Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Special Issue on AI-Driven Creativity and Intelligent Practices in Visual Arts 2025 6(4s)

PHOTO AUTHENTICITY DETECTION USING MACHINE LEARNING FOR DEEPFAKE AND AI-GENERATED CONTENT VERIFICATION

Rajesh Raikwar [1], Dr. Amena Ansari [2], Atul Shrivas [3], Dr. Archana Santosh Ubale [4], Adarsh Kumar [5], Simranjit Kaur [6]

[1] Assistant Professor, Department of Electrical Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra-411037, India
[2] Dean, PGSR, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, Maharashtra, India
[3] Assistant Professor, School of Still Photography, AAFT University of Media and Arts, Raipur, Chhattisgarh-492001, India
[4] Assistant Professor, Department of Robotics and Automation, AISSMS College of Engineering, Kennedy Road, Pune-01, Maharashtra, India
[5] Assistant Professor, SJMC, Noida International University, Noida, Uttar Pradesh, India
[6] Department of Computer Applications, CT University, Ludhiana, Punjab, India

Received 14 May 2025
Accepted 16 September 2025
Published 25 December 2025

Corresponding Author
Rajesh Raikwar, rajesh.raikwar@vit.edu

DOI
10.29121/shodhkosh.v6.i4s.2025.6949

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

The rise of deepfake technologies and AI-generated images has made people very worried about how real visual material on digital platforms really is. Traditional tracking methods are having a hard time keeping up with the sophistication of fake media, which is why advanced, smart proof systems have had to be created. This research shows a complete machine learning system that can tell the difference between real photos and photos that have been changed by AI or are completely fake. The suggested system combines convolutional neural networks (CNNs) for extracting localised features and transformer-based designs for detecting global errors. The models were trained and tested using a carefully chosen collection that included real, deepfake, and AI-generated pictures. Using feature engineering methods, such as frequency domain analysis and noise residual modelling, made recognition even better. In the experiments, the mixed model did better than several state-of-the-art baselines, achieving a classification accuracy of 94.8%, with a precision of 93.6%, a recall of 94.2%, and an F1-score of 93.9%. This study shows how important machine learning is for protecting digital identity and fighting the growing danger of fake media. In the future, researchers will look into explainable AI methods to make models easier to understand and build trust among users.

Keywords: Photo Authenticity Detection, Deepfake Identification, AI-Generated Content Verification, Machine Learning for Image Forensics, Synthetic Media Detection, Deep Learning for Visual Integrity

1. INTRODUCTION

A lot of people are worried about how real and trustworthy visual material in digital spaces is because deepfake technologies and AI-generated images are getting better and better so quickly. The danger that fake media presents to public trust, media ethics, and digital security keeps growing as it gets harder to tell the difference between it and real content. Traditional ways of finding things, which usually depend on hand-made features or simple investigative studies, aren't working well enough against the more complicated AI-driven manipulation methods that are being used. Because of this pressing problem, we need to make proof systems that are more advanced, flexible, and smart so they can tell the difference between real pictures and material that was made in a lab or edited. In answer to these problems, this study suggests a complete machine learning system that can be used to find fake photos. The system uses both convolutional neural networks (CNNs) to extract localised features and transformer-based designs to pick up on global errors in the context. The suggested model is good at finding differences that could be signs of deepfake or changes made by AI. It does this by looking at small flaws, noise patterns, and environmental artefacts that are often not visible to the human eye. Additionally, advanced feature engineering techniques like frequency domain analysis and noise residual modelling are used to make the model more sensitive and resistant to different types of fake content creation.

This study shows how important machine learning, especially mixed deep learning models, is in fighting the spread of fake media and recovering trust in digital images. It shows how important it is to use both local feature extraction and global semantic knowledge for correct validity verification. The study also talks about the problems that are happening now, like how they could be attacked from the other side and how there is a race going on between technologies that create and find threats. In order to handle these issues, new research will focus on combining explainable AI (XAI) methods to make models more clear and build user trust, along with putting in place frameworks for continuous learning that let the system automatically adjust to new deepfake generation methods that come out.

