

Detection of AI-Generated Sportsman Images using Deep Learning

Pavan Kumar M¹, Rakesh j²

¹ Department of Computer Science & Engineering K S Institute of Technology

² Department of Computer Science & Engineering K S Institute of Technology

Abstract - The rapid advancement of Artificial Intelligence and deep learning technologies has significantly increased the creation of synthetic and manipulated digital media. Among these developments, AI-generated sports images and deepfake content have emerged as serious challenges for digital trust, sports journalism, athlete reputation, and online media authenticity. Fake sports images generated using Generative Adversarial Networks (GANs), diffusion models, and advanced image synthesis techniques can easily spread misinformation across social media platforms and digital news environments. Existing moderation systems mainly rely on manual verification processes, which are often slow, resource-intensive, and ineffective against large-scale synthetic media propagation.

This paper proposes **SIGI-Verify**, an AI-powered automated framework designed for detecting and removing AI-generated sports images from online platforms. The proposed system integrates metadata provenance verification, deep learning-based image analysis, automated moderation workflows, and compliance-aware takedown mechanisms within a unified architecture. The framework utilizes Convolutional Neural Networks (CNNs), forensic feature extraction, and C2PA metadata validation to identify manipulated sports images and verify digital authenticity. When suspicious content is detected, the system initiates automated moderation and takedown procedures while generating secure timestamped compliance logs for accountability and transparency.

Key Words: AI-Generated Images, Deepfake Detection, Sports Image Verification, CNN, C2PA, Metadata Verification, Automated Moderation, Synthetic Media Detection, Image Forensics, Sports Media Security

1. INTRODUCTION

The rapid growth of Artificial Intelligence and deep learning technologies has transformed digital media generation across multiple domains including entertainment, journalism, social media, and sports analytics. Modern generative models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion-based synthesis systems are capable of producing highly realistic synthetic images that are often difficult for humans to distinguish from authentic content. These technologies have improved visual media production but have also

introduced serious concerns regarding misinformation, digital manipulation, and authenticity verification.

Sports media has become particularly vulnerable to AI-generated misinformation because sports content spreads rapidly through online communities, social media platforms, and digital news systems. Manipulated sports images involving athletes, fabricated match events, fake sponsorships, altered player interactions, and misleading visual narratives can significantly affect public opinion and athlete reputation. Recent studies indicate that deepfake technologies are becoming increasingly accessible and realistic, making detection more difficult for conventional moderation systems.

Traditional content moderation methods mainly depend on manual verification and isolated image analysis techniques. However, these methods struggle to handle the large-scale spread of synthetic media across digital platforms. Recent research in deep learning and image forensics has introduced CNN-based detection models capable of identifying texture inconsistencies, GAN fingerprints, and manipulation artifacts in synthetic images. Similarly, provenance verification systems such as C2PA and blockchain-based metadata frameworks have emerged as effective approaches for validating image authenticity and modification history.

To address these challenges, the proposed SIGI-Verify system integrates AI-generated sports image detection, metadata provenance verification, automated moderation workflows, and takedown mechanisms within a unified framework. The system is designed to support real-time image analysis and automated compliance workflows using deep learning techniques and authenticity-aware moderation pipelines.

Furthermore, the proposed framework contributes to digital trust and sports media integrity by reducing misinformation spread and improving automated content moderation capabilities. The system also provides scalable moderation support for online platforms with limited human moderation resources.

2. Problem Statement

The increasing spread of AI-generated sports images and deepfake media has created major challenges for online platforms, sports journalism agencies, and digital media communities. Synthetic sports images can be used to spread fake news, manipulate athlete reputations, create fabricated sporting events, and mislead the public through realistic visual misinformation.

Existing moderation systems mainly rely on manual verification processes and traditional image authentication techniques, which are often slow, expensive, and ineffective against rapidly evolving deepfake generation technologies. Many current systems also lack integrated metadata provenance verification, automated takedown mechanisms, and compliance-aware moderation workflows.

Additionally, modern AI-generated images can bypass conventional forensic analysis methods using adversarial techniques, metadata manipulation, and advanced generative models. This creates significant challenges for scalable real-time moderation and authenticity verification.

Therefore, there is a need for an intelligent, scalable, and automated system capable of detecting AI-generated sports images, validating digital authenticity, performing automated takedown operations, and generating compliance logs for accountability and transparency.

