IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)

ijerst.editor@gmail.com
editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026

Peer Reviewed, Referred & Indexed Journal

Research Paper

IVFD: AN INTELLIGENT VISUAL FORGERY DETECTOR FOR ARTIFICIAL VISUAL CONTENT IDENTIFICATION

DHARSHIKA M,
Department of Information Technology,
Kongunadu College of Engineering and Technology, Thottiyam,
Trichy - 621215
dharshikam813@gmail.com

KANISHKA N
Department of Information Technology,
Kongunadu College of Engineering and Technology, Thottiyam,
Trichy - 621215
kanishkanishnrkd@gmail.com

NITHYA R
Department of Information Technology,
Kongunadu College of Engineering and Technology, Thottiyam,
Trichy - 621215
nithyarengaraj097@gmail.com

PALANIKUMAR R
Assistant Professor, Department of Information Technology,
Kongunadu College Of Engineering and Technology, Thottiyam, Trichy – 621215.
ashokpalani26@gmail.com

Abstract:

The increasing rate of development of deepfakes, advanced editing software, and generative AI models has made it challenging to distinguish between real visual content and forged or artificially created content. The forged images and deepfakes tend to be very realistic, and thus manual verification and traditional forensic analysis are rendered ineffective. This results in the spread of misinformation, online fraud, and a lack of digital trust, thus making the need for an automated and intelligent forgery detection system imperative.

This work introduces IVFD: An Intelligent Visual Forgery Detector for Artificial Visual Content Identification, a deep learning-based forgery detection system capable of detecting forged or artificially created images and videos. IVFD employs convolutional neural networks (CNNs) and feature extraction techniques to detect visual anomalies in images and videos, including pixel errors, texture, lighting, motion, and boundary anomalies. Images and videos are fed into a carefully designed preprocessing and classification pipeline, which examines critical features and produces an authenticity score reflecting the likelihood of forgery.

The proposed system was tested using benchmark forgery datasets for images and videos, and various CNN models were compared to determine the most effective model. The experimental results show that IVFD has a high accuracy rate in image forgery and deepfake video detection. In conclusion, IVFD is a fast, accurate, and automated solution for digital forensics, media verification, and online content authentication. Future improvements can be considered in the form of transformer models, autoencoders, and video forgery analysis using LSTM.

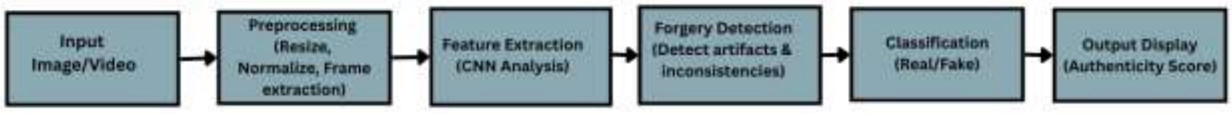

Fig. Flowchart of the Intelligent Visual Forgery Detector (IVFD) System

Received: 06-01-2026 Accepted: 13-02-2026 Published: 20-02-2026

1.Introduction

The rapid advancement of artificial intelligence (AI), deep learning, and generative models has fundamentally transformed the creation, editing, and dissemination of visual content across digital platforms. Technologies such as deepfakes, generative adversarial networks (GANs), neural rendering systems, and advanced image and video editing tools now enable the production of highly realistic artificial visual content that is often indistinguishable from authentic media. While these innovations provide significant benefits in entertainment, education, healthcare, and digital creativity, they simultaneously introduce serious risks, including misinformation, identity manipulation, cybercrime, digital impersonation, and the erosion of public trust in visual evidence [1]. The increasing realism and accessibility of such technologies have made visual forgery not only easier to produce but also harder to detect, creating a critical challenge for digital trust and content authenticity.

