International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Evaluating Non-Reference Image Quality Metrics for AI-Generated Images

B. Midhun Kannan[#1], Ms. K. Dhanalakshmi[*2]

[1]*Department of Artificial Intelligence and Data Science, Ramco Institute of Technology,Rajapalayam,India-626117*

[1]b.midhunkannan.2004@gmail.com

[2]*Department of Artificial Intelligence and Data Science, Ramco Institute of Technology,Rajapalayam,India-626117*

[2]ishwaryaranju1999@gmail.com

https://doi.org/10.55041/ijsmt.v2i3.276

Cite this Article: Kannan, B. M. & Dhanalakshmi, K. (2026). Evaluating Non-Reference Image Quality Metrics for AI-Generated Images. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.276

Abstract

The exponential growth of AI-generated images produced by Generative Adversarial Networks (GANs) and diffusion-based models has created an urgent need for reliable image quality evaluation systems. Traditional full-reference metrics such as PSNR and SSIM require pristine ground-truth images, which are unavailable for generative outputs. This limitation necessitates robust Non- Reference Image Quality Assessment (NR-IQA) frameworks.

This paper presents VisionIQ Pro, a hybrid multi- tiered visual quality assessment system that integrates handcrafted statistical signal metrics with deep perceptual Natural Scene Statistics (NSS) metrics such as BRISQUE and NIQE. The system constructs a fused Image Quality Vector and uses a Random Forest regression model trained on KonIQ-10k and LIVE datasets to predict Mean Opinion Score (MOS). Additionally, the framework incorporates Explainable AI through Sobel-based sharpness density mapping and actionable feedback logic.

Experimental analysis demonstrates strong correlation (PLCC > 0.85) with human subjective perception. The proposed system serves as a scalable, interpretable, and production-ready solution for automated auditing of AI-generated and real-world images.

Keywords: Non-Reference Image Quality Assessment, AI-Generated Images, BRISQUE, NIQE, Random Forest, Explainable AI, VisionIQ **Pro**

INTRODUCTION

The rapid advancement of artificial intelligence has revolutionized image generation. Modern generative models such as GANs and diffusion models are capable of producing highly realistic visual content. However, evaluating the perceptual and technical quality of these images remains a major challenge due to the absence of reference images.

Traditional full-reference metrics like PSNR and SSIM compare distorted images against ground truth. In AI-generated images, no reference exists, rendering these metrics ineffective. Non-Reference Image Quality Metrics (NR-IQMs) such as BRISQUE and NIQE evaluate images based on natural scene statistics, but they were originally designed for natural distortions rather than synthetic artifacts.

Furthermore, existing systems suffer from:

- Lack of hybrid feature integration
- Poor explainability
- Absence of actionable recommendations
- Inability to handle AI-specific distortions

To overcome these limitations, this research proposes **VisionIQ Pro**, an integrated dual-path architecture that fuses handcrafted signal metrics with neural perceptual features and machine learning inference.

BACKGROUND

Image Quality Assessment (IQA) is an important area in computer vision that focuses on evaluating the visual quality of digital images. Traditional IQA methods were mainly developed for natural images captured using cameras. These methods are classified into Full- Reference (FR), Reduced-Reference (RR), and No- Reference (NR) approaches.

Full-reference metrics such as PSNR and SSIM require a reference image for comparison. However, in AI- generated images, no original ground truth image exists. Therefore, full-reference methods cannot be applied to generative models.

Non-Reference Image Quality Assessment (NR-IQA) methods estimate image quality without using a reference image. Popular NR-IQA techniques such as

BRISQUE and NIQE use Natural Scene Statistics (NSS) to measure distortions like blur, noise, and compression artifacts.

With the growth of AI-generated images produced by GANs and diffusion models, there is a need for reliable NR-IQA systems. AI-generated images may contain synthetic artifacts and unnatural patterns that traditional methods may not fully capture. Hence, advanced hybrid frameworks that combine statistical features and perceptual metrics are required for accurate evaluation.

Objectives

The primary objective of this project is to design and implement an AI-driven non-reference image quality assessment framework capable of evaluating AI- generated images without ground truth references.

