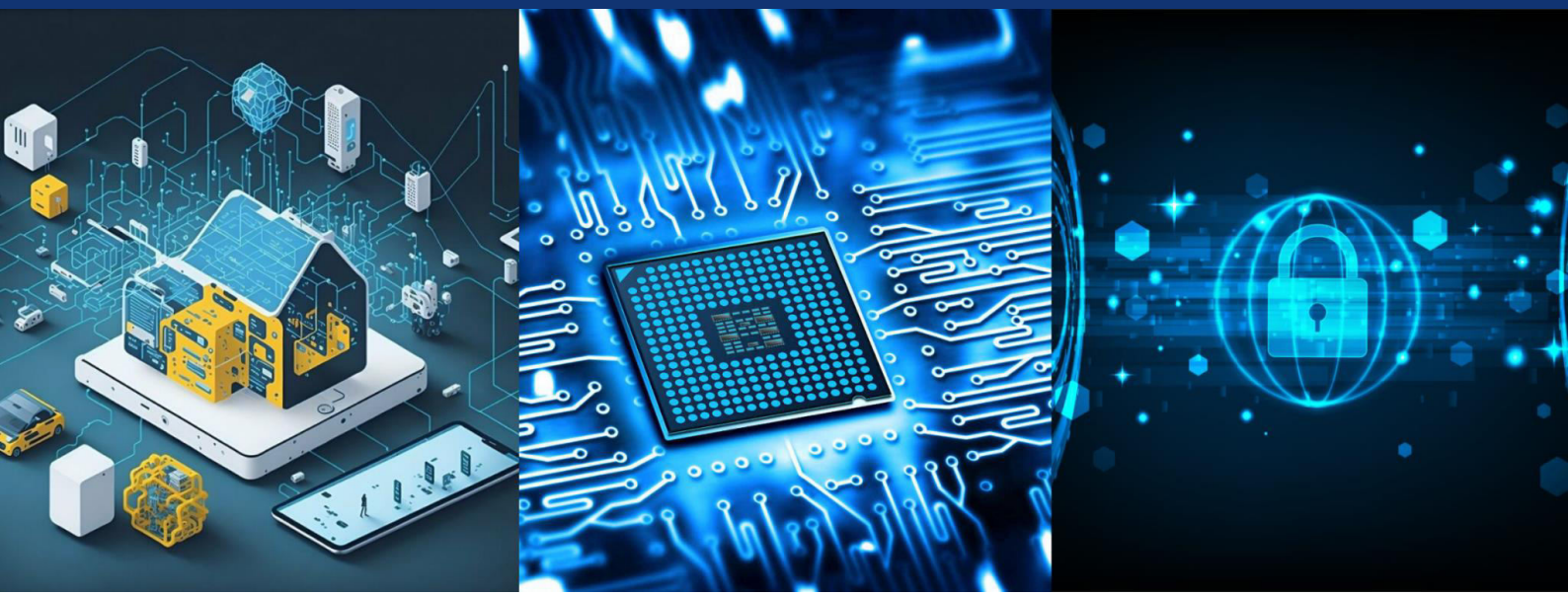
IJIRCCE

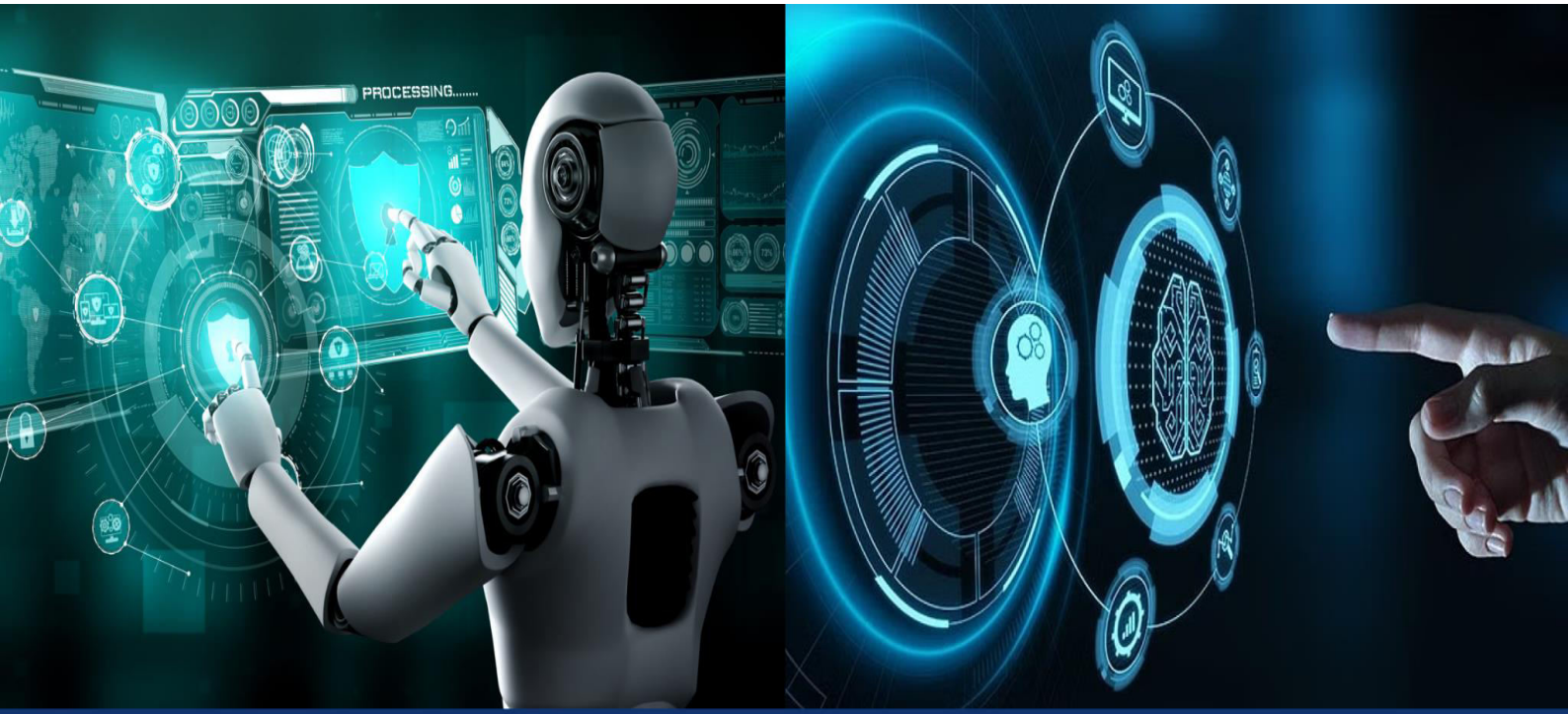
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304222

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Vision Guard Precise Vision for the Digital Age

V.Jahnavi Devi, V.Lohitha, V. Siva Durga, V. Sai Vignesh, V.Ethirajulu

UG Student, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

UG Student, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

UG Student, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

UG Student, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Abstract: With the rise of AI-generated images, distinguishing between real and synthetic visuals has become a significant challenge. This study presents a deep learning-based approach to classify AI-generated and real images using a convolutional neural network (CNN) model. The dataset consists of AI-generated and real artwork, pre-processed and augmented using transformations like resizing and normalization. A custom CNN architecture is implemented, featuring dual feature extraction paths for detail and pattern recognition, followed by a combined classification network. The model is trained using the Binary Cross-Entropy loss function and optimized with Adam.Experimental results indicate that the model effectively learns discriminative features between real and AI-generated images. Training and validation accuracies are monitored to assess performance, and the best model is saved for future inference. This research contributes to AI-driven media authentication, enabling better detection of artificially generated content in digital media and forensic applications.This study proposes a deep learning-based framework for classifying images as either real or AI-generated. The system utilizes convolutional neural networks (CNNs) and transformer-based architectures to detect subtle differences between authentic and synthetic images. In addition to visual pattern analysis, metadata examination and statistical modeling are incorporated to enhance detection accuracy. The goal of this research is to improve digital forensics, ensuring the credibility of online media and mitigating the risks associated with AI-generated visual content.