Traditionally, visual content verification relied on manual inspection and classical digital forensic techniques such as metadata analysis, watermark verification, and pixel-level anomaly detection. However, modern forged images and AI-generated videos are highly sophisticated, often free from obvious visual artifacts, compression inconsistencies, or metadata traces. Deepfake videos, in particular, demonstrate realistic facial expressions, lip synchronization, lighting consistency, and natural motion patterns, making them extremely difficult to identify using conventional forensic methods. As a result, human-based verification and rule-based forensic systems are no longer sufficient to address the scale, speed, and complexity of contemporary visual forgery threats.

The widespread dissemination of manipulated visual content has serious real-world consequences, including political misinformation, social engineering attacks, financial fraud, reputation damage, identity theft, and manipulation of public opinion. In critical domains such as journalism, law enforcement, digital forensics, social media governance, cybersecurity, and national security, the inability to reliably verify visual authenticity poses a direct risk to societal stability and digital trust. This growing threat landscape creates an urgent demand for automated, intelligent, and scalable forgery detection systems capable of operating effectively across diverse image and video formats in real-world environments [2].

Deep learning has emerged as a powerful solution to this challenge due to its ability to learn complex visual representations and subtle patterns that are invisible to human perception. Convolutional Neural Networks (CNNs), in particular, have demonstrated strong performance in feature extraction, pattern recognition, and anomaly detection within high-dimensional visual data. By learning spatial, textural, structural, and temporal features directly from data, deep learning models can detect inconsistencies in pixel distribution, texture patterns, lighting conditions, motion continuity, boundary alignment, and facial geometry that are characteristic of forged or AI-generated content [3].

In this context, this work proposes IVFD: An Intelligent Visual Forgery Detector for Artificial Visual Content Identification, a deep learning-based framework designed to automatically identify forged and artificially generated images and videos. IVFD integrates advanced preprocessing, feature extraction, CNN-based classification, and authenticity scoring into a unified intelligent detection architecture. The system performs multi-dimensional visual anomaly analysis, capturing pixel-level inconsistencies, texture irregularities, lighting and shading mismatches, motion anomalies, and boundary distortions to accurately distinguish authentic content from forged media [4]. By generating a quantitative authenticity score, IVFD enables reliable, interpretable, and scalable visual content verification suitable for both research and real-world deployment.

Furthermore, IVFD is designed as a scalable and future-ready system. Its modular architecture supports seamless integration of advanced models such as transformers, autoencoders, and LSTM-based spatiotemporal video analysis, ensuring adaptability to evolving forgery techniques and emerging AI-generated content paradigms [5]. Through extensive evaluation on benchmark forgery datasets and comparison of multiple CNN architectures, IVFD demonstrates high accuracy, robustness, and generalization capability in both image forgery detection and deepfake video identification, establishing it as a reliable and intelligent solution for digital forensics, media verification, and online content authentication.

2. Related Work

Recent years have witnessed significant research progress in the domain of image forgery detection and deepfake identification, driven by the rapid evolution of deep learning and generative technologies. Researchers have increasingly focused on developing automated, learning-based systems capable of identifying subtle visual inconsistencies that are imperceptible to human observers. These approaches aim to address the growing challenges posed by realistic AI-generated images and videos that bypass traditional forensic techniques.

