IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 04, April 2026

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1204.08

A Hybrid Deep Learning Approach for Fine- Grained Cavity Detection Using ESRGAN and SwinUNet

Manoj G
Assistant Professor, Department of CSE,
Guru Nanak Institute of Technology, Hyderabad, India
https://orcid.org/0009-0004-6555-4271
Devarakonda Sidda Maheswar,Ganti Aditya Srinivas
Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, India
manoratz@gmail.com, siddamaheshwar2451@gmail.com,
reach.adityasrinivas@gmail.com

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.12/Issue04/ISAP26.ISAP10087
Research Article | Open Access | Double-Blind Peer-Reviewed| ArticleID: IJIRAE/RS/Vol.12/Issue04/ISAP26.ISAP10087
Received:02,March 2026, Revised: 29, March 2026, Accepted: 10,April 2026,Published Online: 22, April 2026.
https://www.ijiris.com/volumes/Vol12/iss-04/08.ISAP26.ISAP10087.pdf
Article Citation:Manoj,Devarakonda,Ganti(2026), A Hybrid Deep Learning Approach for Fine- Grained Cavity Detection Using ESRGAN and SwinUNet.IJIRIS: International Journal of Innovative Research in Information Security, Volume 12, Issue 04 of 2026 pages 309-316 **Doi:>** https://doi.org/10.26562/ijiris.2026.v1204.08
BibTeX Key: Manoj@2026Hybrid
IJIRIS papers should be cited as IJIRIS(International Journal of Innovative Research in Information Security, AM Publications, India 2026, ISSN 2349-7017, https://doi.org/10.26562/ijiris.2026.v1204.08 The journal's official abbreviation is IJIRIS. Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: The presence of cavities in nuclear materials can become a serious concern because such defects may develop gradually, remain unnoticed for a period, and eventually affect the structural integrity, safety, and performance of the material. Conventional inspection through manual observation of imaging data is effective only on a limited scale, but it becomes slow, expensive, and less reliable when large volumes of material or high-resolution microstructural images must be analyzed. This paper proposes a practical framework for cavity detection in nuclear materials using a dual-stage deep learning architecture based on ESRGAN and SwinUNet. In the proposed approach, ESRGAN is first used to enhance low-resolution or degraded images by recovering fine structural details and improving cavity visibility, after which SwinUNet performs accurate pixel-level segmentation of cavity regions through transformer- based feature extraction. A labeled dataset consisting of cavity and non-cavity microstructural images is used for training and evaluation, with additional preprocessing and augmentation applied to improve model robustness. The proposed model aims to achieve reliable detection performance in terms of precision, recall, F1-score, and segmentation accuracy. The final implemented system is provided through an intuitive browser-based interface where researchers or inspectors can upload an image, obtain annotated detection results, and view detailed analytical outputs without requiring advanced technical expertise. This makes the system suitable for research laboratories, inspection environments, and future intelligent nuclear material monitoring applications.
Keywords: ESRGAN, SwinUNet, Cavity Detection, Nuclear Materials, Deep Learning, Image Segmentation, Super-Resolution, Transformer Networks, Computer Vision, Defect Detection, mAP, Precision Analysis

INTRODUCTION

The structural integrity of nuclear materials plays a critical role in ensuring the safety and efficiency of nuclear systems. However, these materials are highly susceptible to radiation-induced defects such as cavities and voids, which can develop gradually and remain undetected until they significantly impact performance. Cavities formed due to prolonged radiation exposure can lead to material degradation, swelling, and potential failure. Delayed detection of such defects can result in serious safety risks and economic losses, making early and accurate identification extremely important. Conventional methods for cavity detection rely on manual analysis of imaging data such as Scanning Electron Microscopy (SEM), X-ray, or CT scans, where experts visually inspect images to identify defects. Although effective on a small scale, this process is time-consuming, labor-intensive, and prone to inconsistencies due to human interpretation. Furthermore, the requirement for skilled professionals limits its applicability in large-scale or continuous monitoring environments. Recent advancements in image processing and deep learning have opened new possibilities for automating defect detection in nuclear materials. With the availability of high-resolution imaging systems and rapid progress in deep learning techniques, it is now possible to analyze material structures more efficiently and accurately. In this work, we propose a dual-stage framework for cavity detection in nuclear materials using ESRGAN and SwinUNet. The ESRGAN model enhances low-quality or degraded images by improving resolution and restoring fine structural details, while the SwinUNet model performs precise pixel- level segmentation using transformer-based feature extraction. This integrated approach enables accurate detection and localization of cavities, providing a reliable and scalable solution for advanced nuclear material inspection.