I. INTRODUCTION

Technological advances have made it possible to create images of such high quality that even humans find it difficult to distinguish between real photos and images created using artificial intelligence (AI) technology. Over time, the capabilities of such generative technologies can produce excellent images with fewer and fewer visual flaws. This generative technology can produce high-quality images with customized themes, where the way this technology works is by creating a synthesized image of a writing [15]. The quality of AI-generated images can be proven when images produced by generative models can compete with humans and win art competitions. The ability to distinguish between the original image and the image generated by generative models such as GAN (Generative Adversarial Network) or stable diffusion is important. One reason is that such generative models can generate synthetic images of a person committing a crime.

In addition, it can also provide false evidence as alibi for someone who is somewhere else. False information is a significant modern problem, and high-quality images generated by generative models can be used to manipulate public opinion. Another problem is in the field of digital security, where human faces created by generative models can be used to bypass face verification to gain access to digital systems. In addition to faking face verification, generative models can also create synthetic signatures that can defeat signature verification systems. The extensive variety of state-of-the-art text-to-image generative models, each possessing unique characteristics, poses a significant challenge in developing a model capable of classifying AI-generated images. A model excelling at classifying AI-generated images from a specific generator tends to perform poorly when tested on images from different generators. This outcome reinforces the theory that each generative model has its own set of characteristics, and these differences are substantial, as evidenced by testing results. From a human perspective, the key determinant in classifying AI-generated images lies in the unusual structure of images and colorization that appears overly realistic.

In recent decades, the widespread adoption of social networks has deeply engaged people worldwide. Microblogging platforms have enabled individuals to share their thoughts in real-time on a global scale, providing researchers with

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304222

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

valuable insights into online social dynamics during various events. This freedom of expression has facilitated the exchange of diverse thoughts, emotions, and knowledge among users. However, the digital environment isn't always secure, often becoming a platform for the dissemination of harmful content. Hate speech, a prevalent form of online expression, frequently manifests as prejudice, aggression, racism and other forms of verbal abuse.

The platform is developed using HTML, CSS, and Python (gradio), providing a user-friendly interface that allows individuals to easily upload and receive immediate result of the image provided.

OVERVIEW OF THE PROBLEM

The rise of AI-generated images, powered by models like GANs and diffusion networks, has created a major challenge in verifying the authenticity of digital visuals. These synthetic images are often indistinguishable from real ones, making it difficult to detect misinformation, protect against identity fraud, and uphold trust in media. Traditional detection methods struggle to keep up with the sophistication of AI outputs, especially when metadata is altered or missing. As such images spread rapidly through digital platforms, there is an urgent need for robust, AI-driven detection systems to preserve truth, safety, and transparency in the digital world.

II. LITERATURE SURVEY

1. Detection of Al-Generated Images From Various Generators Using Grated Expert Convolutional Network Yuvang Wang Yizhi Hao Amando Xu Cong.2024,IEEE Transactions on Artificial Intelligence
2. Deepfake Detection Using Deep Learning Methods: A Systematic Review,Muhammad Hussain,Muhammad Adeel,Muhammad Imran,2023,WIREs Data Mining and Knowledge Discovery
3. Performance. Comparison and Visualization of Al-Generated-Image Detection Methods. Y. Wang.Y.Hao A.Cong.2024. JEEE Transactions on Artificial Intelligence.
4. Deeptake Detection and Classification Using Error-Level Analysis and Deep Learnings K. Verma, S. K. Sahay, 2023. Scientific Reports
5. Al-GeneratedImage Detection Using a CrossAttentionEnhancedConvolutionalNeural NetworkPrevious 30 DaysIrrigation System SummaryWalmart Certification SummarAsk anythingCertification for ResumeY. Wang, Y.Hao, A. Cong,2024,IEEE Transactions on Artificial Intelligence.

