aspects such as abnormality in texture and unbalanced mixup of features in deepfakes, EfficientNetB0 can learn discrimination of features between real and synthesized faces with ease in our work. EfficientNetB0 is superior to a considerable number of models, displaying efficiency in face feature alteration at a 0.9366 level of accuracy. Its squeeze-and-excitation blocks and depth-wise convolutions enhance feature representation, making it robust against different levels of false image complexity (easy, mid, and hard). Due to its efficiency, the model is ideally suited for practical applications where correct and computationally efficient deepfake content detection is needed, ensuring consistent detection of forged images.

MobileNetV2

We employed MobileNetV2, a lightweight and low-cost deep network for real face authentication and deepfake detection in our research. MobileNetV2 is optimized for mobile and edge devices and applies depthwise separable convolutions to minimize computational expense at the expense of minimal performance loss. In our experiments, MobileNetV2 applied inverted residual blocks and linear bottlenecks to enhance feature extraction with minimal model size. Although it is optimal, MobileNetV2's comparatively lower accuracy of 52.68% also could not effectively differentiate between real and manipulated faces, as opposed to the deeper networks of ResNet50 and InceptionV3. Lower accuracy would also imply that, even while MobileNetV2 uses less resource and optimizes speed, it might not be able to identify intricate patterns of facial manipulation in deepfake.

InceptionV3

Our proposed model utilizes deep convolutional neural network InceptionV3 to detect deepfakes and real faces. Our model uses auxiliary classifiers, asymmetric kernels, and factorized convolutions to keep computational costs minimal while maintaining maximum accuracy. This structure maintains the subtle face characteristics and image tampering artifacts of images. At 0.9024 accuracy in our testing, InceptionV3 proved to be a powerful tool for separating real images from deepfake images. The network can also identify tiny defects in deepfakes, such as blending artifacts or artificial texture, because it supports multi-scale information analysis. InceptionV3 is a suitable candidate to identify deepfakes because it achieves a balance between accuracy and speed through its deep architecture with improved convolutional layers.

DenseNet121

Our experiment employs DenseNet121 as a genuine face recognition and deepfake detection model of deep learning. Ensured maximum information transfer between layers, this dense connectivity convolutional network provides improved feature propagation. DenseNet121 reduces redundancy and improves gradient flow by connecting each layer to every other layer compared to traditional CNNs. Such a structure is advantageous in detecting inconsistencies in deepfake images since it is useful in preserving fine-grained detail in face images.

DenseNet121 also performed well in our tests, separating real and manipulated faces with 84.39% accuracy. The small size of the model allows it to be very accurate with fewer parameters. The model can learn precise facial representations because of its dense connections, which makes it ideal for separating fake artificial intelligence-generated and real faces.

IMPLEMENTATION

Our approach employs ResNet50, a 50-layer convolutional network, to identify real faces and deepfakes. The general idea behind this architecture is residual learning and skip connections to counter the vanishing gradient issue in deep networks. The skip connections allow the network to learn identity mappings such that deeper layers retain important information from lower layers without losing any information.

In an attempt to leverage its strong feature extraction capabilities, we employ pre-trained ResNet50 within our work with ImageNet weights trained on the same. A binary classification-specific classifier is used in lieu of fully connected layers for adding extra robustness into the model.

Dropout is used to avoid overfitting and Batch Normalization to normalize activations for better generalization. Apart from that, non-linearity is embedded through ReLU activation to facilitate the model in identifying fine facial patterns. Images are resized into ResNet50 size before they are fed into the convolutional and pooling layers, where hierarchical features extracted subject them. The last layer assigns the image label as real or fake based on the probability score produced. The method utilizes ResNet50's capability to learn deep features to provide reliable deepfake detection.

RESULTS OBTAINED

We used multiple deep learning architectures, and they were trained as well as tested on our database to measure how well our system for real face recognition and detection of deepfakes worked. In order to select the top architecture that was capable of classifying between artificially manipulated facial photos and genuine facial images, accuracy of every single model was analyzed (Figures 2 and 3).

