Image Authentication: Differentiating Camera-Captured and AI-Generated Images with Provenance Verification using Vision Transformer (ViT)

K. Yasudha

Department of Computer Science, GITAM School of Science, GITAM (Deemed to be University), Visakhapatnam

DOI: https://doi.org/10.51584/IJRIAS.2026.11060074

Received: 04 June 2026; Accepted: 10 June 2026; Published: 23 June 2026

ABSTRACT

The rapid growth of generative artificial intelligence significantly affects the development of digital image creation. Recently, advances in Generative Adversarial Networks (GAN) and Diffusion-based Generative Networks have made it possible to create synthetic images that are hard to tell apart from real ones. Therefore, verifying images produced by artificial intelligence is essential. This poses a major challenge for many fields, including digital forensics, security, and image verification. There has been a rise in the misuse of AI-generated images to spread false news, impersonate people, and manipulate images. Traditional methods for verifying image authenticity, such as human observation and image metadata analysis, are now unreliable. AI-generated images can be easily altered, and human observation alone cannot confirm an image's authenticity. As a result, there is a pressing need to develop an effective image authentication system to tell apart real images from those created by AI.

This paper proposes an automated image authentication system. It utilizes a Vision Transformer model for classifying images as either camera-captured or AI-generated. The system employs a pre-trained model for feature extraction, which is then fine-tuned for classification. Unlike conventional convolutional neural networks, the Vision Transformer treats an image as a sequence of patches and uses self-attention to capture global dependencies. This method helps to identify subtle differences in AI-generated images. Additionally, the proposed system incorporates a confidence level represented by Softmax probabilities, which helps understand the reliability of the system's results. An explainability feature is also included, using Explainable Artificial Intelligence techniques to highlight areas in the image that influence the results. This system provides a strong solution for the current challenges in image authentication. It can be implemented as a web-based application using the Flask framework. Experimental results demonstrate that the system achieves high accuracy in classifying AI-generated images.

Keywords: Image Authentication, Vision Transformer, AI generated images, Explainable Artificial Intelligence, Deep learning, Image classification, Synthetic image detection

INTRODUCTION

The rapid growth in artificial intelligence has led to significant innovations across various fields. One noteworthy advance involves generating images using generative models. The quality of these images has changed, moving from low-quality synthetic images to realistic and visually acceptable ones. Several methods have been introduced for generating images with generative models, including Generative Adversarial Networks (GANs) and Variational Auto Encoders (VAEs), which have greatly enhanced the quality of synthesized images.

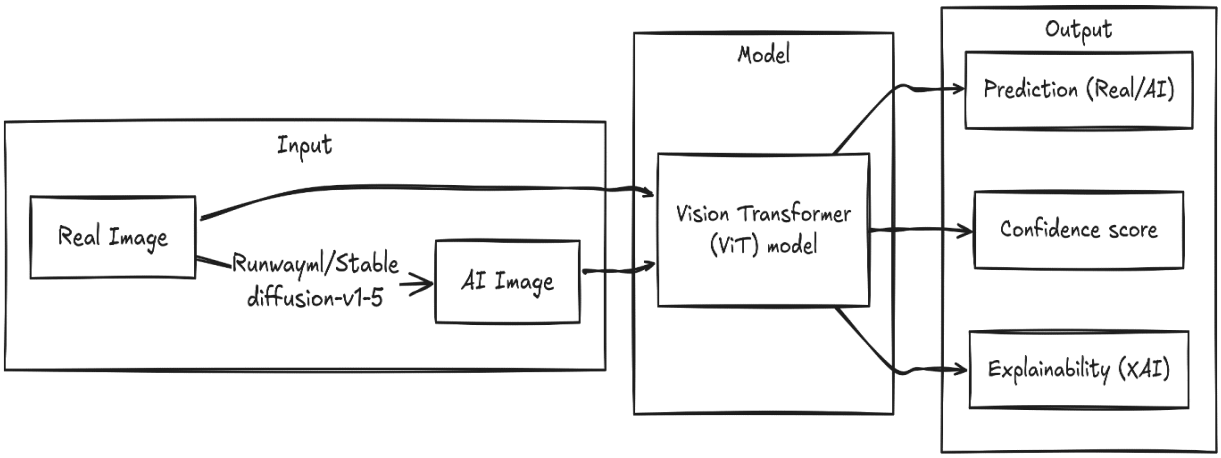
Despite the promise of generative models, the rise in quality of synthetic images has sparked serious worries about the authenticity and trustworthiness of digital media. People often use synthetic images to manipulate public opinion, spread false information, or create fake identities. With the ease of creating realistic fake images using various accessible generative AI models, there is a growing need to develop methods for verifying the authenticity of digital images.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

Traditional methods for image authentication, such as manual inspection and analysis of image metadata, are now insufficient. Manual inspection relies on human observation, which is often unable to catch the subtle differences in images produced by generative models. Likewise, analyzing image metadata is not reliable for image authentication, as it is easy to modify, delete, or forge this information. This highlights the need to create image classification systems that can effectively sort images. Deep learning has proven to be a useful approach for tackling complex image classification challenges. Convolutional Neural Networks (CNN) have been widely applied to image recognition tasks. CNNs excel in this area because they learn hierarchical representations of images, focusing on local features.