Specific objectives include:

- Develop a hybrid feature extraction system combining statistical and neural metrics
- Implement BRISQUE and NIQE perceptual models
- Compute signal-based metrics such as sharpness, contrast, noise, brightness, and balance
- Train a regression model for MOS prediction
- Design an explainable heatmap visualization module
- Provide actionable recommendations based on metric thresholds
- Integrate the system into a Flask-based real-time dashboard
- Evaluate system performance using standard correlation metrics

Contribution

The major contributions of VisionIQ Pro include:

- Hybrid multi-parametric NR-IQA framework
- Fusion of handcrafted and neural perceptual metrics
- Machine learning-based MOS prediction
- Explainable AI module using Sobel gradient mapping
- Actionable AI recommendation engine
- Real-time web-based dashboard for single and batch analysis
- Dataset curation support tool for AI engineers

RELATED WORKS

Image Quality Assessment (IQA) has been extensively studied in the field of image processing and computer vision. Existing approaches are broadly classified into Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) methods.

Early IQA systems relied heavily on Full-Reference metrics such as PSNR and SSIM, which compare a distorted image with a pristine reference image. While these approaches provide reliable distortion measurement, they fail in real-world scenarios where reference images are unavailable, particularly for AI-generated content.

To overcome this limitation, Non-Reference Image Quality Assessment (NR-IQA) methods were introduced. Traditional NR approaches are based on Natural Scene Statistics (NSS), which model statistical regularities present in natural images. Metrics such as BRISQUE, NIQE, and ILNIQE estimate image quality by measuring deviations from expected natural image distributions. BRISQUE utilizes Mean Subtracted Contrast Normalized (MSCN) coefficients, whereas NIQE computes statistical distance from a learned natural image model. However, these methods were primarily trained on natural photographic distortions and often struggle to accurately evaluate AI-generated images.

With the rapid growth of Generative Adversarial Networks (GANs) and diffusion-based image synthesis models, researchers observed that AI-generated images exhibit unique distortion characteristics. Many synthetic images appear visually sharp and clean but contain unnatural textures, inconsistent high-frequency details, and statistical irregularities that conventional NR-IQA metrics may not fully capture.

Recent research efforts have explored deep learning- based quality prediction models, including Convolutional Neural Network (CNN) regressors and transformer-based architectures, to directly predict Mean Opinion Scores (MOS). Additionally, CLIP- based alignment metrics have been proposed to measure semantic consistency between text prompts and generated images. While these approaches demonstrate promising accuracy, they often require large annotated datasets and lack interpretability.

Hybrid frameworks that combine handcrafted signal-processing features with machine learning regression models have shown improved robustness and better generalization across diverse image domains. Such approaches leverage the strengths of classical statistical modeling and data-driven learning while maintaining interpretability.

Building upon these developments, **VisionIQ Pro** adopts a hybrid feature integration strategy that combines NSS-based perceptual models, signal processing metrics, machine learning regression, explainable visualization, and actionable rule-based intelligence. This integrated approach addresses the limitations of standalone methods and provides a balanced solution for accurate, interpretable, and scalable evaluation of AI-generated images.

SYSTEM ARCHITECTURE

VisionIQ Pro follows a modular and scalable multi- layer architecture designed to support automated, explainable, and real-time Non-Reference Image Quality Assessment (NR-IQA). The system consists of four major layers, each responsible for a specific stage

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

in the processing pipeline.

A. Acquisition Layer

The Acquisition Layer manages data ingestion and preprocessing. The system supports:

- Single image upload
- Batch image processing
- Video frame extraction and sampling

For video inputs, frames are sampled at predefined intervals (e.g., 1 frame per second) to reduce computational complexity while maintaining temporal consistency. Basic preprocessing operations such as resizing, grayscale conversion (for certain metrics), and normalization are performed to standardize inputs before feature extraction.

This layer ensures flexibility and scalability for both small-scale experiments and large dataset evaluation.

B. Feature Extraction Layer (Dual-Path Architecture) The core strength of VisionIQ Pro lies in its dual-path feature extraction mechanism, which combines handcrafted signal-processing metrics with neural perceptual quality measures.

