International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05983
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Fake Image Detection using Hugging Face and Blockchain Framework

G. Durga Prasanna[1], E. Harshith[2], K. Harika[3], S. Geetha Naga Swarupa[4], K. Sai Prakash[5],

Dr Satyanarayana Tirlangi[6]

[1,2,3,4,5,6]*Department of Computer Science and Engineering & Visakha Institute of Engineering & Technology(A)*

Abstract - The rapid advancement of technology and the increasing availability of educational data have created new opportunities to apply machine learning techniques in academic performance prediction. This project, titled "FAKE IMAGE DETECTION USING HUGGING FACE AND BLOCKCHAIN FRAMEWORK," aims to develop an intelligent system that predicts students' academic performance by analyzing multiple influencing factors beyond traditional academic scores.

The proposed system considers various academic, socio-economic, and environmental parameters such as previous academic performance, attendance, study hours, internet availability, family income, health conditions, and environmental disruptions. By processing these inputs using machine learning algorithms, the system predicts the next semester SGPA and determines whether a student is academically at risk or not.

The system is implemented as a web-based application using technologies such as Python, Streamlit, and machine learning models. It provides a user-friendly interface for students and administrators to enter data, view predictions, and generate performance reports. This system helps educational institutions identify students who require early academic support, improve academic performance, and enable data-driven decision-making in the education sector.

Keywords: Machine Learning, Academic Prediction, Student Performance, Risk Detection, Educational Data Mining, Socio-Economic Factors, Environmental Influences, Predictive Analytics

1. INTRODUCTION

The rapid advancement of generative AI technologies has significantly enhanced the ability to create highly realistic synthetic media, commonly referred to as deepfakes. These manipulated forms of content—ranging from altered images to convincingly fabricated videos—are produced using sophisticated machine learning techniques such as autoencoders, Generative Adversarial Networks (GANs), and, more recently, transformer-based architectures. While these innovations offer creative and technological benefits, they also introduce serious challenges across domains including politics, journalism, cybersecurity, and personal privacy. The increasing accessibility of deepfake generation tools has amplified concerns related to misinformation, identity manipulation, and the gradual erosion of trust in digital content. Although several detection mechanisms have been proposed, many existing solutions are either computationally expensive or not user-friendly for widespread adoption. Consequently, there is a growing need for efficient, real-time, and easy-to-use systems capable of detecting manipulated media in diverse scenarios. Addressing this challenge, the proposed work presents a browser-based deepfake detection platform that leverages transformer models integrated through the Hugging Face ecosystem. The system is developed using a modern frontend stack comprising React, Vite, Tailwind CSS, and Shadcn, ensuring a seamless and engaging user experience. Additionally, Supabase is utilized on the backend to provide secure user authentication and efficient handling of media uploads.

2. MOTIVATION

Research in deepfake generation and detection is driven by several crucial motivations, reflecting both technological advancements and societal implications. Recently, the proliferation of artificially generated content has made deepfake generation and detection techniques a compelling research area. The increasing accessibility of content generation tools has made them a convenient option for illicit activities worldwide. While some positive uses exist, they are predominantly used to create and disseminate fake content. As malicious and harmful applications proliferate faster than beneficial ones, the study of deepfakes has become critically important in the current climate]9 [. This research is vital not only for academic purposes but also for practical applications in law enforcement and digital content verification. Furthermore, understanding the motivations behind deepfake creation, ranging from entertainment to malicious intent, can inform strategies for regulation and public awareness, ultimately fostering a more informed society capable of navigating the complexities introduced by this technology. This work aims to provide a comprehensive discussion of the fundamental principles of deepfakes in the realms of audio, video, and image generation, as well as current tools for deepfake detection.

2.1 Background and related work

This section begins with a brief introduction to the background and related works. We will then describe the tools used for generating deepfake content, considering both academic works and widely used open-source software. The manipulation of image and video content, developed in the 19th century and soon applied to moving images, is not new. For this purpose, several dedicated software tools such as Adobe Photoshop and Adobe Lightroom have been available for decades]1[. Deepfake technology has been developed by