However, concentrating only on local features may not address the global inconsistencies that can arise in AI-generated images, which occur in different parts of the image and require a comprehensive view. Vision Transformers offer another solution for image recognition issues, serving as an alternative to CNNs. Their operation resembles traditional transformer models used for natural language processing tasks. In Vision Transformers, images are treated as sequences of patches. The self-attention mechanism within Vision Transformers enables the model to examine long-range relationships within an image, allowing it to capture subtle patterns, especially in AI-generated images. This ability to learn both local and global features positions Vision Transformers as an effective method for image authentication. The proposed work presents an image authentication system based on Vision Transformers.

This system can classify images, whether they are taken by a camera or generated through AI. It also boasts advantages like confidence and interpretability, and it can be integrated with a web application for user convenience. The main contributions of this paper include developing a classification model based on a transformer, incorporating confidence through Softmax probabilities, adding Explainable AI (XAI) techniques, and deploying the system as a web application. These contributions collectively create a robust solution for the challenges of image authentication in the digital age.



LITERATURE SURVEY

The rapid evolution and improvement of generative artificial intelligence have led to an increase in the amount of research being conducted to develop effective techniques for the detection of AI-generated images. This is due to the improvement in the quality and realism of generative artificial intelligence. As a result, the complexity of distinguishing between synthetic and real images has increased. As a result, different techniques, including traditional image processing techniques and deep learning techniques, have been used for this purpose.

Traditionally, different techniques have been used for image authentication, which include the use of different handcrafted features such as edge detection, noise, and statistical characteristics. The techniques used for image authentication are based on the detection of inconsistencies within the image, which are usually introduced during the manipulation process. However, with the evolution and improvement of generative artificial intelligence, the traditional techniques were not found to be effective due to the inability of the techniques to detect semantic features.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

However, with the advent of deep learning, convolutional neural networks (CNN) were developed for image classification and detection of image forgery. ResNet, and EfficientNet were some of the CNN architectures developed for image classification. These networks were developed for learning the hierarchical representations of images. The CNN-based techniques were found to be efficient for learning low-level features such as edge detection, as well as learning high-level features such as shape and pattern detection. However, the techniques based on CNN are mostly based on local spatial information, which may not be sufficient for identifying global inconsistencies within AI-generated images. However, recently, different researchers have started adopting different techniques based on the use of transformers, i.e., Vision Transformers, which have been found to be efficient for image classification tasks. Vision Transformers are found to be efficient for learning the relationships between different regions within an image, which enables the model to analyze global inconsistencies within AI-generated images.

Vision Transformers are capable of extracting both local and global features of images. This makes it a suitable candidate for detecting AI-generated images. Apart from frequency domain-based approaches, researchers are also adopting spatial domain-based approaches for detecting AI-generated images. Another significant area of research is explainable artificial intelligence (XAI), which aims at increasing the explainability of deep learning models. Various techniques, including XAI, have been employed for detecting AI-generated images. The areas of images are being considered for prediction using these techniques. One of the significant benefits of using XAI techniques is explainability, which can be achieved using image authentication systems.

Despite all these advancements, there are several challenges that need to be addressed. One of the major issues is that the models are not capable of generalizing across different AI-generated images. Another major problem is that researchers are giving more priority to offline detection, whereas little attention is being paid to real-time detection. One more problem is explainability, which is also being faced.

The proposed system aims at resolving the challenges faced due to a lack of generalization and real-time detection. The proposed system is based on the transformer model, which is capable of providing accurate and efficient detection of AI-generated images. The system is also capable of providing explainability.

SYSTEM ARCHITECTURE

The system architecture of the proposed image authentication system has been developed in a structured format. This ensures the efficient processing of the image provided in the input and the generation of output results.

The image authentication system has been developed with various layers. The first layer is the user interface layer. In this layer, users interact with the image authentication system with the help of a web browser. This interface has been developed in a user-friendly format. This ensures that users can upload images directly from their devices or provide image links. This ensures that the image authentication system supports various types of sources for image input.

The image provided in the input by users is processed in the application layer. In this layer, the image authentication system has been developed with the help of a backend server using Flask. This backend server serves as a bridge between the user interface and the deep learning model. Once the image is received by the backend server, preprocessing of the image takes place. In this step, the size of the image is adjusted to a standard size of 224 x 224 pixels. This ensures that the image authentication system supports the requirements of the Vision Transformer model. The normalization of the image takes place to ensure uniform distribution of pixel values. This ensures efficient performance of the model. Once the preprocessing of the image is complete, the image is converted to tensor form.