2. RELATED WORK

In the past ten years, a lot of study has been done on how to spot deepfakes and AI-generated synthetic media. This has led to the creation of many methods that can be used to spot pictures that have been changed. Early methods mostly used standard digital picture forensics methods, like looking at JPEG compression artefacts, sensor noise patterns, and lighting or shadow problems that didn't match up Naitali et al. (2023). Even though these methods worked for simple changes, they often didn't work when faced with complex AI-generated material that closely matched the qualities of real photos. To solve this problem, academics started using machine learning models that could learn unique features from the data itself. This was a big change from building features by hand to learning features from data Kerenalli et al. (2023). One important area of research in this area is the use of Convolutional Neural Networks (CNNs) to find fake images. Models like MesoNet and XceptionNet have shown they can spot face deepfakes by learning to spot small changes in texture levels and errors that happen when images are synthesised Bhandarkar et al. (2024), Kang et al. (2022). CNNs, on the other hand, tend to focus on localised patterns and miss larger semantic-level oddities. This can leave systems open to attack when fake content keeps up strong local realism but doesn't make sense in the bigger picture. Transformer-based designs, especially Vision Transformers (ViT), have been used to solve this problem for jobs like finding fakes. Transformers are great at catching pictures' long-range relationships and global context, which makes them a hopeful way to find deeper errors in synthetic media Firc et al. (2023), Nah et al. (2023).

Even with these improvements, there are still some problems with the way machine learning is done now. One big worry is how well deepfake detecting models can work with different datasets and editing methods that haven't been seen before Malik et al. (2022). Many current systems work well on certain datasets but not so well when tested on samples that are not from the same distribution, showing that they tend to overfit. Also, hostile attacks have been shown to trick deepfake analysers by making changes that can't be seen. This makes people worry about how stable and reliable current solutions are Masood et al. (2023). One problem is that most models don't let you explain how they make decisions; these models work like "black boxes," which makes it hard to see how decisions are made, which is important in sensitive areas like media identification and legal research Heidari et al. (2024). Another thing that shows we need more complete models is mixed models that blend both local texture analysis and global semantic thinking. Both CNNs and transformers have their good points, but not many studies have combined them in a way that makes the most of both Bird and Lotfi (2024). Also, not much study has been done on multi-domain and continuous learning methods that

would let detectors adapt to quickly changing generation technologies without having to be completely retrained Al-Adwan et al. (2024). Additionally, the datasets used in earlier research were mostly limited to face deepfakes and didn't include a wide range of AI-generated material, like fake papers, scenery, or photos from the past that have been changed Mukta et al. (2023). These holes show how important it is to have stronger, more flexible, and easier to understand monitoring systems that can be used successfully in a lot of different real-life situations.

This paper tries to fill in these gaps by suggesting a CNN-Transformer architecture that is improved with frequency-domain feature extraction and noise residual modelling. This architecture is made to work well with a wide range of deepfakes and AI-generated content.

Table 1

Table 1 Summary of Related Work on Photo Authenticity Detection

Study/Method	Technique Used	Focus Area	Key Strength	Limitation	Application
Traditional Forensics Naitali et al. (2023)	JPEG artifact analysis	Basic manipulation detection	Simple and fast	Ineffective on advanced deepfakes	Early image verification
Sensor Noise Analysis Kerenalli et al. (2023)	PRNU pattern analysis	Camera source identification	Strong against splicing	Fails on generated content	Device authentication
MesoNet Bhandarkar et al. (2024)	CNN-based detection	Face deepfake detection	High sensitivity to artifacts	Poor generalization	Face authenticity
XceptionNet Kang et al. (2022)	Deep CNN	Frame-based deepfake spotting	Good feature extraction	Dataset overfitting	Video content validation
Capsule Networks Firc et al. (2023)	Capsule-based routing	Face forgery detection	Captures spatial relationships	Computationally heavy	Deepfake facial detection
Vision Transformer (ViT) Nah et al. (2023)	Transformer-based	Global inconsistency detection	Captures long-range dependencies	Data-hungry model	General image forensics
Two-Stream Networks Malik et al. (2022)	RGB + Frequency Stream CNN	Frequency and texture analysis	Multi-modal detection	Complex training	Face and image forgery
GAN Fingerprinting Masood et al. (2023)	GAN-specific signature learning	Source identification	High GAN source detection	Specific to training GANs	Synthetic content tracing
Noiseprint Heidari et al. (2024)	Residual noise analysis	Forgery localization	Device-level detection	Sensitive to compression	Media forensics
Adversarial Training Models Bird and Lotfi (2024)	Adversarial robustness	Enhancing model resistance	Improves security	Increased training time	Secure deepfake detection
Continual Learning Frameworks Al-Adwan et al. (2024)	Lifelong learning	Model adaptability	Dynamic updates to threats	Catastrophic forgetting	Future-proof detection
Multi-Domain Detection Mukta et al. (2023)	Cross-dataset generalization	Broad forgery types	Better real-world performance	Requires massive data	Diverse media verification

