

An Optimized Deep Learning Pipeline for Image Authenticity Detection Using Multi-Stage Deep Learning Pipeline

Noorshaba Khatoun¹, Aditi Sharma², Divyanshu Chauhan³

¹Centre for Advanced Studies, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India

²Institute of Engineering and Technology, Lucknow, India.

³Centre for Advanced Studies, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India

DOI: <https://doie.org/10.10399/JBSE.2026880368>

ABSTRACT

The rapid development of generative AI has produced very convincing fake images, making it difficult to tell what's real or fake. This situation presents major challenges for digital forensics, detecting misinformation and verifying media authenticity. Because of this, we've developed a deep learning classification pipeline for distinguishing between natural and computer-generated images, using the DeepDetect dataset with over 1,00,000 images from a variety of categories. We used a transfer learning approach with the VGG16 back-end, and we finetuned the feature extraction through a series of multi-stage fine tunings. The Adam optimizer was used to optimize performance, utilizing multiple different learning rates and a binary cross entropy loss function incorporating label smoothing. Extensive hyperparameter tuning and data augmentation were performed. The model performance was evaluated with several evaluation metrics such as accuracy, precision, recall, F1- score, and confusion matrix. Our average overall accuracy was 99.13%. Our work shows how deep learning pipelines can reliably detect image authenticity.

KEYWORDS

Deep Learning, Image Classification, AI-Generated Images, Real vs Fake Detection, Image Authenticity Detection, Synthetic Image Detection

1. Introduction

We have seen tremendous advancements in Artificial Intelligence as a result of advances in deep learning and generative models that allow us to create very realistic synthetic images that are difficult to distinguish from images taken of real events. Technologies such as Generating Adversarial Networks and diffusion models have vastly increased the quality of the images generated by these methods [1][2]. Each has improved the quality of imagery, semantic coherence, and scalability of the material created over the previous generation. With millions of images being created by AI being available on the web today, there are three areas that raise serious concerns regarding the consequences of generating misinformation; manipulating digital images; and ultimately losing trust in visual media. As simulated visual content is capable of deceiving or fooling our sensory perception, it is important that there are meaningful computational systems available to verify whether or not an image is real. The classification of real vs synthetic imagery will remain a significant issue in both digital forensics and computer vision research.

1.1 Background

Rapid advances in Deep Fake and AI-generated image technology can be attributed to improvements in deep learning architectures and large training datasets. These technologies use models like GANs and Diffusion-based models to create photo-realistic images with little human input [2]. With increased access to such tools, there is now widespread use of these tools in entertainment, advertising, & content creation. However, there are risks associated with the misuse of these tools, including identity manipulation, fraud, & the distribution of false information [3]. Traditional image forensics methods are often not effective against the latest generative techniques; therefore, advanced detection frameworks based on deep learning will need to be developed to identify weaknesses, abnormalities, or statistical differences in synthetic images from authentic ones.

1.2 Motivation for Study

The purpose of this research is to investigate the increasing difficulty of differentiating between authentic and synthetic images. Research has shown that synthetic images possess many characteristics similar to those found in authentic images, thereby making the use of human detection systems subject to human error and thus unreliable [1]. As a result, the problem lies in the existence of misuse through the use of synthetic imagery in social media

platforms, political propaganda, illicit commerce, etc. Therefore, there is a need for a mechanism to be put into place to assist with the detection and identification of synthetic art through machine learning systems. The inability of many current detection methodologies to generalize across the broad array of generative production systems that exist today provides motivation for designing and implementing robust deep learning systems capable of learning effective discriminative features and improving classification performance across various types of imagery when comparing to authentic imagery, thereby ensuring operational reliability and impervious scalability for use within present-day societal environments.

1.3 Aim and Objectives

This research will develop a Deep Learning Model Pipeline in order to accurately classify arrays of actual images to those that have been generated by Artificial Intelligence. For this, we will develop a deep learning model using models that are trained using transfer learning techniques and optimally extracting features via pre-trained convolutional neural networks while overcoming class imbalance challenges during training. The outputs from our model will be compared to established performance measures accuracy, precision, recall, F1-score, and the generalization capability of the model will be determined on multiple dataset splits. Finally, we will implement strategies that improve performance such as augmenting data, fine-tuning to build a strong and dependable detector of synthetic images under real-world conditions.