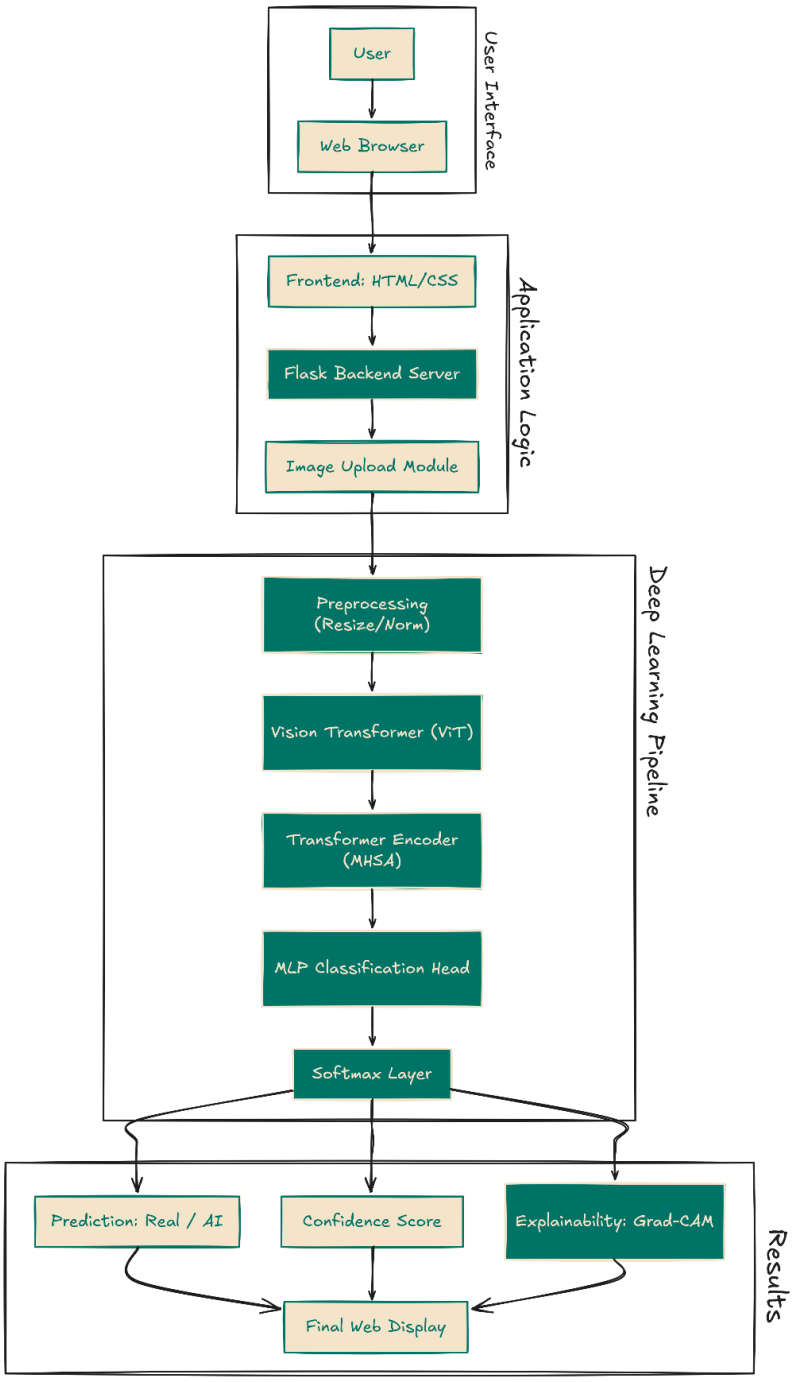
The Vision Transformer model is the core component of the system. This component carries out the process of feature extraction and classification. In this model, the image is segmented into different patches of 16 x 16 pixels. These patches are then flattened and represented in a vector form, which represents an embedding vector. This process of embedding ensures that the model can recognize the spatial relationship between different patches of the image.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

This sequence of embedded patches is then fed into the transformer encoder layers, where multiple head self-attention mechanisms are applied. This process enables the model to learn the different relationships between different regions of the image. This is particularly important in the case of AI-generated images, as the inconsistencies in these images are spread out over the entire image, rather than in a localized manner.

The features learned in the transformer encoder are then fed into a multilayer perceptron classification head. The output of this classification head takes the form of a logit, which is then mapped to a probability using the Softmax function. The class with the highest probability is then selected as the final output, which takes the form of a probability value, which is then displayed as the confidence score.Apart from the classification process, the system also includes a separate module for explainability, which makes use of the XAI Techniques. This module generates a heatmap representing the regions of the image that are considered the most important for the classification process.

The final output, which includes the predicted class and the confidence, is displayed on the web interface, which makes it easy for the user to interpret the results and make an informed decision regarding the authenticity of the image.



ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

PROPOSED METHODOLOGY

The methodology proposed also indicates the whole process involved in the development of the image authentication system. It is important to note that there are several stages involved in the methodology, including data preparation, preprocessing, feature extraction using Vision Transformers, classification, generation of the confidence score, and explainability. Each of these stages is well designed for ensuring the efficiency of AI-generated image detection, along with accuracy in the process.

The methodology for image authentication is designed in a pipeline, whereby several transformations are applied to the image for generating the output. The use of a transformer-based model ensures that local and global features are being captured for detecting even subtle inconsistencies within AI-generated images.

Data Preparation and Dataset Description

The dataset used in this research includes images obtained from real-world scenarios. These images are obtained from publicly available datasets hosted on Kaggle. Along with this, images are also obtained synthetically by utilizing a variety of models based on the concept of a diffusion-based model. The dataset includes a variety of natural images, objects, and scenarios. This category of images includes images of the real world obtained through a camera. This dataset includes a ground truth for images of the “real” category. In order to generate images based on AI technology, a diffusion model-based approach has been used. In this approach, a Stable Diffusion model has been utilized.