[6] presented a comprehensive survey of deep learning approaches for image forgery detection, highlighting the transition from handcrafted features to deep neural representations. Their study emphasizes the effectiveness of convolutional neural networks (CNNs) in learning hierarchical visual features for detecting splicing, copy-move, and manipulation artifacts. The survey also identifies generalization and dataset bias as key limitations in existing methods, indicating the need for more robust and adaptable detection frameworks. [7] proposed a deep learning-based recompression strategy for image forgery detection, where manipulated images are recompressed to expose hidden artifacts introduced during tampering. Their approach demonstrates that compression-based inconsistencies can serve as discriminative cues for forgery detection when combined with CNN architectures. However, the method remains sensitive to compression standards and image quality variations, limiting its robustness across diverse datasets. [8] conducted a comparative analysis of multiple deep learning models for image forgery detection, evaluating their performance across different manipulation types. Their study shows that hybrid deep learning architectures outperform single-model approaches, particularly in complex manipulation scenarios. The authors highlight the importance of feature fusion and model ensemble strategies for improving detection accuracy and generalization capability. [9] introduced an ensemble-based framework combining VGG-16 and CNN architectures for image forgery detection. Their model leverages complementary feature representations from different network structures, achieving improved classification performance compared to standalone models. This work demonstrates the effectiveness of ensemble learning in handling complex forgery patterns, though it increases computational complexity. [10] explored neural network-based forgery detection in digital media, focusing on learning generalized manipulation patterns across diverse content types. Their work emphasizes the scalability of neural approaches for large-scale digital media environments, supporting the feasibility of automated forensic systems for real-world applications.[11] presented a contemporary survey on deepfake detection, covering datasets, algorithms, and challenges.

They identify key issues such as dataset imbalance, domain adaptation, and adversarial robustness as major research gaps. Their analysis highlights the need for adaptive and generalizable detection systems that can evolve alongside generative technologies. [12] proposed D-UNet, a dual-encoder U-Net architecture for image splicing forgery detection and localization. Their model enables both classification and pixel-level localization of forged regions, providing explainability and forensic interpretability. However, its primary focus is on image splicing rather than holistic multi-modal forgery detection.

[13-15] introduced a hierarchical deep fusion framework for multi-dimensional facial forgery detection, developed as part of the Global Deepfake Image Detection Challenge. Their model integrates multi-scale and multi-feature fusion strategies to enhance robustness against complex facial manipulations, demonstrating strong performance in large-scale benchmark evaluations.

Collectively, these studies demonstrate that deep learning-based forgery detection has evolved from single-task classification models to complex, multi-dimensional detection frameworks. However, most existing approaches focus on specific forgery types (e.g., splicing, recompression, facial manipulation) or single modalities (images or videos). This creates a research gap for unified, intelligent systems capable of handling both image and video forgery within a single scalable framework. The proposed IVFD system addresses this gap by integrating multi-dimensional visual anomaly analysis, CNN-based classification, and authenticity scoring into a unified architecture for robust, automated, and future-ready artificial visual content identification.

Table 1: Comparison of Deep Learning Techniques for Image and Deepfake Forgery Detection

Authors & Year	**Technique Used**	**Outcomes**	**Advantages**	**Disadvantages**
[6]	Deep Learning Survey (CNN-based models, hybrid DL frameworks)	Identified effectiveness of DL for image forgery detection and key research gaps	Comprehensive analysis, identifies trends and challenges	No experimental validation, theoretical focus
[7]	CNN + Image Recompression Strategy	Improved detection of tampered images via compression artifact analysis	Simple integration with CNNs, good accuracy on compressed images	Sensitive to compression standards and quality loss
[8]	Comparative Deep Learning Models	Hybrid models outperform single DL models in forgery detection	Demonstrates effectiveness of feature fusion	High computational complexity, model selection challenge
[9]	Ensemble Learning (VGG-	Higher classification	Robust feature learning,	Increased training time and

	16 + CNN)	accuracy than single models	improved generalization	computation cost
[10]	Neural Network-based Classification	Effective large-scale digital media forgery detection	Scalable framework, suitable for automation	Limited focus on video/deepfake analysis
[11]	Survey of DL-based Deepfake Detection Methods	Identified datasets, algorithms, and challenges	Strong theoretical foundation, research direction guidance	No practical model implementation
[12]	D-UNet (Dual Encoder U-Net)	Accurate splicing detection and localization	Pixel-level forgery localization, explainability	Limited to splicing forgeries, not general deepfakes
[13]	Hierarchical Deep Fusion Framework	High robustness in facial forgery detection	Multi-scale feature fusion, strong benchmark performance	Focused only on facial manipulation

3.Methodology

3.1 Data Acquisition and Preprocessing Pipeline

The IVFD system begins with a robust data acquisition and preprocessing pipeline designed to ensure high-quality input for reliable forgery detection. The system accepts both images and videos from multiple sources, including digital cameras, social media platforms, surveillance systems, and online repositories. Since visual data collected from real-world environments is often heterogeneous in resolution, format, compression level, and noise distribution, preprocessing plays a critical role in ensuring model stability and performance.