1.4 Contribution of this work

This study offers a new contribution to the area of digital image forensics through development of a structured deep learning architecture designed to differentiate between authentic and AI-generated images. The proposed multi-stage end-to-end deep learning transfer learning strategy enhances representation of the image features and improves classification results. The approach addresses the class imbalance problem and develops appropriate training techniques to achieve consistent and robust performance. Finally, the proposed method was evaluated against a diverse dataset confirming the method's ability to be used across multiple applications. This research provides a solution to issues related to generalization and robustness of detection of synthetic images. Therefore, it contributes in an effort to combat on-going issues related to the spread of misinformation and to improve the authenticity of digital content.

2. LITERATURE REVIEW

2.1 Related Work

2.1.1 CNN-based detection approaches

CNNs are useful for AI image detection due to their ability to learn spatial and hierarchical features. CNNs are able to detect pixel-level artifacts, inconsistencies with texture and variations in structure from image generation. There have been many studies on CNN-based classifiers, which indicate that when trained with very large datasets and when transfer learning is utilized, CNN-based classifiers can achieve a very high level of accuracy. Hybrid frameworks that incorporate GAN's have demonstrated improved performance by using both discriminative and generative features through the use of this method [5]. One of the limitations of CNN-based classifiers is that they often rely primarily on dataset specific features to generalize to other domains and to detect generative models that have not been seen before [6].

2.1.2 Transformer-based models

Detecting whether an image is authentic through transformer-based architectures has been a recent focus, as the architecture is capable of modelling both long-range dependencies and global contextual relationships in images. CNNs work primarily with local receptive fields whereas transformers work with self-attention mechanisms using an entire image to control global structures and maintain semantic coherence. This ability provides a benefit when identifying inconsistencies that are not limited to a specific portion of the image but can be related to the entire image in AI-generated photos. Recent research has concluded that models that are attention-based and hybrid forms of CNN-transformer based architecture perform better in terms of robustness and generalisability; however, the implementation of these types of models typically requires large amounts of training data and greater computational power, which makes them less practical for applications that are real-time or which are resource constrained.

2.1.3 Frequency-domain and hybrid methods

Frequency domain methods evaluate images based on their spectra to help detect artifacts created in the process of creating them. Methods such as the Fourier Transform and co-occurrence matrix analysis can be used to identify non-conforming frequency patterns in synthetic images. Studies have shown that images generated through GANs tend to have noticeable discrepancies in the distribution of frequencies and color channels [8]. There are hybrid

procedures that have been developed by using both spatial and frequency-based attributes to improve the accuracy and robustness of the detection. This improves the performance of the two methods combined because they each provide complementary information. However, most recently developed diffusion-based models have significantly decreased these types of artifacts, thus making it more difficult for a frequency-based method to identify these types of images.

Table 1. Comparative Analysis of Deep Learning Pipelines for Synthetic Image Detection

S.No.	Author	Year	Technology Used	Performance
1	Khan & DangNguyen [9]	2023	CNN + Transformer	AUC up to 0.99
2	Yeeshu Ralhen et al [10]	2025	CNN Ensemble (ResNet + EfficientNet)	Accuracy: 90.58%, F1: 94.69%
3	Wang et al [11]	2020	CNN (GAN detection)	99% accuracy
4	Sohail Ahmed Khan et al [12]	2021	VGG16 + InceptionV3 + Xception (Fusion CNN)	96.5% (DFDC), 99% (TIMIT)
5	Sumedha Arya [13]	2025	CNN + Vision Transformer	92.56%–97.16% Accuracy
6	Van-Nhan Tran et al [14]	2021	CNN + Attention	~97% Accuracy
7	Sreeraj Ramachandran et al [15]	2021	Face Recognition + CNN	AUC: 0.99
8	Subhram Dasgupta et al [16]	2025	CNN + SE Attention	94.14% Accuracy
9	Vrizlynn Thing [17]	2023	CNN + Transformer	Accuracy up to 99.73%