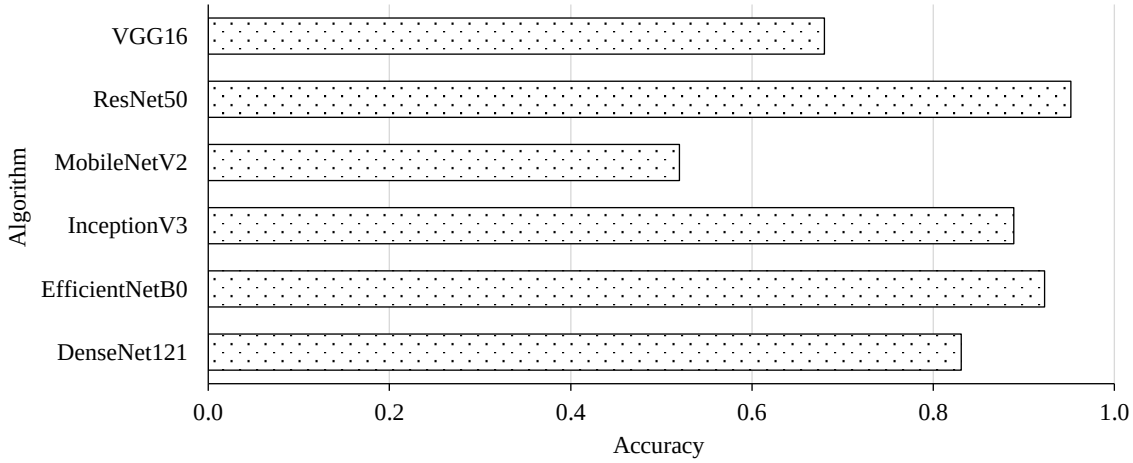

Figure 2. Accuracy comparison chart.


```
VGG16 --> 0.6926829218864441
ResNet50 --> 0.9657558736801152
MobileNetV2 --> 0.5268292427062988
InceptionV3 --> 0.9024389982223511
EfficientNetB0 --> 0.9365853667259216
DenseNet121 --> 0.8439024686813354
```

Figure 3. Accuracy values.

Table 1. Accuracy comparison in percentage.

Algorithm	Accuracy
VGG19	23.33%
ResNet50	86.19%
EfficientNetB3	90.95%
MobileNetV2	59.04%
IncecptonV3	90.23%
DenseNet121	48.33%

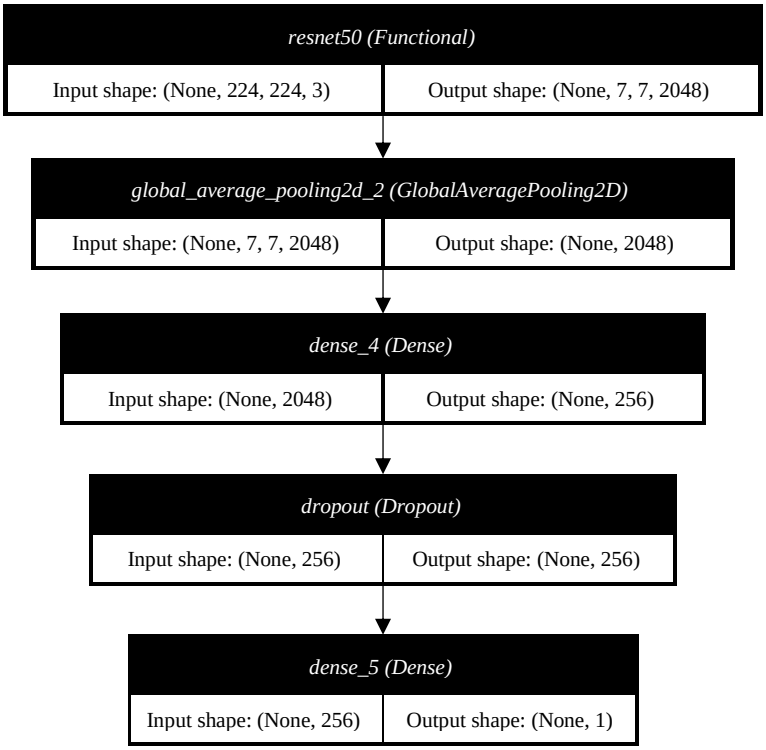

Figure 4. Model structure.

ResNet50 is the best-performing model in detecting deepfake images with the highest accuracy of 96.58% among all architectures tested. The discovery infers that ResNet50 was able to effectively discriminate between real faces and forged faces by easily extracting fine facial features and high-level facial features.

Table 1 provides a thorough accuracy comparison of the several models that were employed in our project. The high performing model's architecture is uncovered by the model structure visualization, which comprises its convolutional, pooling, and fully connected layers (Figure 4). Classification report confirms the model's strength across most classes through providing information regarding precision, recall, and F1-score (Figure 5).

	precision	recall	f1-score	support
Fake	0.9902	0.9439	0.9665	107
Real	0.9412	0.9897	0.9648	97
accuracy			0.9657	204
macro avg	0.9657	0.9668	0.9657	204
weighted avg	0.9669	0.9657	0.9657	204

Figure 5. Classification Report.