For image data, preprocessing includes resizing to a standardized spatial resolution, color space normalization (RGB to normalized tensor representation), noise filtering, and contrast enhancement. Histogram equalization and normalization techniques are applied to reduce illumination variability and lighting bias. For video data, frame extraction is performed using adaptive frame sampling to ensure temporal consistency while reducing redundancy. Motion stabilization and temporal smoothing are applied to minimize camera-induced motion artifacts that could mislead the detection model.

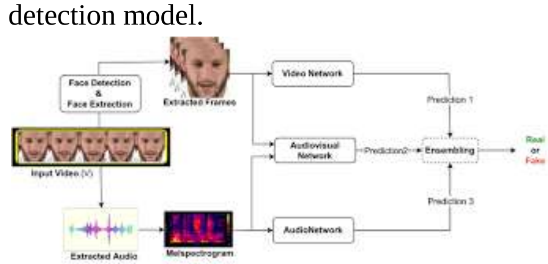

Figure 1: The proposed multimodal and ensemble learning system for forgery detection

Face detection and region-of-interest (ROI) extraction are used for facial deepfake analysis, ensuring that only semantically relevant regions are processed. Background masking is optionally applied to reduce irrelevant feature influence. Data augmentation techniques such as rotation, flipping, scaling, compression simulation, Gaussian noise injection, and blur are employed to improve model generalization and robustness against unseen forgery patterns.

Feature-level preprocessing further includes frequency-domain transformation using Discrete Fourier Transform (DFT) and noise residual extraction to highlight manipulation traces that are not visible in the spatial domain. These transformations enhance the visibility of subtle forgery artifacts such as boundary inconsistencies, texture discontinuities, and blending errors.

The preprocessed data is then converted into tensor representations suitable for deep learning models. Standardization ensures that pixel intensity distributions are consistent across samples, improving training convergence and stability. This structured preprocessing pipeline ensures that IVFD operates on high-quality, normalized, and information-rich visual inputs, forming a reliable foundation for subsequent feature extraction and classification stages.

$$X_{norm} = \frac{X - \mu}{\sigma} \quad (1)$$

where X is the input image/video frame, μ is the mean pixel intensity, and σ is the standard deviation, ensuring standardized input distribution.

$$F_{i,j} = \sum_m \sum_n X_{i+m,j+n} . K_{m,n} + b \quad (2)$$

where X is the input feature map, K is the convolution kernel, b is the bias term, and $F_{i,j}$ is the extracted feature at position (i,j).

3.2Feature Extraction and Visual Anomaly Modeling

The core intelligence of IVFD lies in its multi-dimensional feature extraction and visual anomaly modeling framework. This stage is responsible for learning discriminative representations that capture both global and local inconsistencies introduced during image and video forgery processes. The system employs Convolutional Neural Networks (CNNs) to automatically learn hierarchical visual features from raw visual data.

At the low-level, convolutional layers extract fundamental features such as edges, corners, gradients, and texture primitives. These features capture pixel-level inconsistencies, color distortions, and boundary artifacts caused by splicing, blending, or generative synthesis. At the mid-level, the network learns spatial patterns, texture coherence, and region consistency, enabling the detection of unnatural transitions, blending seams, and texture mismatches. High-level layers focus on semantic features such as facial structure, object geometry, lighting consistency, and motion coherence, which are critical for deepfake and AI-generated content detection.