2.2 Research Gaps Identified

2.2.1 Limitations: Dataset Bias Lack of Robustness to Unseen Generators

Advances have been made in both generating images and detecting them; however, many current methods are still quite limited. One of the biggest problems is that there is often a bias in datasets for specific types of images. Thus, training a model from a dataset, results in the possibility of a poor performance when evaluating datasets from generative models. When cross-dataset evaluation is performed for detection models, results indicate a significant decrease in the ability of the model to detect compared to its original trained dataset. As new generative techniques have developed, they have produced less detectable artifacts, furthering the difficulty of using today's detection models. The majority of the current detection models are supervised and require a relatively large sample size of labelled images; however, there is not an adequate sample size available for all image generation techniques. Therefore, these limitations create a pressing need for developing generalized, adaptive, and domaininvariant detection frameworks that can accommodate multiple methods of image generation, as well as continuously improving methods of generating images [18].

2.2.2 Limitations of CNN-based detection approaches

The CNN-based methods discussed in this section have performed well under benchmarking conditions; however, there are shortcomings when it comes to expanding beyond that. In the first case, most CNN network models do not extend to implementing generative architecture and applying them to untrained datasets which is the datasets that were not used to develop and train a network well. This is because they develop dataset-specific artefacts instead of demonstrating the true characteristics of synthetic images and thus do not exhibit invariance to them. In the second case, CNN network models rely significantly on spatial features for learning and do not generally have the ability to detect both fine detail or local inconsistencies and large, global inconsistencies. The third issue relates to how susceptible CNN models are to standard image transformations such that compression, resizing, and noise, all of which have adverse effects on their performance. Additionally, there are substantial deficits in terms of the interpretability of current CNN approaches, preventing us from effectively understanding the decision boundaries that exist and why they exist. Solving these issues necessitates new feature integration methodologies, improved generalisation methodologies and better training methodologies to develop robust networks that can operate effectively in the real-world [19][20].

2.2.3 Limitations of Transformer-based models

Transformers have successfully distinguished between honest images and deceptive photographs by making effective use of their ability to recognize the connection between images via "self-attention," but there are many constraints to this type of architecture. For example, transformer-based architectures have greatly reduced effectiveness when evaluating images created by novel image generators. They mostly can recognize the hidden features of the generators, as opposed to common indicators of forgery. Prior research shows that the performance of transformers drops off significantly with regard to images produced using some of the more advanced image generation technologies such that diffusion models [21]. Additionally, the high resources required by transformer-based architectures due to the need for large datasets and compute capabilities create challenges with real-time application of these models. Finally, the challenges associated with multi-domain feature extraction such that spatial, frequency, wavelet domains, remain unresolved because current methods for performing feature fusion do not deliver sufficient robust results in practice.

2.2.4 Limitations of Frequency-Domain and Hybrid Methods

Methods that use spectral analysis to evaluate textures in newly created images provide an alternative technique of detecting generated textures through combining frequency based and hybrid designs. These methods will often identify some latent artifacts, such as high frequency content above normal density ranges; however, there are many limitations to their use practically due to the fixed nature of the transforms that are available for detecting generated images and these limitations are further exacerbated by the evolution of generative models with the introduction of GANs and Diffusion Models which have less high frequency content than they did when first produced, thus making detection based on low frequency characteristics less effective. Hybrid design approaches can partially offset the loss of detection due to the take advantage of both spatial and frequency feature will often require more computational resources and redundant models. Moreover, hybrid design approaches are often not able to generalize across multiple domains of images, thus there is a need to develop adaptive domain invariant detection methodologies [22][23].