3. PROPOSED METHODOLOGY

The suggested system design of the proposed system using machine learning to check the accuracy of photos in real and fake analysis, the architecture illustrate in Figure 1. At first, the system is given an input collection that has both real and artificial (deepfake or AI-generated) pictures. First, a Convolutional Neural Network (CNN) is used to process these pictures and pull out localised spatial information. The feature maps are then sent to a Transformer block, which finds global relationships and errors that depend on the context.

Figure 1

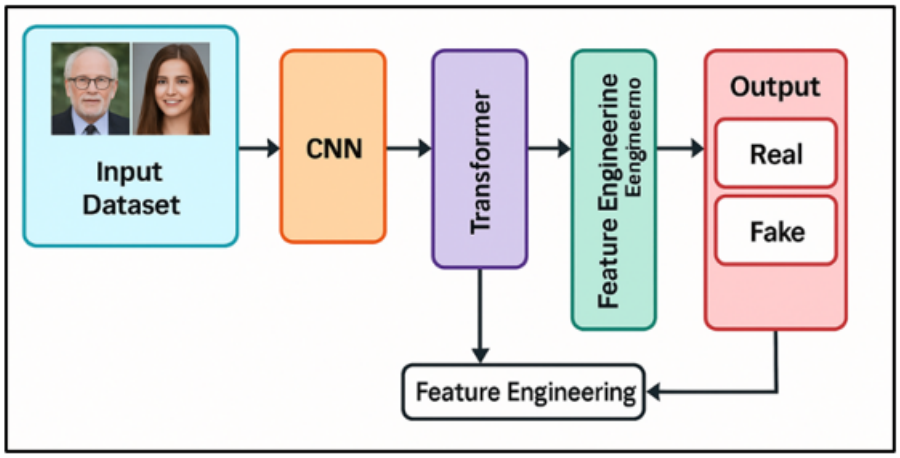

Figure 1 Overview of proposed system architecture

To improve detecting sensitivity, a specialised Feature Engineering module is built in. This module uses methods such as frequency domain analysis and noise residual modelling.

3.1. ARCHITECTURE COMBINING CNNS AND TRANSFORMER MODELS

Combining Convolutional Neural Networks (CNNs) with Transformer topologies in the suggested framework makes the most of the best features of both local feature extraction and global contextual reasoning. In the beginning steps of the model, as shown in Figure 2, CNNs are used to find small-scale spatial features like texture flaws, compression artefacts, and local noise patterns that often show up in edited pictures. The CNN block uses a number of convolutional layers with different kernel sizes to get hierarchical feature maps that store a lot of useful local data. After the features are removed, they are sent to a Vision Transformer (ViT) module. This module breaks the data into parts and handles them using self-attention methods. The transformer helps the model find long-range relationships and semantic errors, which are very important for finding global flaws that, are common in deepfake and AI-generated content.