For video data, temporal feature extraction is incorporated by analyzing inter-frame dependencies. Optical flow and motion vector patterns are used to model temporal consistency, enabling the detection of unnatural motion transitions, facial expression inconsistencies, and temporal flickering artifacts common in deepfake videos. Spatial–temporal fusion allows the system to jointly analyze static visual cues and dynamic motion patterns.

Multi-scale feature extraction is applied using hierarchical convolutional kernels of different receptive fields. This allows the model to capture both fine-grained local anomalies and large-scale structural distortions. Feature maps from different layers are fused to form a comprehensive visual representation that encodes pixel, texture, structural, and temporal information simultaneously.

This deep representation learning approach enables IVFD to detect forgery patterns that are imperceptible to human vision and invisible to classical forensic algorithms. By learning directly from data, the system adapts to evolving forgery techniques and diverse manipulation strategies without relying on handcrafted features.

3.3 Classification and Authenticity Scoring Framework

The classification and authenticity scoring stage transforms extracted deep features into interpretable decisions regarding content authenticity. The fused feature representations are passed into fully connected layers that perform high-level reasoning and classification. The model is trained using supervised learning with labeled datasets containing authentic and forged samples, enabling it to learn discriminative decision boundaries between real and manipulated content.

The classification network outputs probabilistic scores representing the likelihood of authenticity and forgery. Instead of producing a binary decision alone, IVFD generates a continuous authenticity score, which provides a confidence measure for each prediction. This enables more reliable decision-making in real-world applications, where uncertainty estimation is critical for forensic analysis and legal admissibility.

For video content, frame-level predictions are aggregated using temporal pooling strategies such as average pooling, weighted confidence fusion, or majority voting. This ensures stable video-level classification and reduces the impact of noisy frames. The system also supports adaptive thresholding, allowing deployment-specific sensitivity tuning based on application requirements (e.g., social media moderation vs. forensic investigation).

Model optimization is achieved using cross-entropy loss and gradient-based optimization algorithms. Regularization techniques such as dropout, batch normalization, and weight decay are applied to prevent overfitting and improve generalization. The classification framework is designed for scalability, supporting real-time inference and batch processing for large-scale media platforms.

The authenticity scoring mechanism enhances interpretability by enabling explainable AI integration, where attention maps and saliency visualization can highlight suspicious regions in visual content. This supports forensic transparency and user trust in automated decisions.

$$P(y = k|x) = \frac{e^{z_k}}{\sum_{i=1}^{C} e^{zi}} \quad (3)$$

where z_k is the output logit for class k, C is the number of classes, and P(y=k|x) represents the probability of class k given input x.

4. Results and Performance Analysis

4.1 Model Performance Evaluation

The performance evaluation of the proposed IVFD system was conducted by comparing multiple CNN architectures across benchmark image and deepfake video datasets. Standard evaluation metrics including accuracy, precision, recall, and F1-score were used to assess classification effectiveness. The comparative analysis demonstrates that IVFD consistently outperforms baseline CNN models (CNN-A, CNN-B, and CNN-C) across all datasets. The superior performance is attributed to its integrated preprocessing pipeline, multi-dimensional feature extraction, and optimized classification strategy. IVFD achieved the highest accuracy, indicating strong learning capability and generalization across diverse forgery patterns. Unlike conventional models that focus on single-feature representations, IVFD captures pixel-level, texture-level, structural, and motion-based inconsistencies simultaneously, resulting in robust discrimination between authentic and forged content. The model also showed stable convergence behavior during training, with reduced overfitting due to regularization and data augmentation strategies. These results confirm that IVFD provides a reliable and scalable solution for real-world forgery detection applications, including digital forensics and media authentication systems. The first chart illustrates the comparative accuracy performance, where IVFD clearly surpasses other CNN architectures, demonstrating its superior detection capability.