3. MATERIALS AND METHODS

3.1 Dataset Description

The DeepDetect-2025 dataset [24] is a comprehensive and vetted resource of both authentic images and artificially produced images developed to aid researchers with image verification, deepfake verification and analysis, and adversarial resilience testing. The complete dataset contains over 100,000 images comprised of about 60,000 actual and 55,000 synthetic images. The collection provides image representation across various visually identifiable types e.g., human portraits, animals, natural scenes, cityscapes as well as different styles e.g., photos, artworks, abstract. The authentic images were retrieved from public repositories while the synthetic images were generated using some of today's most advanced AI models e.g., StyleGAN3, DALL·E3, Midjourney, Stable Diffusion. The images contain binary classifications i.e., real or fake along with metadata that is the output of the generation technique, image resolution, and semantic category. The dataset provides an opportunity for examination of model generalization, bias detection, and performance assessment of deep learning model techniques across multiple visual categories.

3.2 Data Pre-Processing

The goal of data preprocessing is to create an identical representation of input data to enhance the generalization of the Neural Network body. For example, images will be resized to 224×224 pixels according to the input requirements of the VGG16 NN architecture. Prior to passing through the VGG16 NN architecture, the preprocessing function of VGG16 will be performed on the image data in order to ensure pixel normalisation is applied and image features are compatible with the pre-trained weights of the VGG16 NN architecture that were obtained by training on ImageNet. In order to diversify the data used for training the model and to reduce overfitting to avoid bias, data augmentation techniques will be employed, such as rotating the image by upto 25°, shifting the image horizontally or vertically by 15%, zooming into the image by upto 25% increase, flipping the image horizontally, and brightness is randomly adjusted within the range of 0.8 to 1.2 to account for varying lighting conditions. When analysing the effect of different data distribution split percentages on the performance of the NN model, four distinct training-validation ratios of 90:10, 80:20, 70:30, and 60:40 will be used. The number of samples images processed by the NN model's body before weights are updated will be limited to either 16 or N (batch-size). This maximises computing power while also enabling stable convergence. The overall preprocessing procedures will enable improved representation of the data.

3.3 Model Utilized

An investigation of the VGG16 architecture will be carried out. The VGG16 is a deep Convolution Neural Network architecture that has been highly utilised in a variety of different image classification tasks due to its ease of use and effectiveness. The VGG16 architecture consists of cascading convolutional layers that have small sizes for their receptive fields and additional fully-connected layers that allow for hierarchical feature extraction.

A pretrained VGG16 model with ImageNet weights will be used to build the feature extractor and all of the fullyconnected layers will be removed from the VGG16 model. New classification layers will be added to fit the VGG16 model to binary classification, including global average pooling, batch normalisation and dense layers. The model will also utilise transfer learning, by taking advantage of the previously learned visual representations, to increase performance and decrease training time.

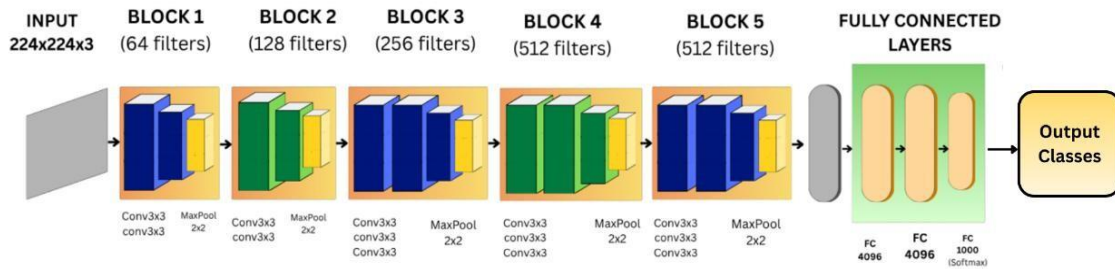


Fig. 1. VGG16-Based Deep Learning Architecture for Real vs AI-Generated Image Classification

3.4 Proposed Methodology

The multi-stage approach to robust classification utilizes transfer learning to improve the overall performance of models through subsequent adjustments of the VGG16 architecture. Initially, VGG16 is frozen with an unfrozen classifier head, the last layer trained first at a learning rate of 1×10^{-4} in Stage 1. Next, most of the ConvNet (Conv4 and Conv5) was adapted for better feature extraction through re-training with relatively low learning rates of 1×10^{-5} in Stage 2. Finally, unfrozen the entirety of Conv4 and Conv5 with a higher learning rate of 5×10^{-6} . When training the models, the use of computed weighting of the classes will address any class imbalance during training. Furthermore, to improve the generalization of the trained model, various forms of regularization, including dropout and label smoothing, will be incorporated. By taking this multi-stage approach to transfer learning, applying progressive layer adjustments in conjunction with other regularization techniques, the overall performance of the model and preventing overfitting while progressing towards stable convergence throughout the modeling process will be increased.