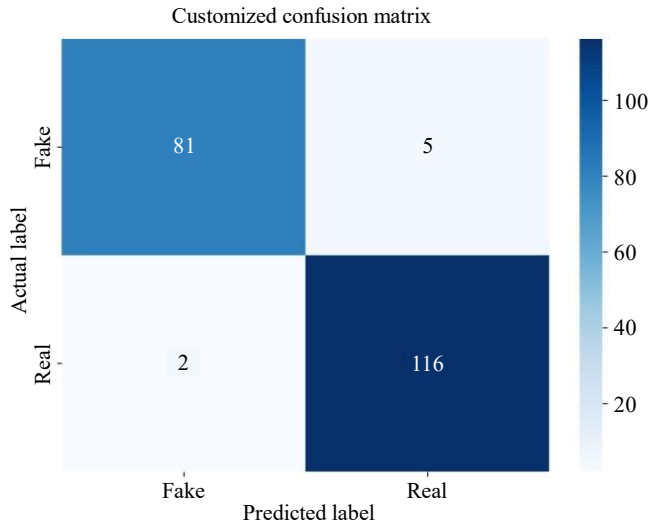

Figure 6. Confusion matrix.

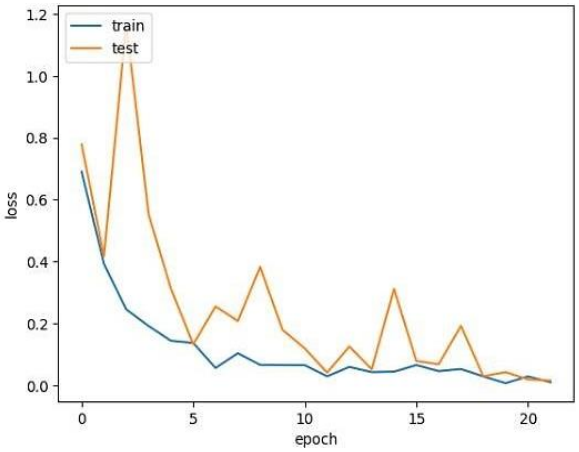

Figure 7. Model loss graph.

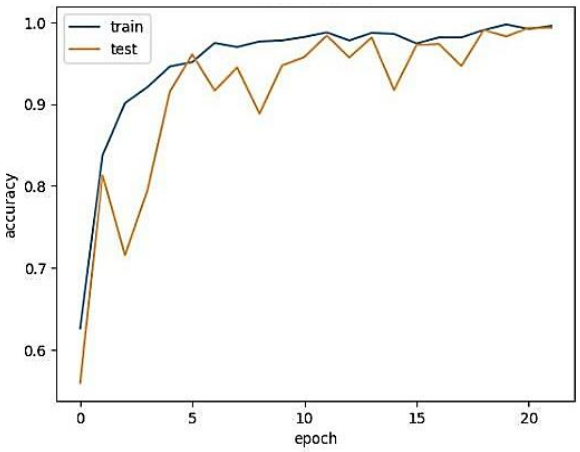

Figure 8. Model accuracy graph.

The distinction between right and wrong predictions is also brought out by the confusion matrix, which indicates a remarkable ability to correctly classify true and false images (Figure 6). Convergence and generalization are confirmed by plots of accuracy and model loss for monitoring the process of learning. That our training process is perfectly effective is substantiated by the accuracy plot that shows a gradual rise and the loss plot, which illustrates an abrupt decline in error (Figures 7 and 8).

The ability of the model to distinguish accurately between real and deepfake images is substantiated via a prediction confidence analysis. The outputs provide predicted labels along with their respective confidence scores, whereby it reflects the high degree of confidence of the model in providing probabilities, particularly in distinguishing between high-level deepfake manipulations and real faces (Figure 9).

