Artificial Intelligence Generated Image Detection

V ROSHAN KUMAR*[1], SYPU SUSMITHA[2], VARI NAGA PUJITHA [3], PASAM GOPI CHAND[4], YARRAREDDI LAKSHMI SURENDRA [5]

[1]*Assistant Professor, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP,India*
[2]*Student, Department of CSE(AIML),Bapatla Engineering College, Bapatla 522101,AP, India*
[3]*Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India*
[4]*Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India*
[5]*Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India*

Abstract—This project introduces an intelligent system for detecting whether an image is real or AI-generated by combining deep learning with image feature analysis. The approach is based on a Convolutional Neural Network (CNN) model that predicts image authenticity, supported by additional handcrafted features such as noise score, edge density, brightness, contrast, and color variance. These features help capture subtle visual differences between natural images and synthetic outputs.

A key contribution of this system is the integration of explainable AI, where the model not only provides a classification result but also generates clear, human-readable reasons for its decision. This improves transparency and user trust in the system. The framework also includes an automated report generation module that produces detailed PDF reports containing the prediction, confidence level, feature analysis, and explanation.

The implementation utilizes TensorFlow for model inference, OpenCV for feature extraction, and Gradio to create an interactive interface supporting batch image uploads. This combined approach enhances detection accuracy while ensuring usability and interpretability. The system can be applied in digital forensics, content verification, and detection of AI-generated media, contributing to more reliable and trustworthy image analysis solutions

Keywords:Image detection, Convolutional Neural Networks (CNNs), Local features, Share weights, Pooling mechanisms, Innovative approach, ShortCut3-ResNet, Residual Network (ResNet), Feature extraction, PDF Report Generation.

II. INTRODUCTION

With the rapid advancement of artificial intelligence, the generation of highly realistic synthetic images has become increasingly common. Technologies such as deep learning and generative models can now produce images that are often indistinguishable from real photographs. While these innovations have opened new possibilities in fields like entertainment, design, and media, they also raise serious concerns regarding misinformation, digital manipulation, and authenticity verification. As a result, the need for reliable systems to distinguish between real and AI-generated images has become critical.

This project addresses this challenge by developing an AI-based image authenticity detection system. The proposed approach combines the power of Convolutional Neural Networks (CNNs) with traditional image processing techniques to improve detection accuracy. In addition to classification, the system extracts meaningful visual features such as noise, edges, brightness, contrast, and color variation to support the decision-making process.

A significant aspect of this work is the focus on explainability, where the system provides clear reasons for its predictions, enhancing user trust and understanding. Furthermore, the integration of an automated report generation feature ensures that results are well-documented and easy to interpret. This makes the system useful for applications in digital forensics, content verification, and media analysis.

III. LITERATURE SURVEY

Literature Review: AI-Generated Image Detection Using Neural Networks

The rapid development of artificial intelligence (AI) in recent years have led to the expansion of AI-generated images, raising concerns about them. can be misused. affects accuracy and stability. To solve this problem, researchers have proposed various methods that use neural networks to analyze images created by artificial intelligence. This research paper presents the main findings and methods of relevant research in this field.

1. Tian (2020) [1]: Tian introduced a new convolutional neural network (CNN) algorithm and enhanced it with a recurrent neural network (RNN) to improve image learning. This combination aims to increase convergence speed and recognition accuracy by simultaneously using the advantages of CNN and RNN architectures!

2. Karandika et al. (2020) [2]:Karadika et al. A deep image detection model using CNN together with transformation learning is proposed. By analyzing video frames, the model can identify discrepancies between facial features and the compression value shown during the creation of the deep video, thus achieving a powerful result in analyzing the content?

3. Yadav and Balabantaray (2022) [3]: Yadav and Balabantaray investigate deep learning techniques using CNN architecture, with a special focus on the EfficientNetB4 variant with tracking. Their research demonstrates the effectiveness of integrated tracking techniques to improve the accuracy of depth detection, especially in identifying areas of interest.

4. Jiang et al. (2023) [4]: Jiang et al. Focusing on the analysis of animation in CNNs for image classification. Their proposed method, which includes regular vector conditions, histogram analysis, and residual image analysis, increases the robustness of the CNN model against attacks and thus actually improves its detection?

5. Pertigkiozoglou and Maragos (2018) [5]: Pertigkiozoglou and Maragos review various methods for analyzing deep neural networks created by artificial neural networks (GANs). They said that in the face of increasing problems caused by fake images, it is necessary to create a suitable search tool that will protect the integrity of the content.

6. Levi et al. (2023) [6]: Levy et al. A two-stream system designed to analyze smart vision-generated images is proposed, with special focus on the text-to-image conversion process. Their method combines residual and content streaming with multi-head integration and better processing techniques in AI-generated content analysis?

