Lung Cancer Detection Using Attention-Enhanced Hybrid CNN–ViT Models for CT Scan Classification

R R Shantha Spandana [1], V Bharath Sanjay [2], C Venkatesh [3]

[1] Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: shanthaspandana@gmail.com, ORC-ID: https://orcid.org/0009-0003-4236-1250

[2] P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: bharathsanjayvelayudham@gmail.com , ORC-ID: https://orcid.org/0009-0000-1688-1226

[3] Assistant Professor, Department of CSE(AI & ML), Sri Venkatesa Perumal College of Engineering & Technology, Puttur, E-mail: chevireddyvenkatesh22@gmail.com, ORC-ID: https://orcid.org/0009-0007-1038-5978

Abstract: Lung cancer is the foremost cause of cancer-related mortality, requiring prompt and precise diagnosis to enhance patient outcomes. Manual interpretation of computed tomography (CT) scans and traditional deep learning techniques frequently inadequately identify multi-scale features and accurately localize lesions. Experiments are performed using two publicly accessible datasets: the IQ-OTH/NCCD Lung Cancer Dataset and the Chest CT-Scan Images Dataset. The suggested attention-enhanced hybrid CNN–ViT framework amalgamates ResNet50, DenseNet169, EfficientNetV2-Medium, ConvNeXt-Base, InceptionNeXt-Base, MobileViT-Small, ConViT-Base, Swin-Base, MaxViT-Base, and DeiT3-Base for classification, in conjunction with YOLOv5, YOLOv8, YOLOv9, and YOLOv11 for detection. Preprocessing encompasses image resizing to 299×299, data augmentation, tensor normalization, and stratified data partitioning, whereas YOLO datasets are organized with bounding box annotations. GradCAM produces heatmaps that emphasize significant areas, while a Flask-based interface facilitates comprehensive user interaction. ConvNeXt-Base attains the maximum classification accuracy of 99.09% on the IQ-OTH/NCCD dataset, whereas InceptionNeXt-Base achieves 99.01% accuracy on the chest CT-scan dataset. YOLOv5 attains the highest mean average precision (mAP) of 72.8% for detection. The approach exhibits enhanced resilience, equitable performance, and interpretable predictions by the integration of categorization and detection within a cohesive system.

"Index Terms: *Lung Cancer Detection, Computed Tomography, Hybrid CNN–Vision Transformer, Attention Mechanisms, Deep Learning, Explainable AI".*

1. INTRODUCTION

Lung cancer continues to be one of the most common and lethal cancers globally, dramatically impacting mortality rates. The ailment arises from atypical cellular proliferation in pulmonary tissues and frequently advances swiftly, resulting in significant respiratory difficulties. Notwithstanding progress in healthcare, lung cancer persists in impacting millions of individuals each year, with survival rates predominantly reliant on early diagnosis and prompt intervention. Epidemiological data reveals an increasing incidence of diagnosed cases and related deaths, underscoring the necessity for the development of better diagnostic techniques to enhance clinical decision-making and patient care [1]. The categorization of lung cancer into principal

subgroups underscores the intricacy of precise diagnosis and treatment strategy formulation [2].

Traditional diagnostic methods, including imaging-based screening and invasive clinical procedures, are essential for detecting lung problems. Nonetheless, these techniques frequently entail substantial expenses, prolonged analytical durations, and possible patient distress. Furthermore, early-stage malignancies and diminutive pulmonary nodules may go unnoticed or are misconstrued owing to their inconspicuous visual attributes. The manual assessment of imaging data relies heavily on specialist expertise and is susceptible to discrepancies and human mistake, which consequently undermines diagnostic reliability. Despite improvements in technology enhancing imaging capabilities, the demand for automated and accurate diagnostic assistance systems continues to pose a significant barrier in medical imaging analysis [3], [4], [5].