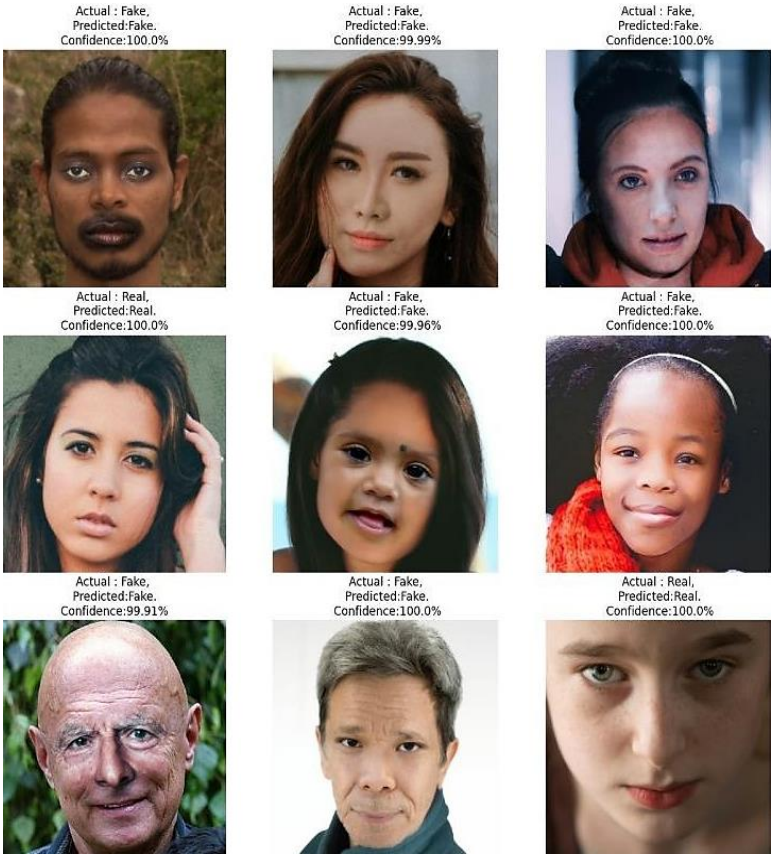

Figure 9. Prediction made by model with confidence.

CONCLUSION

In this study, we tried and employed a lot of deep learning models for real face recognition and deepfake detection. Our main goal was to create a successful model that would be able to differentiate between real and fake facial images. To find the best model for this, we compared six well-known CNN-based models: VGG16, ResNet50, MobileNetV2, InceptionV3, EfficientNetB0, and DenseNet121.

Following thorough testing, ResNet50 was the most effective model, which reflected its high feature extraction capability towards identifying deepfake patterns with 96.58% accuracy. Model performance and model generalizability were also improved by applying dropout layers, batch normalization, and fine-tuning operations. Moreover, our confusion matrix analysis, accuracy trends, and model loss supported the efficacy of our approach.

This study indicates that deep learning models are robust against deepfake attacks, which pose a significant risk to misinformation and online safety. Deep learning models, such as ResNet50, can potentially detect manipulated images effectively, as indicated in the study, which can be used in real-world applications in digital forensics, cybersecurity, and media verification. To counter the next-generation deepfake generation techniques, future research can work on enhancing detection accuracy even more with the help of adversarial training and ensemble learning.

REFERENCES

1. Cozzolino D, Verdoliva L. Noiseprint: A CNN-based camera model fingerprint. IEEE Trans Inf Forensics Secur. 2019 May 13; 15: 144–59.
2. Zhou P, Han X, Morariu VI, Davis LS. Two-stream neural networks for tampered face detection. In 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). 2017 Jul 21; 1831–1839.

3. Li Y, Chang MC, Lyu S. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In 2018 IEEE International workshop on information forensics and security (WIFS). 2018 Dec 11; 1–7.
4. Afchar D, Nozick V, Yamagishi J, Echizen I. Mesonet: a compact facial video forgery detection network. In2018 IEEE international workshop on information forensics and security (WIFS). 2018 Dec 11; 1–7.
5. Güera D, Delp EJ. Deepfake video detection using recurrent neural networks. In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS). 2018 Nov 27; 1–6.
6. 11Li L, Feng X, Boulkenafet Z, Xia Z, Li M, Hadid A. An original face anti-spoofing approach using partial convolutional neural network. In 2016 IEEE Sixth international conference on image processing theory, tools and applications (IPTA). 2016 Dec 12; 1–6.
7. 12Wen D, Han H, Jain AK. Face spoof detection with image distortion analysis. IEEE Trans Inf Forensics Secur. 2015 Feb 4; 10(4): 746–61.
8. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019; 4401–4410.
9. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020; 8110–8119.
10. Tolosana R, Vera-Rodriguez R, Fierrez J, Morales A, Ortega-Garcia J. Deepfakes and beyond: A survey of face manipulation and fake detection. Inform Fusion. 2020 Dec 1; 64: 131–48.
11. Agarwal S, Farid H, Gu Y, He M, Nagano K, Li H. Protecting world leaders against deep fakes. In CVPR workshops. 2019 Jun 16; 38–45.
12. Ekman P, Friesen WV. Measuring facial movement. Environmental psychology and nonverbal behavior. 1976 Sep; 1(1): 56–75.
13. Kim Y, Yoo JH, Choi K. A motion and similarity-based fake detection method for biometric face recognition systems. IEEE Trans Consum Electron. 2011 May; 57(2): 756–62.