7. Xi et al. (2023) [7]: Xi et al. Introduce a framework for analyzing images generated by GANs using CNN and transfer learning. Their method includes Class Activation Maps (CAM) and contributes to the identification of real images by achieving accuracy in distinguishing between real and synthetic images!!

8. Baraheem and Nguyen (2023) [8]: Baraheem and Nguyen examine the ability of AI to capture AI-generated images and highlight the importance of deep connections in image recognition. They discuss the challenges of analyzing AI-generated content due to its literal nature:

9. Li and Lv (2021) [9]: Li and Lv present a comprehensive review of CNN-based image classification algorithms, detailing their modifications and applications in various fields. Their review highlights the progress CNN architectures have made in image classification!!

10. Chen et al. (2021) [10]: Chen et al. A specially designed CNN-based target detection model for high-speed scene image recognition is proposed. Their model combines preprocessing and normal L2 processing to improve recognition accuracy and generalization capabilities beyond the performance of simple CNN algorithms?

11. Wang et al. (2022) [11]: Wang et al. An improved CNN-based image recognition system based on the integration of multiple CNN groups and batch normalization is proposed. Their research contributes to improving knowledge about performance and general capabilities in scene image recognition operations!!

current research on AI-based image detection using neural networks covers a variety of approaches, including CNN architectures, tracking systems, transfer learning, and prioritization procedures. Together, these studies support the development of robust detection methods to reduce the risks associated with AI-generated content and ensure the integrity and security of content in the digital space.

IV. EXISTING SYSTEM
Traditional Convolutional Neural Network (CNN)-based image recognition systems rely on large annotated datasets to learn features from images. They are widely used in tasks like object detection and classification but face several limitations.

These systems require high annotation effort and cost, and often suffer from limited data diversity, which affects performance in real-world conditions. They also have difficulty generalizing to new environments and may produce biased results due to imbalanced training data.

Additionally, CNN models demand high computational resource**s**, making them expensive and less efficient. Overall, these challenges limit the practical effectiveness of traditional CNN-based systems.

V. PROPOSED SYSTEM
To overcome the limitations of traditional CNN-based image recognition systems, the proposed system introduces a hybrid approach that combines deep learning with image feature analysis for improved accuracy and reliability. The system utilizes a Convolutional Neural Network (CNN) model to classify images as real or AI-generated, while also extracting important visual features such as noise score, edge density, brightness, contrast, and color variance using image processing techniques.

A key enhancement of the proposed system is the integration of explainable AI, where the model provides clear, human-readable reasons for its predictions. This improves transparency and helps users understand the decision-making process. Additionally, the system supports batch image processing, allowing multiple images to be analyzed efficiently at once.

Another important feature is the automated PDF report generation, which includes the classification result, confidence score, feature analysis, and explanation. This makes the output well-structured and easy to interpret.

The system is implemented using TensorFlow for model prediction, OpenCV for feature extraction, and Gradio for an interactive user interface. Overall, the proposed approach enhances accuracy, interpretability, and usability, making it suitable for applications such as digital forensics and AI-generated image detection.

VI. SYSTEM ARCHITECTURE
1. Input Module
• The system accepts image files as input from the user.
• Users can upload single or multiple images through the interface.
• The uploaded images are sent to the system for processing.
• Clear and high-quality images are important for accurate results.

2. Pre-Processing Module
• The input images are resized to a fixed dimension.
• Images are converted into a suitable format for processing.
• Noise and unwanted variations are minimized.
• This step improves the overall performance of the system.

3. Feature Extraction Module
• The system extracts features such as noise, edges,

brightness, contrast, and color variance.
• These features help in distinguishing real and AI-generated images.
• Each feature is calculated using image processing techniques.
• This improves the accuracy and interpretability of the system.

4. CNN Prediction Module
• The processed image is passed to the CNN model.
• The model analyzes image patterns and textures.
• It predicts whether the image is real or AI-generated.
• A confidence score is also generated for the prediction.

5. Decision Explanation Module
• The system generates reasons based on extracted features.
• It explains why the image is classified as real or AI-generated.
• The explanations are simple and easy to understand.
• This improves transparency and user trust.

6. Report Generation Module
• A PDF report is generated automatically for each image.
• The report includes prediction, confidence, features, and reasons.
• It provides a structured summary of the analysis.
• This helps in easy documentation and review.

7. Batch Processing Module
• The system supports processing multiple images at once.
• Each image is analyzed separately.
• Reports are generated for all images.
• Multiple reports can be combined into a ZIP file.