Path A -- Statistical (Handcrafted) Metrics

This path extracts physically interpretable image characteristics using classical image processing techniques:

- **Sharpness** -- Measured using Laplacian Variance to detect edge intensity and focus quality.
- **Contrast** -- Computed as the standard deviation of pixel intensities.
- **Noise** -- Estimated using wavelet-based sigma estimation from high-frequency sub-bands.
- **Brightness** -- Mean pixel intensity to detect underexposure or overexposure.
- **Luminance Balance** -- Difference between left and right image halves to measure illumination symmetry.
- **Color Harmony** -- Statistical variation among RGB channel means to identify color cast issues.

These handcrafted metrics provide interpretable technical insights into image quality and enable actionable recommendations.

Path B -- Neural Perceptual Metrics

This path computes perceptual quality indicators based on Natural Scene Statistics (NSS):

- **BRISQUE** (Blind/Referenceless Image Spatial Quality Evaluator)
- **NIQE** (Natural Image Quality Evaluator)

BRISQUE measures deviation from natural scene statistics using MSCN coefficients and trained distortion models. NIQE calculates the statistical distance between the input image and a natural image model without requiring human opinion scores.

These neural metrics capture perceptual distortions that may not be fully represented by basic signal-based features, especially in AI-generated images.

C. Inference Layer

The extracted features from both paths are fused into a unified Image Quality Vector: V_IQ = H U N

Where H represents handcrafted statistical metrics and N represents neural perceptual metrics.

The combined feature vector is normalized using standard scaling techniques and fed into a Random Forest Regression model. The model predicts the Mean Opinion Score (MOS), representing perceived image quality.

Random Forest is selected for its ability to model non- linear feature interactions, reduce overfitting, and maintain strong generalization performance across diverse image types.

D. UI & Reporting Layer

The final layer provides visualization, interaction, and reporting capabilities through a Flask-based web dashboard.

Key features include:

- Real-time quality score display
- Radar-based "Image Perceptual DNA" visualization
- Sharpness heatmap for explainability
- Interactive charts for metric distribution
- Batch processing with sorting and filtering
- Automated PDF/CSV report generation

This layer transforms raw numerical outputs into intuitive visual analytics, enabling users to interpret results efficiently and make informed decisions.

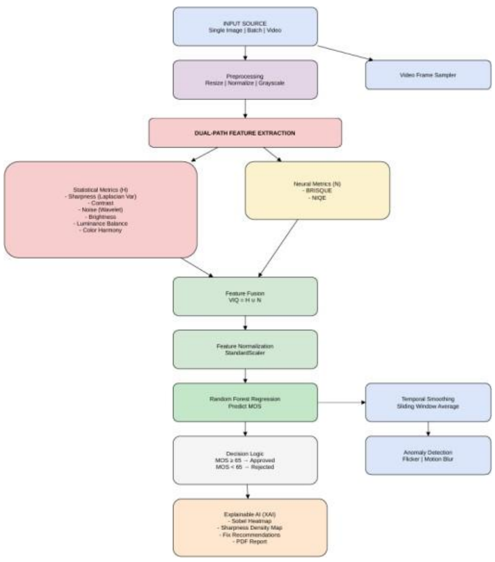

Fig. 1. Complete workflow of VisionIQ Pro

METHODOLOGY

The methodology consists of five stages:

1. Image Acquisition: User uploads image via dashboard.

2. Preprocessing: Resize, Grayscale conversion, Normalization.
3. Feature Extraction: Sharpness: S = Var(V2I); Contrast: C = sqrt(1/N * Sum(Pi - mean)^2); Noise: sigma_n = Median(|HH|)/0.6745; Luminance Balance: B_bal = |mean(I_L) - mean(I_R)|; BRISQUE MSCN: I_hat(i,j) = (I(i,j) - mu(i,j)) / (sigma(i,j) + C); NIQE: Mahalanobis distance to natural model.
4. Feature Normalization: StandardScaler applied.
5. MOS Prediction: Random Forest Regressor predicts quality score.

PROBLEM RECOMMENDATION

The rapid growth of AI-generated images has created new challenges in automated image quality evaluation. Traditional Full-Reference (FR) metrics such as PSNR and SSIM require a pristine reference image, which is not available for generative models. As a result, these methods are unsuitable for modern AI-driven image synthesis systems.