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1204.08

A. Objective

The primary objective of this work is to develop an end-to-end framework for detecting cavities in nuclear materials using advanced deep learning techniques. The proposed system integrates ESRGAN for enhancing low-quality or degraded images and SwinUNet for accurate pixel-level segmentation of cavity regions, enabling precise detection even in challenging conditions such as noise, low resolution, uneven contrast, and very small defects. The model is designed to provide efficient and near real-time analysis while maintaining high accuracy, reducing reliance on manual inspection by experts. To ensure robustness, it is trained on diverse datasets representing different imaging conditions and material structures, allowing it to generalize effectively across various scenarios. Additionally, the system includes a user-friendly interface where users can upload images, visualize annotated results, and access analytical insights such as confidence scores and cavity distribution, making it suitable for both research and practical inspection applications. A key focus of the project is to provide an intuitive and user-friendly interface that allows users to easily upload images, visualize detection results, and analyze outputs without requiring deep technical knowledge. The interface presents annotated images along with detailed analytical insights such as confidence scores, cavity distribution, and performance metrics. This enhances interpretability and usability, making the system suitable for both research and practical inspection environments. Overall, the proposed solution aims to deliver an accurate, scalable, and accessible approach for advanced cavity detection in nuclear materials.

B. Problem Statement

Many of the existing cavity detection systems for nuclear materials struggle to deliver reliable results under practical imaging conditions, especially when dealing with noise, low contrast, and varying resolution in SEM, X-ray, or CT images. Traditional detection approaches, including earlier deep learning models, often fail to accurately identify small or early- stage cavities and are unable to capture irregular defect shapes with high precision. Additionally, many existing methods depend on complex model configurations or require extensive manual tuning for different datasets, making them less accessible and difficult to generalize across diverse material conditions. These limitations highlight the need for a more advanced and efficient system that can accurately detect and segment cavity regions in real time, handle variations in imaging quality, and operate effectively across different datasets while maintaining high accuracy and computational efficiency.

C. Scope of the Project

The project focuses on the design, development, and evaluation of an intelligent cavity detection framework for nuclear materials using advanced deep learning techniques, along with a web-based interface for practical inspection applications. The primary objective is to accurately detect and localize cavity regions in imaging data such as SEM, X-ray, or CT scans obtained under varying conditions. By integrating ESRGAN for image enhancement and SwinUNet for precise segmentation, the system is capable of identifying defects with high accuracy regardless of variations in image quality, noise levels, or structural complexity. The proposed framework ensures reliable detection even in challenging scenarios involving low-resolution or degraded images. Additionally, the system is designed to be deployable on research workstations, cloud platforms, and advanced inspection environments, making it suitable for large-scale analysis and monitoring.

LITERATURE SURVEY

Volety et al. [1] proposed a hybrid model combining object detection with Generative Adversarial Networks (GANs) to address data imbalance and improve detection of subtle defects under varying imaging conditions such as noise and low contrast.

Liu et al. [2] developed a modified detection architecture optimized for identifying small-scale features by incorporating preprocessing techniques like contrast enhancement and noise reduction, achieving improved sensitivity and near real-time performance.

Kumar et al. [3] introduced lightweight deep learning models for classification tasks, demonstrating that efficient architectures can be deployed on resource-constrained systems while maintaining reliable detection performance.

Dixit and Nema [4] presented a comprehensive review of machine learning approaches, highlighting key challenges such as detecting small irregular defects, handling complex backgrounds, and ensuring model generalization across diverse datasets.

Zhang et al. [5] enhanced detection accuracy by integrating attention mechanisms into deep learning models, allowing the network to focus on relevant regions while suppressing background noise.

Wang et al. [6] proposed a hybrid architecture using deformable convolutional layers, enabling better modeling of irregular and complex defect shapes through adaptive feature extraction.

Sladojevic et al. [7] demonstrated the effectiveness of deep convolutional neural networks in large-scale image-based classification, establishing a strong foundation for modern detection and segmentation techniques.

Xue et al. [8] introduced a domain-specific detection model with customized data augmentation strategies to improve robustness under varying imaging conditions.

Huang et al. [9] proposed a knowledge distillation approach to transfer knowledge from large models to lightweight ones, achieving high accuracy with reduced computational cost.

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 04, April 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1204.08

SYSTEM DESIGN

A. System Architecture

The proposed system takes nuclear material images, such as SEM, X-ray, or CT scans, as input for cavity detection. Initially, the images are resized to a standard resolution and passed through preprocessing steps including normalization, noise reduction, and contrast enhancement to improve input quality. In the first stage, ESRGAN is applied to enhance the image by reconstructing high-frequency details and improving the visibility of fine structural features such as micro- cavities. This super-resolution process helps in overcoming limitations caused by low-quality or degraded imaging conditions. In the second stage, the enhanced images are processed using the SwinUNet architecture, which performs accurate pixel- level segmentation of cavity regions. The model utilizes transformer-based encoding to capture both local texture details and long-range dependencies within the material structure, enabling precise identification of irregular and small cavities. Multi-scale feature extraction and fusion allow the system to detect cavities of varying sizes effectively. Finally, the output is refined through post-processing techniques to generate clear segmented maps and annotated results, along with confidence scores and quantitative metrics, ensuring reliable and interpretable cavity detection.