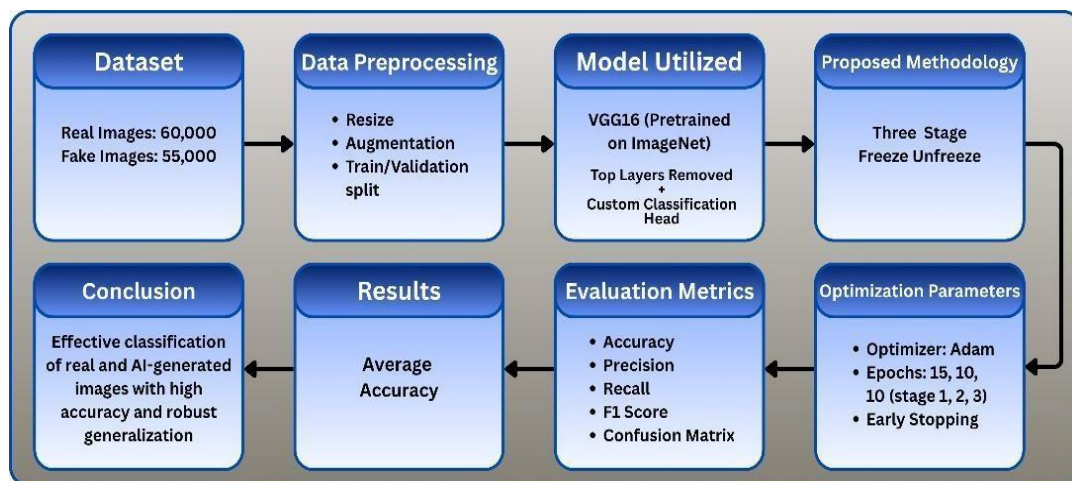


Fig. 2. Workflow of the Proposed Multi-Stage Transfer Learning Approach

3.5 Optimization Algorithms and Hyperparameters

The model has been trained using an Adam optimizer which has been chosen due to its ability to adaptively adjust the learning rate and converge efficiently. The learning rate used for training will vary across different training stages that is 10^{-4} for stage 1, 10^{-5} for stage 2 and 5×10^{-6} for stage 3, which is intended to stabilize the training phase during fine-tuning. The Binary Cross-Entropy loss function with Label Smoothing of 0.1 for stage 1 and

0.05 for subsequent stages, was used to reduce the amount of overconfidence when making predictions. The model has been trained using 16 sample batches and 15, 10 and 10 epochs (respectively) for each training stage. Additional training optimizations employed include early stopping (patience = 5), ReduceLROnPlateau (factor = 0.3; patience = 3) and ModelCheckpoint for saving the best model.

3.5 Evaluation Metrics

In order to provide a complete evaluation model performance are assessed against a number of different measures. Overall classification performance is determined using accuracy as the primary metric and a classification report will then be generated that contains precision, recall and F1-score to evaluate performance from each class perspective. A confusion matrix will then be created that will include true positive, false positive, true negative, and false negative counts to give a visual description of the classification errors associated with the training dataset. Finally, predictions will be made by applying a threshold value of 0.5 to the sigmoid output values for each test image. These evaluation metrics will allow for a thorough assessment and verification of the model's ability to differentiate between images created by human and artificial intelligence.