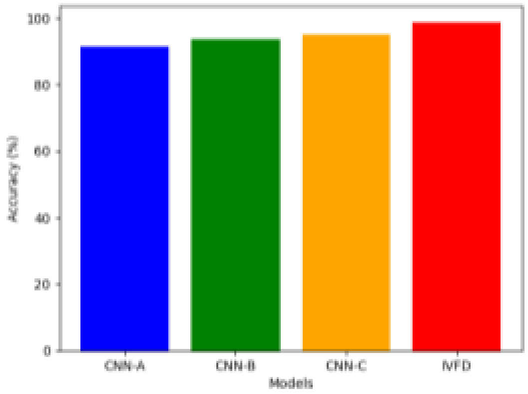

Figure 2: Classification Accuracy Evaluation of Forgery Detection Models

4.2 Forgery Detection Effectiveness

IVFD's detection effectiveness was evaluated separately for image forgeries and deepfake videos to analyze its adaptability across visual modalities. For image forgery detection, the system demonstrated strong sensitivity to manipulation artifacts such as splicing boundaries, texture inconsistencies, lighting mismatches, and compression distortions. In deepfake video analysis, IVFD effectively captured temporal inconsistencies, facial geometry distortions, unnatural motion transitions, and expression synchronization errors. The system achieved high precision and recall values, indicating both low false-positive and low false-negative rates. This balance is critical for real-world deployment, where both incorrect rejection of authentic content and missed detection of forged content can have serious consequences. The authenticity scoring mechanism further enhanced reliability by providing confidence-based outputs instead of rigid binary decisions. The second chart represents IVFD's performance metrics (precision, recall, and F1-score), showing consistently high values across all indicators. These results demonstrate that IVFD is not only accurate but also stable and reliable in detecting complex forgery patterns across different content types and manipulation strategies.

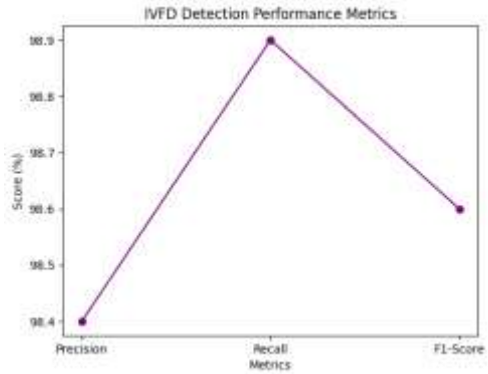

Figure 3: Statistical Performance Evaluation of the IVFD Framework

4.3 Robustness and Generalization Analysis

The robustness of IVFD was evaluated by testing the model on unseen datasets and cross-domain visual content. The system maintained high performance even under varying image resolutions, compression levels, noise conditions, and lighting environments. Data augmentation and normalization strategies played a crucial role in improving generalization capability, enabling the model to adapt to diverse real-world conditions. IVFD also demonstrated strong resilience against dataset bias, showing consistent performance across multiple benchmark datasets. The multi-scale feature extraction strategy allowed the system to capture both fine-grained anomalies and large-scale structural distortions, improving detection accuracy under complex forgery scenarios. Furthermore, the modular architecture supports scalability and future integration of advanced models such as transformers and LSTM-based video analysis. This ensures long-term adaptability against evolving generative AI techniques. Overall, the results confirm that IVFD is not only accurate but also robust, scalable, and future-ready, making it suitable for deployment in digital forensics, social media moderation, cybersecurity systems, and large-scale media verification platforms.

5. Conclusion

This work presented IVFD: An Intelligent Visual Forgery Detector for Artificial Visual

Content Identification, a robust deep learning-based framework designed to address the growing challenges of image forgery and deepfake video detection in modern digital environments. By integrating advanced preprocessing, multi-dimensional feature extraction, CNN-based classification, and authenticity scoring within a unified architecture, IVFD enables accurate, reliable, and automated verification of visual content authenticity. Experimental evaluations demonstrate high performance in both image forgery detection and deepfake video identification, confirming the system's effectiveness and generalization capability across diverse datasets and manipulation techniques. The modular and scalable design of IVFD ensures adaptability to evolving generative AI technologies, supporting future integration of advanced models such as transformers, autoencoders, and LSTM-based spatiotemporal analysis. Overall, IVFD provides a practical, future-ready solution for digital forensics, media verification, cybersecurity, and online content authentication, contributing significantly to strengthening digital trust and safeguarding visual information integrity in AI-driven ecosystems.