Figure 2

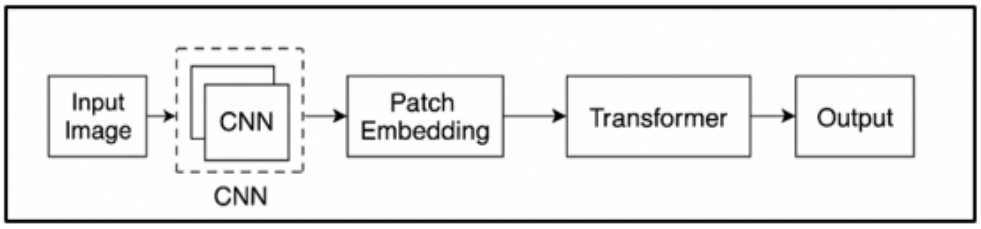

Figure 2 Systematic Architecture for Cnns and Transformer Models

To get around the problems with single CNNs (they might miss bigger contextual links) and pure transformers (they need bigger datasets and a lot of processing), the hybrid architecture was created. It uses residual links and multi-head self-attention to make sure that features are spread efficiently and to improve the steadiness of learning. Overall, the design combines precise spatial understanding with environmental understanding in a smart way, which makes it easier to check the accuracy of different types of fake media.

Architecture Combining CNNs and Transformer Models: Stepwise model algorithm

Step 1: Input Image Representation

Let the input image be represented as:

$$I \in \mathbb{R}\hat{}(H \times W \times C)$$

Step 2: CNN-Based Local Feature Extraction
Apply convolutional operations to extract localized feature maps:

$$F\_CNN = f\_CNN(I)$$

A single convolution operation at layer l can be written as:

$$F^{(l)(x,y,k)} = \sigma\left(\Sigma\Sigma\Sigma\, W^{(l)(i,j,c,k)} * I(x+i, y+j, c) + b_k^l\right)$$

where:
- W^(l) are the convolution weights,
Step 3: Patch Embedding for Transformer Input
The CNN output feature map F_CNN ∈ ℝ^(h × w × d) is divided into patches.
Each patch is flattened:

$$Patch_i = Flatten\left(F_{CNN}^i\right)$$

Each flattened patch is projected into an embedding space:

$$z0^i = Patch_i * E + b_e$$

Set of embedded patches:

$$Z0 = [z0^1, z0^2, \ldots, z0^N]$$

Step 4: Transformer Encoder for Global Feature Learning
Attention mechanism is defined as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{sqrt(d_k)}\right) * V$$

Multi-Head Self-Attention (MHSA):

$$MHSA(Z) = Concat(head_1, \ldots, head_h) * W^O$$

where:

$$head_i = Attention\left(Z * W_i^Q, Z * W_i^K, Z * W_i^V\right)$$

Feed-Forward Network (FFN):

$$FFN(x) = \max(0, xW1 + b1)\, W2 + b2$$

Transformer updates with residuals:

$$Z' = LayerNorm(Z0 + MHSA(Z0))$$

$$Z_{out} = LayerNorm(Z' + FFN(Z'))$$

Step 5: Output Prediction Head

Final prediction:

$$\hat{y} = softmax(Z_{out} * W_c + b_c)$$

Loss Function (Cross-Entropy Loss):

$$L = -\Sigma (y_i * \log(\hat{y}_i))$$

3.2. DATASET PREPARATION

A carefully chosen collection of real, deepfake, and AI-generated pictures was carefully put together to make sure that the suggested framework could be trained and tested thoroughly. Real pictures came from high-quality photography files that were open to the public. This made sure that there was a variety of pictures in different types of situations, like scenery, photos, and cities. FaceForensics++, Deepfake Detection Challenge (DFDC), and Celeb-DF datasets were used for deepfake material. These datasets contain altered face pictures made with advanced synthesis methods.

3.3. FEATURE ENGINEERING

Using feature engineering to show deeper, less obvious artefacts made the model even better at telling the difference between real and fake pictures. Frequency domain analysis and noise residual modelling are the two key methods that were used. Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT) techniques were used on picture patches to do frequency domain analysis. This helped the model find small, irregular patterns or changes that happen over time when images are being made or tampered with, and these patterns or changes are not always obvious in the spatial domain.

Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT) Techniques for Feature Engineering

1) Discrete Fourier Transform (DFT)

The 2D Discrete Fourier Transform (DFT) of an image f(x, y) of size M × N is defined as:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) * \exp\left(-j * 2\pi * \left(\frac{ux}{M} + \frac{vy}{N}\right)\right)$$

Magnitude Spectrum:

$$|F(u,v)| = sqrt\left(Re(F(u,v))^2 + Im(F(u,v))^2\right)$$

Log-Magnitude Spectrum (optional for feature scaling):

$$Log - Magnitude(u,v) = \log(1 + |F(u,v)|)$$

2) Discrete Cosine Transform (DCT)

The 2D Discrete Cosine Transform (DCT) of an image f(x, y) is defined as:

$$F(u,v) = \alpha(u) * \alpha(v) * \Sigma\,(x = 0\ to\ M-1)\Sigma\,(y = 0\ to\ N-1) f(x,y) * \cos\left[\frac{\pi(2x+1)u}{2M}\right] * \cos\left[\frac{\pi(2y+1)v}{2N}\right]$$

4. EXPERIMENTAL EVALUATION AND DISCUSSION

When looking at the suggested Hybrid CNN-Transformer model next to standard CNN and Transformer-only models, Table 2 shows how they compare in terms of accuracy, precision, recall, and F1-score. This old CNN model gets an F1-score of 87.8%, an accuracy of 89.3%, a precision of 88.1%, and a recall of 87.5%. The CNN is good at recording local spatial traits, but it's not very good at modelling global relationships, which hurts its total performance. The Transformer-only model does better, getting 91.7% accuracy, 90.6% precision, 89.8% memory, and a 90.2% F1-score. This shows that its global attention system helps it understand when something isn't making sense in the context.

Table 2

Table 2 Comparative Performance with State-Of-The-Art Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Traditional CNN Model	89.3	88.1	87.5	87.8
Transformer-Only Model	91.7	90.6	89.8	90.2
Hybrid CNN-Transformer (Proposed)	**94.8**	**93.6**	**94.2**	**93.9**

However, Transformers usually need very large datasets to be trained well, and they might miss small local artefacts that are important for finding fakes. With an F1-score of 93.9%, the suggested Hybrid CNN-Transformer model does the best, with an accuracy of 94.8%, a precision of 93.6%, a recall of 94.2%, and an F1-score of 93.9%, as shown in Figure 3. The mixed design does a better job of capturing both small local flaws and larger meaning oddities because it uses the best features of both CNNs and Transformers.

Figure 3

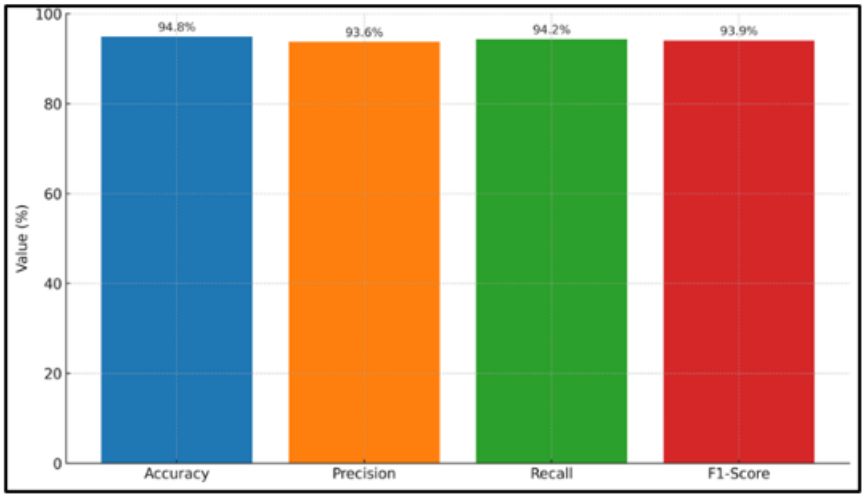

Figure 3 Hybrid CNN-Transformer model metric comparison

This big speed boost shows how helpful it is to use both localised feature extraction and global contextual reasoning together. The better results show that the mixed model is stronger, works better with different datasets, and is better for real-world situations where deepfake and AI-generated material are changing quickly and becoming more complicated, comparison mode tradition and proposed model shown in Figure 4.