OBJECTIVE

To develop a Convolutional Neural Network (CNN)-based model capable of accurately classifying images as either AI-generated or real. With the rapid advancement of AI-generated media, distinguishing between authentic and synthetic images has become increasingly crucial in combating misinformation, ensuring digital content integrity, and supporting forensic analysis. This project aims to train a deep learning model using a diverse dataset of real-world images and AI-generated images from tools like DALL·E, Stable Diffusion, and GAN-based models. By optimizing the model through data augmentation, hyperparameter tuning, and performance evaluation using metrics such as accuracy, precision, and recall, we seek to improve the robustness and reliability of the classification system.

Accuracy and Efficiency

To ensure high accuracy in Ai or Real predictions by using advanced machine learning algorithms, improving prediction capabilities with each data input and enhancing the system's overall efficiency.

Real-Time Predictions and Feedback

To offer instant, real-time predictions and, empowering individuals to take prompt action regarding the information they are consuming.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304222

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. EXISTING SYSTEM

Existing systems for detecting AI-generated images primarily rely on traditional image forensics, statistical analysis, and early deep learning models such as Convolutional Neural Networks (CNNs). These approaches analyze texture, lighting inconsistencies, or compression artifacts to spot manipulations. Some tools also use metadata verification to detect tampered or synthetic images by checking inconsistencies in EXIF data. However, many of these systems struggle to detect images created by advanced models like StyleGAN3 or diffusion-based generators, which produce highly realistic results with minimal visible flaws. Additionally, if metadata is stripped or altered, these detection methods become less reliable, highlighting the need for more robust and adaptive techniques.

Current detection systems employ a combination of machine learning classifiers, image analysis tools, and forensic algorithms to identify AI-generated content. Tools like Microsoft's Video Authenticator and media forensics platforms can analyze visual artifacts, inconsistencies in pixel patterns, and signs of manipulation. While these systems are effective against older generative models, they often fall short when facing newer AI techniques that generate near-perfect visuals. Moreover, most existing systems are not fully automated, lack real-time processing capabilities, and are limited in scalability, making them less suitable for handling the vast volume of digital content shared daily across the internet.

IV. PROPOSED SYSTEM

The proposed system introduces an advanced AI vs. real image detection framework that combines Convolutional Neural Networks (CNNs), transformer-based architectures, and metadata analysis to identify synthetic images with high accuracy. The system is trained on a large and diverse dataset containing real and AI-generated images from various sources, including StyleGAN, Big GAN, and diffusion models. By leveraging CNNs, the system can detect fine-grained texture inconsistencies, while transformers capture global context and spatial relationships within the image.

In addition to visual analysis, the system also examines image metadata (EXIF data) to identify unusual patterns such as missing camera details or mismatched timestamps, which often indicate synthetic origins. The model uses a multi-layer classification approach, integrating both pixel-level features and metadata insights to improve prediction reliability. The goal is to develop a robust, scalable, and explainable detection system that can generalize across different types of AI-generated content and adapt to emerging generative models. This hybrid approach enhances detection accuracy and supports real-time deployment in applications like media verification, cybersecurity, and digital forensics.

- **C**ustom CNN Model Dual-path feature extraction network for detailed texture analysis.
- Fully connected layers for final classification.
- Training & Optimization Dataset preprocessing (resizing, normalization, augmentation).
- Binary Cross-Entropy loss for effective learning.
- Adam optimizer for stable training.
- Performance Evaluation Metrics: Accuracy, Confusion Matrix, ROC Curve.

User-Friendly Interface

A web-based interface will allow users to easily input their symptoms and medical history.

The platform will provide real-time predictions and recommendations to users based on their inputs.

