Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Review on Transfer Learning Approaches for Deep Learning- Based AI-Generated Image Detection

Shivani kamodiya
Department of Information Technology,
Mahakal Institute of Technology
Ujjain , India

Yashovardhan kelkar
Department of computer science and Engineering
Mahakal Institute of Technology
Ujjain , India

***Abstract*- The fast development of generative artificial intelligence (AI) between 2022 and 2025 has facilitated the creation of very realistic artificial images, increasing the anxieties about misinformation, internet fraud, identity, and content authenticity. It is a systemic review of AI-generated image detection methods using deep learning and a particular focus on transfer learning models. The main goals included assess the efficiency of transfer learning in the discriminative tasks of AI-generated versus real images, determine so-called pre-trained architectures that are more efficient on various datasets, determine robustness against the changing synthetic generation approaches, and determine approaches to finding a balance between detection and efficiency of AI-generated imaging.A literature review that covers the past five years (20202025) was conducted on the topics of evaluating the effectiveness of transfer learning, identifying pre-trained architectures that perform better in different datasets, evaluating the robustness under the changing approach to synthesizing Comparative analysis reveals that methodological tendencies, dataset contributions, and fundamental gaps in generalization and scalability are present.The results consist of the fact that transfer learning greatly improves detection accuracy, training time, in comparison to models trained fresh, and adaptability to limited labeled data. The pre-trained models with fine-tuning are always more successful than the traditional methods, but cross-domain generalization and real-world application issues still remain, transfer learning provides a flexible and more realistic scope of the development of a strong AI-generated image detection.**

Keywords- Transfer Learning, AI-Generated Image Detection, Deep Learning, Convolutional Neural Networks (CNNs), Synthetic Media Forensics

I. INTRODUCTION

The active development of the technologies of artificial intelligence (AI) and deep learning has radically changed the digital content production environment. Some of the most influential ones include generative models like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), diffusion models, and transformer-based networks that can be used to generate extremely realistic synthetic images. These can be AI-created images also known as deepfakes or synthetic media, which can reproduce human faces, objects, and other complex scenes with impressive visual accuracy [1], [2], [3], [4], [5]. Although such innovations present the world with the multiple benefits in both entertainment, virtual reality, art, advertising, and medical imaging, these innovations also present the world with severe ethical, social, and security challenges. The negative consequences associated with the misuse of AI-generated images to carry out misinformation, identity manipulation, fraud, and reputational damage have presented the necessity of a scalable and real-time detection system. The classic methods of image forensics are based on hand-crafted features, statistical anomalies, metadata processing or ad hoc methods of detecting manipulated images. As successful as these methods have proved to be in identifying simple editing operations, they frequently fail to identify images produced by modern AI systems that have low visible artifacts. Newer and newer generative models get closer to the statistical distribution of real images thus diminishing the usefulness of traditional detectors. This is a changing threat environment that requires the creation of intelligent, adaptive and high-level detection systems.

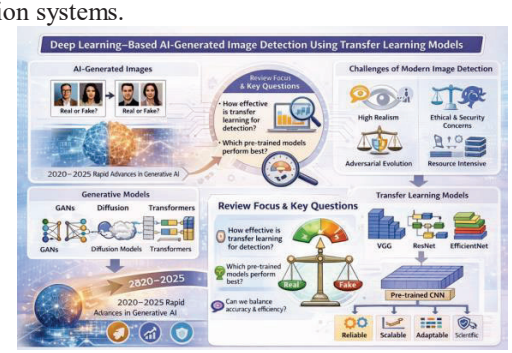

Fig. 1 Deep learning for AI image Detection

Deep learning has been a promising research to solve the problem of AI-generated image detection because it explicitly learns hierarchical and discriminative features automatically as a result of learning on data. CNNs, specifically, have performed well in the tasks of image classification and in forensic analysis. Nonetheless, deep neural networks that are trained in a standard way demand huge labeled datasets and a lot of computation and time. In most realistic cases, particularly in fast changing synthetic media settings, the process of gathering and labeling large volumes of data is costly and time-intensive. Furthermore, the models that are trained on small data commonly have overfitting and ineffective generalization in the face of unknown generative approaches [6], [7], [8], [9], [10].