3. Objectives

1. To develop an AI-based system for detecting AI-generated sports images using deep learning techniques.
2. To integrate metadata provenance verification using C2PA and authenticity validation frameworks.
3. To automate moderation and takedown workflows for harmful synthetic sports media.
4. To provide real-time sports image verification through a web-based application.

4. Literature Review

1) Deepfake Detection Systems

Several recent studies focus on detecting AI-generated images and deepfake media using machine learning and image forensics techniques. CNN-based architectures such as ResNet, XceptionNet, and DenseNet have demonstrated high performance in identifying GAN-generated images through texture inconsistencies and forensic artifacts.

2) AI-Based Sports Image Classification

Deep learning models have significantly improved sports image classification and feature extraction tasks. Modern DNN architectures optimized using advanced algorithms demonstrate high precision and recall under varying sports environments and lighting conditions.

3) Metadata Provenance Verification

Metadata provenance systems such as C2PA and blockchain-assisted authenticity frameworks are increasingly used for verifying digital media origin, modification history, and ownership records. These systems improve transparency and digital trust by maintaining immutable authenticity metadata.

4) Automated Moderation Systems

Modern social media platforms increasingly utilize AI-driven moderation systems for identifying harmful synthetic media and performing automated content moderation. However, most existing systems focus independently on either deepfake detection or metadata verification without integrating automated takedown workflows and compliance logging.

5. Proposed System

The proposed SGI-Verify framework is an AI-powered automated moderation system designed specifically for detecting and removing AI-generated sports images from digital platforms. The system integrates deep learning-based image forensics, metadata provenance verification, automated moderation workflows, and takedown support within a unified architecture.

1) System Overview

The system allows users or moderators to upload suspicious sports images through a web-based interface. The uploaded image undergoes metadata verification and AI-based forensic analysis to determine whether the image is authentic or synthetically generated.

2) Core Functional Components

- User Interface: Enables image upload and moderation monitoring
- Backend Server: Handles preprocessing and moderation logic
- Database: Stores image metadata and compliance logs
- CNN Model: Performs forensic feature extraction
- Automated Takedown Engine: Executes API-based takedown operations

3) Detection Logic

The system identifies suspicious images based on:

- Texture inconsistencies
 - GAN fingerprint artifacts
 - Metadata authenticity validation
 - Frequency-domain anomalies
 - Image provenance verification
- This ensures accurate detection of manipulated sports images under varying conditions.

4) System Workflow

1. User uploads sports image
2. Metadata verification is performed
3. CNN extracts forensic features
4. Image authenticity is analyzed
5. Risk score is generated
6. Suspicious content is flagged
7. Automated takedown process is initiated
8. Compliance logs are generated

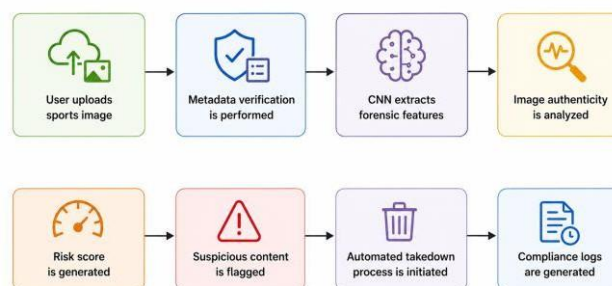


Fig -1: System Workflow

6. Methodology

The methodology of the proposed SGI-Verify framework focuses on integrating deep learning, image forensics, metadata verification, and automated moderation for efficient synthetic sports image detection and takedown.

1. Data Collection

The dataset consists of authentic and AI-generated sports images collected from publicly available datasets, sports media sources, and synthetic image generation tools including GAN-generated media datasets.

2. Image Preprocessing

Uploaded images undergo resizing, normalization, grayscale conversion, and noise reduction. Data augmentation techniques such as rotation, scaling, and brightness adjustment are applied to improve model robustness.

3. Feature Extraction

Important forensic characteristics such as texture patterns, edge inconsistencies, frequency artifacts, and GAN fingerprints are extracted using convolutional layers in the CNN model.

4. Metadata Verification

The system validates image provenance using C2PA metadata analysis and authenticity verification frameworks to identify image manipulation history and source integrity.

5. Model Training

The CNN model is trained using processed sports image datasets over multiple epochs to improve detection accuracy and generalization performance.