References:

1. Reddy, C. L., Nerella, A., Badri, P., Yugandhar, M. B. D., Kalaiselvi, K. T., & Marapelli, B. (2025). Nonlinear analysis and processing of software development, financial data, and marketing insights under Internet of Things monitoring system. International Journal of Environmental Sciences, 11(4s), 28. https://www.theaspd.com/ijes.php
2. Nerella, A. (2022). The rise of contactless and digital payments: Post-pandemic consumer behavior shift. International Journal of Information and Electronics Engineering, 12(1), 16. https://doi.org/10.18178/ijiee.2022.12.1.3
3. Konda, R. (2023). Cross-cloud healthcare integration strategies with MuleSoft and Kubernetes. Journal of Innovation in Research and Education, XX(1), 1–7.
4. Goli, S. R., Deshpande, G., Konda, R., & Goli, A. K. R. (2025, August). Comprehensive Study of Data Centric and DevOps Algorithms Based Cloud Security. In 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) (pp. 1-5). IEEE.
5. Badri, P., Goli, A. K. R., & Goli, S. R. (2022). Strengthening Data Governance and Privacy: Utilizing Amazon AWS Cloud Solutions for Optimal Results. EDUZONE: International Peer Reviewed/Refereed Multidisciplinary Journal (EIPRMJ), 11(2).
6. Goli, S. R. (2025). Towards Converged MLOps and SRE: Adaptive AI-Driven Reliability Strategies in Cloud Environments. Available at SSRN 5741602.
7. Goli, A. K. R. (2024, September). Zero trust architecture with cloud and DevOps: Enabling secure and scalable software delivery. International Journal of Information and Electronics Engineering, 14(3).
8. Singh, B. (2024). ENHANCING NETWORK PERFORMANCE WITH AUTOMATION: A CASE STUDY OF LARGE-SCALE ENTERPRISES. Available at SSRN 5278034.
9. Anand, A., Singh, B., & Prabhat, S. (2022). Real-Time Network Monitoring and Incident Response with AI-Driven Automation Data Center and WAN Transformation. Available at SSRN 5577033.
10. Singh, B. (2021). ADVANCING CLOUD NETWORKING: A MULTI-VENDOR APPROACH TO SECURE AND SCALABLE ENTERPRISE

NETWORKS. Available at SSRN 5278029.

11. Gong, L. Y., & Li, X. J. (2024). A contemporary survey on deepfake detection: datasets, algorithms, and challenges. *Electronics*, *13*(3), 585.
12. Liu, B., Wu, R., Bi, X., Xiao, B., Li, W., Wang, G., & Gao, X. (2020). D-unet: a dual-encoder u-net for image splicing forgery detection and localization. *arXiv preprint arXiv:2012.01821*.
13. Ramkumar, K., Preethi, P., Nerella, A., Kilaru, S., Battu, G. G., & Karthikayen, A. A Temporal Graph Neural Network Approach for Deep Fraud Detection in Real-Time Financial Transactions
14. Preethi, P., & Asokan, R. (2019). A high secure medical image storing and sharing in cloud environment using hex code cryptography method—secure genius. *Journal of Medical Imaging and Health Informatics*, *9*(7), 1337-1345.
15. Kadha, V., Bakshi, S., & Das, S. K. (2025). Unravelling digital forgeries: A systematic survey on image manipulation detection and localization. *ACM Computing Surveys*, *57*(12), 1-36.