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Transfer learning deals with these issues by relying on the information learned on a deep learning model previously trained on a large-scale dataset, e.g. ImageNet. In contrast to constructing models by starting from the ground up, transfer learning allows researchers to repurpose learned feature representations and fine-tune them to particular tasks, e.g. AI-generated image detection. This method is less computational, convergence is quicker, and performance is enhanced even in the case of limited labeled data. Yimpopular end-of-the-pipe models such as VGG, ResNet and EfficientNet have strong feature extraction properties, which can be modified to discriminate between real and synthetic images. Although there is an increasing interest in deep learning-based detection frameworks, a number of important questions are not answered yet. First, one has to conduct an assessment of the efficiency of transfer learning models to identify AI-generated images compared to conventional methods and models trained in a manner of the classical approach. Second, it is essential to find out the pre-trained architectures that can best optimize detection frameworks due to their excellent performance on a wide range of datasets and generative methods. Third, the resilience of the transfer learning models to the emergent and previously unexplored AI-generation techniques is not fully studied yet. As the generative technologies evolve at a very high pace, detection systems have to be generalized to no longer rely on known manipulation patterns in order to be relevant. Also, it is necessary to decrease the computational needs without the accuracy loss to allow real-time application in resource-limited settings. Lastly, the ability of transfer learning models to generalize the work of other image areas, facial images, natural scenes, or domain-specific data, needs to be systematically studied. Available studies reveal that there are major gaps that constrain the practical use of AI-generated image detectors. There is little knowledge of cross-domain and cross-model generalization because many studies work with a narrow dataset or specific GAN-based manipulations. Detection systems are prone to adversarial evolution, with the robustness to generative architectures that have been developed typically ignored. In addition, large labeled datasets and computationally intensive architectures are limiting to scalability and implementation in operational settings. The limitations also support the significance of thorough assessments and improved transfer learning plans that can integrate the performance, efficiency, and adaptability.

This is a systematic review of the AI image detection based on deep learning with emphasis on transfer learning models. It assesses the usefulness of pre-trained architectures, investigates how to enhance robustness and generalization and evaluates the trade-off between accuracy and computational efficiency. The literature review synthesis reveals essential gaps in research and gives an overview of the current developments and perspectives in synthetic image detection. With the increasing realism of AI-generated content, it is necessary to deploy robust detection mechanisms to ensure digital trust, information integrity, and cybersecurity. This review will aid in the formulation of scalable, flexible, and efficient detection frameworks in future applications.

II. RELATED WORK

Zhang et al., 2025[11] introduces the AI-created visual content detection technologies that have arisen over the last few years, split into two categories, namely AI-generated image detection and deepfake detection. The AI-generated image detection section begins by presenting the existing generative models and fundamental detection systems and surveys the existing detection system based on the perspective of unimodal and multimodal. The section on deepfake detection gives the overview of the existing classification of the techniques of deepfake generation, widely used datasets and after it an overview of some of the most popular evaluation metrics in the field. The technical properties of existing methods are also analyzed through the various feature information that they use and summarized and categorized. At last, we suggest the ways of future research and conclusions, to which we provide recommendations on how AI-generated visual content detection technologies can be developed.

Plested et al., 2025[12] describe deep transfer learning and the problem it is attempting to solve in terms of image classification in detail. We also examine the present position of the field, and where we have gotten in the recent past. We demonstrate the knowledge gaps that exist in the existing knowledge, and we provide the recommendations on the ways of advancing the field in order to cover these knowledge gaps. We introduce a novel taxonomy of the transfer learning application to image classification. The following taxonomy allows easier observation of the overall trends in the locations where transfer learning has worked and in those cases where it has not reached the promise it holds. This will also enable us to recommend where the issues are and where it will be utilized in a better way. We demonstrate that with this new taxonomy, several of the applications, in which transfer learning has been either demonstrated to be ineffective or even counterproductive to performance, are not surprising when these factors, the source and target datasets and methods applied, are considered.

