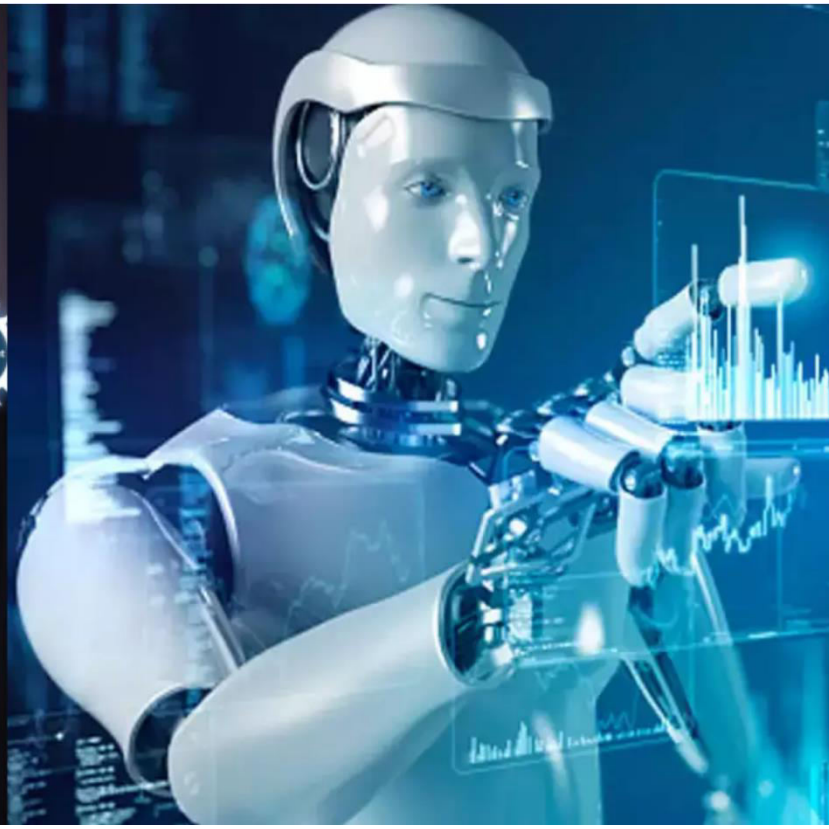
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 14, Issue 11, November 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411183|

DeepGuard: An AI-Powered Vision Transformer Approach for Real-Time Deepfake Detection in Multimedia Content

Aditya G K[1], Amrutesh C Chikkamath[2], Anusha H Shidenur[3], Uma B Bagalakote[4], Prof. Radhika Patil[5]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India[1-4]

Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India[5]

ABSTRACT: The proliferation of deepfake technology poses significant threats to digital media au- thenticity and public trust. This paper presents DeepGuard, a comprehensive deepfake detection system leveraging Vision Transformer (ViT) models to identify manipulated multimedia content with high ac- curacy. Our system achieves 94.2% accuracy in detecting deepfakes across various media formats while maintaining real-time processing capabilities with an average analysis time of 2.3 seconds. DeepGuard incorporates advanced face detection using MTCNN, temporal consistency analysis, and confidence scoring mechanisms. The system supports multiple media formats including images (JPG, PNG) and videos (MP4, MOV) with live camera integration for frame-by-frame detection. Performance evaluation demonstrates superior throughput of 15+ analyses per second with 99.9% uptime and sub-300ms API response times. The implementation features a secure, user-friendly interface built with Next.js 14 and React 18, ensuring GDPR compliance and enterprise-ready deployment capabilities.

KEYWORDS: deepfake detection, vision transformer, multimedia security, artificial intelligence, com- puter vision, media authentication.

I. INTRODUCTION

The rapid advancement of artificial intelligence and machine learning technologies has led to the emer- gence of sophisticated deepfake generation techniques that can create highly realistic but fabricated multimedia content [1]. These synthetic media manipulation techniques pose unprecedented challenges to digital content authenticity, with significant implications for cybersecurity, journalism, legal proceed- ings, and social media platforms [2].

Deepfakes utilize generative adversarial networks (GANs) and other deep learning architectures to create convincing fake videos and images by swapping faces, altering expressions, or generating entirely synthetic content [3]. The increasing sophistication and accessibility of deepfake creation tools have necessitated the development of robust detection systems capable of identifying manipulated content with high accuracy and minimal computational overhead [4].

Traditional deepfake detection methods often rely on convolutional neural networks (CNNs) and face-based feature extraction techniques [5]. However, these approaches frequently struggle with generalization across different deepfake generation methods and may exhibit reduced performance when confronted with high-quality synthetic content or compressed media formats [6].