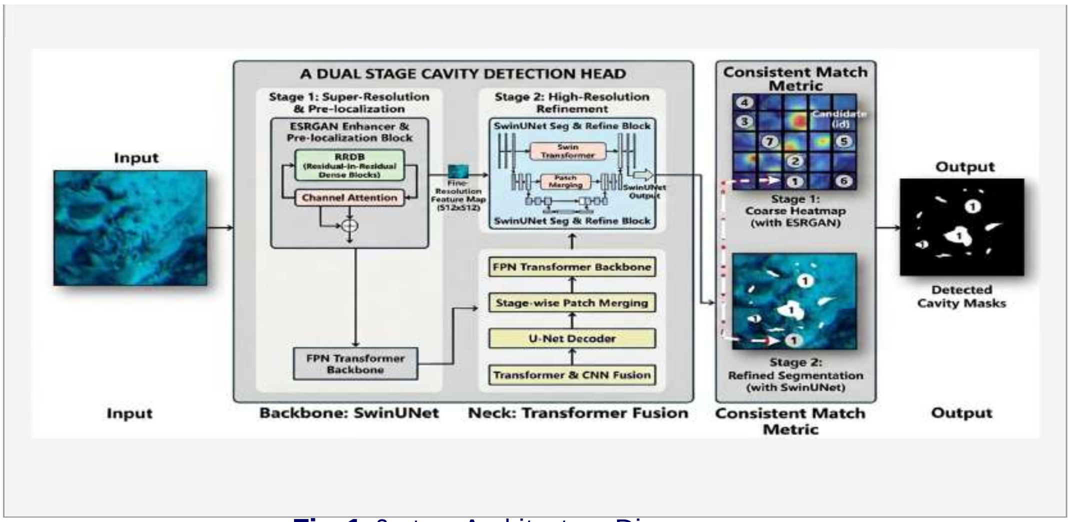

Fig. 1: System Architecture Diagram

B. Methodology

1. Data Acquisition: The system takes nuclear material images such as Scanning Electron Microscopy (SEM), X-ray, or CT scan images as input. These images may come from laboratory equipment or stored datasets. The system is designed to handle variations in image quality, including noise, low contrast, and degraded resolution, which are common in real- world inspection scenarios.
2. Data Preparation: All input images are resized to a standard resolution and normalized to ensure consistency across the dataset. Preprocessing steps such as noise reduction, contrast enhancement, and intensity normalization are applied to improve image quality. Data augmentation techniques like rotation, flipping, brightness adjustment, and scaling are used during training to increase dataset diversity and improve model generalization under different imaging conditions.
3. Image Extraction: In the first stage, ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) is used to enhance low-resolution or degraded images. This model reconstructs high-frequency details, sharpens edges, and improves the visibility of fine structural features such as micro-cavities. The enhancement step plays a crucial role in enabling better feature extraction in the subsequent stage.
4. Feature Extraction and Segmentation (SwinUNet Stage): The enhanced images are passed to the SwinUNet model, which performs pixel-level segmentation of cavity regions. The model uses transformer-based encoding to capture both local textures and global contextual information. Multi-scale feature extraction allows the system to detect cavities of varying sizes, while the encoder-decoder architecture ensures accurate localization of defect regions.
5. Model Training: The model is trained using pre-trained weights to accelerate convergence and improve performance. Optimization is performed using advanced optimizers such as Adam or AdamW with learning rate scheduling. Loss functions such as Dice Loss and Binary Cross-Entropy are used to improve segmentation accuracy. The training process includes validation monitoring, early stopping, and saving the best-performing model based on evaluation metrics.
6. Model Deployment: After training, the model is deployed as a web-based application using frameworks such as Flask or FastAPI. Users can upload images and receive real-time detection results along with visualizations and analytical insights. For scalable and high-performance deployment, the model can be exported to formats like ONNX and optimized for GPU inference using tools such as TensorRT.

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 04, April 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1204.08

C. Modules

The architecture is divided into five logical modules, each designed as an independently testable software component. The Dataset Preparation module handles the collection of nuclear material images such as SEM, X-ray, or CT scans, along with annotation of cavity regions, verification of dataset balance, and splitting of data into training, validation, and testing sets in a 70:20:10 ratio. The Preprocessing and Augmentation module manages image resizing, normalization of pixel values, conversion into suitable formats for model input, and application of augmentation techniques such as rotation, flipping, scaling, and contrast adjustment during the training phase. The Model Configuration module provides options for selecting and configuring the deep learning components, including ESRGAN for image enhancement and SwinUNet for segmentation, along with hyperparameter tuning and loading of pre- trained weights to enable transfer learning. The Training and Validation module oversees the complete training process, including forward propagation, loss computation, backpropagation, epoch-wise logging, early stopping, and saving of the best-performing model.