Wang et al., 2025 [13] As they continue to evolve at a very fast pace, AI-generated images have become highly realistic and pose some serious implications that could be extremely challenging to determine the authenticity of digital content. The existing deepfake detection approaches regularly rely on the datasets of limited generation models and content diversity which do not match the complexity of the AI-generated material and its growing realism. Larger multimodal models (LMMs) which are popular in many vision tasks have shown great properties in zero-shot classification, but their capabilities in deepfake detection are barely investigated. To address this gap we introduce a large-scale DeepFake Benchmark, DFBench, (i) with great diversity, comprising 540,000 images of real, AI-edited, and AI-generated content, (ii) with the latest model, the fake images are generated by 12 state-of-the-art generation models, (iii) bidirectional benchmarking and evaluating both the detection accuracy of deepfake detectors and the evasion capability of generative models. According to DFBench, we have suggested MoA-DF, Mixture of Agents to DeepFake detection, which is using a combined probability approach of a group of LMMs. MoA-DF can deliver state-of-the-art results, which additionally prove the usefulness of the use of LMMs in detecting deepfakes. Publicly available are database and codes. at https://github.com/IntMeGroup/DFBench.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Y. Cao et al., 2025[14] is a review of the history of generative models and modern developments in AIGC, both in unimodal and multimodal interaction. In the view of unimodality, we present the text and image generation tasks and relative models. The cross-application between the above modalities is introduced as far as multimodality is concerned. Last but not least, the survey dwells on the current open issues and expected future challenges in AIGC. As a whole, this survey can be a good recommendation to find out the background and secrets of the amazing performance of AIGC techniques by people who show interest in its background and secrets.

Huang et al., 2025 [15] present RU-AI, a novel large-scale multimodal dataset, on robust and effective machine-generated content detection in text, image and voice. We build our dataset upon three massive publicly available datasets Flickr8K, COCO and Places205 by adding their respective AI counterparts, which yields a total of 1, 475, 370 instances. We also duplicated an extra variant of the noise in the dataset to evaluate the soundness of detection models. We did a lot of experimentation on the existing SOTA detection models on our data. The findings indicate that state of the art models are yet to perform accurate and robust detection on our dataset. We believe that such new data set can be used to advance the research in the area of machine-generated content detection and the responsible use of generative AI.

III. Background On AI-Generated Image Technologies

A. Evolution of Generative Models

The research area of AI-generated image synthesis has become much more advanced in the last ten years. Earlier methods of image generation used probabilistic graphical models and simple autoencoders which did not perform well in generating realistic results. Deep learning became a breaking point and allowed models to learn the complexes of data distribution based on large-scale data. Some of the earliest deep generative models that can learn latent representations and sample novel samples include Variational Autoencoders (VAEs). Though VAEs had an orderly latent space and stable training, images they generated were typically not sharp and detailed.

With the development of Generative Adversarial Networks (GAN), image synthesis was transformed with a new adversarial training process between two neural networks, a generator and a discriminator. This antagonistic paradigm allowed producing very realistic and high-resolution images. Advanced versions of GANs including DCGAN, StyleGAN, StyleGAN2, and BigGAN have enhanced the quality of visuals and the ability to control generated content in the future [16], [17], [18], [19], [20]

More recently, diffusion models and transformer-based generative architectures have become more fidelity- and diversity-dense, as compared to traditional GANs. The contemporary models have the ability to produce realistic images in the form of photographs which at times are hard to tell the difference between a real photograph and the produced photo. With the development of generative technologies, synthetic images become harder to detect and more realistic, which is why it is becoming more important to conduct a strong forensic analysis.

B. GAN-Based Image Generation

Generative Adversarial Networks (GANs) are still among the most powerful models in the generation of images by AI. A GAN is made up of two parts: a discriminator and a generator which is used to generate synthetic images out of random noise. Adversarial training allows the generator to learn to generate outputs which resemble real-world outputs more closely, and the discriminator to become more skilled at detecting small inconsistencies [21].

With time, there have been many variants of GAN that have been created to enhance training stability, resolution, and controllability. As an example, Chapter StyleGAN has proposed a style-based architecture and adaptive instance normalization, which allows to control face features at a fine level. Conditional GANs (cGANs) are based on the label information to create the images of a particular class. CycleGAN and Pix2Pix also support image-image translation, which further enhances synthetic manipulation [22], [23], [24].

Although the results of the GAN-based images are impressive, they usually carry a slight trace of artifact in texture, frequency distributions or a lack of color consistency. These artifacts were used in early detection methods to employ either frequency-domain analysis or handcrafted statistical features. But with the evolution of GAN architecture, these artifacts diminish and other conventional forensic methods are no longer as useful [25].

C. Diffusion Models and Transformer-Based Generators

Diffusion models are a major step in generative modeling. In contrast to GANs, diffusion models have two-step functionality: they are trained by adding noise to real images and trained to undo this noise in the creation of images. These models can be used to produce high quality images with better stability and diversity because they generate images through denoising random noise samples. Popular diffusion-based systems have shown outstanding performance in the area of text-to-image synthesis where they generate very detailed images as well as those which are contextually consistent.