4. RESULTS AND DISCUSSION

4.1 Experimental Results

Table 2. Performance comparison of VGG16 model when train and validated on different training and validation ratios

Training Validation Split	Base Model Used	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
90:10	VGG16	99.11	99.00	99.00	99.00
80:20		99.18	99.00	99.00	99.00
70:30		99.10	99.00	99.00	99.00
60:40		99.13	99.00	99.00	99.00
Average		99.13	99.00	99.00	99.00

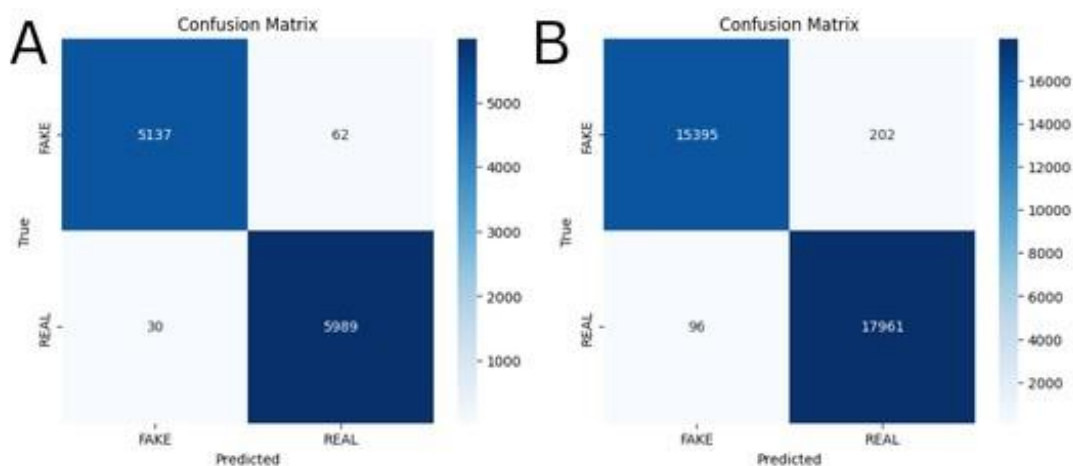


Fig. 3. Confusion Matrix generated on VGG16 model when Train and validated on (A) 90:10 (B) 70:30 dataset split ratios

4.2 Comparative Analysis

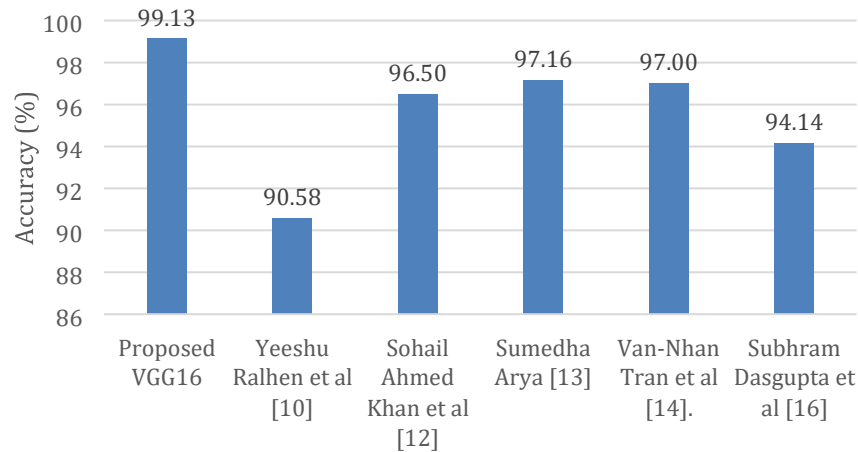


Fig. 4. Comparative analysis of Proposed Pipeline with existing methodologies

4.3 Limitations

There are numerous limitations to the proposed deep learning pipeline, despite demonstrating sound end-results. To begin with, the model is based on supervised learning, and therefore any bias within the underlying dataset such as an image's category or as a result of the use of generative models will result in poor generalization. Second, this system will not perform well when analyzing previously unseen or newly developed generative methods, especially advanced diffusion-based generative methods, which produce outputs that are highly aesthetically pleasing indistinguishable from real jpg images. Third, using a fixed resolution of 224×224 pixels as an input to the model may result in the significant loss of important details that will be needed for performing forensic analyses. Fourth, the amount of computational resources required to implement a multi-stage fine-tuning process and deep model architectures will most likely prohibit real-time deployment. For all of these reasons, there is an urgent need for the development of more adaptable, less resource-intensive detection frameworks.