In this approach, a model named “runwayml/stable-diffusion-v1-5” has been utilized. This model includes a Stable Diffusion Img2Img model. This model helps in transforming images from a real-world dataset to a synthetic dataset. This dataset includes a variety of images based on a similar structure. The generation process takes place by passing the input images along with the prompts, which in turn aids in the generation of realistic synthetic images. This process also ensures the maintenance of semantic similarity between the original and generated images, along with the addition of small artifacts that are associated with AI-generated images. The safety checker of the diffusion pipeline is disabled during the generation process, as it is used for academic purposes. The dataset is maintained in a balanced state, with an equal number of actual and AI-generated images, which in turn prevents bias during the training process. The dataset is then split into training and testing sets, with a ratio of 75% and 25% respectively, in order to ensure that the model learns meaningful patterns during the training process. One of the significant advantages of the dataset construction methodology lies in the generation of synthetic images in a controlled environment.

The images generated by the AI are based on real images, and as a result, a dataset of these pairs exists, which in turn aids in the learning of finer differences, such as the presence of inconsistent textures, unnatural smoothness, etc., which are associated with images generated by AI-based models. The dataset acts as a robust base for training the Vision Transformer model, which in turn aids in the learning of key features that can help in differentiating between images taken by a camera and images generated by AI-based models.

Image Preprocessing

Before the image is input into the model, a series of preprocessing activities take place to ensure consistency and compatibility of the image with the Vision Transformer model. To begin with, the image is resized to a standard size of 224 x 224 pixels. This is important to ensure that the images are of a standard size, a prerequisite before batch processing.

Next, the image is subjected to normalization, whereby the pixel intensity values of the images are normalized. This is important to ensure consistency and stability of the model and enhance the training speed. The image is then converted to a tensor format. This format is important as it ensures the image can be read by the framework of the deep learning model.Another important step in the preprocessing of the images involves partitioning, whereby an image is divided into smaller patches of 16 x 16 pixels. The patches are then subjected to flattening, whereby the patches are converted into a vector format and fed into the Vision Transformer

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

model as a token. This step is important as it enables the model to read the image as a sequence rather than a single entity.

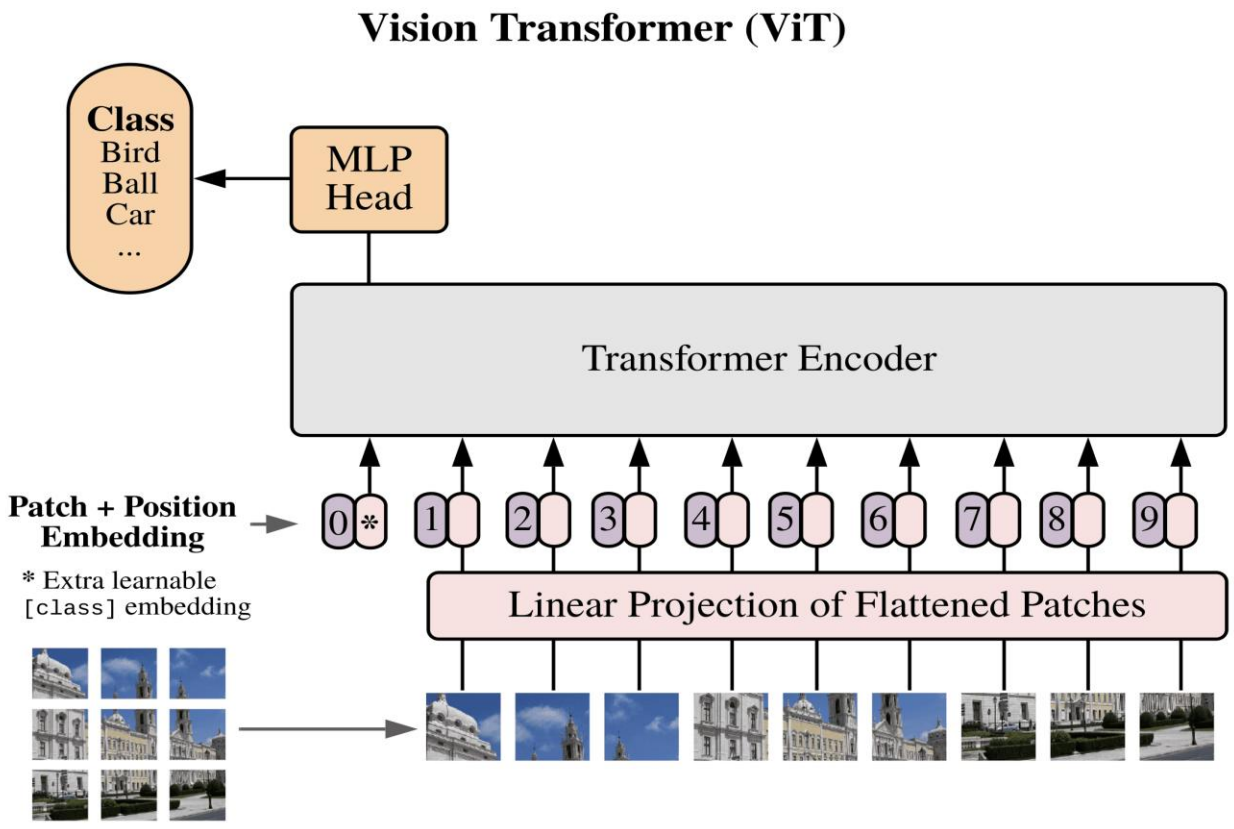
Vision Transformer Architecture

The main component of this proposed system is the Vision Transformer. Unlike CNNs, where an image is passed through various convolutional filters, the Vision Transformer works by passing an image through an attention mechanism based on a transformer.