Recent advancements in sophisticated computational methods have facilitated the automated interpretation of medical imaging data, resulting in significant enhancements in disease diagnosis and classification precision. Advanced learning frameworks have proven capable of extracting intricate visual features and detecting tiny anomalies that may be missed through manual analysis. These methods have garnered heightened interest for their ability to improve diagnostic efficiency and uniformity across several imaging modalities. The main goal is to create a strong and effective diagnostic framework that can precisely detect lung abnormalities in medical imaging while overcoming the limits of current automated systems [6], [7], [8].

The incorporation of sophisticated computer analysis in lung cancer diagnosis possesses considerable therapeutic and societal significance. Precise and prompt detection can facilitate early treatment interventions, enhance survival rates, and alleviate healthcare burdens linked to late-stage illness management. Moreover, improved diagnostic reliability might aid healthcare workers in making educated treatment decisions and diminish variability in clinical evaluations. The implementation of intelligent image analysis technologies facilitates scalable screening initiatives and enhances access to diagnostic services, especially in areas with restricted medical proficiency. These innovations possess the capacity to revolutionize diagnostic workflows and enhance patient outcomes worldwide [9], [10].

2. LITERATURE REVIEW

Recent breakthroughs in artificial intelligence have profoundly impacted automated lung cancer detection using medical imaging. A thorough evaluation in [11] underscores the swift advancement of deep learning methodologies in lung cancer diagnosis, highlighting their capacity to enhance accuracy and minimize manual involvement. The survey in [12] examines diverse deep learning architectures utilized in CT imaging, showcasing enhanced feature extraction and classification efficacy relative to conventional machine learning methods. While these evaluations offer useful insights into current approaches, they predominantly emphasize summarizing architectures instead of tackling issues associated with multi-scale feature representation and lesion localization.

Numerous studies have presented customized attention-based and neural network models to improve diagnostic efficacy. The MorphAttnNet framework introduced in [13] combines morphological feature extraction with attention

processes to enhance subtype classification accuracy, demonstrating effective results in capturing structural tumor features. A different method in [14] presents a bidirectional recurrent neural network improved by a bio-inspired algorithm to improve classification accuracy and stability. The GoogLeNet-AL framework in [15] offers an adaptive automated detection approach that enhances model adaptation to diverse imaging situations. Although these methods exhibit robust classification performance, they frequently depend on single-model designs, constraining their capacity to generalize across varied datasets and imaging complexities.

Ensemble learning methodologies have garnered interest for enhancing detection robustness and classification dependability. The ensemble-based deep learning methodology outlined in [16] integrates various neural architectures to augment diagnostic precision from thoracic CT images, showcasing enhanced performance via model variety. The interpretable architecture in [17] presents explainable artificial intelligence to enhance transparency in diagnostic predictions, fostering improved clinical trust and interpretability. Simultaneously, the integration of federated learning with blockchain technology in [18] improves data security and the efficiency of distributed training, effectively tackling privacy issues in the sharing of medical data. Notwithstanding these advances, ensemble and explainable frameworks frequently elevate computational complexity and may necessitate considerable computer resources, hence constraining real-time clinical implementation.

Further research has investigated sophisticated network architectures for enhanced feature learning and classification. The ensemble model based on capsule networks presented in [19] optimizes spatial feature representation and improves detection performance by maintaining hierarchical linkages in imaging data. Lung-EffNet, as presented in [20], employs the EfficientNet architecture to attain superior classification accuracy while optimizing computing economy. Despite the robust performance of these methodologies, issues persist in attaining a balance of computing efficiency, multi-scale feature learning, and precise lesion localization across diverse datasets.

Notwithstanding considerable advancements in automated lung cancer detection, current methodologies frequently encounter difficulties in concurrently capturing both local and global imaging characteristics while preserving computing efficiency and model interpretability. Numerous frameworks emphasize categorization while neglecting thorough lesion detection and visualization functionalities. This inquiry tackles these problems by presenting an integrated hybrid learning system aimed at augmenting multi-scale feature extraction, enhancing detection reliability, and offering explainable diagnostic insights. This method seeks to enhance resilience and provide balanced performance across various CT imaging datasets, thus facilitating dependable clinical decision-making.