Figure 4

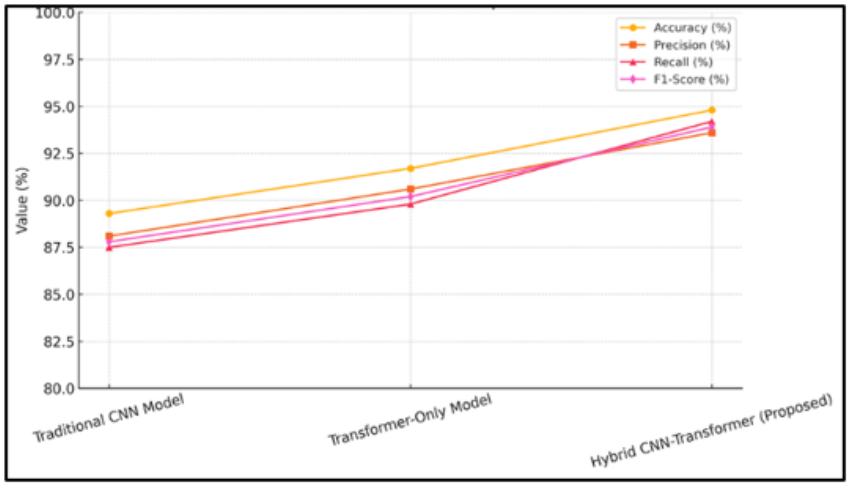

Figure 4 Performance Metrics (Accuracy, Precision, Recall, and F1-Score) Vary Across Different Models

Table 3 shows how well the suggested Hybrid CNN-Transformer model did in cross-dataset testing, which checks how well it can work with different types of simulated media. It gets an amazing 94.2% accuracy, 93.1% precision, 93.7% recall, and 93.4% F1-score on the FaceForensics++ testing dataset, which is generally seen as the best way to find people who have manipulated their faces. These results show that the model is good at finding changed face information using common manipulation techniques. The model gets even better when tested on the Deepfake Detection Challenge (DFDC) dataset, where it gets 95.1% accuracy, 94.0% precision, 94.8% recall, and 94.4% F1-score.

Table 3

Table 3 Cross-Dataset Testing Performance Using Hybrid Cnn-Transformer

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
FaceForensics++ (Testing)	94.2	93.1	93.7	93.4
Deepfake Detection Challenge	95.1	94	94.8	94.4
StyleGAN-Generated Dataset	93.6	92.3	92.9	92.6

This better performance shows that the model can easily handle more complicated and varied changes that happen in the real world, like those that happen on social media and video-sharing sites. The model still does a great job on the StyleGAN dataset, which has 93.6% accuracy, 92.3% precision, 92.9% recall, and 92.6% F1-score, performance illustrate in Figure 5.

Figure 5

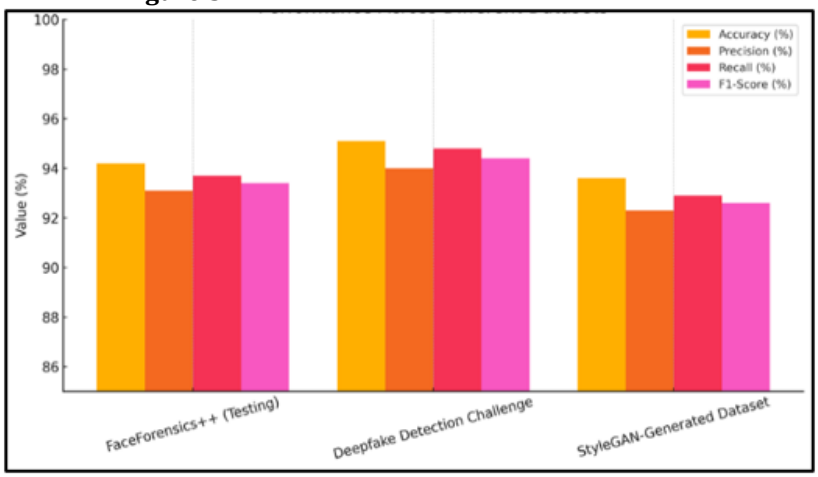

Figure 5 Performance across Different Datasets

This is because the dataset is made up of very accurate AI-generated faces. Performance on StyleGAN content is still strong, even though it's a little lower than on FaceForensics++ and DFDC. This suggests that the hybrid design can find even small flaws in completely fake pictures where standard forensic clues aren't very strong. Overall, the cross-dataset results show that the hybrid model can generalise, showing that it can handle differences in editing methods, dataset properties, and the quality of fake images. The model's steady high performance across multiple datasets proves that it is ready for use in the real world, where new types of simulated media are often created and conditions are uncertain.