Transformer-based architectures, which were originally created in the domain of natural language processing, have been applied to image generation. Vision Transformers (ViT) and multimodal transformer models are models that use self-attention-based mechanisms to achieve long-range dependencies and global contextual information. Transformer-based diffusion models and autoregressive transformers also support highly controlled style, composition, and semantics in generative tasks.

These new generative methods go a long way in minimizing traditional artifacts in GAN based images thus complicating the detection. They require more sophisticated detection techniques that can detect more minor distributional changes than the blatant visual anomalies that their capability to generate globally consistent textures and natural lighting effects requires.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

D. Challenges in Detecting Synthetic Images

The fast development of generative models poses considerable challenges to AI-generated image detection. With increasingly realistic images, artifacts become less visible, and rule-based and hand-made techniques are less effective. One of the most critical problems is the cross-model generalization where models, which have been trained on a set of techniques, tend to fail on unknown techniques.

Other issues are adversarial robustness, where the generative models reduce observable artifacts and domain variability, such as resolution, compression, lighting, and content differences. It has high computational requirements that further restrict real-time deployment. These challenges highlight the need for adaptive, scalable, and generalizable detection frameworks.

IV. Fundamentals Of Transfer Learning In Image Classification

A. oncept and Types of Transfer Learning

Transfer learning is knowledge transfer that takes the knowledge of a previously trained model and uses it to enhance performance on a similar task. ImageNet-trained models, or models trained with large datasets, acquire general visual features, like edges, textures, and patterns, and can be re-purposed in other tasks, such as AI-generated image detection. It encompasses three major techniques which include feature extraction, fine-tuning and domain adaptation. The pre-trained models are used as fixed feature extractors in feature extraction and fine-tuning is used to update the model layers with task-specific learning. Domain adaptation is interested in knowledge transfer between varying data distributions [25], [26], [27], [28], [29].

B. eature-Based vs Fine-Tuning Approaches

In feature-based transfer learning, a pre-trained model is frozen at the convolutional level, and the extra classification layers are only trained. The method is less expensive in computations and is appropriate when the target population is small. Nevertheless, it can inhibit flexibility to domain patterns.

Fine-tuning however, permits partial or complete updating of model weights. The network can be unfreezed by learning more task-specific features that are useful in identifying synthetic artifacts. Fine-tuning often produces more accurate results although hyperparameters need to be carefully chosen to avoid overfitting [30].

C. Advantages Over Training from Scratch

There are a number of benefits of transfer learning over deep-network training. To begin with, it saves much training time and computation. Second, it enhances performance whereby labelled data is small. Third, pre-trained models give strong low-level and mid-level features representations that improve generalization in a wide variety of datasets. Transfer learning is especially appropriate in cases where AI-generation image detection is the task at hand, and, on the one hand, diversity of the data and patterns of its changes are the key issues of interest.

D. Limitations of Transfer Learning

Transfer learning has its limitations although it has its benefits. The representations of features that are trained using natural images sets might not be able to grasp the fine forensic clues that are embedded in synthetic images. Negative transfer may take place in case the source and target domains are very different. Secondly, even large pre-trained models might demand large computational resources, restricting use to edge devices or real-time systems.

V. Transfer Learning–Based AI-Generated Image Detection

This section presents AI-generated image detection and transfer learning by using the pre-trained CNNs, including VGG, ResNet, EfficientNet, and Vision Transformers. It addresses binary and multi-classification, hybrid, and multimodal methods that utilize spatial, frequency, and semantic features in order to make the methods more robust, better at generalisation, and at detecting objects [30].

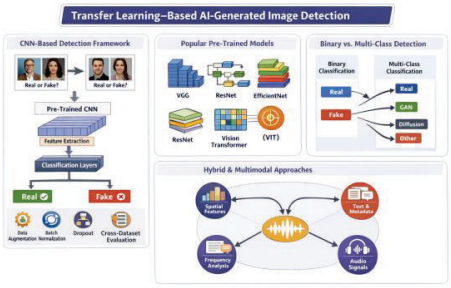

Fig. 2 Transfer learning–based AI-generated image detection

A. NN-Based Detection Frameworks

Most AI-based image detection systems are based on Convolutional Neural Networks (CNNs). CNNs automatically learn the hierarchical features of images, and they reveal the spatial patterns, textures and structural inconsistencies [31], [32], [33], [33], [34]. With transfer learning models, trained CNNs can be used as feature extractors, and then trained classification layers can be used to differentiate real and synthetic images. In order to augment data and increase robustness, detection frameworks commonly include data augmentation, batch normalization, dropout regularization, and early stopping. Generalization capability is usually evaluated by using cross-dataset [35].