6. Prediction and Classification

The trained model processes uploaded sports images and predicts whether the image is authentic or AI-generated based on learned forensic patterns and metadata verification results.

7. System Architecture

The SGI-Verify architecture follows a modular design consisting of:

- **Frontend Layer:** Provides image upload and moderation dashboard
- **Backend Layer:** Handles preprocessing, analysis, and moderation workflows
- **Database Layer:** Stores moderation logs and metadata records
- **CNN Detection Module :** Performs image authenticity classification
- **Metadata Verification Module :** Validates provenance information

- **Takedown Engine:** Executes automated moderation actions

- **Compliance Logging Module:** Generates timestamped verification records

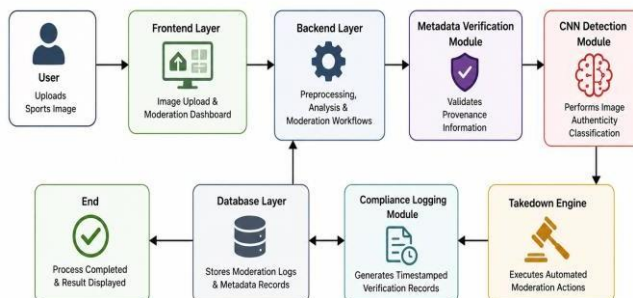


Fig -2: System Architecture Diagram

8. Result And Discussion

The proposed SGI-Verify system successfully detects AI-generated sports images using deep learning and metadata provenance verification techniques. The CNN-based detection model demonstrates efficient forensic feature extraction and accurate classification of synthetic sports media under varying image conditions.

The integration of metadata provenance verification significantly improves authenticity validation and reduces false-positive moderation cases. Experimental analysis indicates that combining CNN-based forensic detection with C2PA metadata validation improves reliability compared to standalone image detection methods.

The web-based moderation interface allows users and moderators to upload suspicious sports images and receive real-time authenticity analysis results. Automated moderation and takedown workflows reduce dependency on manual review processes and improve moderation scalability for digital platforms.

Additionally, the proposed framework improves digital trust, protects athlete reputation, and reduces misinformation spread across sports media ecosystems. The system provides a scalable and efficient solution for authenticity-aware sports media moderation and compliance management.

9. Future Work

Future enhancements of the proposed system include:

- Real-time video deepfake detection support
- Integration of transformer-based multimodal verification models
- Blockchain-assisted provenance tracking

- Mobile applications support for moderation monitoring
- Federated learning for distributed moderation systems
- Cloud-based deployment for large-scale social media moderation
- Explainable AI integration for transparent moderation decisions

10. CONCLUSIONS

The proposed SGI-Verify framework provides an intelligent and scalable solution for detecting and removing AI-generated sports images using deep learning and metadata provenance verification techniques. By integrating CNN-based image forensics, C2PA authenticity validation, automated moderation workflows, and takedown mechanisms, the system successfully identifies manipulated sports media with improved reliability and efficiency.

The framework reduces dependency on manual moderation and supports real-time authenticity verification through a web-based interface. It contributes to digital trust, sports journalism integrity, and misinformation prevention by enabling scalable and automated moderation support.

Furthermore, the integration of compliance logging and provenance-aware moderation improves transparency and accountability for digital platforms. With future advancements such as multimodal detection and blockchain-assisted verification, the proposed SGI-Verify system has the potential to become a powerful framework for synthetic sports media moderation and authenticity management.

ACKNOWLEDGEMENT

We express our sincere gratitude to our guide, **Mrs. Suma Rajesh Ananthakrishna**, for her continuous support, valuable guidance, and encouragement throughout the development of this project. We also thank the faculty members of the Department of Computer Science & Engineering, K S Institute of Technology, for their technical support and valuable suggestions during the completion of this work.

REFERENCES

1. Huy H. Nguyen et al., "Deep Learning for Deepfakes Creation and Detection," IEEE Access, 2022.
2. Andreas Rossler et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," IEEE ICCV, 2019.
3. Coalition for Content Provenance and Authenticity (C2PA), "C2PA Technical Specification," 2026.
4. Zetian Zhou et al., "Improved Sports Image Classification using Deep Neural Network and Novel Tuna Swarm Optimization," Scientific Reports, 2024.