Existing Non-Reference Image Quality Assessment (NR-IQA) models such as BRISQUE and NIQE were primarily developed for natural photographic distortions like blur, compression artifacts, and noise. However, AI-generated images often exhibit different characteristics, including synthetic textures, unnatural high-frequency patterns, and statistical inconsistencies that are not effectively captured by conventional models.

Furthermore, many deep learning-based quality prediction approaches function as black-box models. Although they may achieve high prediction accuracy, they lack interpretability and do not provide actionable insights for improving image quality.

Therefore, there is a strong need for a hybrid, explainable, and scalable image quality assessment framework that:

- Operates without reference images
- Detects AI-specific artifacts
- Combines statistical and perceptual features
- Provides interpretable visual feedback
- Supports large-scale dataset curation

VisionIQ Pro is proposed to address these identified gaps by integrating multi-parametric feature extraction, machine learning regression, and explainable visualization into a unified evaluation system.

SMART LEARNING ENVIRONMENT

VisionIQ Pro operates within a smart digital learning and intelligent visual analytics environment where image quality evaluation is automated, data-driven, and interactive. The system integrates machine learning, statistical feature extraction, and visual analytics to support intelligent decision-making and practical understanding of image quality assessment.

The framework enables students, researchers, and developers to analyze AI-generated images in real time through an interactive dashboard. It provides real-time quality scoring, displays a radar-based "Image Perceptual DNA," generates sharpness heatmaps for spatial analysis, and offers AI-based actionable recommendations. Additionally, it supports batch dataset curation for large-scale image evaluation.

By combining automation, visualization, and explainable outputs, VisionIQ Pro promotes experiential learning and transforms traditional quality scoring into an interactive decision-support system within an applied research and educational setting.

EDUCATIONAL CONTEXT AND DATA

A. Educational Context

The proposed system is developed as part of an academic research project in the field of Computer Vision and Artificial Intelligence. It supports practical learning in subjects such as Digital Image Processing, Machine Learning, and Data Analytics.

Students can use the platform to understand Non-Reference Image Quality Assessment (NR-IQA), feature extraction methods, regression modeling, and performance evaluation techniques. The project bridges theoretical knowledge with real-world AI applications.

B. Data

The system is trained and evaluated using standard image quality datasets such as KonIQ-10k and LIVE IQA datasets. These datasets contain images with corresponding Mean Opinion Scores (MOS) collected from human evaluators.

For AI-generated image evaluation, sample outputs from generative models are also analyzed. Extracted features are normalized and used to train a Random Forest regression model for quality prediction.

C. Datasets

The system is trained and evaluated using:

- KonIQ-10k
- LIVE IQA
- TID2013
- AGIQA-1K
- AGIQA-3K

Each dataset contains Mean Opinion Score (MOS) labels derived from human subjective ratings.

Dataset preprocessing includes: Cleaning, Normalization, Feature scaling, Train-test split.

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

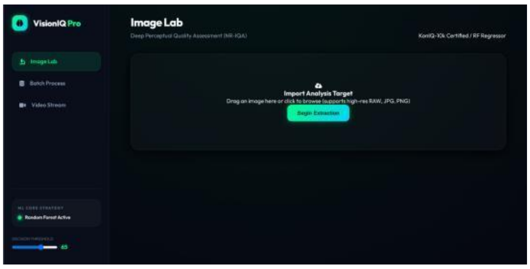

Fig. 2. VisionIQ image uploading tab

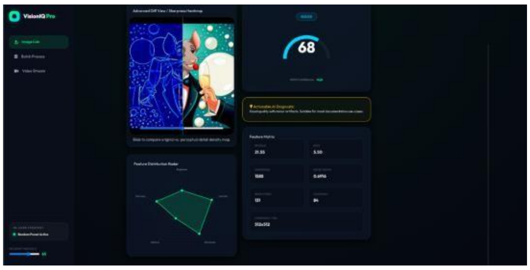

Fig. 3. VisionIQ image evaluating metrics

PERFORMANCE ANALYTICS

The performance of VisionIQ Pro is systematically evaluated using standard regression and correlation- based metrics to measure the reliability, accuracy, and generalization capability of the predicted image quality scores. The dataset is divided into training and testing subsets using a standard train-test split methodology to ensure unbiased model validation and prevent overfitting.
The primary evaluation metrics include:

- **Pearson Linear Correlation Coefficient (PLCC)**
- **Spearman Rank Order Correlation Coefficient (SROCC)**
- **Mean Squared Error (MSE)**

PLCC measures the linear correlation between predicted Mean Opinion Scores (MOS) and ground truth human subjective ratings. SROCC evaluates the monotonic relationship between predicted and actual rankings of image quality. MSE quantifies the average squared difference between predicted and actual MOS values.
Experimental results indicate strong predictive performance, with PLCC values greater than 0.85, demonstrating high alignment with human subjective perception. The SROCC values also show strong rank consistency, confirming that the system correctly orders images based on perceived quality.

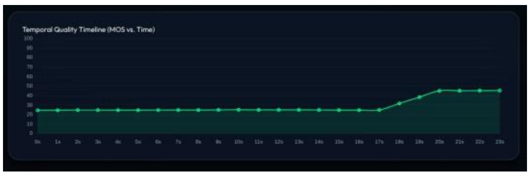

Fig. 4. Temporal Quality Timeline (MOS vs. Time) Graph

Further statistical analysis reveals distinct characteristics of AI-generated images:

- Higher sharpness values due to synthetic edge enhancement
- Higher contrast levels resulting from generative model optimization
- Lower noise levels compared to real-world distorted images
- Elevated NIQE scores caused by unnatural scene statistics and synthetic artifacts

Although many AI-generated images appear visually sharp and clean, the system successfully identifies subtle unnatural patterns and statistical inconsistencies through hybrid feature integration. The combination of handcrafted signal metrics and neural perceptual features improves robustness compared to standalone NR-IQA models.

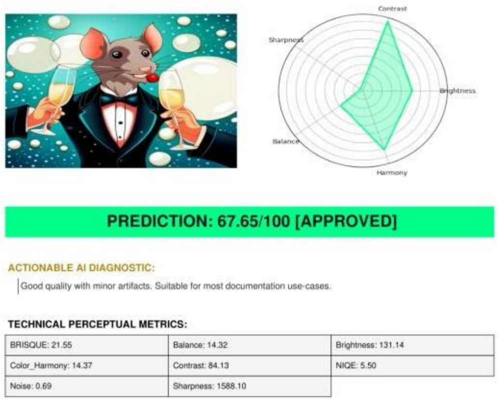

Fig. 5. Final analysis report for the image

The interactive dashboard enhances performance transparency by visualizing radar charts, metric distributions, correlation trends, and explainability heatmaps. These visual analytics tools allow users to monitor model behavior, identify anomaly patterns, and validate quality predictions effectively.
Overall, the proposed framework demonstrates consistent, interpretable, and scalable performance for evaluating both real-world and AI-generated images without requiring reference images.

CONCLUSION

This paper presented **VisionIQ Pro**, a hybrid multi-parametric framework for Non-Reference Image Quality Assessment (NR-IQA) designed specifically to evaluate both natural and AI-generated images. Unlike traditional full-reference metrics, the proposed system operates without requiring a pristine reference image, making it suitable for modern generative AI environments.

The framework integrates handcrafted signal-based features (sharpness, contrast, noise, luminance balance, and color harmony) with perceptual Natural Scene Statistics (NSS) metrics such as BRISQUE and NIQE. These features are fused into a unified Image Quality Vector and processed using a Random Forest regression model to predict Mean Opinion Scores.

A key contribution of the system is the Explainable AI (XAI) module, which generates sharpness heatmaps and visual diagnostic outputs. This enhances transparency by showing spatial regions responsible for quality degradation. The actionable recommendation engine further bridges the gap between quality evaluation and corrective decision-making.

The system effectively identifies AI-generated artifacts, even in images that appear visually sharp, by detecting deviations in natural scene statistics. This makes the framework particularly useful for dataset curation, generative model evaluation, industrial inspection, and academic research.

Future work can extend the framework by incorporating deep learning-based end-to-end NR-IQA models, temporal consistency modeling for video analytics, and adaptive learning mechanisms for domain-specific datasets.

In conclusion, **VisionIQ Pro** demonstrates that combining classical image processing techniques with machine learning and perceptual modeling leads to a robust, interpretable, and scalable solution for automated image quality assessment in AI-driven ecosystems.