B. opular Pre-Trained Architectures

1) VGG

VGG networks have a characteristic of being simple but efficient and are composed of consecutive convolutional layers that have small receptive fields. The fact that they are uniform in their structure is why they are likely to be transferred to learning, but they are quite computationally intensive because of the high number of parameters.

2) ResNet

ResNet provides residual connections that alleviate the vanishing gradient issues, and deeper architectures are

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

possible. It has a high capability to learn complicated hierarchies of features and this enables it to be very effective in synthetic image detecting activities.

3) EfficientNet
EfficientNet scales network width, depth and resolution through a compound scaling strategy. It is also highly accurate with minimal parameters in contrast to the traditional CNNs, which makes it appropriate in computationally efficient detection systems.

4) Vision Transformers (ViT)
Vision Transformers use the self-attention mechanisms on image patches by capturing the global contextual relations. ViT models are very successful in image classification and have a potential to be useful in the detection of globally consistent synthetic artifacts.

C. Binary vs Multi-Class Detection Strategies
Majority of AI-generated image detection models use binary classification (real vs synthetic). Nonetheless, multi-class strategies seek to discover particular generative techniques, including alternative GAN or diffusion models. Multi-class classification gives more forensic details and would need more extensive labeled data whereas binary detection would make it easier to deploy.

D. Hybrid and Multimodal Approaches
Recent studies discuss hybrid systems of detection based on spatial features, frequency-domain features, and semantic features. Multi modal methods incorporate image data and metadata, textual descriptions or audio cues in order to enhance robustness. The objective of these systems is to increase the detection accuracy and become more resilient to the changing generative technologies.

VI. Comparative Analysis

The section will provide a systematic comparison of recent research on AI-generated image and multimedia detection with respect to the variations of methodologies, datasets, interpretability, and overall generalization abilities. It is noted in the analysis that the shift of the traditional convolutional neural network (CNN)-based methods towards the more advanced multimodal, explainable, and large-scale models is quite obvious.

Recent studies are also incorporating Multimodal Large Language Models (MLLMs), hybrid systems, and instruction-conditioned systems to boost detection performance and interpretability. These methods show effective ability to learn more detailed visual-semantic relations and enhance the robustness in different data sets. Also, explainability has become an important point of research, which allows making decisions in a human-verifiable manner and enhances the confidence in detection systems.

Nonetheless, there are a number of difficulties. A large number of modern models are based on large datasets and computationally expensive models, making them less applicable in real-time or when resources are limited. Moreover, there is no uniform benchmarking systems, and thus it is hard to carry out fair comparisons on various studies. Although the multimodal methods enhance generalization and interpretability, they can create a higher level of complexity and training costs.

Conversely, classical CNN-based models are still computationally efficient and simpler to implement, but have weaknesses including low cross-dataset generalization, and reliance on particular generative artifacts. In general, the analysis indicates a trade-off between accuracy, interpretability, and computational efficiency, and some of the main gaps in research such as the necessity of lightweight models, better generalization, andard evaluation benchmarks, and practical deployment plans.