5. CONCLUSION AND FUTURE SCOPE

5.1 Conclusions

Through the implementation of the proposed approach, it is clear that using deep transfer learning as an approach to differentiating between real and artificial images is viable. In addition, a multi-stage tuning process of the model's weights enhances the representation of relationship features enabling the model to effectively capture both low-level and high-level patterns of discrimination between real and simulated images respectively. Evaluation through experiments demonstrates the stability of convergence through the use of class weighting, dropout regularization, and adaptive learning rate, improving the model's ability to classify accurately. By using a pretrained VGG16 model as the base of the system the time required to train significantly decreases while still achieving nearly 100% real and simulated image accuracy post-training. Therefore, through proper preprocessing, optimization, and evaluation methods throughout the entire training process, it has been validated that deep learning models can be utilized as reliable methods of image validation.

5.2 Future Scope

Future studies can help make detection systems more resilient and agile when working with new types of generative systems. For instance, by using new methods for building detection systems such as transformer architecture or combining CNNs with transformers will provide additional capability to detect features that span large areas across multiple images taken over an extended period of time i.e., global features. Furthermore, using both frequency feature analysis and spatial feature analysis in conjunction with each other could help identify subtle differences between "real" and "fake" images produced by generative methods. To help in detecting a wider range of generative methods, it can also be helpful to expand datasets beyond those currently available to include new and previously unseen generative methods. Finally, the development of lightweight architectures to handle real-time deployment will help to facilitate the use of detection systems in this manner. In addition to developing these types of detection systems, the use of explainable AI techniques may help with model interpretability, thereby improving user trust in model results. Overall, this type of development within the field of detection systems will lead to improved reliability and scalability when identifying AI-created visual media in real-life examples.