8. Output Module
• The final result is displayed as REAL or AI-GENERATED.
• Confidence score is shown along with the result.
• Users can download the generated report(s).
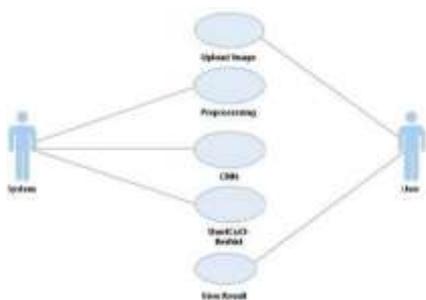• This completes the interaction between the user and the system.
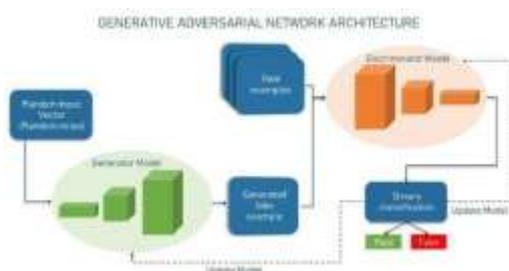

Figure1: Use Case Diagram

Figure2: System Architecture

VII. METHODOLOGY

1. Environment
• The project is developed using Python programming language and TensorFlow framework.
• It supports GPU acceleration for faster computation during training and prediction.
• Libraries like OpenCV, Gradio, and FPDF are used for image processing, interface design, and report generation.
• This environment ensures efficient model development and execution.

2. Data Processing and Pre-Processing
• Image datasets are loaded using TensorFlow functions.
• Images are resized to a uniform size of 32×32 pixels.
• Pixel values are normalized to the range [0, 1].
• This step ensures consistency and improves model performance.

3. Dataset Preparation
• The dataset is divided into training, validation, and testing sets.
• Typically, 70% is used for training, 20% for validation, and 10% for testing.
• TensorFlow Dataset API is used for efficient data handling and shuffling.
• This helps in better generalization of the model.

4. Neural Network Architecture
• A CNN model is built using TensorFlow's Sequential API.
• It includes convolutional and max-pooling layers for feature extraction.
• Dense layers with ReLU activation are used for learning patterns.
• A sigmoid activation function is used for binary classification.

5. Model Training
• The model is trained using the Adam optimizer.
• Binary cross-entropy is used as the loss function.
• Training is performed over multiple epochs.
• TensorBoard is used to monitor training progress and performance.

6. Model Evaluation
• The trained model is tested using unseen data.
• Performance metrics like accuracy, precision, recall, and F1-score are evaluated.
• This helps in measuring the effectiveness of the model.

7. Model Saving and Loading
• The trained model is saved using TensorFlow's model.save() method.
• It is stored in HDF5 format for future use.
• The model can be reloaded using load_model() for predictions.

8. Prediction on New Images
• Users can upload new images for classification.
• Images are preprocessed before prediction.

• The model classifies images as REAL or AI-GENERATED.

• Predictions are based on a confidence threshold of 0.5.

9. Report Generation Module

• The system automatically generates a PDF report for each analyzed image.

• The report includes the prediction result, confidence score, and extracted features.

• It also provides clear reasons explaining the model's decision.

• Multiple reports can be combined into a ZIP file for batch processing.

• This feature helps in easy documentation and understanding of results.

VIII. IMPLEMENTATION

The system is implemented using Python with the support of TensorFlow, OpenCV, Gradio, and FPDF libraries. Initially, the trained CNN model is loaded, and the user provides input images through an interactive interface that supports both single and multiple uploads. The input images are preprocessed by resizing them to a fixed size of 32×32 pixels and normalizing pixel values to ensure consistency and improve prediction accuracy.

Once preprocessing is complete, the system extracts important visual features such as noise, edge density, brightness, contrast, and color variance using image processing techniques. These features help in understanding the image characteristics. The processed image is then passed to the CNN model, which classifies it as either REAL or AI-GENERATED and provides a confidence score based on the prediction probability.

To enhance transparency, the system generates human-readable explanations based on the extracted features, describing why a particular decision was made. Additionally, a detailed PDF report is automatically created for each image, including the prediction result, confidence score, feature analysis, and explanation. The system also supports batch processing, allowing multiple images to be analyzed at once, with reports available for download individually or as a combined ZIP file.
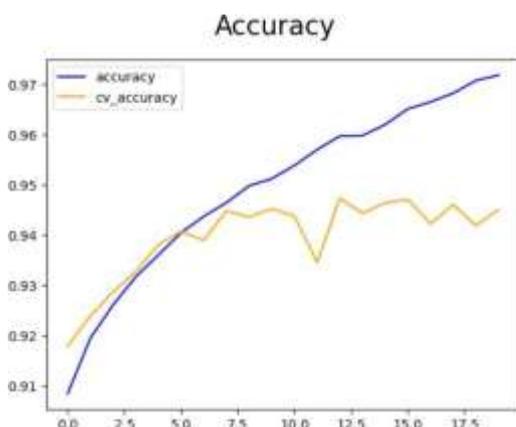
IX. RESULTS