Table I. Comparative Analysis

Study	Year	Focus Area	Methodology / Model	Dataset Contribution	Key Strengths	Limitations / Research Gap
Zhou et al. (AIGI-Holmes) [36]	2025	Explainable AI-generated image detection	Multimodal Large Language Models (MLLMs) with Holmes pipeline (visual expert pre-training, SFT, DPO, collaborative decoding)	Holmes-Set with explanation-based annotations	High interpretability; strong generalization; visual-semantic reasoning	High computational cost; dependency on large models; limited focus on lightweight deployment
L. Cao (Survey on AIGC Detection) [37]	2025	Detection and mitigation of AI-generated content (text, image, audio)	Observation-based, statistical, watermarking, model-based and ensemble methods	No new dataset (survey study)	Comprehensive coverage; includes ethical, policy, and robustness aspects	Lack of experimental validation; no quantitative comparison; limited emphasis on transfer learning
Li et al. (FakeScope) [38]	2025	Interpretable AI-generated image forensics	Multimodal expert model (LMM) with token-based probability estimation and instruction tuning	FakeChain + FakeInstruct (2M multimodal instructions)	Strong interpretability; state-of-the-art performance; good real-world generalization	High training complexity; resource-intensive; limited efficiency analysis
Lin et al. (LAIM Survey) [39]	2025	Detection of large AI model-generated multimedia	Taxonomy-based systematic survey	Compilation of public datasets and tools	Comprehensive taxonomy; strong conceptual framework; societal relevance	No experimental validation; limited model-level comparison
Ashani et al. [40]	2025	Deepfake image detection	CNN-based models (VGG16, VGG19, ResNet50)	1,200-image dataset (FaceApp + real images)	High accuracy; practical comparison of CNN models	Small dataset; limited diversity; weak generalization; no cross-dataset validation

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

V. Conclusion

This review provides a detailed discussion of deep learning-based methods in AI-generated image detection, especially transfer learning methods. The paper analyzes the performance of the pre-trained models to differentiate between fake and real images, their flexibility to the quickly developed generative models, and the trade-off between detection and computational efficiency.

The results confirm the claim that transfer learning leads to much better detection performance than models that are trained on their own. VGG, ResNet, EfficientNet, and Vision Transformers are pre-trained architectures that offer good feature extraction abilities, shorter training periods, and higher accuracy, especially on smaller volumes of labeled data. Fine-tuning techniques also promote the adaptability of the model, as they allow learning task-specific characteristics applicable to the detection of synthetic images.

Although there are these benefits, there are a number of challenges that are yet to be addressed. Limitations in cross-model and cross-domain generalization, susceptibility to invisible generative methods, instability to adversarial manipulation, and challenges with real-time deployment due to high computational costs are all key issues. Moreover, most of the current models are tested on small datasets, which restricts their practical application. In order to overcome these shortcomings, future studies need to consider coming up with lightweight and efficient architectures, enhancing cross-dataset benchmarking, and combining hybrid and multimodal designs in order to improve robustness and scalability. Explainability and adaptability should be further prioritized as well to guarantee reliable performance under dynamic conditions, transfer learning provides a scalable and efficient solution to AI-generated image detection, but additional work on model design, dataset diversity, and evaluation is necessary to provide robust and reliable real-world performance.

- Future Directions

Future studies must be aimed at enhancing cross-model and cross-domain generalization to be resistant to novel generative methods, especially those of diffusion and transformer-based models. Real time and edge deployment will also be essential because of the development of lightweight and energy-efficient transfer learning architectures. Benchmark data of large scale, diverse and constantly updated datasets is required to test in the real world. Also, enhance interpretability and trust can be achieved by adding multimodal cues, frequency-domain analysis, and explainable AI methods. The concept of adversarial robustness and lifelong learning mechanisms also needs to be investigated in order to evolve detection systems to generative models that change very fast and guarantee its sustainability and preservation of authenticity of digital content in the long term.

References

[1] N. Manakitsa, G. S. Maraslidis, L. Moysis, and G. F. Fragulis, "A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision," *Technologies*, vol. 12, no. 2, 2024, doi: 10.3390/technologies12020015.

[2] M. Z. Khaliki and M. S. Başarslan, "Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN," *Sci. Rep.*, vol. 14, no. 1, pp. 1–10, 2024, doi: 10.1038/s41598-024-52823-9.

[3] K. Sun *et al.*, "DiffusionFake: Enhancing Generalization in Deepfake Detection via Guided Stable Diffusion," *Adv. Neural Inf. Process. Syst.*, vol. 37, no. NeurIPS, pp. 1–24, 2024, doi: 10.52202/079017-3218.

[4] L. Zhang, X. Liu, A. V. Martin, C. X. Bearfield, Y. Brun, and H. Guan, "Attack-Resilient Image Watermarking Using Stable Diffusion," *Adv. Neural Inf. Process. Syst.*, vol. 37, no. NeurIPS, 2024, doi: 10.52202/079017-1215.

[5] S. Li *et al.*, "Introduction to the Special Issue on AI-Generated Content for Multimedia," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 6809–6813, 2024, doi: 10.1109/TCSVT.2024.3427488.