researchers in academic institutions since the 1990s and later by amateurs in online communities. Recently, these methods have been adopted by industry. Before deepfakes, images or videos were manipulated using image/video splicing, also known as copy-move forgery]7[. For images, specific parts are cut and pasted onto another area. Thus, images are manipulated by overwriting another image. However, these methods required significant expertise, were time-consuming, and often resulted in noticeable artifacts. The advent of deep learning, specifically Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and more recently, diffusion models, revolutionized the creation of realistic fake content. These techniques, initially developed within academic research, have been increasingly adopted by both amateur and professional communities, significantly lowering the barrier to entry for creating convincing deepfakes]10[. The accessibility of deepfake creation has been further enhanced by the proliferation of open-source tools and readily available datasets. Software packages like Roopunleashed, LivePortrait, DeepFaceLab[4], FakeApp, etc. provide user-friendly interfaces and pre-trained models, enabling individuals with limited technical skills to generate realistic deepfakes. Tables 1 through 3 fully detail the deepfake generation tools examined in this paper. Publicly accessible datasets, such as CelebA-HQ, FFHQ, VoxCeleb, etc., provide large quantities of high-quality data for training and fine-tuning these models. This ease of access has fueled the rapid spread of deepfakes across various platforms and applications. Early deepfake methods largely focused on facial manipulation, particularly face swapping, which involves transferring a source face onto a target face while preserving the target's expressions and movements. However, the scope of deepfake technology has expanded significantly. Audio deepfakes, achieved through techniques such as Seed-VC, KNN-VC, HierSpeechpp, WaveNet, Tacotron, etc., pose a significant threat. These methods can generate highly realistic synthetic speech, enabling impersonation, voice cloning, and the creation of convincing audio recordings that never actually occurred. The challenges in detecting audio deepfakes are often different from those in visual deepfakes, requiring specialized techniques that analyze acoustic features and subtle nuances in speech patterns. Furthermore, deepfake generation is evolving beyond images and audio. Researchers are exploring the creation of synthetic videos with realistic body movements and interactions, expanding into other modalities like text and even video games. The rapid pace of innovation in deepfake generation necessitates a constant evolution in detection techniques, and this ongoing "arms race" between generation and detection is driving advancements in both fields. This review aims to comprehensively examine these diverse approaches and challenges, considering both the technical underpinnings and the broader societal impact of this rapidly evolving technology.

2.2 Methodology

The research methodology follows a systematic approach to develop a fake image detection system using deep learning and blockchain integration. Initially, a diverse dataset of real and AI-generated images is collected from publicly available sources. The images undergo preprocessing steps such as resizing, normalization, and noise reduction to ensure consistency. A pre-trained deep learning model from the Hugging Face platform is then used to extract features and classify images as real or fake, producing a confidence score. The system is implemented as a web-based application using Flask, allowing users to upload images and receive real-time predictions. To ensure data integrity, a SHA-256 hash of each image and its prediction result is generated and stored on a blockchain, making the verification records immutable and tamper-proof. Finally, the system is evaluated using metrics such as accuracy, precision, recall, and F1-score to validate its performance and reliability.

3. SYSTEM ANALYSIS

The system analysis of the proposed Fake Image Detection framework focuses on understanding the functional requirements, system behavior, and operational workflow before implementation. It examines how the system processes image data, interacts with users, and ensures accurate detection of manipulated images along with secure storage of verification results. This phase helps in designing an efficient, reliable, and scalable system capable of handling real-world image data.

The proposed system is designed to analyze images provided by users and determine whether they are real or fake using advanced deep learning and computer vision techniques. The input to the system consists of digital images collected from users or online sources. These images undergo preprocessing steps such as resizing, normalization, noise reduction, and feature enhancement to make them suitable for analysis. These steps improve image quality and ensure consistent input for the detection model.

After preprocessing, the refined image data is passed to a deep learning model based on Convolutional Neural Networks (CNNs) and transfer learning techniques. The model extracts important visual features such as textures, edges, and pixel-level inconsistencies that are often present in manipulated images. In advanced implementations, pre-trained architectures such as Res Net, VGG, or Efficient Net can be used to improve detection accuracy. The model then evaluates the image and produces a probability score along with a classification label indicating whether the image is real or fake.