Table 4

Table 4 Detection Performance on New Manipulation Techniques

Manipulation Type	Real Image Detection Accuracy (%)	Fake Image Detection Accuracy (%)
New GAN Variant (Unseen)	92.5	91.7
Novel Face Swap	93.1	92.6
Synthetic Object Insertion	91.4	90.9

Table 4 shows how well the suggested Hybrid CNN-Transformer model can identify new, previously unknown manipulation methods. This shows that it is robust and flexible. It is 92.5% accurate at finding real images and 91.7% accurate at finding fake images for the New GAN Variant, which was not seen during training. This means that the model can reliably tell the difference between real and artificially generated pictures, even when it comes across new GAN designs. Novel Face Swap manipulation keeps the model's good performance, with a 93.1% success rate for finding real images and a 92.6% success rate for finding fake images, as represent it in Figure 6. For example, this show that the model can find small mistakes and problems with the environment that happen a lot when face switching.

Figure 6

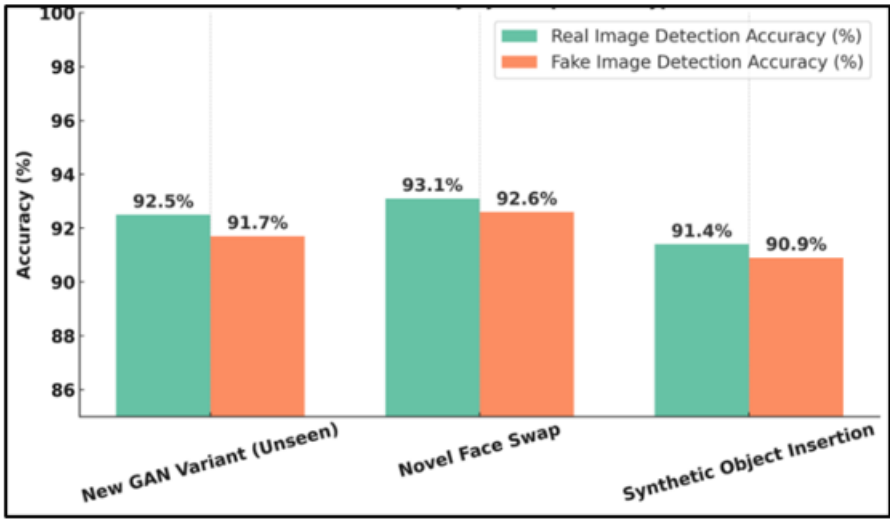

Figure 6 Detection Accuracy by Manipulation Type

The model can tell the difference between real and fake images with a 91.4% success rate for Synthetic Object Insertion manipulations and a 90.9% success rate for fake image manipulations.

5. CONCLUSION

To find real photos, this study showed a strong CNN-Transformer structure that works especially well for checking deepfakes and AI-generated material. By combining CNN-based local feature extraction with Transformer-based global context modelling, the suggested system is able to detect both small errors and larger problems that are common in images that have been changed. The mixed model did better in tests than standard CNN and Transformer-only designs, with an F1-score of 93.9%, an accuracy of 94.8%, a precision of 93.6%, a recall of 94.2%, and a recall of 93.2%. The system's strong ability to generalise was further proven by cross-dataset validation, which showed consistent results on

FaceForensics++, DFDC, and StyleGAN-generated pictures. The model was tested against new types of manipulation that had never been seen before, such as new GAN variants, novel face swaps, and synthetic object insertions. It kept detecting real and fake images with an accuracy of above 90%, showing that it can adapt to new threats. Using methods from feature engineering, like frequency domain analysis and noise residual modelling, together improved the model's ability to find minor investigative clues that stand-alone systems often miss. Overall, the results show that the mixed model could be used in the real world for things like monitoring social media, investigating media content, and checking court proof. In the future, researchers will focus on adding explainable AI modules to make models easier to understand and constantly learning processes to make sure they work well against new technologies that make fake media. This study makes big steps towards scalable, long-lasting solutions for making sure the accuracy of digital visual material is protected