3. MATERIALS AND METHODS

The system intends to deliver precise lung cancer classification and anomaly detection utilizing the IQ-OTH/NCCD Lung Cancer Dataset and the Chest CT-Scan Images Dataset. The framework handles CT images via a systematic pipeline that includes data ingestion, quality enhancement, transformation, and equitable dataset partitioning to facilitate dependable model training and validation. Various baseline deep learning architectures are assessed to set performance benchmarks, succeeded by a streamlined hybrid CNN–Vision Transformer

model augmented with InceptionNeXt blocks and combined grid and block attention mechanisms to enhance multi-scale feature extraction and contextual representation. The framework enhances diagnostic capability by integrating YOLO-based object detection models for the automated localization of questionable areas, while GradCAM-based explainable visualization emphasizes key features that affect classification decisions. A Flask-based interface facilitates effortless deployment and user engagement, guaranteeing accessibility and scalability. The integrated system improves diagnostic precision, interpretability, and resilience, offering a thorough and clinically beneficial approach for automated lung cancer analysis across diverse CT imaging datasets.

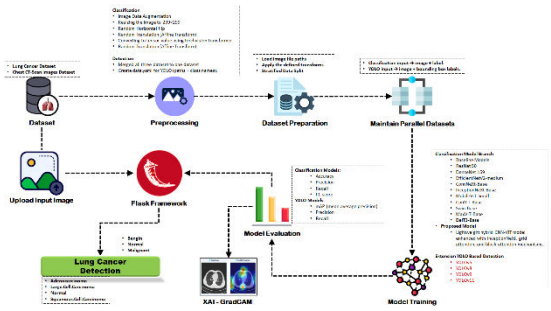


Fig.1 Proposed Architecture

The system architecture has a hybrid deep learning framework intended for lung cancer classification and anomaly detection. The input CT image is subjected to incremental feature extraction through various phases that incorporate hybrid blocks, which merge convolutional and attention-based processing. Inception depthwise convolution captures multi-scale spatial characteristics, whereas grid and block attention modules improve contextual representation and emphasize discriminative regions. Hierarchical feature refinement is executed by iterative hybrid stages, succeeded by pooling and fully linked layers. The final output layer categorizes photos into various lung cancer classifications, guaranteeing precise and reliable diagnosis.

a) Dataset Collection:

The experimental study employs the IQ-OTH/NCCD Lung Cancer Dataset and the Chest CT-Scan Images Dataset sourced from publically accessible medical imaging repositories. The IQ-OTH/NCCD dataset comprises 1,190 CT scan slices derived from 110 cases, classified into normal, benign, and malignant categories, obtained by standardized clinical imaging techniques. The Chest CT dataset comprises images of adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal cases in JPG/PNG formats, along by designated training, validation, and testing divisions. The integrated datasets offer varied imaging attributes, facilitating dependable model training and assessment across several lung cancer classifications.