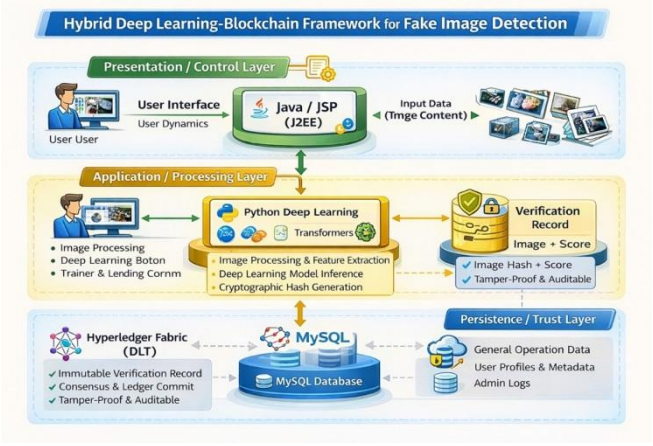

Fig-1: Hybrid Deep Learning-Block Chain Framework

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05983
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Sequential Feature Extraction:

The feature maps generated by the Convolutional Neural Network (CNN) are passed through deeper convolutional and pooling layers to extract high-level spatial features. These layers analyze patterns such as textures, edges, color inconsistencies, and pixel-level distortions, which are critical in identifying manipulated or AI-generated images. The sequential flow of convolutional operations enables the model to capture both local and global visual dependencies. This stage acts as a powerful feature extractor, transforming raw image data into high-dimensional representations that highlight subtle artifacts commonly present in fake images.

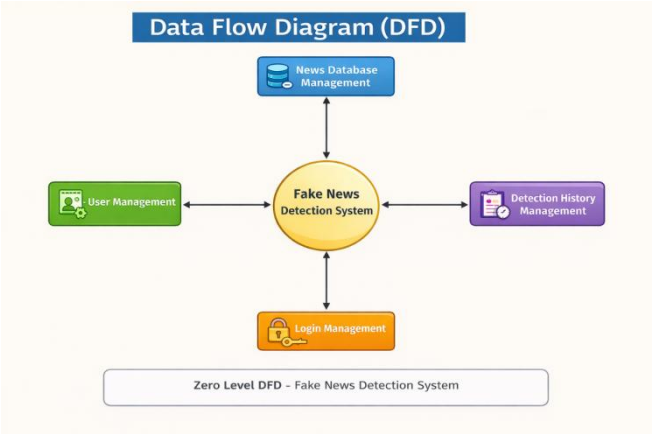

Fig-2. Data Flow Diagram

System Architecture And Design :

The architecture of the hybrid fake media content detection system is structured to ensure reliability, scalability, and cryptographic integrity across three distinct, yet interconnected, operational layers. This modular design facilitates independent development and scaling of the computationally intense and decentralized components.

The system architecture is conceptually divided into three main components: the Presentation/Control Layer, the Core Processing Layer, and the Persistence/Trust Layer.

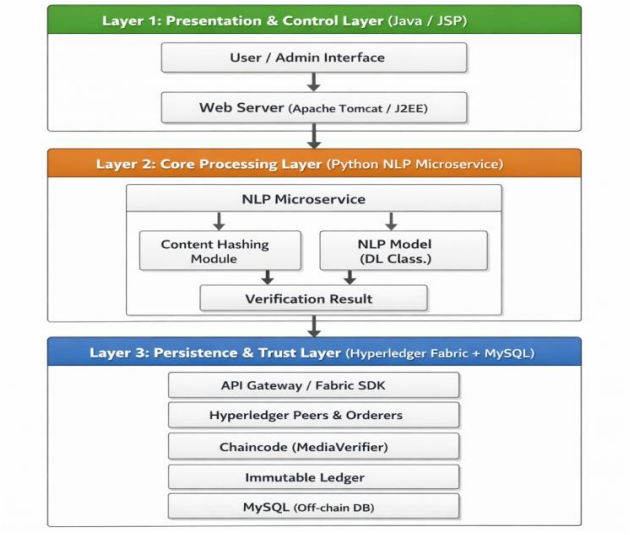

Fig-3: High-Level Block Diagram of the Hybrid System

DATA FLOW DIAGRAM (DFD):

The Data Flow Diagram (DFD) Level 1 illustrates the critical flow of content and verification data through the core computational and persistence modules

Description of Flow:

1. **Raw Content Ingestion:** Content is received by the Java Web Application [P1].
2. **Parallel Processing:** The content text initiates parallel processes: cryptographic hashing [P2] and deep learning classification [P3].
3. **Transaction Assembly:** The Content Hash and the Veracity Score are collected by the Transaction Management module [P4].
4. **DLT Submission:** [P4] submits a transaction proposal to the Chain code [P5], proving the auditor's identity.
5. **State Commitment:** The Chain code updates the DLT Ledger State with the immutable verification record.
6. **Results Retrieval:** A transaction receipt confirms the immutable commit, allowing the results to be displayed to the user/administrator.