The patches of an image are projected onto an embedding vector by passing them through a linear projection. To maintain the spatial positions of these patches, positional encoding is added to these embedding vectors.

The embedding vectors are then passed through multiple layers of the transformer encoder. Each of these encoder layers contains two main components: the multi-head self-attention mechanism and the feed-forward neural network. The self-attention mechanism allows the model to learn how to associate different patches of an image by assigning different weights to these patches.

The capacity of the Vision Transformer to look at the entire context of an image makes it very effective in detecting AI-generated images. This is due to the fact that CNNs only look at local features of an image and may not be able to detect inconsistencies present throughout an image.



Feature Extraction and Representation

The features produced by the transformer encoder are a mix of local and global features. The features produced by the transformer encoder represent the patterns present in the image. The patterns may include textures and edges. In the case of camera-captured images, the noise patterns and textures are present. In the case of AI-generated images, smooth textures and irregularities are present. The features produced by the transformer encoder are used for the differentiation of the patterns. The final feature vector produced by the Vision Transformer is used for further processing by the classification head.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

Classification and Decision Making

At the classification stage, the image is checked to see if it was taken by a camera or if it was generated by an AI. This is achieved by passing the features to the multilayer perceptron (MLP) classification head, which has fully connected layers.

The output of the classification head is the logits output, which represents the predictions of the model. The logits are converted to probabilities by the Softmax function:

$$P(y = i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

The class with the highest probability is considered as the final prediction. The probability value for this class is displayed as a confidence score, which represents the confidence level for the prediction.

Model Training and Optimization

The model is trained using the "CrossEntropyLoss" function. This function calculates the difference between the predicted probabilities and the actual labels. This loss function assists the model in learning to minimize the classification errors.

AdamW is the optimizer used for the model. This optimizer assists the model in converging better and avoids overfitting. During the training process, the classification head is fine-tuned, while the backbone is frozen. This makes the model more efficient and less complex.

The model is subjected to various epochs. This allows the model to learn the unique features between the real and AI-generated images.

This assists the user in understanding the rationale for the model's classification of the image as AI-generated or camera-captured. The model may be concentrating on areas where the texture, lighting, and patterns are unnatural.