Fig.2 IQ-OTH/NCCD - Lung Cancer Dataset

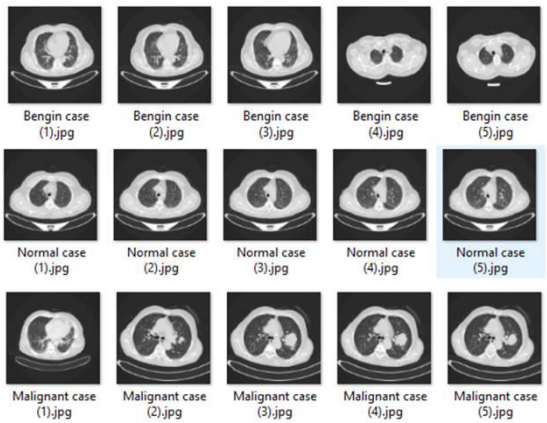
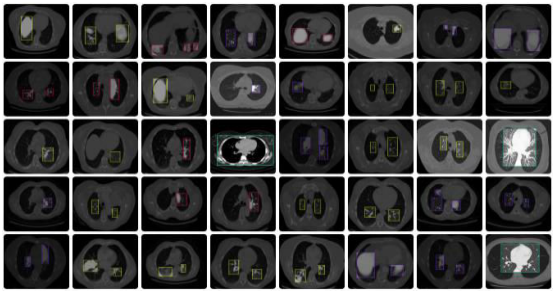


Fig.3 Chest CT-Scan images Dataset



b) Pre-Processing:

Efficient preprocessing is crucial for augmenting data quality, enhancing model generalization, and guaranteeing dependable learning. This preprocessing pipeline standardizes imaging data, balances dataset distribution, and generates structured inputs for classification and detection tasks.

Image Data Augmentation: Image data augmentation and standardization are executed to augment dataset variety and enhance model generalization. Input photos are uniformly scaled to a predetermined resolution to ensure uniformity throughout the dataset. Supplementary augmentation methods, such as horizontal flipping and spatial translation, are employed to replicate variations in image orientation and location. The photos are subsequently transformed into organized numerical formats appropriate for computational analysis. These changes mitigate overfitting, enhance resilience to spatial fluctuations, and facilitate the model's ability to learn discriminative visual patterns successfully.

Detection Dataset Integration and Configuration: For the detection job, various picture datasets are amalgamated into a singular dataset to enhance sample diversity and representation of distinct lung disorders. The consolidated collection is arranged with structured annotation definitions that link image locations to their respective class labels. A uniform dataset configuration is established to delineate data pathways and category descriptions. This stage guarantees compatibility with object identification frameworks, promotes training efficiency, and improves the model's capacity to reliably identify and localize suspicious areas across diverse imaging sources.

Dataset Loading: A systematic workflow for dataset loading and transformation is established to organize and prepare imaging data for model training. This stage entails obtaining picture file locations and doing certain adjustments to ensure uniform data formatting and quality. The transformation pipeline guarantees consistent preprocessing for all samples, minimizing variability and improving training stability. Automating data preparation enhances processing performance, facilitates reproducibility, and guarantees the structural integrity of the dataset during the model development process.

Stratified Data Partitioning: Stratified data partitioning is executed to segment the dataset into training and evaluation subsets, maintaining class distribution across all categories. This method guarantees that each subset reflects the properties of the whole dataset, hence mitigating class imbalance issues that could adversely impact model training. Balanced partitioning boosts model dependability, improves assessment fairness, and facilitates generalization to unseen data sets. Ensuring representative distributions within subsets facilitates consistent performance and mitigates bias during model validation and testing.

Parallel Dataset Structuring for Classification: A parallel dataset structure is established for classification tasks by associating each processed image with its respective diagnostic label. This organized framework guarantees effective data

retrieval and precise correlation between input photos and output categories. The parallel configuration facilitates effortless connection with classification models and streamlines the management of training workflows. This phase promotes data accessibility, improves training accuracy, and maintains stable performance across various classification architectures by keeping constant input-label connections.

c) Algorithms:

Classification:

ResNet-50: ResNet-50 improves classification efficacy by utilizing residual connections that stabilize the training of deep networks and maintain hierarchical feature representations. It facilitates strong visual pattern extraction by preserving gradient flow across layers, hence enhancing generalization and ensuring accurate recognition of intricate structural differences in imaging data.

DenseNet-169: DenseNet-169 enhances feature learning via dense connectivity, facilitating efficient feature reuse and superior gradient propagation. This design improves classification stability and parameter efficiency while collecting both detailed and abstract image representations, enhancing discriminative performance and minimizing overfitting.