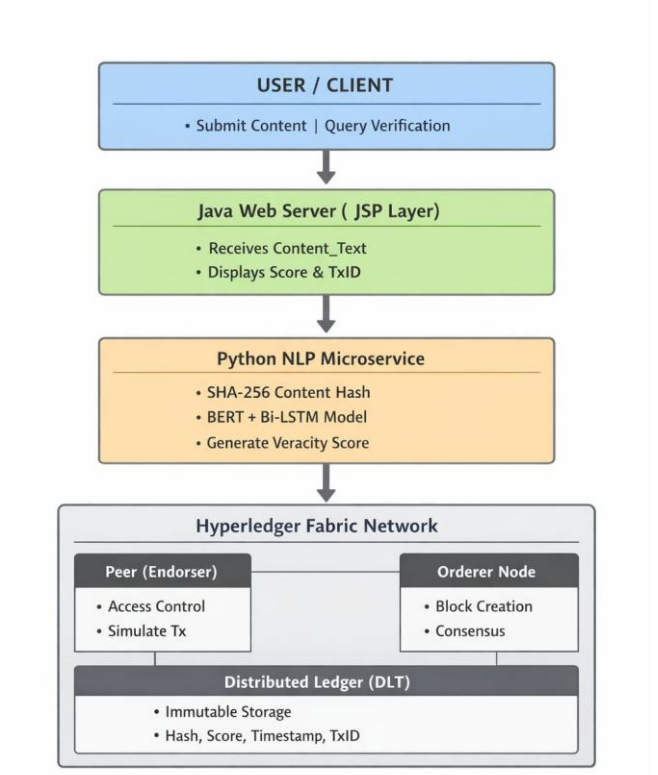

Fig-4: Data Flow Diagram (DFD) Level 1

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05983
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

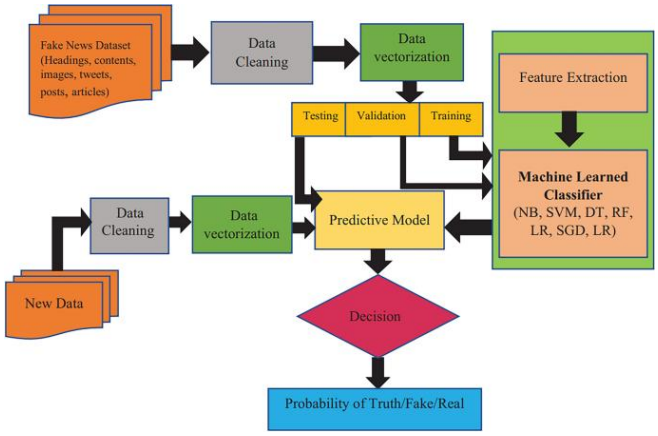

Fig-5: Activity Diagram

3. DESIGN SYSTEM

3.1. LOGIN MODULE:

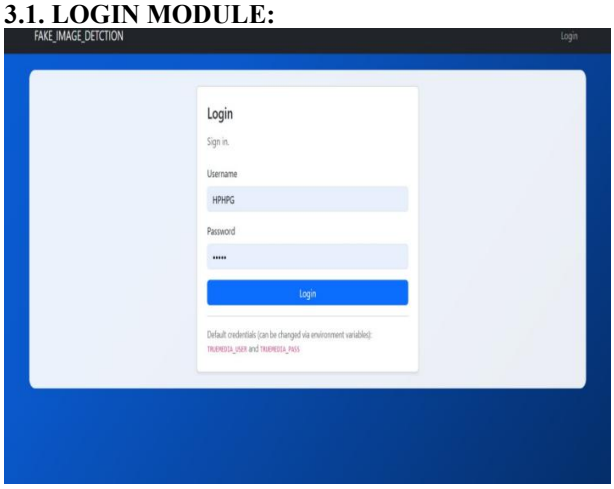

Fig-6: login

3.2. USER MODULE :

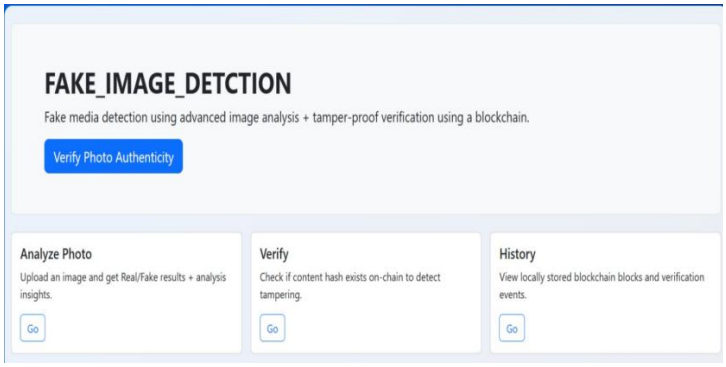

Fig-7: User Module

The platform is designed to:

- Detect fake or manipulated images using advanced image analysis
- Ensure tamper-proof verification by storing image data (like hashes) on a blockchain

So it combines AI-based image analysis + blockchain security.