EXISTING SYSTEM VS PROPOSED SYSTEM

A. Existing System

Traditional approaches for cavity detection in nuclear materials often rely on conventional image processing techniques or standard deep learning models that perform detection using bounding boxes. These methods typically use convolutional neural networks (CNNs) for feature extraction and basic segmentation or detection heads to localize defect regions. While such systems are effective to some extent, they face significant limitations when applied to real-world nuclear imaging data. One major issue is their dependence on image quality, as low-resolution, noisy, or low-contrast images make it difficult to identify small or early-stage cavities. Additionally, bounding box-based detection lacks the precision required to accurately represent irregular cavity shapes, leading to incomplete or inaccurate localization of defects. Another limitation of existing systems is their inability to capture fine-grained structural details and contextual relationships within the material. Many models struggle with detecting very small or subtle cavities due to limited spatial resolution and inadequate feature representation. Furthermore, traditional approaches often require manual tuning of parameters and are not easily adaptable to different datasets or imaging conditions. These systems also fail to provide detailed quantitative insights, such as exact cavity boundaries or severity, restricting their usefulness in critical inspection scenarios. As a result, there is a need for a more advanced and integrated approach that can enhance image quality and perform precise segmentation for reliable cavity detection.

B. Proposed System — ESRGAN + SwinUNet

The proposed system addresses the limitations of existing methods by introducing a dual-stage deep learning framework that combines ESRGAN for image enhancement and SwinUNet for accurate segmentation. Unlike traditional approaches, ESRGAN improves the resolution and quality of input images by reconstructing fine details and enhancing structural features, making small cavities more visible and easier to detect. This enhancement stage significantly improves the effectiveness of subsequent processing by providing clearer and more informative input data. Following enhancement, the SwinUNet model performs pixel-level segmentation using transformer-based architecture, which captures both local and global features within the image. This allows the system to accurately detect cavities of varying sizes and shapes, including very small or irregular defects. The use of multi-scale feature extraction and attention mechanisms improves sensitivity to subtle variations while maintaining high accuracy. Additionally, the integrated framework eliminates the need for complex manual tuning and provides detailed outputs such as segmented cavity maps, confidence scores, and quantitative analysis. The system is also designed to be efficient and scalable, enabling deployment in research environments and advanced inspection systems for real-world nuclear material analysis.

IMPLEMENTATION

A. Development Environment

The entire project is developed using Python and implemented in Visual Studio Code from the initial stages to completion. The deep learning models, including ESRGAN for image enhancement and SwinUNet for segmentation, are implemented using the PyTorch framework, which manages model training, inference, and overall functionality. Image processing operations, along with visualization of detection results such as segmented cavity regions and annotated outputs, are handled using OpenCV. Data handling and result storage are managed using NumPy and Pandas, while performance metrics and analytical graphs are generated using Matplotlib. The web-based application interface is developed using Flask to enable user interaction and result visualization. Version control throughout the development process is maintained using Git and GitHub. System Requirements: NVIDIA GPU with min 8GB VRAM, 16GB of system RAM is needed for training with reasonable speed. Any system that has at least 4GB RAM would be needed for CPU only (non-real-time, batch processing), but the GPU is required for live detection on site since there's no time for it to take forever to process. System is compatible with Windows 10/11 and Ubuntu 20.04 LTS. Provided are a pip requirements file for quick replication on any other machine.

B. System Workflow

The system follows a structured multi-stage pipeline designed to provide a seamless experience for users without requiring technical expertise. Initially, the user uploads a nuclear material image, such as an SEM, X-ray, or CT scan, through the web interface.

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 04, April 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1204.08

The input image is then preprocessed by resizing, normalization, and enhancement to ensure consistency and quality before being passed to the processing pipeline. In the first stage, ESRGAN enhances the image by improving resolution and restoring fine structural details. The enhanced image is then processed by the SwinUNet model, which performs pixel-level segmentation to identify and localize cavity regions. The output is generated in the form of segmented masks and annotated images, along with confidence scores indicating detection reliability. All inference results, including detected cavity regions, spatial information, confidence values, and timestamps, are stored in a structured format for further analysis and record-keeping. The system also provides a visualization dashboard that displays performance metrics such as accuracy trends, loss curves, and distribution of detected cavities. This design ensures that the system can be easily integrated into research workflows and advanced inspection environments, providing an efficient and scalable solution for cavity detection in nuclear materials.