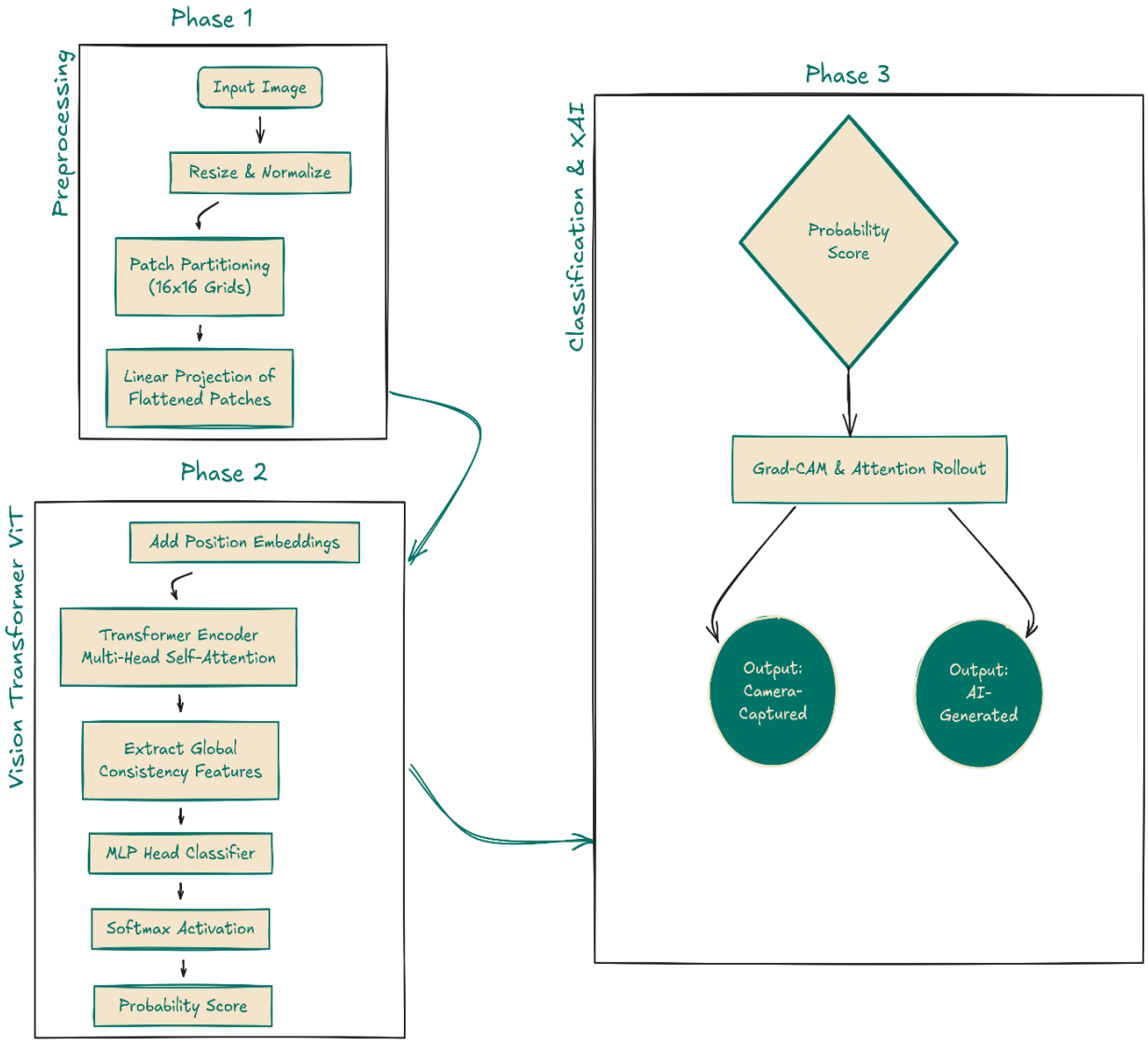
Workflow of the Proposed System

The overall workflow of the proposed system is divided into three main phases, as depicted in the above diagram. In Phase 1 (Preprocessing), the input image is provided by the user through the web interface, either by uploading an image or providing a link. The image is then resized and normalized for consistency, after which it is divided into smaller patches of size 16 x 16. These patches are then flattened and converted into vectors through linear projection.

In Phase 2 (Vision Transformer Processing), the patch representations are combined with positional embeddings, which help retain spatial information. The embeddings are then passed through a series of transformer encoder layers, which make use of the multi-head self-attention mechanism. This allows the model to capture global consistency features from the input image, thereby establishing relationships between different regions of the image. The extracted features are then passed through a multilayer perceptron (MLP) classifier, followed by a Softmax layer that generates a probability score for each class.

In Phase 3 (Classification and Explainability), the probability score is used to make a prediction, classifying the image as either camera-captured or AI-generated. In addition, the model also employs Explainable Artificial Intelligence (XAI) techniques that help identify and highlight the regions of the image that are responsible for the model’s decision.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026



MODEL EVALUATION

The performance of the proposed Vision Transformer-based image authentication system is evaluated by taking into account various metrics and analyzing the performance of the proposed system. In addition to the accuracy of the proposed system, various factors such as the confidence level and qualitative and comparative assessments of the proposed system are taken into account while evaluating the performance of the proposed system.

Evaluation Metrics

For the evaluation of the effectiveness of the proposed model, the evaluation metrics are considered. The accuracy of the model is considered the primary evaluation metric, which is the ratio of correctly classified images to the total number of images. However, the accuracy of the model is not the only factor that needs to be considered for the evaluation of the effectiveness of the system.

Apart from the accuracy of the model, the confidence levels generated by the Softmax function are also considered for the evaluation of the effectiveness of the proposed model. The confidence levels generated by the Softmax function provide a clear idea about the level of certainty that the model possesses about the classification of the images.

Moreover, the effectiveness of the proposed model is also evaluated qualitatively by considering the behavior of the model with different types of images. The behavior of the model with different types of images includes the behavior of the model with different textures and lighting conditions and the complexity of the background.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

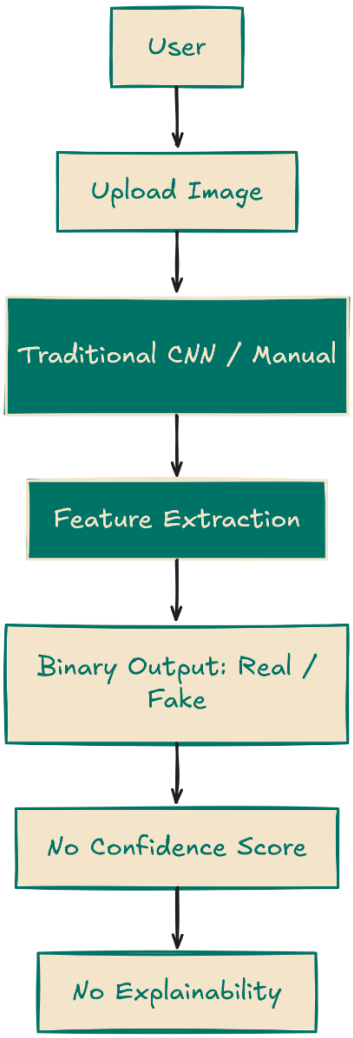
Model Performance Analysis

The proposed Vision Transformer model has shown promising results for the differentiation between the camera-captured and AI-generated images. The confidence score for the proposed model lies between 80% and 90%. This shows the proposed model has a high degree of confidence for the predictions. It has been observed that the proposed model gives a high confidence score for the images where differentiation is possible.The performance of the proposed model depends on the diversity of the dataset. A diverse dataset is required for the proposed model to perform effectively. The performance of the proposed model may be affected if the dataset is limited.