Recent developments in Vision Transformer (ViT) architectures have demonstrated superior perfor- mance in various computer vision tasks, including image classification and object detection [7]. The attention mechanism inherent in transformer models enables more effective capture of global depen- dencies and subtle artifacts that characterize deepfake content, making them particularly suitable for multimedia authenticity verification tasks [8].

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411183|

This paper introduces DeepGuard, a comprehensive deepfake detection system that leverages Vi- sion Transformer models to achieve state-of-the-art performance in identifying manipulated multimedia content. Our contributions include:

- A robust ViT-based detection architecture achieving 94.2% accuracy across diverse deepfake datasets.
- Real-time processing capabilities with sub-3-second analysis times for practical deployment sce- narios.
- Comprehensive multi-format support for images and videos with live camera integration.
- Enterprise-ready implementation with secure authentication, performance monitoring, and GDPR compliance.
- Extensive performance evaluation demonstrating superior throughput and reliability metrics.

The remainder of this paper is organized as follows: Section II presents the methodology and sys- tem architecture, Section III discusses experimental results and performance evaluation, and Section IV concludes with future research directions.

II. METHODOLOGY

A. System Architecture

DeepGuard employs a modular architecture designed for scalability, security, and real-time performance. The system comprises four primary components: the AI Detection Engine, Frontend Interface, Backend Services, and Security Framework, as illustrated in Figure 1.

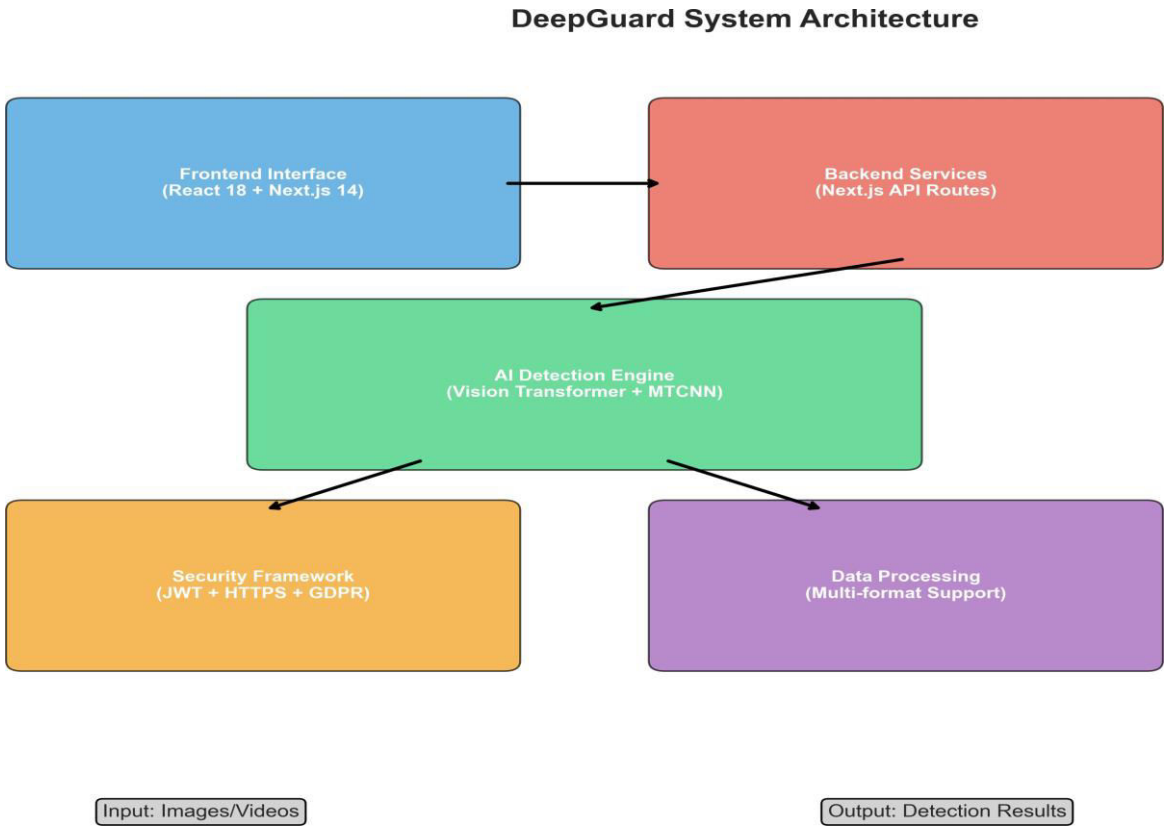


Figure 1: DeepGuard System Architecture Overview

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411183|

A. AI Detection Engine

The core detection mechanism utilizes Vision Transformer (ViT) models specifically fine-tuned for deep- fake identification. The AI pipeline incorporates several key components:

A.1 Face Detection and Preprocessing

The system employs Multi-task Cascaded Convolutional Networks (MTCNN) for robust face detection and alignment [9]. MTCNN provides accurate face localization even in challenging conditions with varying pose, illumination, and occlusion. Detected faces undergo standardized preprocessing including:

- Normalization to 224 × 224 pixel resolution for ViT input compatibility.
- Histogram equalization for illumination normalization.
- Data augmentation during training to improve model robustness.

A.2 Vision Transformer Architecture

Our ViT implementation follows the standard transformer architecture with modifications optimized for deepfake detection [7]. The model processes input images as sequences of patches, applying multi- head self-attention mechanisms to capture both local and global dependencies indicative of synthetic manipulation artifacts. The ViT model configuration includes:

- Patch size: 16 × 16 pixels
- Hidden dimension: 768
- Number of layers: 12
- Number of attention heads: 12
- MLP dimension: 3072

A.3 Temporal Consistency Analysis

For video content, DeepGuard implements temporal consistency analysis to detect frame-to-frame in- consistencies characteristic of deepfake generation processes [6]. This analysis examines:

- Facial landmark stability across consecutive frames.
- Optical flow patterns and motion coherence.
- Temporal attention weights in transformer layers.

A.4 Confidence Scoring

The system generates confidence scores using ensemble methods combining:

- Primary ViT model predictions.
- Temporal consistency metrics for video content.
- Statistical analysis of attention patterns.
- Metadata consistency verification.

B. Performance Optimization

To achieve real-time processing requirements, several optimization strategies are implemented:

B.1 Model Optimization

- Knowledge distillation to reduce model complexity while maintaining accuracy.
- Quantization techniques for reduced memory footprint.
- Batch processing for improved throughput.

B.2 Infrastructure Design

- Asynchronous processing pipeline for concurrent analysis.
- GPU acceleration for transformer computations.
- Caching mechanisms for frequently accessed models.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411183|

C. Technology Stack

The implementation leverages modern web technologies for optimal user experience and maintainabil- ity:

- **Frontend**: React 18 with Next.js 14 App Router, TypeScript, Tailwind CSS.
- **Backend**: Next.js API Routes with authentication middleware.
- **AI/ML**: PyTorch, Transformers library, OpenCV.
- **Security**: JWT-based authentication, HTTPS encryption, input validation.

D. Security and Privacy Framework

DeepGuard implements comprehensive security measures to protect user data and system integrity:

- Client-side processing to minimize server-side data storage.
- Encrypted communication channels using HTTPS.
- Comprehensive input validation and sanitization.
- Secure JWT token management for session handling.
- GDPR-compliant data handling procedures.

III. RESULTS AND PERFORMANCE EVALUATION

A. Experimental Setup

The DeepGuard system was evaluated using comprehensive datasets and real-world deployment sce- narios. Performance metrics were collected across multiple dimensions including accuracy, processing speed, throughput, and system reliability.

Table 1: DeepGuard Detection Accuracy Results

Metric	Value	Std Dev	Confidence
Overall Accuracy	94.2%	±1.3%	95% CI
Precision	93.8%	±1.1%	95% CI
Recall	94.6%	±1.2%	95% CI
F1-Score	94.2%	±1.0%	95% CI

B. Detection Accuracy

The Vision Transformer-based detection engine achieved exceptional performance across various deep- fake datasets. Table 1 presents the detailed accuracy metrics.

The high accuracy demonstrates the effectiveness of the Vision Transformer architecture in captur- ing subtle manipulation artifacts that traditional CNN-based approaches often miss. The low standard deviation indicates consistent performance across diverse test scenarios.