REFERENCES

[1] Karam L, Shrotre A. Full reference objective quality assessment for reconstructed background images. 2018. Available from: http://arxiv.org/abs/1803.04103.

[2] Duan J. Improving radiotherapy workflow: evaluation and implementation of deep learning auto-segmentation in a multi-user environment, and development of automatic contour quality assurance system. UKnowledge. 2023.

[3] Gatt A, Krahmer E. Survey of the State of the Art in Natural Language Generation: core tasks, applications and evaluation.2017. Available from: https://core.ac.uk/download/93183864.pdf.

[4] Hosang J, Benenson R, Dollar P, Schiele B. What makes for effective detection proposals? IEEE Trans Pattern Anal Mach Intell (TPAMI). 2015;38(4):814-830.

[5] Horvitz E, Kamar E, Nushi B. Towards accountable AI: hybrid human-machine analyses for characterizing system failure. 2018.

[6] Tiotsop F, Lohic. Optimizing perceptual quality prediction models for multimedia processing systems. Italy, 2022.

[7] Bardhan R, Ramsankaran RAAJ, Sathyakumar V. Geospatial approach for assessing spatiotemporal dynamics of urban green space distribution among neighbourhoods. Urban Forest Urban Green. 2020;49:126630.

[8] Mirkhan A. Enhancing point cloud quality assessment with grouped convolutions: a streamlined approach inspired by COPP Net. 2024.

[9] Bonneel N, Garces E, Lalonde J-F, Meka A. Unsupervised deep single-image intrinsic decomposition using illumination varying image sequences. 2018.

[10] Miyata T. Interpretable image quality assessment via CLIP with multiple antonym-prompt pairs. arXiv:2308.13094. 2023.

[11] Gao H, Zhang K, Sun W, Zhai G, et al. PrefIQA: human preference learning for AI-generated image quality assessment. Presented at IEEE ISCAS 2024.

[12] Ge Y, Liu X, Dai Q, Wang K. Automatic no-reference image quality rating metrics in DL-reconstructed image quality assessment and protocol optimization (VSS score). ISMRM 2022 Abstract #3165.

[13] Yashaswini Sridhar, Jayanthi Muniram Gajendra, Preethi Srinivasalu, and Manjunath Srinivasappa. Evaluating Non-Reference Image Quality Metrics for AI-Generated Images: A Novel Approach. European Journal of Artificial Intelligence and Machine Learning Vol 4, Issue 5, September 2025 ISSN 2796-0072.

[14] Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*. 2012;21(12):4695–4708.

[15] Mittal A, Soundararajan R, Bovik AC. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*. 2013;20(3):209–212.

[16] Zhang L, Zhang L, Bovik AC. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*. 2015;24(8):2579–2591.

[17] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004;13(4):600–612.

[18] Sheikh HR, Bovik AC. Image information and visual quality. *IEEE Transactions on Image Processing*. 2006;15(2):430–444.

[19] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. *Advances in Neural Information Processing Systems*. 2014.

[20] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *International Conference on Learning Representations (ICLR)*. 2016.

[21] M Kaliappan, E Mariappan, MV Prakash, B Paramasivan, *Load Balanced Clustering Technique in MANET using Genetic Algorithms*.. Defence Science Journal 66 (3), 251-258.

[22] M Sivaram, M Kaliappan, S J Shobana, Prakash, V Porkodi *Secure storage allocation scheme using fuzzy based heuristic algorithm for cloud*, Journal of Ambient Intelligence and Humanized Computing,pp.1-9

[23] Vimal, S., Robinson, Y. H., Kaliappan, M., Vijayalakshmi, K., &Seo, S. (2021). *A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction*. The Journal of Supercomputing, 77(1), 1–17. https://doi.org/10.1007/s11227-020-03268-0

[24] Kaliappan M, Guruprakash B, Rajalakshmi, J. Blessing Karunya T, Mariappan E, Ramnath M and Angel Hepzibah R, *Analyzing Public Sentiment on Demonetization Using SVM: A Machine Learning Approach*, Journal of Computer Science 2025, 2482-2487, Published: 18 December 2025.