Comparative Analysis with Existing Models

In order to understand the efficiency of the proposed system, a comparative analysis is performed based on the traditional convolutional neural network-based models. CNN models like ResNet and EfficientNet are popular models for image classification tasks.

CNN models are quite efficient for extracting local features like edges and textures. However, it is also possible that these models might not be able to identify global inconsistencies, like those present in AI-generated images. Unlike CNN models, Vision Transformer is capable of analyzing the whole image using self-attention mechanisms.The whole image is analyzed because it is quite efficient in extracting features from the whole image, unlike CNN models.



Confidence Score Evaluation

The importance of the confidence score in the reliability of the predictions made by the model can never be emphasized enough. By examining the results provided by the Softmax function, a user can get a clear idea about the confidence level with which the model has made the predictions.For example, high confidence scores

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

indicate that the model has identified strong features to distinguish the objects, while low scores indicate a lack of clarity in this regard, which can be used in situations where important decisions are to be made, as in the case of digital forensics.The presence of the confidence score has made the system more user-friendly by providing more information.

Explainability Evaluation

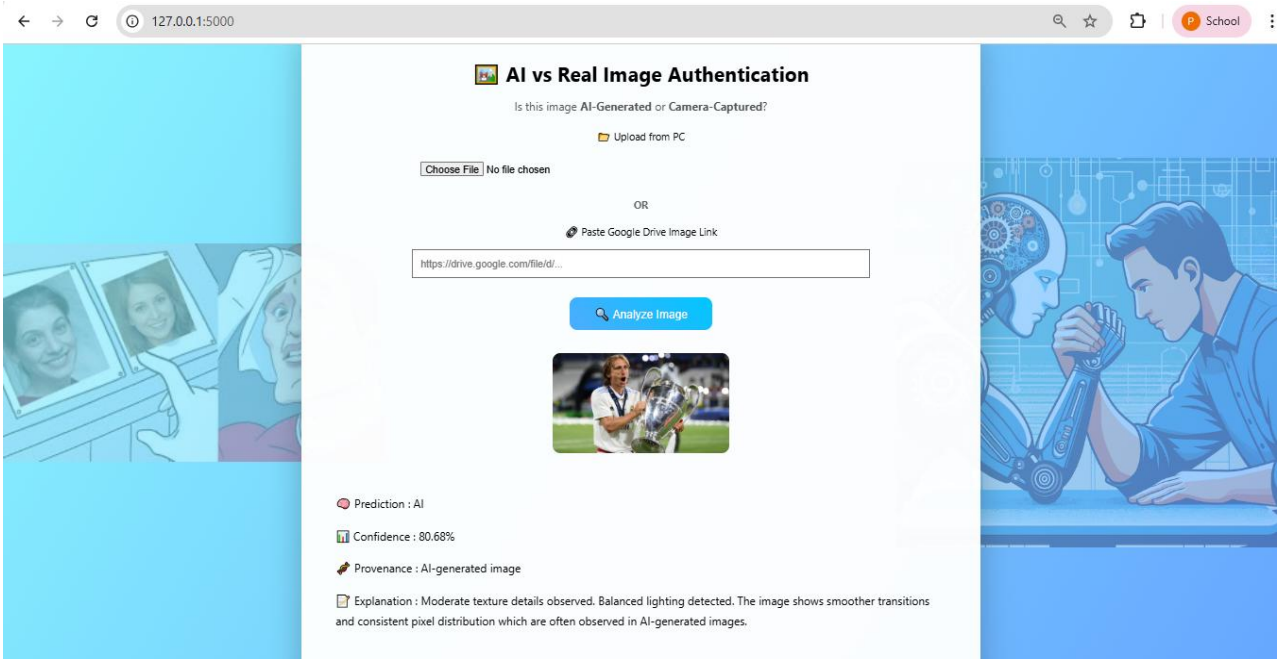
Explainability of the model can be determined based on the visualizations obtained through XAI. Based on the visualizations, a clear understanding of the regions of the image used by the model can be obtained. The heatmaps generated provide a clear understanding of the regions of the image that contribute significantly to the output generated by the model.

It has been observed that the model focuses more on the regions of the image that contain the background, textures, and regions that contain irregularities in the arrangement of the pixels. This is due to the presence of artifacts, which are characteristic of AI-generated images. The usage of explainability not only provides a clear understanding of the model but also validates the output. This is achieved by focusing on the regions of the image that are highlighted by the model.

RESULTS

The results obtained through the proposed system clearly indicate the effectiveness of the Vision Transformer in differentiating between camera-captured and AI-generated images. The capability of the model to comprehend the global relationships between the pixels enables it to recognize the inconsistencies that may not be clearly visible through the naked eye.One of the interesting observations is that the AI-generated images have smoother textures and do not contain the inherent noise that is present in the camera-captured images. The model is able to recognize this difference and use this as a differentiating factor. The inconsistencies in lighting and background regions are also recognized by the model. Another interesting aspect is the significance of explainability in building trust with the users. The XAI visualization gives a clear idea about the decision-making process of the model. The highlighting of the regions gives a clear idea about the authenticity of the image.