Data Preprocessing

Data preprocessing will be done to clean the input data (such as missing values, outliers, and data normalization) to ensure better model performance.

Feature selection will be employed to focus on the most relevant features to predict Ai or Real image.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304222

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

SYSTEM ARCHITECTURE

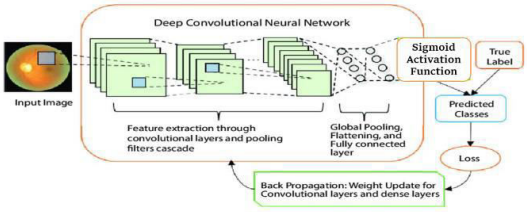


Figure1. System Architecture

V. METHODOLOGY

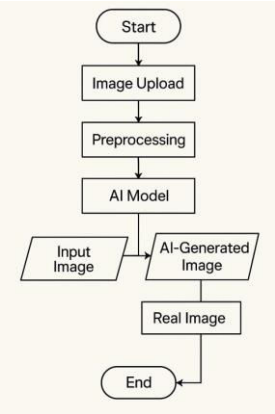


Figure2. Methodology

VI. EXPLANATION OF THE DESIGN METHODOLOGY IN THE GIVEN FLOWCHART

1. Start

This is the initiation point of the system where the process begins. The user triggers the system to analyse an image to determine its authenticity—whether it's AI-generated or real.

2. Image Upload

The user uploads an image through the application interface. Accepted formats may include JPEG, PNG, or WebP. The image is temporarily stored and forwarded for preprocessing.

3. Image Preprocessing

Before the image is fed into the AI model, it undergoes preprocessing to standardize its format. This includes:

- Normalization of pixel values
- Colour format conversion if required (e.g., RGB to grayscale)
- Noise removal and enhancement to improve analysis

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304222

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4. Feature Extraction

The processed image is then passed into a Convolutional Neural Network (CNN) or equivalent deep learning architecture. Here, the model extracts key visual features such as:

- Texture inconsistencies
- Compression artifacts
- Unnatural edge blending
- Irregular lighting or facial symmetry (for face-based detection)

5. Classification using Trained Model

The extracted features are then evaluated using a pre-trained binary classification model, which categorizes the image as either:

AI-Generated (Fake)

Real (Authentic)

Popular models used could include:

EfficientNet

ResNet

Xception (used in deepfake detection tasks)

Custom CNN built specifically for GAN artifact detection

6. Output Prediction

The system displays the final prediction to the user, such as:

"This image is 92% likely to be AI-generated."

"Authentic image detected with 84% confidence."

7. End

The process is completed, and the user is offered options such as:

- Uploading another image
- Downloading the report
- Reporting misinformation (if used in a journalistic or forensic setting)

VII. EXPERIMENTAL RESULTS

we developed a web-based application that can accurately classify whether an uploaded image is **real or AI-generated**. The system was implemented using **Python** with frameworks such as **PyTorch**, **OpenCV**, and **Gradio** for the user interface. The experimental setup was carried out both on **Google Colab** and a local machine equipped with an local machine macbook air m2, which significantly accelerated inference time. The software requirements included Python 3.8+, along with essential libraries like torch, torchvision, opencv-python, numpy, and gradio. During testing, users were allowed to upload images through a simple web interface. Once uploaded, the image was pre- processed (resized, normalized) and passed to a trained CNN-based model. The model then predicted whether the image was real or AI-generated and displayed the result with a confidence percentage.