EfficientNetV2-Medium: EfficientNetV2-Medium enhances classification accuracy by optimizing network scaling in terms of depth, width, and resolution. It facilitates efficient feature extraction with less computing complexity while preserving contextual and spatial information, hence promoting accelerated training convergence and dependable generalization across varied visual patterns.

ConvNeXt-Base: ConvNeXt-Base enhances convolutional learning by integrating transformer-inspired design elements while preserving convolutional efficiency. It improves spatial feature representation and contextual comprehension, hence enhancing classification robustness and consistency across intricate visual systems through optimized architectural designs.

InceptionNeXt-Base: InceptionNeXt-Base improves multi-scale feature extraction by integrating varied receptive fields through concurrent processing pathways. This architecture successfully captures localized details and global contextual data, enhancing classification accuracy and adaptability across diverse imaging characteristics and intricate structure patterns.

MobileViT-Small: MobileViT-Small combines lightweight convolutional operations with transformer-based attention mechanisms to achieve a compromise between computational efficiency and contextual learning. It captures long-range dependencies while maintaining spatial inductive biases, facilitating scalable implementation with dependable classification performance in resource-limited settings.

ConViT-Base: ConViT-Base enhances feature learning by integrating convolutional inductive biases with transformer-based attention mechanisms. It dynamically equilibrates local spatial representations with global contextual information, thereby improving convergence stability, interpretability, and classification reliability across diverse visual patterns.

Swin-Base: Swin-Base utilizes hierarchical transformer architecture with shifted window attention to effectively capture multi-scale spatial interdependence. It improves feature aggregation and contextual representation while minimizing

computing complexity, hence enhancing classification accuracy and ensuring robust generalization across high-resolution visual inputs.

MaxViT-Base: MaxViT-Base amalgamates convolutional processes with dual-axis attention methods to concurrently collect local details and global context. This hybrid formulation augments feature representation capacity, enhancing classification resilience and facilitating effective modeling of intricate visual interactions across various image structures.

DeiT3-Base: DeiT3-Base enhances classification efficacy via data-efficient transformer learning and optimized attention mechanisms. It emphasizes the extraction of distinctive global characteristics while ensuring training stability, facilitating consistent classification performance and enhanced generalization across diverse visual recognition tasks.

Lightweight Hybrid CNN–ViT with Attention: The lightweight hybrid CNN–ViT architecture integrates convolutional feature extraction with transformer-based contextual modeling, augmented by sophisticated attention mechanisms. It efficiently captures multi-scale spatial information and global dependencies, enhancing classification robustness, interpretability, and scalability while preserving parameter efficiency.

Detection:

YOLOv5: YOLOv5 executes real-time object detection by concurrently predicting object location and classification inside a cohesive framework. It improves detection precision and inference velocity via enhanced spatial feature learning, facilitating dependable identification and localization of anomalies in various visual contexts.

YOLOv8: YOLOv8 enhances detection accuracy with anchor-free prediction and improved feature aggregation techniques. It facilitates strong localization among objects of diverse scales and shapes, enhancing generalization performance while ensuring efficient inference speed and detection reliability.

YOLOv9: YOLOv9 augments detection stability through the enhancement of feature representation learning and the optimization of gradient flow across network layers. It facilitates precise localization and classification of intricate visual patterns while preserving equilibrium in detection precision and recall across various detection contexts.

YOLOv11: YOLOv11 enhances single-stage detection efficacy via optimized feature extraction and improved prediction methodologies. It enhances localization precision and computational efficiency, facilitating scalable and resilient object identification while preserving real-time processing capabilities across intricate visual tasks.

e) Integration of XAI & Flask:

The incorporation of Explainable Artificial Intelligence (XAI) improves model transparency by offering visual elucidations of prediction results. GradCAM-based visualization is utilized to produce heatmaps that emphasize significant image areas affecting classification and detection outcomes. This interpretability mechanism aids in comprehending model reasoning by pinpointing discriminative elements that contribute to the recognition of abnormalities. The visualization enhances clinical dependability by allowing verification of model attention regions, hence increasing trust and aiding the confirmation of automated diagnostic results through intuitive visual feedback.