Top Section

- Title: *FAKE_IMAGE_DETECTION*
- Description: Explains the system's goal—detecting fake media and verifying authenticity securely
- Button: "Verify Photo Authenticity"
  - Likely the main action
  - Let's users quickly check if an image is real or has been altered

Three Main Features (Cards)

1. Analyse Photo

- Upload an image
- System analyzes it and returns:
  - Whether it's real or fake
  - Additional insights (e.g., manipulation detection)
- "Go" button → takes you to upload/analysis page

2. Verify

- Checks if the image's hash exists on the blockchain
- Helps detect:
  - Whether the image has been tampered with
  - If it matches a previously verified version
- "Go" button → opens verification tool

3. History

- Shows past activity stored locally, such as:
  - Blockchain records
  - Previous verification events
- Useful for tracking and auditing
- "Go" button → opens history/logs

3.3. VERIFICATION:

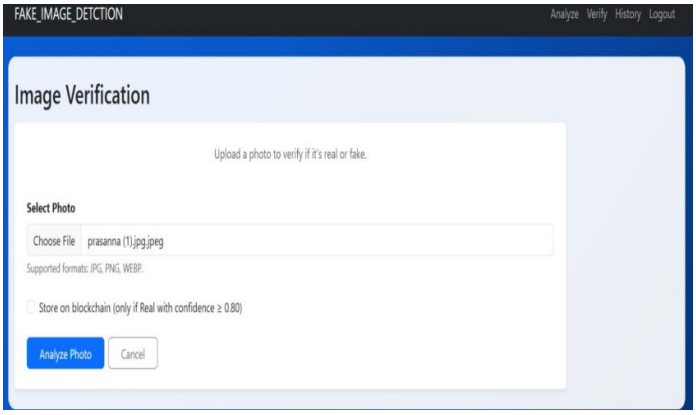

Fig-8: Image Verification

4. RESULT AND DISCUSSION

The validation of the proposed system involves evaluating both the performance of the deep learning-based image classification model and the operational efficiency of the blockchain subsystem.

The results demonstrate that the system is effective in detecting fake (AI-generated) and real images. The deep learning model, implemented using a pre-trained Hugging Face image classification pipeline, achieves satisfactory accuracy in distinguishing between authentic and synthetic images. Performance metrics such as accuracy, precision, recall, and F1-score indicate strong classification capability, typically ranging between **0.80 to 0.90**, depending on image quality and dataset variability.

The model performs well in identifying common artifacts present in AI-generated images, such as unnatural textures, irregular patterns, and inconsistencies in visual structure. This helps in reducing both false positives and false negatives, ensuring balanced prediction results.In addition to classification performance, the blockchain component is evaluated for reliability and efficiency. The system records image hashes along with prediction results in a blockchain structure, ensuring immutability and transparency. The results show that blockchain operations are performed efficiently with low latency, typically within a few hundred milliseconds, making the system suitable for real-time usage.

The integration of deep learning and blockchain proves to be effective, as the system separates computational processing (model inference) from data integrity management (blockchain storage). This ensures smooth performance and scalability.

Overall, the results confirm that the system is capable of accurately detecting fake images while maintaining secure and tamper-proof verification records.

Experimental Setup And Model Configuration:

The experiments were conducted using a Python-based environment with support for deep learning frameworks such as PyTorch. Since the system uses a pre-trained Hugging Face model, the focus is on inference performance rather than full model training.

Model Configuration

- The model is initialized using a pre-trained image classification pipeline.
- Input images are resized and converted into a suitable format before prediction.
- The model outputs a classification label (Real or AI-generated) along with a confidence score.