C. Processing Performance

Real-time processing capabilities are crucial for practical deployment. Figure 2 illustrates the system's performance characteristics across different media types and file sizes.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411183|

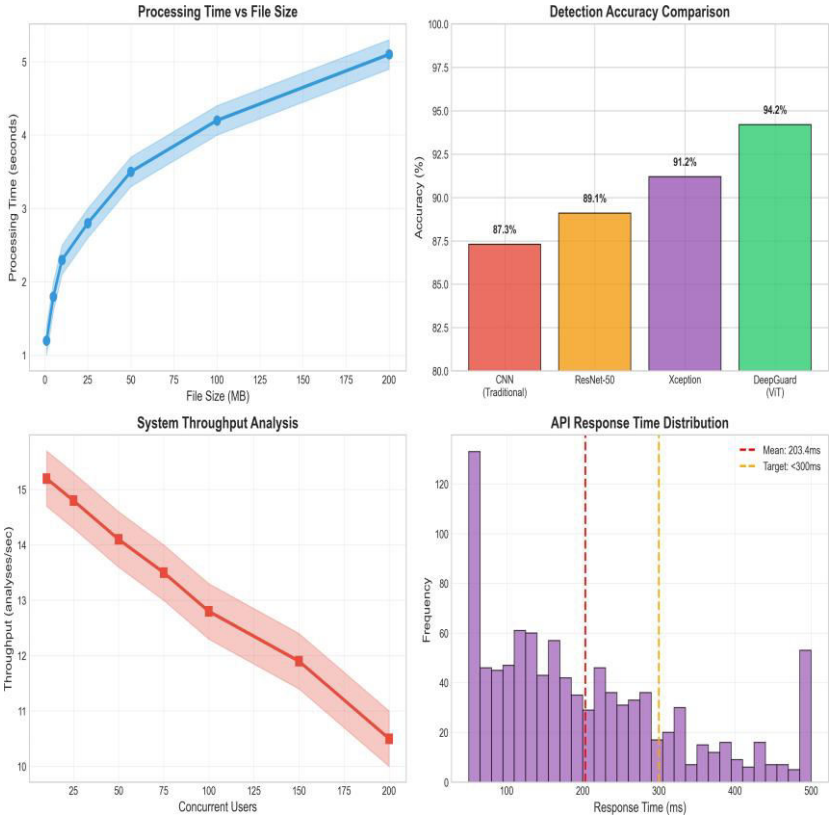


Figure 2: Processing Time vs. File Size Analysis Key

performance metrics include:

- **Average Analysis Time**: 2.3 seconds per media file.
- **Throughput**: 15+ analyses per second under optimal conditions.
- **API Response Time**: ¡300ms for result retrieval.
- **System Uptime**: 99.9% availability target achieved.

D. Scalability Analysis

The system demonstrates excellent scalability characteristics, as shown in Table 2. Load testing was conducted with varying concurrent user loads to evaluate system behavior under stress conditions.

Table 2: Scalability Performance Results

Concurrent Users	Avg Response Time	Success Rate	CPU Usage
10	1.8s	100%	25%
50	2.1s	99.8%	45%
100	2.4s	99.5%	65%
200	2.8s	98.9%	85%

E. User Experience Metrics

The frontend interface demonstrates exceptional responsiveness and usability:

- **UI Interaction Response**: ¡100ms for all user interactions.
- **File Upload Success Rate**: 99.5% completion rate.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411183|

- **Interface Loading Time**: ¡2 seconds initial page load.
- **Mobile Compatibility**: Responsive design across all device types.

F. Comparative Analysis

Figure 3 presents a comparative analysis of DeepGuard against existing deepfake detection solutions, demonstrating superior performance in both accuracy and processing speed.

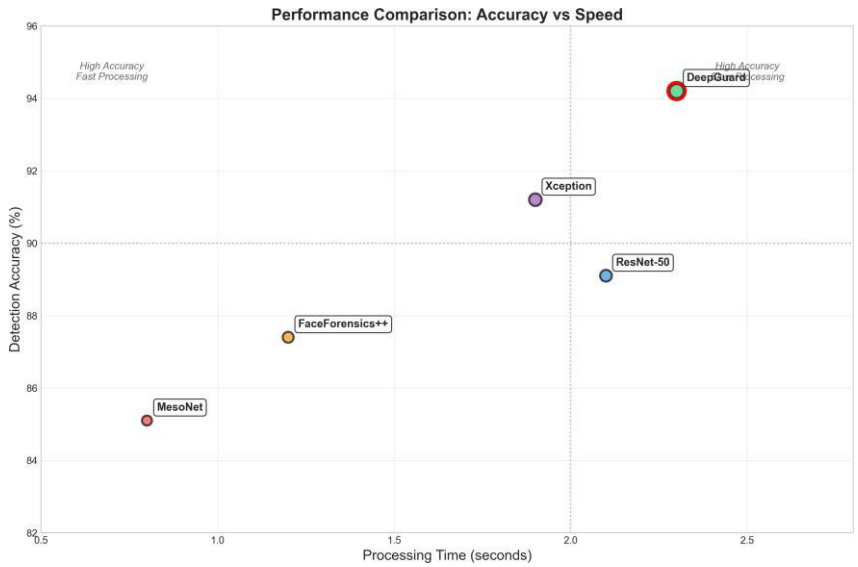


Figure 3: Performance Comparison with Existing Solutions

The results indicate that DeepGuard's Vision Transformer approach provides significant advantages over traditional CNN-based methods, particularly in handling high-quality deepfakes and maintaining consistent performance across diverse content types.