The Flask framework is incorporated to offer an intuitive deployment interface that facilitates smooth interaction with trained models. It facilitates the submission of picture inputs, the processing of predictions, and the viewing of displays within a cohesive environment. This lightweight web interface improves accessibility, enabling users to conduct real-time analysis without necessitating specialist technical skills. The deployment framework guarantees scalability, effective model integration, and seamless communication between backend prediction modules and frontend visualization elements.

4. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test refers to its capacity to correctly distinguish between patient and healthy cases. To assess the accuracy of a test, one must calculate the ratio of true positives and true negatives across all assessed cases. This can be expressed mathematically as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} (1)$$

Precision: Precision assesses the proportion of accurately classified cases among those identified as positive. Consequently, the formula for calculating precision is expressed as:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)$$

Recall: Recall is a metric in machine learning that assesses a model's capacity to recognize all pertinent instances of a specific class. The ratio of accurately predicted positive observations to the total actual positives, offering insights into a model's efficacy in identifying occurrences of a specific class.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} (3)$$

F1-Score: The F1 score is a metric for evaluating the accuracy of a machine learning model. It integrates the precision and recall metrics of a model. The accuracy metric quantifies the frequency of true predictions generated by a model throughout the entire dataset.

$$F1\ Score = 2 * \frac{Recall\ \mathrm{X}\ Precision}{Recall + Precision} * 100 (4)$$

mAP: Mean Average Precision (MAP) is a statistic for evaluating ranking quality. It evaluates the quantity of pertinent recommendations and their placement within the list. MAP at K is determined as the arithmetic mean of the Average Precision (AP) at K for all users or queries.

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \quad (5)$$

Table.1 Performance Evaluation Table – IQ-OTHNCCD

ML Model	Accuracy	Precision	Recall	F1-Score
Proposed Hybrid	0.98643	0.98851	0.95833	0.97193
ResNet50	0.97738	0.95925	0.95040	0.95468
DenseNet169	0.98190	0.97322	0.95437	0.96317
EfficientNetV2-Medium	0.98643	0.98851	0.95833	0.97193
ConvNeXt-Base	**0.99095**	**0.99225**	**0.97222**	**0.98159**

InceptionNext-Base	0.98190	0.97322	0.95437	0.96317
MobileViT-Small	0.98643	0.98851	0.95833	0.97193
ConViT-Base	0.96833	0.96264	0.93254	0.94623
Swin-Base	0.87330	0.82051	0.88889	0.81053
MaxViT-Base	0.95023	0.94194	0.85714	0.88527
DeiT3-Base	0.97285	0.95419	0.93651	0.94476

Table 1 delineates the performance assessment of various deep learning models. ConvNeXt-Base attains the greatest accuracy (0.99095) and F1-score (0.98159), however the Proposed Hybrid model exhibits robust and competitive performance overall.

Table.2 Performance Evaluation Table – Chest-CT

ML Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Proposed Hybrid	0.97015	0.96954	0.97480	0.97193
ResNet50	0.98507	0.98416	0.98697	0.98534
DenseNet169	0.98010	0.98226	0.97908	0.98061
EfficientNetV2-Medium	0.98507	0.98255	0.98897	0.98547
ConvNeXt-Base	0.98010	0.97852	0.98329	0.98078
Inception Next-Base	**0.99005**	**0.99064**	**0.99064**	**0.99064**
MobileViT-Small	0.99005	0.98870	0.99064	0.98956
ConViT-Base	0.97512	0.97240	0.97961	0.97560
Swin-Base	0.96020	0.97080	0.95487	0.96140
MaxViT-Base	0.92040	0.94394	0.92339	0.92769
DeiT3-Base	0.98010	0.97991	0.98329	0.98143

Table 2 encapsulates the performance evaluation outcomes of diverse deep learning models. InceptionNext-Base attains the greatest accuracy and F1-score (0.99064), closely followed by MobileViT-Small and EfficientNetV2-Medium, indicating exceptional classification performance overall.