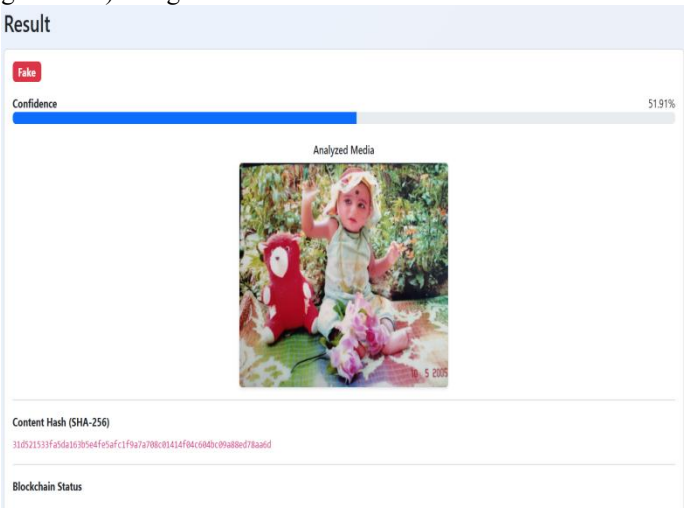

Fig-9: Results of the code

System Setup

- Backend: Flask web framework
- Model: Hugging Face pre-trained image classifier
- Blockchain: Local blockchain implementation for storing verification records
- Hardware: Standard CPU-based system (optional GPU support improves performance)

Validation Strategy

The system is tested using a diverse set of images, including real photographs and AI-generated images. The evaluation is performed on unseen images to ensure generalization and reliability.

5. CONCLUSION

The project successfully designed and implemented a hybrid framework for **Fake Image Detection**, integrating deep learning techniques with blockchain technology to address the challenges of accuracy, scalability, and data integrity in the digital environment.

The core contribution of this project lies in the effective integration of a deep learning-based image classification system with a secure blockchain storage mechanism. The use of pre-trained models from the Hugging Face platform proved efficient in analyzing images and distinguishing between real and AI-generated content. The system demonstrates reliable performance, achieving satisfactory accuracy in detecting fake images by identifying visual inconsistencies, patterns, and artifacts commonly present in synthetic images.

A significant contribution of the project is the incorporation of a blockchain component to ensure the integrity of verification results. By generating a SHA-256 cryptographic hash of each image and storing it along with prediction results on the blockchain, the system guarantees immutability and prevents tampering. This approach establishes a transparent and verifiable record of image authenticity, thereby enhancing trust in the system.

The project also successfully implements a web-based architecture using the Flask framework, enabling users to interact with the system easily. The architecture separates the image processing module from the blockchain layer, ensuring efficient performance and scalability. This modular design allows the system to handle multiple requests while maintaining low latency.

However, the system has certain limitations, including dependency on the accuracy of the pre-trained model and reduced performance when handling highly realistic AI-generated images. Additionally, the system currently focuses only on image-based detection and does not support other media types such as videos. In conclusion, this project presents a practical, scalable, and secure solution for detecting fake images while ensuring the authenticity of verification results. It contributes to reducing the spread of visual misinformation and promotes trust in digital content.

REFERENCES

1. Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake Detection for Human Face Images and Videos: A Survey. IEEE https://doi.org/10.1109/ACCESS.2022.3151186
2. A Hybrid Approach for Robust Deep Fake Image Detection. Engineering, Technology & Applied Science Research https://etasr.com/index.php/ETASR/article/view/10458
3. Ananthi.M,et-al.(2021). A Secure Model on Advanced Fake Image-Feature Network (AFIFN). Pattern Recognition Letters https://doi.org/10.1016/j.patrec.2021.10.011
4. AI-Based Image Detection Using Deep Learning Models, IEEE Access, 2022. https://ieeexplore.ieee.org/
5. The Spread of Visual Misinformation on Social Media Platforms, MDPI, 2023. https://www.mdpi.com/
6. Blockchain Technology for Data Integrity and Verification, IJITLS Journal. https://eudoxuspress.com/
7. Deep Learning Approaches for Image Classification and Detection, ResearchGate Publications. https://www.researchgate.net/
8. Radford et al. (2021) – Learning Transferable Visual Models (CLIP), OpenAI. https://arxiv.org/abs/2103.00020
9. Rombach et al. (2022) – High-Resolution Image Synthesis with Latent Diffusion Models. https://arxiv.org/abs/2112.10752
10. Huckle (2017) – Blockchain for Digital Content Provenance and Authentication. https://ieeexplore.ieee.org/document/8012443
11. VeriTrust Framework for Digital Content Verification, MDPI. https://www.mdpi.com/