RESULTS AND DISCUSSION

This section presents the outcomes obtained from the proposed ESRGAN–SwinUNet framework, including the working web interface, model performance evaluation, training behavior, and analytical outputs provided to the end users. The developed system demonstrates effective cavity detection through enhanced image processing and precise segmentation, along with intuitive visualization of results

A. Application Interface

The web application is designed with a strong focus on simplicity and ease of use, ensuring accessibility for users without technical expertise. The home page clearly conveys the purpose of the system by displaying sample nuclear material images along with highlighted cavity regions, providing users with an immediate understanding of the application. A navigation bar allows easy access to features such as login, image upload, and result visualization. Users are required to register once to create an account, enabling them to securely log in and maintain a record of their previous analyses. After logging in, users can access the upload interface, where they can select and submit nuclear material images such as SEM, X-ray, or CT scans with a single action. The uploaded image is then processed by the backend system, where ESRGAN enhances the image quality and SwinUNet performs cavity segmentation.

Fig. 2a: Home Page

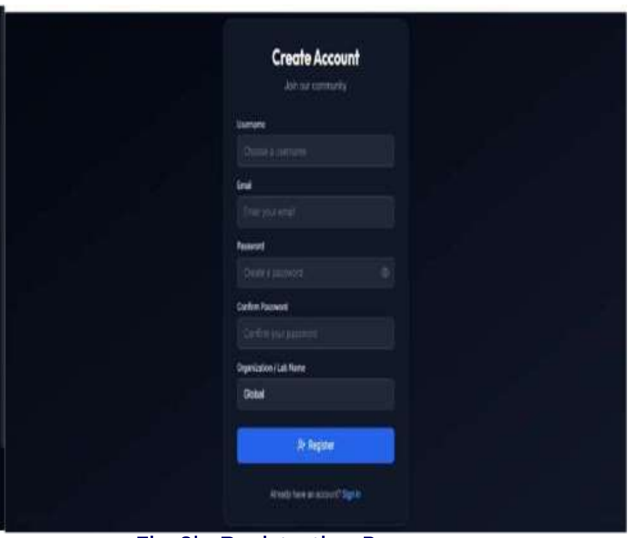

Fig. 2b: Registration Page

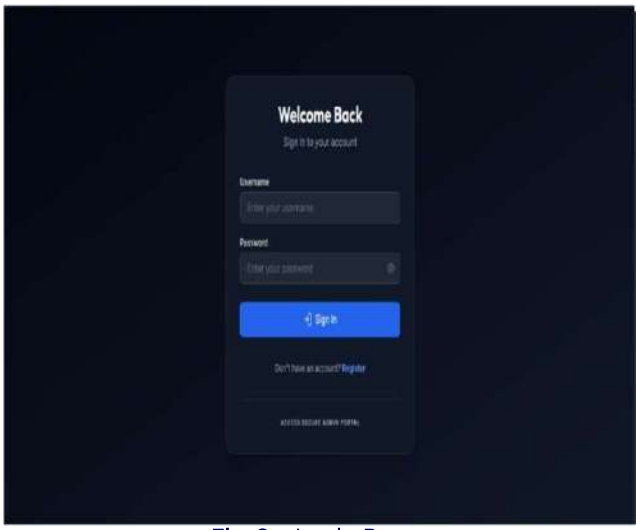

Fig. 2c: Login Page

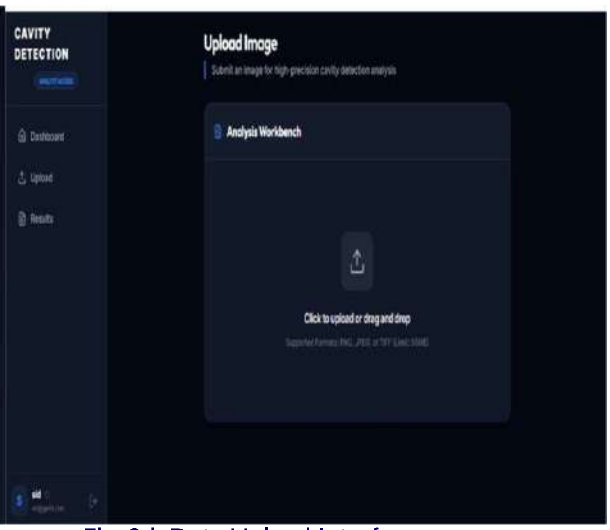

Fig. 2d: Data Upload Interface

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 04, April 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1204.08