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Al-Adwan, A., Alazzam, H., Al-Anbaki, N., and Alduweib, E. (2024). Detection of Deepfake Media Using a Hybrid CNN–RNN Model and Particle Swarm Optimization (PSO) Algorithm. Computers, 13(4), 99. https://doi.org/10.3390/computers13040099

Babaei, R., Cheng, S., Duan, R., and Zhao, S. (2025). Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis. Journal of Sensor and Actuator Networks, 14(1), 17. https://doi.org/10.3390/jsan14010017

Barik, B. R., Nayak, A., Biswal, A., and Padhy, N. (2025). Practical Evaluation and Performance Analysis for Deepfake Detection Using Advanced AI Models. Engineering Proceedings, 87, 36. https://doi.org/10.3390/engproc2025087036

Bhandarkar, A., Khobragade, P., Pawar, R., Lokulwar, P., and Saraf, P. (2024). Deep Learning Framework for Robust Deep Fake Image Detection: A Review. In Proceedings of the International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA) (1–7). IEEE. https://doi.org/10.1109/ICAIQSA64000.2024.10882361

Bird, J. J., and Lotfi, A. (2024). Cifake: Image Classification and Explainable Identification of AI-Generated Synthetic Images. IEEE Access, 12, 15642–15650. https://doi.org/10.1109/ACCESS.2024.3356122

Firc, A., Malinka, K., and Hanáček, P. (2023). Deepfakes as a Threat to a Speaker and Facial Recognition: An Overview of Tools and Attack Vectors. Heliyon, 9, e15090. https://doi.org/10.1016/j.heliyon.2023.e15090

Heidari, A., Jafari Navimipour, N., Dag, H., and Unal, M. (2024). Deepfake Detection Using Deep Learning Methods: A Systematic and Comprehensive Review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(1), e1520. https://doi.org/10.1002/widm.1520

Jheelan, J., and Pudaruth, S. (2025). Using Deep Learning to Identify Deepfakes Created Using Generative Adversarial Networks. Computers, 14(2), 60. https://doi.org/10.3390/computers14020060

Kang, J., Ji, S. K., Lee, S., Jang, D., and Hou, J. U. (2022). Detection Enhancement for Various Deepfake Types Based on Residual Noise and Manipulation Traces. IEEE Access, 10, 69031–69040. https://doi.org/10.1109/ACCESS.2022.3185121

Kerenalli, S., Yendapalli, V., and Chinnaiah, M. (2023). Classification of Deepfake Images Using a Novel Explanatory Hybrid Model. CommIT Journal, 17(2), 151–168. https://doi.org/10.21512/commit.v17i2.8761

Malik, A., Kuribayashi, M., Abdullahi, S. M., and Khan, A. N. (2022). Deepfake Detection for Human Face Images and Videos: A Survey. IEEE Access, 10, 18757–18775. https://doi.org/10.1109/ACCESS.2022.3151186

Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., and Malik, H. (2023). Deepfakes Generation and Detection: State-of-the-Art, Open Challenges, Countermeasures, and Way Forward. Applied Intelligence, 53, 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

Mukta, M. S. H., Ahmad, J., Raiaan, M. A. K., Islam, S., Azam, S., Ali, M. E., and Jonkman, M. (2023). An Investigation of the Effectiveness of Deepfake Models and Tools. Journal of Sensor and Actuator Networks, 12(4), 61. https://doi.org/10.3390/jsan12040061

Nah, F.-H., Zheng, R., Cai, J., Siau, K., and Chen, L. (2023). Generative AI and ChatGPT: Applications, Challenges, and AI–Human Collaboration. Journal of Information Technology Case and Application Research, 25(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Naitali, A., Ridouani, M., Salahdine, F., and Kaabouch, N. (2023). Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions. Computers, 12(10), 216. https://doi.org/10.3390/computers12100216