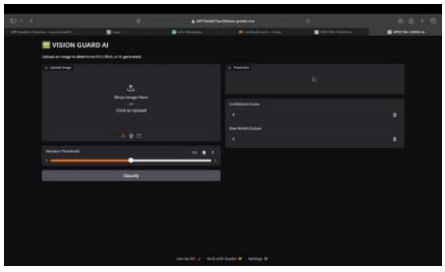


Figure 3: Vision guard interface

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304222

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

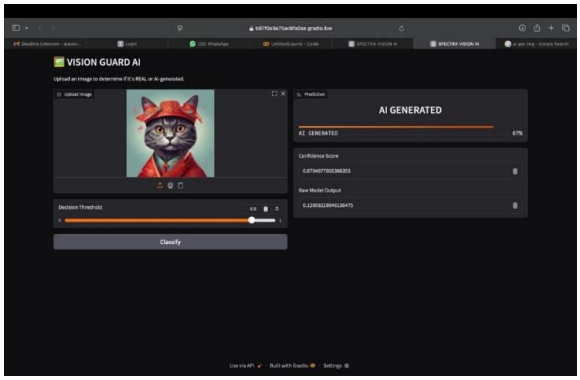


Figure 4: Result

The predicted output is displayed at the right side of the screen with a confidence score that supports the output such as "REAL" or "AI-GENERATED". Based on the image features it achieved.

VIII. CONCLUSION

In today's digital era, the rise of deepfake technology has made it increasingly difficult to distinguish between real and manipulated images. These AI-generated deepfakes pose serious threats in areas such as media, security, and social trust, making robust detection methods essential.To address this challenge, we developed **DeepDetect: A Robust Model for Deepfake Image Classification**. Our system leverages a **Vision Transformer (ViT) model** to analyze images, detect subtle artifacts, and classify them as **real or fake** with high accuracy. The model is trained on a diverse dataset to improve its ability to recognize deepfake patterns, and it also provides a **confidence score** for each prediction, ensuring transparency in its results.With an accuracy of **93%**, our system is designed for real-time classification and continuous learning, adapting to evolving deepfake techniques. Future enhancements will focus on improving detection accuracy by expanding the dataset, incorporating adversarial training, and refining the model's efficiency. **DeepDetect** serves as a crucial tool in combating digital misinformation, reinforcing trust in visual content.

IX. FUTURE SCOPE

Expand Dataset and Data Quality

Collect a broader and more diverse dataset, covering various AI generation techniques (GANs, Diffusion Models, Deepfakes) and real-world variations across different lighting,

Incorporate Advanced Algorithms

Upgrade the model using advanced architectures like Vision Transformers (ViT), Efficient Net, or ensemble learning, and employ techniques like contrastive learning or adversarial

training to improve detection robustness.

Real-Time Implementation

Optimize the model for real-time processing to enable instantaneous detection in high-load environments such as content moderation platforms, digital media verification, or security

Multimodal Forensics Integration

Enhance the system by integrating metadata analysis, reverse image search, and other forensic indicators (e.g., inconsistencies in shadows, textures, and noise patterns) to support more comprehensive image evaluation.

Enhanced System Integration

Work towards seamless integration of VISION GUARD with platforms such as fact-checking tools, news.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

X. ACKNOWLEDGMENT

We express our heartfelt gratitude to our esteemed Chairman, Dr.S.Jagathrakshakan, M.P., for his unwavering support and continuous encouragement in all our academic endeavors. We express our deepest gratitude to our beloved President Dr.J.Sundeep Aanand, President and Dr.E.Swetha Sundeep Aanand, Managing Director for providing us the necessary to complete our project. We take great pleasure in expressing sincere thanks to Dr.K.Vijaya Baskar Raju ProChancellor, Dr.M.Sundararajan Vice Chancellor (i/c), Dr.S.Bhuminathan Registrar and Dr.R.Hariprakash Additional Registrar, Dr. M. Sundararaj Dean Academics for moulding our thoughts to complete our project. We thank our Dr.S.Neduncheliyan Dean, School of Computing for his encouragement and the valuable guidance. We record indebtedness to Dr.S.Maruthuperumal Head of the Department, Computer Science and Engineering for his immense care and encouragement towards us throughout the course of this project. We also take this opportunity to express a deep sense of gratitude to our guide Mr.V.ETHIRAJULU and our Project Coordinator Dr. K.V.Shiny for their cordial support, valuable information, and guidance, they helped us in completing this project through various stages. We thank our department faculty, supporting staff and friends for their help and guidance to complete this project.