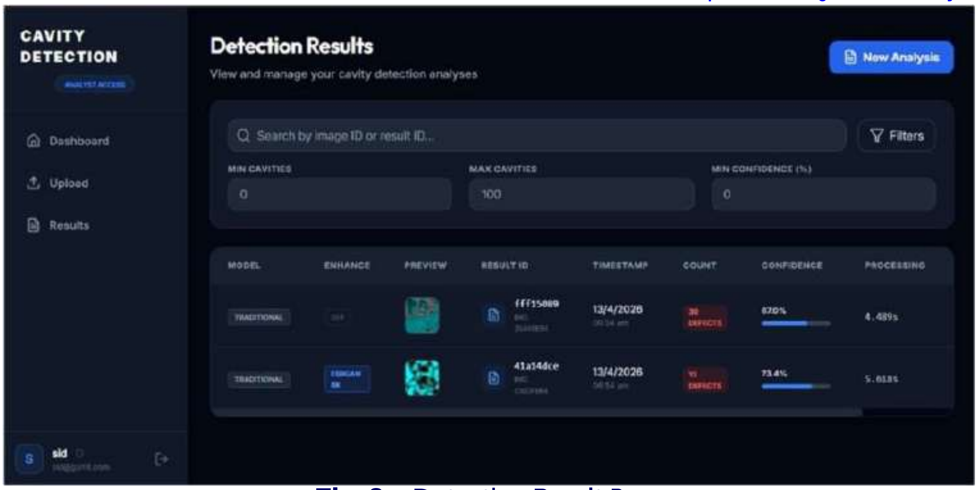

Fig. 2e: Detection Result Page

B. Performance Metrics

To evaluate the performance of the proposed model in detecting and segmenting cavity regions, several standard metrics were used, including confusion matrix, precision–confidence curve, recall–confidence curve, F1-score curve, and precision–recall curve. The confusion matrix, presented in both normalized and non-normalized forms, provides insight into how accurately the model distinguishes between cavity and non-cavity regions. The results indicate that clearly visible and well-defined cavities are detected with high accuracy, as their structural patterns are more distinguishable from the surrounding material.However, the model shows relatively lower performance when detecting very small or low-contrast cavities. In some cases, these subtle defects are misclassified as background due to their similarity with normal material textures or imaging noise. This is expected, as early-stage or micro-scale cavities often lack strong visual features, making them harder to detect even after enhancement. Cavities with moderate visibility fall between these two extremes, where the model achieves balanced performance in terms of precision and recall. Overall, the results demonstrate that the proposed ESRGAN–SwinUNet framework significantly improves detection capability, especially for clearly defined cavities, while still presenting opportunities for further improvement in detecting extremely subtle defects.