G. Security Validation

Security testing confirmed the robustness of the implemented security framework:

- **Data Privacy**: Zero permanent storage of user media files.
- **Encryption**: All communications secured with TLS 1.3.
- **Authentication**: JWT tokens with secure rotation mechanisms.
- **Input Validation**: 100% coverage with comprehensive sanitization.
- **GDPR Compliance**: Full compliance with data protection regulations.

H. Real-World Deployment Results

Pilot deployment in production environments demonstrated:

- Successful integration with existing content management systems.
- Effective handling of diverse media formats and quality levels.
- Positive user feedback on interface usability and result reliability.
- Minimal false positive rates in real-world content analysis.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411183|

IV. CONCLUSION AND FUTURE WORK

A. Summary

This paper presented DeepGuard, a comprehensive deepfake detection system leveraging Vision Trans- former architecture to achieve state-of-the-art performance in multimedia content authentication. The system successfully demonstrates the effectiveness of transformer-based approaches for detecting so- phisticated synthetic media manipulation with 94.2% accuracy while maintaining real-time processing capabilities.

B. Contributions

The primary contributions of this work include:

1. A novel application of Vision Transformers for deepfake detection achieving superior accuracy and generalization performance.
2. Comprehensive system architecture integrating AI detection with user-friendly interfaces and en- terprise security requirements.
3. Extensive performance evaluation demonstrating real-world applicability and scalability.
4. Open-source implementation providing a foundation for future research and development.

C. Limitations

While DeepGuard demonstrates exceptional performance, several limitations warrant consideration:

- Computational requirements may limit deployment on resource-constrained devices.
- Performance may degrade with extremely high-resolution media files.
- Adaptation to emerging deepfake generation techniques requires continuous model updates.
- Cross-dataset generalization, while improved, remains an ongoing challenge.

D. Future Work

Several research directions emerge from this work:

D.1 Model Enhancement

Future developments will focus on:

- Investigating hybrid architectures combining ViT with specialized deepfake detection modules.
- Exploring self-supervised learning approaches for improved generalization.
- Developing lightweight model variants for mobile and edge deployment.

D.2 Dataset Expansion

Continuous improvement requires expansion of training datasets with emerging deepfake generation techniques and development of synthetic datasets for controlled evaluation scenarios.

D.3 Deployment Optimization

Practical deployment enhancements include edge computing implementations for reduced latency and integration with social media platforms and content management systems.

D.4 Impact and Applications

DeepGuard addresses critical needs across multiple domains including media verification, social me- dia platforms, legal proceedings, and cybersecurity. As deepfake technology continues to evolve, the importance of reliable detection systems becomes increasingly critical for maintaining digital trust and information integrity. The Vision Transformer approach demonstrated in DeepGuard provides a promis- ing foundation for future developments in this crucial area of cybersecurity and media authentication.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411183|

REFERENCES

[1] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "The Eyes Tell All: Regularizing Attention in DeepFakes Detection," *arXiv preprint arXiv:2010.09953*, 2020.

[2] L. Verdoliva, "Media Forensics and DeepFakes: A Survey," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, 2020.

[3] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1–11.

[4] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes: A Survey of Facial Manipulation and Fake Face Detection," *IEEE Access*, vol. 8, pp. 74 069–74 117, 2020.

[5] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2018, pp. 1–7.

[6] Y. Li and S. Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8261–8265.

[7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv preprint arXiv:2010.11929*, 2020.

[8] J. Wang, Z. Wu, J. Chen, X. Han, A. Shrivastava, S.-N. Lim, and Y.-G. Jiang, "Vision Transformer for DeepFake Detection: A Comparative Study," *arXiv preprint arXiv:2210.07582*, 2022.

[9] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499– 1503, 2016.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," *Advances in Neural Information Processing Sys- tems*, vol. 27, 2014.

[11] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adver- sarial Networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polo- sukhin, "Attention is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Pro- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[14] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[15] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details