REFERENCES

1. WANG, Y., HAO, Y., & CONG, A. X. (2024). DETECTION OF AI-GENERATED IMAGES FROM VARIOUS GENERATORS USING GATED EXPERT CONVOLUTIONAL NEURAL NETWORK. IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE. HTTPS://IEEEXPLORE.IEEE.ORG/DOCUMENT/10696938
2. HUSSAIN, M., USAMA, M., ADEEL, M., & IMRAN, M. (2023). DEEPFAKE DETECTION USING DEEP LEARNING METHODS: A SYSTEMATIC REVIEW. WIRES DATA MINING AND KNOWLEDGE DISCOVERY. HTTPS://WIRES.ONLINELIBRARY.WILEY.COM/DOI/10.1002/WIDM.1520
3. WANG, Y., HAO, Y., & CONG, A. X. (2024). PERFORMANCE COMPARISON AND VISUALIZATION OF AI-GENERATED-IMAGE DETECTION METHODS. IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE. HTTPS://IEEEXPLORE.IEEE.ORG/DOCUMENT/10508937
4. VERMA, S. K., & SAHAY, S. K. (2023). DEEPFAKE DETECTION AND CLASSIFICATION USING ERROR-LEVEL ANALYSIS AND DEEP LEARNING. SCIENTIFIC REPORTS. HTTPS://WWW.NATURE.COM/ARTICLES/S41598-023-34629-3
5. WANG, Y., HAO, Y., & CONG, A. X. (2024). AIGENERATED IMAGE DETECTION USING A CROSS-ATTENTION ENHANCED CONVOLUTIONAL NEURAL NETWORK. IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE. HTTPS://IEEEXPLORE.IEEE.ORG/DOCUMENT/10317126
6. PATEL, Y., & JAIN, M. (2024). DEEPFAKE IMAGE DETECTION USING MACHINE LEARNING AND DEEP LEARNING. EDUCATIONAL ADMINISTRATION: DEEP_LEARNING HTTPS://WWW.RESEARCHGATE.NET/PUBLICATION/384788048_DEE PFAKE_IMAGE_DETECTION_USING_MACHINE_LEARNING_AND_
7. WANG, Y., HAO, Y., & CONG, A. X. (2024). ONLINE DETECTION OF AI-GENERATED IMAGES. IEEE TRANSACTIONSON ARTIFICIAL INTELLIGENCE. HTTPS://IEEEXPLORE.IEEE.ORG/DOCUMENT/10350523
8. WANG, Y., HAO, Y., & CONG, A. X. (2024). DETECTING AI-GENERATED IMAGES WITH CNN AND INTERPRETATION USING GRAD-CAM. IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENC HTTPS://IEEEXPLORE.IEEE.ORG/DOCUMENT/10649158
9. Aragani, Venu Madhav and Maroju, Praveen Kumar and Mudunuri, Lakshmi Narasimha Raju, Efficient Distributed Training through Gradient Compression with Sparsification and Quantization Techniques (September 29, 2021). Available at SSRN: https://ssrn.com/abstract=5022841 or http://dx.doi.org/10.2139/ssrn.5022841
10. WANG, Y., HAO, Y., & CONG, A. X. (2024). DEEP LEARNING-BASED MODEL FOR DEEPFAKE IMAGE DETECTION.IEEETRANSACTIONSONARTIFICIALINTELLIGENCEHTTPS://IEEEXPLORE.IEEE.ORG/DOCUMENT/10426561
11. WANG, Y., HAO, Y., & CONG, A. X. (2024). A GAN-BASED APPROACH TO DETECT AI-GENERATED IMAGES.IEEETRANSACTIONSONARTIFICIALINTELLIGENCE. HTTPS://IEEEXPLORE.IEEE.ORG/DOCUMENT/10223798

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details