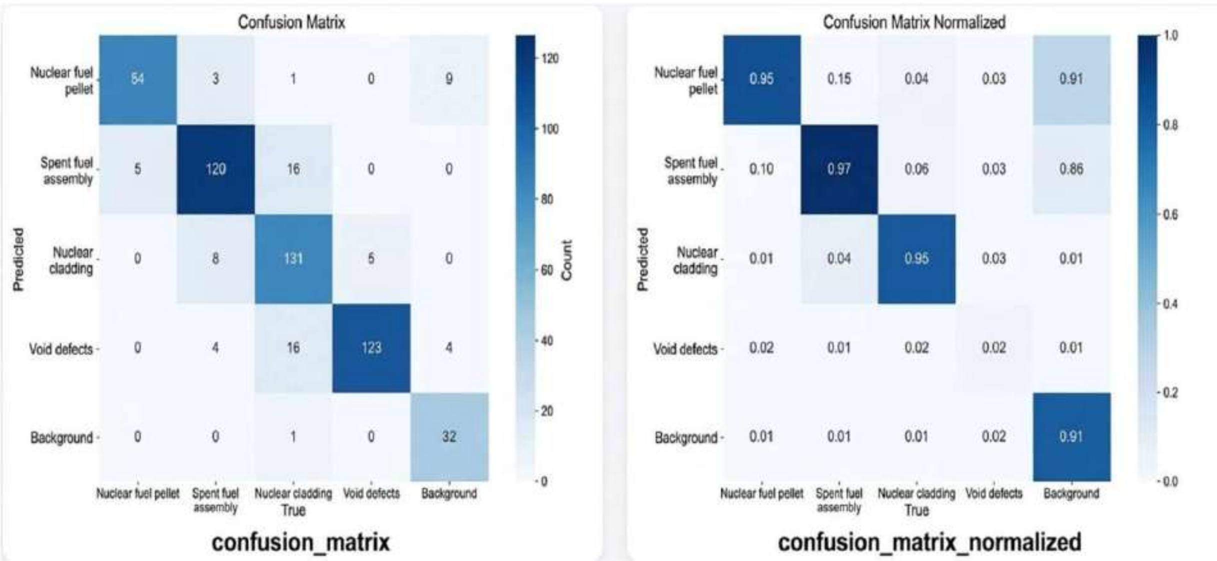

Fig. 3: Confusion Matrix and Normalised Confusion Matrix

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1204.08

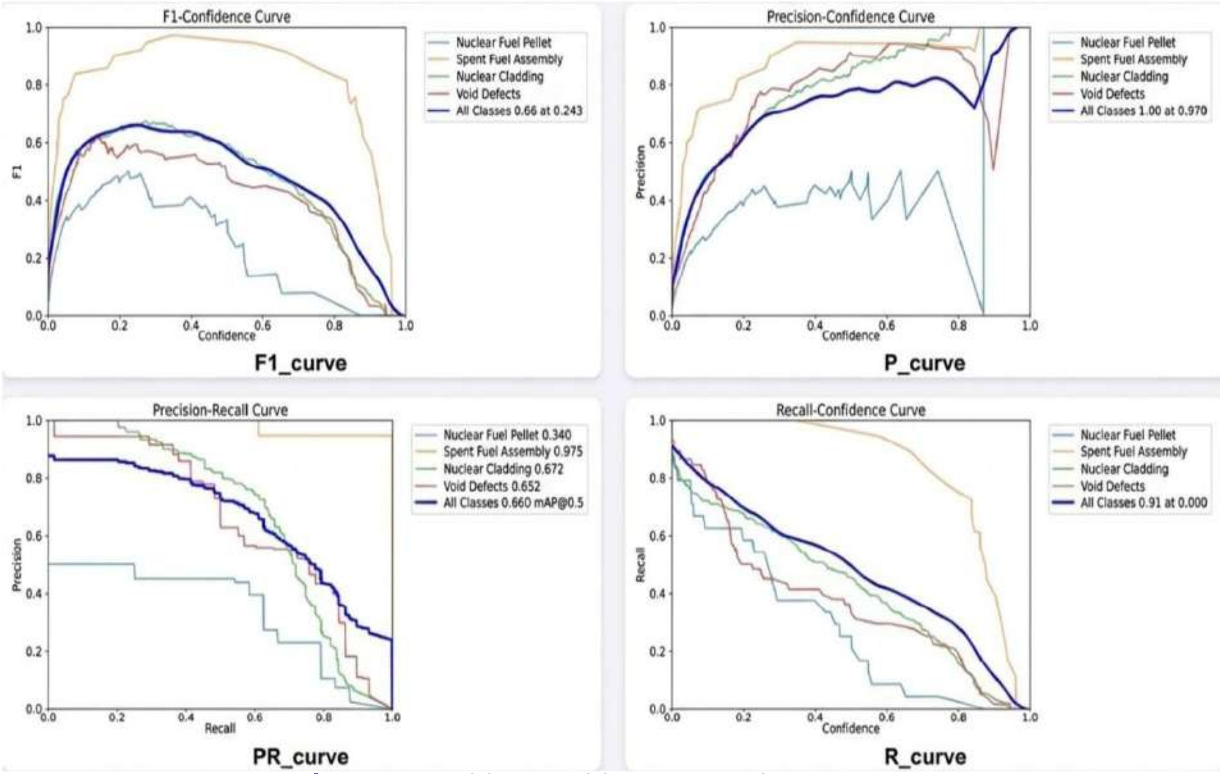

Fig. 4: F1, Precision, Precision-Recall, and Recall Curves

The performance curves obtained from the validation dataset illustrate the effectiveness of the proposed model across different evaluation metrics. The class-averaged F1-score reaches its peak at an optimal confidence threshold, indicating a balanced trade-off between precision and recall where the model achieves reliable detection without being overly strict or lenient. At higher confidence thresholds, the precision approaches near-perfect values, meaning the model produces very few false positives when it is highly confident about its predictions. This demonstrates the model's ability to maintain high reliability in critical detection scenarios. In terms of segmentation performance, the model achieves strong results for clearly visible cavity regions, with higher accuracy and overlap scores, while performance is relatively lower for very small or low-contrast cavities. This variation is expected, as subtle cavities often lack distinct structural boundaries and may resemble background textures, making them harder to detect even after enhancement. Overall, the model achieves a strong average performance, demonstrating its capability to accurately segment and localize cavity regions.

C. Training Results

Fig. 5 illustrates the training progress of the proposed model over multiple epochs, highlighting the learning behavior and convergence of the system.