[6] O. A. H. H. Al-Dulaimi and S. Kurnaz, "A Hybrid CNN-LSTM Approach for Precision Deepfake Image Detection Based on Transfer Learning," *Electron.*, vol. 13, no. 9, pp. 1–22, 2024, doi: 10.3390/electronics13091662.

[7] A. Zulfiqar, S. Muhammad Daudpota, A. Shariq Imran, Z. Kastrati,
M. Ullah, and S. Sadhwani, "Synthetic Image Generation Using Deep Learning: A Systematic Literature Review," *Comput. Intell.*, vol. 40, no. 5, pp. 1–22, 2024, doi: 10.1111/coin.70002.

[8] J. J. Bird and A. Lotfi, "CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images," *IEEE Access*, vol. 12, no. December 2023, pp. 15642–15650, 2024, doi: 10.1109/ACCESS.2024.3356122.

[9] G. Cazenavette, A. Sud, T. Leung, and B. Usman, "FakeInversion: Learning to Detect Images from Unseen Text-to-Image Models by Inverting Stable Diffusion," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 10759–10769, 2024, doi: 10.1109/CVPR52733.2024.01023.

[10] S. M. Abdullah *et al.*, "An Analysis of Recent Advances in Deepfake Image Detection in an Evolving Threat Landscape," *Proc.
- IEEE Symp. Secur. Priv.*, pp. 91–109, 2024, doi: 10.1109/SP54263.2024.00194.

[11] Y. Zhang, Z. Pang, S. Huang, C. Wang, and X. Zhou, "Unmasking AI-created visual content: a review of generated images and deepfake detection technologies," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 37, no. 6, 2025, doi: 10.1007/s44443-025-00154-8.

[12] J. Plested, M. Phiri, and T. Gedeon, "Deep transfer learning for image classification: a survey," pp. 1–50, 2025, [Online]. Available: http://arxiv.org/abs/2205.09904

[13] J. Wang *et al.*, "DFBench: Benchmarking Deepfake Image Detection Capability of Large Multimodal Models," *MM 2025 -Proc. 33rd ACM Int. Conf. Multimedia, Co-Located with MM 2025*, no. February, pp. 12666–12673, 2025, doi: 10.1145/3746027.3758204.

[14] Y. Cao *et al.*, "A Survey of AI-Generated Content (AIGC)," *ACM Comput. Surv.*, vol. 57, no. 5, 2025, doi: 10.1145/3704262.

[15] L. Huang, Z. Zhang, Y. Zhang, X. Zhou, and S. Wang, "RU-AI: A Large Multimodal Dataset for Machine-Generated Content Detection," *WWW Companion 2025 - Companion Proc. ACM Web Conf. 2025*, vol. 2025, no. May 2025, pp. 733–736, 2025, doi: 10.1145/3701716.3715306.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

[16] R. Huang, L. Dugan, Y. Yang, and C. Callison-Burch, "MiRAGeNews: Multimodal Realistic AI-Generated News Detection," *EMNLP 2024 - 2024 Conf. Empir. Methods Nat. Lang. Process. Find. EMNLP 2024*, pp. 16436–16448, 2024, doi: 10.18653/v1/2024.findings-emnlp.959.

[17] A. Panda, D. Panigrahi, S. Mitra, S. Mittal, and S. Rahimi, "Transfer Learning Applied to Computer Vision Problems: Survey on Current Progress, Limitations, and Opportunities," vol. 1, no. 1, pp. 1–16, 2024, [Online]. Available: http://arxiv.org/abs/2409.07736

[18] X. Yu *et al.*, "Fake Artificial Intelligence Generated Contents (FAIGC): A Survey of Theories, Detection Methods, and Opportunities," vol. 3162, pp. 0–3, 2024, [Online]. Available: http://arxiv.org/abs/2405.00711

[19] Ans Ibrahim Mahameed, "Transfer Learning-Based Models for Comparative Evaluation for the Detection of AI-Generated Images," *J. Electr. Syst.*, vol. 20, no. 6s, pp. 2570–2578, 2024, doi: 10.52783/jes.3244.

[20] Z. Wang *et al.*, "DIRE for Diffusion-Generated Image Detection," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 22388–22398, 2023, doi: 10.1109/ICCV51070.2023.02051.