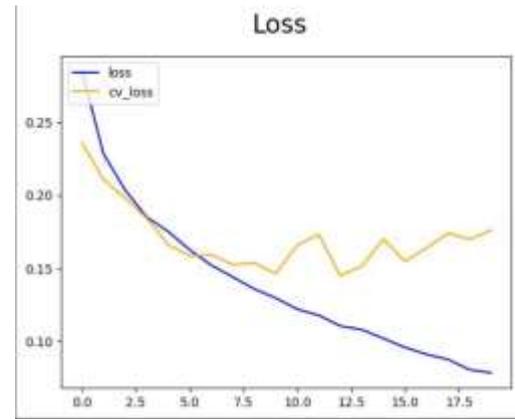**Figure 3: This figure shows the accuracy of our model**

Figure 4: This shows the loss of our model
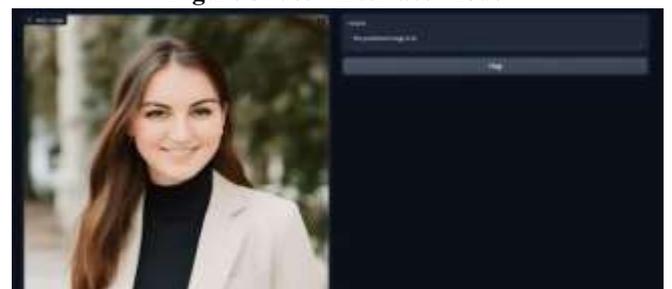
Figure 5: user Interface Model

Figure 6: Predicting AI Image

CONCLUSION

In this project, an AI-based image authenticity detection system was successfully developed and implemented. The system uses a CNN model along with image processing techniques to classify images as real or AI-generated. It analyzes features such as noise, edges, brightness, contrast, and color variation to support its predictions. Based on these features, the system also provides clear reasons for its decisions, improving transparency.

The system works efficiently for both single and multiple image inputs and generates detailed PDF reports with results and analysis. Although the accuracy depends on the trained model and image quality, it provides a simple and effective solution for identifying AI-generated images.

REFERENCES

[1] Tian, Y. (2020). Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. IEEE Access, 8, 115235-115243. Published June 30, 2020. DOI: 10.1109/ACCESS.2020.3006097.

[2] Karandikar, A., Deshpande, V., Singh, S., Nagbhidkar, S., & Agrawal, S. (2020). Deepfake Video Detection Using Convolutional Neural Network. , International Journal of

Advanced Trends in Computer Science and Engineering, 9(2), ISSN 2278-3091 . Published March-April 2020.

[3] Yadav, S. S., & Balabantaray, B. K. (2022). Analysis of image forgery detection using convolutional neural network. International Journal of Applied Systemic Studies, Vol.9, 240 . Published January 2022. DOI: 10.1504/IJASS.2022.124085

[4] Jiang, W., Guo, Z., & Liang, R. (2023). Using ensemble models to detect deepfake images of human faces. In Proceedings of the MICML 2023 (pp. 32). December 15–17, 2023, Chengdu, China. Published [MICML 2023, December 15–17, 2023]. https://doi.org/10.1145/3638264.3638272

[5] Pertigkiozoglou, S., & Maragos, P. (8 Dec 2018). Detecting Adversarial Examples in Convolutional Neural Networks. arXiv.org, Published [8 Dec 2018]. arXiv:1812.03303v1.

[6] Revi K., R., Vidya K. R., & Wilscy, M. Detection of Deepfake Images Created Using Generative Adversarial Networks –A Review. J. 1Saintgits College of Engineering, Kottayam, India, remyarevik123@gmail.com

[7] Xi, Z., Huang, W., Wei, K., Luo, W., & Zheng, P. (9 Nov 2023). AI-Generated Image Detection using a Cross-Attention Enhanced Dual-Stream Network. In Proceedings of the 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) . Published [9 Nov 2023].

[8] Baraheem, S. S., & Nguyen, T. V. (: 28 September 2023). AI vs. AI: Can AI Detect AI-Generated Images? J. Imaging 2023, 9, 199, 18 of 18. Published [: 28 September 2023]. https://doi.org/10.3390/jimaging9100199

[9] Li, X., & Lv, X. (2021). Research on Image Recognition Method of Convolutional Neural Network with Improved Computer Technology. Journal of Physics: Conference Series, 1744, 042023.doi:10.1088/1742-6596/1744/4/042023

[10] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sens., 13(24), 4712. DOI: 10.3390/rs13224712.

[11] Wang, P., Qiao, J., & Liu, N. (2022). An Improved Convolutional Neural Network-Based Scene Image Recognition Method. Computational Intelligence and Neuroscience, Volume 2022, Article ID 3464984, 10 pages. DOI: 10.1155/2022/3464984.