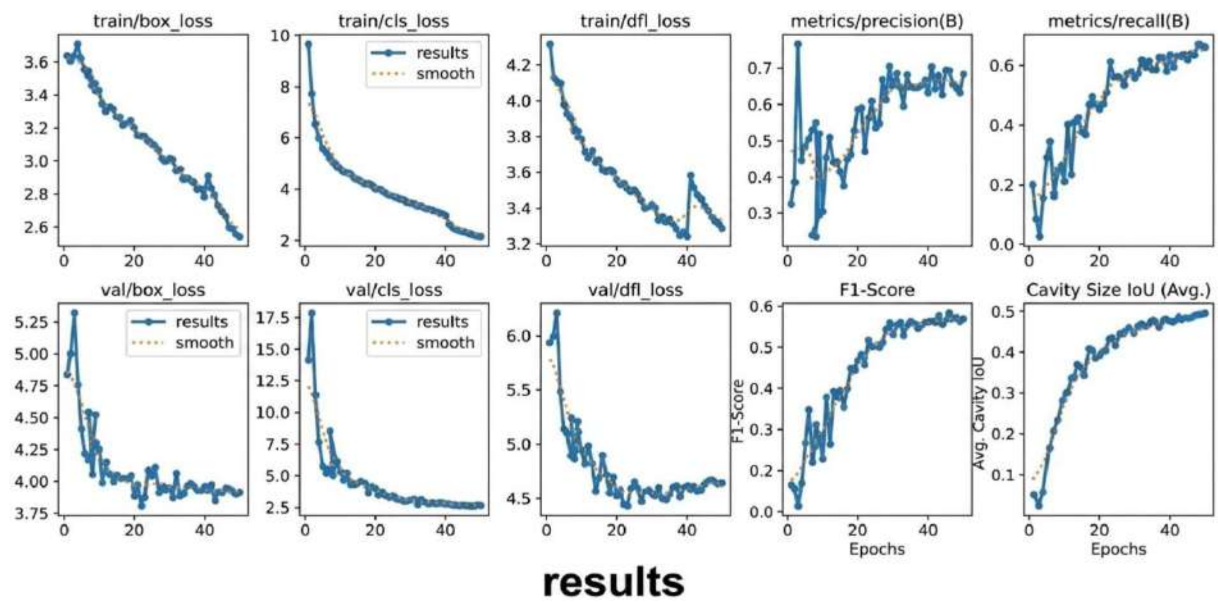

Fig. 5: Training and Validation Results Visualization

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1204.08

The loss curves, including segmentation loss and total training loss, show a consistent decrease throughout the training process, indicating effective learning and optimization. Similarly, the validation loss follows a closely aligned trend without significant divergence, suggesting that the model generalizes well and does not suffer from overfitting.

CONCLUSION

The objective of this project was to develop a practical and reliable system for cavity detection in nuclear materials, and the results indicate that this goal has been successfully achieved. By implementing a dual-stage framework combining ESRGAN for image enhancement and SwinUNet for segmentation, the system demonstrates strong performance in detecting and localizing cavity regions across varying imaging conditions. The model achieves high precision at optimal confidence levels and maintains strong recall, indicating its ability to accurately identify most cavity instances while minimizing false detections. The integration of super-resolution and transformer-based segmentation significantly improves detection accuracy compared to traditional approaches, while also simplifying the overall process by reducing the need for complex manual tuning. Additionally, the user-friendly web interface allows non-expert users to easily upload images and obtain meaningful analytical results, making the system practical for real-world inspection scenarios. Despite these promising outcomes, certain limitations remain, particularly in detecting very small or low-contrast cavities. Such defects often lack distinct visual characteristics and can resemble background textures, making them challenging even for advanced models. This highlights the need for further improvements, including the use of more diverse and high- quality datasets that capture variations in imaging conditions and defect characteristics. Future work can explore advanced transformer-based architectures for improved sensitivity to subtle features, as well as the integration of temporal analysis for monitoring cavity progression over time. Additionally, deploying lightweight and optimized versions of the model for edge devices, along with exploring multi-modal imaging techniques such as hyperspectral analysis, could further enhance the system's capability and applicability in real-world nuclear material inspection environments.

ACKNOWLEDGMENT

The authors express heartfelt gratitude to Mr G. Manoj, Assistant Professor, Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad, for her unwavering support and mentorship during this project. They also express gratitude to Dr. B. Santhosh Kumar, Head of Department, for his valuable guidance, and Mr. D. Srinivas, Project Coordinator, for his useful recommendations throughout the system development process. Appreciation is also given to the faculty and laboratory personnel of the CSE Department, GNIT, and to the families of the authors for their support and encouragement

REFERENCES

1. X.Wang, K.Yu,S.Wu,J. Gu,Y.Liu, C.Dong, Y.Qiao, and C.C.Loy,"ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in Proc. ECCV Workshops, 2018.
2. H.Cao, Y.Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation," arXiv:2105.05537, 2021.
3. Z.Liu et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," in Proc. ICCV, 2021.
4. O.Ronneberger, P.Fischer, and T.Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. MICCAI, 2015.
5. K.He, X.Zhang, S.Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. CVPR, 2016.
6. J.Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," in Proc. CVPR, 2016.
7. TY. Lin et al., "Focal Loss for Dense Object Detection," in Proc. ICCV, 2017.
8. D.P.Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980, 2014.
9. A.Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," in Proc. ICLR, 2021.
10. S.Zhang et al., "Deep Learning-Based Defect Detection in Industrial Applications: A Survey," IEEE Trans. Ind. Informat., 2020.
11. PyTorch Foundation, "PyTorch: An Open-Source Machine Learning Framework," 2024.
12. OpenCV, "Open Source Computer Vision Library," 2023.
13. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
14. NumPy Developers, "NumPy: Fundamental Package for Scientific Computing in Python," 2023.
15. J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in Proc. CVPR, 2015.