Another interesting aspect is the comparison with CNN-based models. It is evident that the Vision Transformer model has a clear advantage over the CNN-based models. The CNN-based models may not be able to recognize the patterns if they are present throughout the image. This limitation of the CNN-based models is addressed in the proposed model by considering the whole image context.The proposed system can be considered a reliable solution for image authentication.



ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

CONCLUSION

In this paper, we proposed a Vision Transformer-based image authentication system to differentiate between images captured by cameras and those generated by AI systems. The quick development of generative AI technology makes it difficult to verify the authenticity of images. This poses significant challenges for applications like digital forensics, media verification, and cybersecurity. The proposed image authentication system leverages the transformer architecture to learn global context and identify subtle differences between images. Unlike traditional CNN models that focus on local features, the Vision Transformer uses a self-attention mechanism to analyze the entire context of an image and detect discrepancies that might be missed by the human eye. By using a pre-trained model and fine-tuning the classification head, the system learns from the data efficiently and achieves high accuracy. The system provides not only classification results but also confidence levels in the form of Softmax probabilities. These confidence levels are important for assessing accuracy, allowing users to make informed decisions.

Additionally, the inclusion of explainability through XAI techniques enhances the system's transparency by highlighting the areas of images considered for the final prediction, thereby increasing trust in the system. Deploying the system as a web application boosts its usability and practicality. It allows users to upload images or provide image links, making it easy and convenient to use.Experimental results show the model's accuracy and effectiveness as it can successfully differentiate between camera-captured and AI-generated images. The capacity to capture global dependencies makes the Vision Transformer well-suited for this task. However, the system does have limitations. Its performance relies on the diversity and quality of the dataset. The system might also encounter challenges with images produced by various or unknown AI models. Furthermore, image quality can impact classification results.

Overall, the proposed image authentication system based on the Vision Transformer model is a strong and effective solution for authenticating images and detecting AI-generated content. The use of deep learning, confidence scoring, and explainability makes the system a dependable tool for tackling image authentication challenges in the digital age. There is potential for improvement and adaptation to different and complex scenarios.

REFERENCES

1. D. Karageorgiou et al., “Any-resolution AI-generated image detection by spectral learning,” arXiv preprint, 2025.
2. Y. Zhang et al., “Unmasking AI-created visual content: A review of generated images and deepfake detection technologies,” Journal of King Saud University - Computer and Information Sciences, 2025.
3. S. Tang et al., “Towards extensible detection of AI-generated images via content-agnostic adapter-based category-aware incremental learning,” IEEE Transactions on Information Forensics and Security, 2025.
4. J. J. Bird and A. Lotfi, “CIFAKE: Image classification and explainable identification of AI-generated synthetic images,” IEEE Access, vol. 12, pp. 156379–156393, 2024.
5. N. Carlini et al., “Extracting training data from diffusion models,” in Proc. USENIX Security Symposium, 2023.
6. R. Rombach et al., “High-resolution image synthesis with latent diffusion models,” in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10684–10695.
7. A. Ramesh et al., “Hierarchical text-conditional image generation with CLIP latents,” OpenAI, 2022.
8. Z. Liu et al., “Swin transformer: Hierarchical vision transformer using shifted windows,” in Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV), 2021, pp. 10012–10022.
9. X. Zhang et al., “Detecting AI-synthesized images using frequency domain analysis,” IEEE Access, vol. 9, pp. 124256–124268, 2021.
10. B. Tondi et al., “Detection of GAN-generated fake images over social networks,” in Proc. IEEE ICASSP, 2021, pp. 3010–3014.
11. J. Frank et al., “Leveraging frequency analysis for deep fake image recognition,” in Proc. ICML Workshops, 2020.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

12. S. Wang et al., “CNN-generated images are surprisingly easy to spot… for now,” in Proc. IEEE/CVF CVPR, 2020, pp. 8695–8704.
13. H. Dang et al., “On the detection of digital face manipulation,” in Proc. IEEE/CVF CVPR Workshops, 2020.
14. S. Verdoliva, “Media forensics and deepfakes: An overview,” IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, 2020.
15. P. Korshunov and S. Marcel, “Deepfake detection: A survey,” IEEE Signal Processing Magazine, vol. 36, no. 1, pp. 85–95, 2019.
16. Y. Nirkin et al., “Inverting face embeddings with convolutional neural networks,” in Proc. IEEE/CVF CVPR, 2019, pp. 1070–1078.
17. M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. ICML, 2019, pp. 6105–6114.
18. H. Nguyen et al., “Capsule-forensics: Using capsule networks to detect forged images and videos,” in Proc. IEEE ICIP, 2019, pp. 2307–2311.
19. M. Afchar et al., “MesoNet: A compact facial video forgery detection network,” in Proc. IEEE WIFS, 2018, pp. 1–7.
20. F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proc. IEEE CVPR, 2017, pp. 1251–1258.