REFERENCES

- [1] Zhang, Y., Pang, Z., Huang, S., Wang, C., & Zhou, X. (2025). Unmasking AI-created visual content: A review of generated images and deepfake detection technologies. *Journal of King Saud University - Computer and Information Sciences*, 37, 148. <https://doi.org/10.1007/s44443-025-00154-8>
- [2] Raza, A., Basit, A., Amin, A., Arfeen, Z. A., Masud, M. I., Fayyaz, U., & Jumani, T. A. (2026). A comprehensive review of deepfake detection techniques: From traditional machine learning to advanced deep learning architectures. *AI*, 7(2), 68. <https://doi.org/10.3390/ai7020068>
- [3] Wadhwa, I., Choudhary, Y., Chauhan, P., Singh, A., Sathua, S. K., & Bains, S. (2025). Deepfake and AI-generated image detection system. *International Journal for Multidisciplinary Research*, 7(2). <https://doi.org/10.36948/ijfmr.2025.v07i02.42245>
- [4] Soundarya, B. C., & Gururaj, H. L. (2026). Deepfake detection: Critical review of state-of-the-art approaches and future perspectives. *Discover Applied Sciences*, 8, 201. <https://doi.org/10.1007/s42452-025-08174-9>
- [5] Sadhana, P., Ravishankar, N., Ashok, A., Ravichandran, R., Paul, R., & Murali, K. (2024). Enhancing fake image detection: A novel two-step approach combining GANs and CNNs. *Procedia Computer Science*, 235, 810–819. <https://doi.org/10.1016/j.procs.2024.04.077>
- [6] Kumar, K., Satya Karthik, R., Jena, S. K., Joshi, S. S., & Rao, S. H. (2025). A survey on detection of deep fake images using CNN model. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*. <https://doi.org/10.17148/IJARCCE.2025.14623>
- [7] Tiwari, M. R., & Patil, S. S. (2026). Deep-fake image detection using machine learning techniques: A comprehensive review. *International Journal on Advanced Electrical and Computer Engineering*, 15(1S), 137–145. <https://doi.org/10.65521/ijaece.v15i1S.1350>
- [8] Barni, M., Kallas, K., Nowroozi, E., & Tondi, B. (2020). CNN detection of GAN-generated face images based on cross-band co-occurrences analysis. *arXiv*. <https://arxiv.org/abs/2007.12909>
- [9] Khan, S. A., & Dang-Nguyen, D.-T. (2023). Deepfake detection: A comparative analysis. *arXiv*. <https://arxiv.org/abs/2308.03471>
- [10] Ralhen, Y., & Sharma, S. (2025). Enhanced deepfake detection using an ensemble of convolutional neural networks. *IAES International Journal of Artificial Intelligence*, 14(5), 4043–4049. <https://doi.org/10.11591/ijai.v14.i5.pp4043-4049>
- [11] Wang, Sheng-Yu, et al. “CNN-Generated Images Are Surprisingly Easy to Spot... for Now.” *arXiv*, 2020. <https://arxiv.org/abs/1912.11035>
- [12] Khan, Sohail Ahmed, Alessandro Artusi, and Hang Dai. “Adversarially Robust Deepfake Media Detection Using Fused Convolutional Neural Network Predictions.” *arXiv*, 2021. <https://arxiv.org/abs/2102.05950>
- [13] Arya, S. (2026). Deepfake video detection using CNN-based architectures and vision transformer model. *International Journal on Science and Technology*, 17(1). <https://doi.org/10.71097/IJSAT.v17.i1.10337>
- [14] Tran, V.-N., Lee, S.-H., Le, H.-S., & Kwon, K.-R. (2021). High performance deepfake video detection on CNN-based with attention target-specific regions and manual distillation extraction. *Applied Sciences*, 11(16), 7678. <https://doi.org/10.3390/app11167678>
- [15] Ramachandran, S., Nadimpalli, A. V., & Rattani, A. (2021). An experimental evaluation on deepfake detection using deep face recognition. *arXiv*. <https://arxiv.org/abs/2110.01640>
- [16] Dasgupta, S., Mason, J., Yuan, X., Odeyomi, O., & Roy, K. (2024). Enhancing deepfake detection using SE block attention with CNN. In *Proceedings of the 2024 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1–6). IEEE. <https://doi.org/10.1109/icABCD62167.2024.10645262>
- [17] Thing, V. L. L. (2023). Deepfake detection with deep learning: Convolutional neural networks versus transformers. *arXiv*. <https://arxiv.org/abs/2304.03698>
- [18] Tiwari, M. R., & Patil, S. S. (2026). Deep-fake image detection using machine learning techniques: A comprehensive review. *International Journal on Advanced Electrical and Computer Engineering*, 15(1S), 137–145. <https://doi.org/10.65521/ijaece.v15i1S.1350>
- [19] Zanardelli, M., Guerrini, F., Leonardi, R., et al. (2023). Image forgery detection: A survey of recent deep learning approaches. *Multimedia Tools and Applications*, 82, 17521–17566. <https://doi.org/10.1007/s11042-02213797-w>
- [20] Cozzolino, D., Gragnaniello, D., Poggi, G., & Verdoliva, L. (2021). Towards universal GAN image detection. *arXiv*. <https://arxiv.org/abs/2112.12606>

- [21] Man, Q., Gee, S.-J., & Cho, Y.-I. (2026). Multi-domain perception transformer for generalized forgery image detection. *Applied Sciences*, 16(1), 533. <https://doi.org/10.3390/app16010533>
- [22] Luo, X., & Wang, Y. (2025). Frequency-domain masking and spatial interaction for generalizable deepfake detection. *Electronics*, 14(7), 1302. <https://doi.org/10.3390/electronics14071302>
- [23] Say, T., Alkan, M., & Kocak, A. (2025). Advancing GAN deepfake detection: Mixed datasets and comprehensive artifact analysis. *Applied Sciences*, 15(2), 923. <https://doi.org/10.3390/app15020923>
- [24] Datta, A. (2025). DeepDetect-2025 [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/11658923>