[21] Y. Patel *et al.*, "An Improved Dense CNN Architecture for Deepfake Image Detection," *IEEE Access*, vol. 11, no. March, pp. 22081–22095, 2023, doi: 10.1109/ACCESS.2023.3251417.

[22] N. Ullah *et al.*, "An Effective Approach to Detect and Identify Brain Tumors Using Transfer Learning," *Appl. Sci.*, vol. 12, no. 11, 2022, doi: 10.3390/app12115645.

[23] A. A. Mukhlif, B. Al-Khateeb, and M. A. Mohammed, "An extensive review of state-of-the-art transfer learning techniques used in medical imaging: Open issues and challenges," *J. Intell. Syst.*, vol. 31, no. 1, pp. 1085–1111, 2022, doi: 10.1515/jisys-2022-0198.

[24] T. Tamagusko, M. G. Correia, M. A. Huynh, and A. Ferreira, "Deep Learning applied to Road Accident Detection with Transfer Learning and Synthetic Images," *Transp. Res. Procedia*, vol. 64, no. C, pp. 90–97, 2022, doi: 10.1016/j.trpro.2022.09.012.

[25] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," *Technologies*, vol. 11, no. 2, pp. 1–14, 2023, doi: 10.3390/ technologies11020040.

[26] A. W. Salehi *et al.*, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustain.*, vol. 15, no. 7, 2023, doi: 10.3390/su15075930.

[27] R. Rafique, R. Gantassi, R. Amin, J. Frnda, A. Mustapha, and A. H. Alshehri, "Deep fake detection and classification using error-level analysis and deep learning," *Sci. Rep.*, vol. 13, no. 1, pp. 1–13, 2023, doi: 10.1038/s41598-023-34629-3.

[28] M. Zhu *et al.*, "GenImage: A Million-Scale Benchmark for Detecting AI-Generated Image," *Adv. Neural Inf. Process. Syst.*, vol. 36, no. NeurIPS, 2023.

[29] P. Fernandez, G. Couairon, H. Jégou, M. Douze, and T. Furon, "The Stable Signature: Rooting Watermarks in Latent Diffusion Models," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 22409–22420, 2023, doi: 10.1109/ICCV51070.2023.02053.

[30] C. G. Simhadri and H. K. Kondaveeti, "Automatic Recognition of Rice Leaf Diseases Using Transfer Learning," *Agronomy*, vol. 13, no. 4, 2023, doi: 10.3390/agronomy13040961.

[31] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00652-w.

[32] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Med. Imaging*, vol. 22, no. 1, pp. 1–13, 2022, doi: 10.1186/s12880-022-00793-7.

[33] J. Gupta, S. Pathak, and G. Kumar, "Deep Learning (CNN) and Transfer Learning: A Review," *J. Phys. Conf. Ser.*, vol. 2273, no. 1, 2022, doi: 10.1088/1742-6596/2273/1/012029.

[34] A. Raza, K. Munir, and M. Almutairi, "A Novel Deep Learning Approach for Deepfake Image Detection," *Appl. Sci.*, vol. 12, no. 19, 2022, doi: 10.3390/app12199820.

[35] B. B. Sapkota *et al.*, "Use of synthetic images for training a deep learning model for weed detection and biomass estimation in cotton," *Sci. Rep.*, vol. 12, no. 1, pp. 1–18, 2022, doi: 10.1038/s41598-022-23399-z.

[36] Z. Zhou *et al.*, "AIGI-Holmes: Towards Explainable and Generalizable AI-Generated Image Detection via Multimodal Large Language Models," 2025, [Online]. Available: http://arxiv.org/abs/2507.02664

[37] L. Cao, "A Practical Synthesis of Detecting AI-Generated Textual, Visual, and Audio Content," 2025, [Online]. Available: http://arxiv.org/abs/2504.02898

[38] Y. Li *et al.*, "FakeScope: Large Multimodal Expert Model for Transparent AI-Generated Image Forensics," vol. 1, pp. 1–14, 2025, [Online]. Available: http://arxiv.org/abs/2503.24267

[39] L. Lin *et al.*, "Detecting Multimedia Generated by Large AI Models: A Survey," pp. 1–38, 2025, [Online]. Available: http://arxiv.org/abs/2402.00045

[40] Z. N. Ashani *et al.*, "Comparative Analysis of Deepfake Image Detection Method Using VGG16 , VGG19 and ResNet50," vol. 1, no. 1, pp. 16–28, 2025.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)