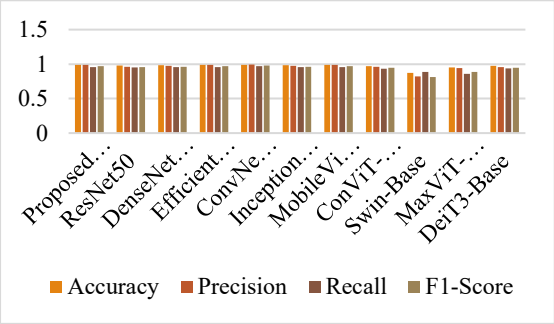
Table.3 Performance Evaluation Table – Detection

ML Model	**Precision**	**Recall**	**mAP**
Yolo v5	0.732	0.684	0.728
Yolo v8	**0.767**	**0.686**	**0.727**

Yolo v9	0.746	0.557	0.650
Yolo v11	0.415	0.497	0.454

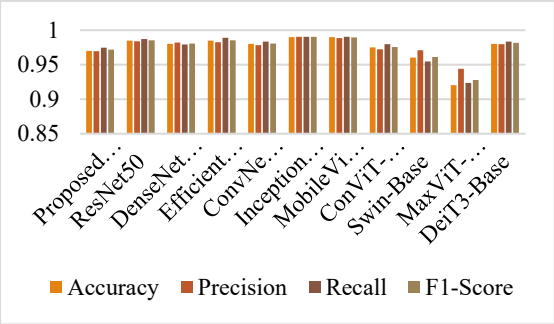
Table 3 delineates the comparative performance of YOLO models. YOLOv8 attains the best precision (0.767), although YOLOv5 registers the highest mean Average Precision (mAP) (0.728). YOLOv11 exhibits relatively inferior overall detection performance.

Graph.1 Comparison Graph – IQ-OTHNCCD



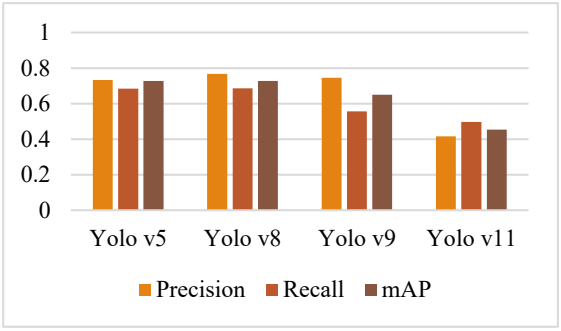
Graph 1 depicts the comparative efficacy of different deep learning models according to Accuracy, Precision, Recall, and F1-Score. Most models exhibit consistently elevated metrics, with minor variances underscoring performance discrepancies among architectures.

Graph.2 Comparison Graph – Chest-CT



Graph.2 juxtaposes the efficacy of various deep learning models utilizing Accuracy, Precision, Recall, and F1-Score as metrics. InceptionNext-Base and MobileViT-Small provide greater performance, whilst MaxViT-Base exhibits comparatively lower metrics.

Graph.3 Comparison Graph – Detection



Graph 3 depicts the comparative efficacy of YOLO models utilizing Precision, Recall, and mAP metrics. YOLOv8 attains superior precision, although YOLOv5 has competitive overall detection efficacy relative to its counterparts.

Fig.4 Choose File

Figure 4 depicts the interface of the chest CT categorization system, wherein users submit CT scan images for automatic evaluation. The platform accommodates many file types and delivers AI-driven diagnostic results effectively.

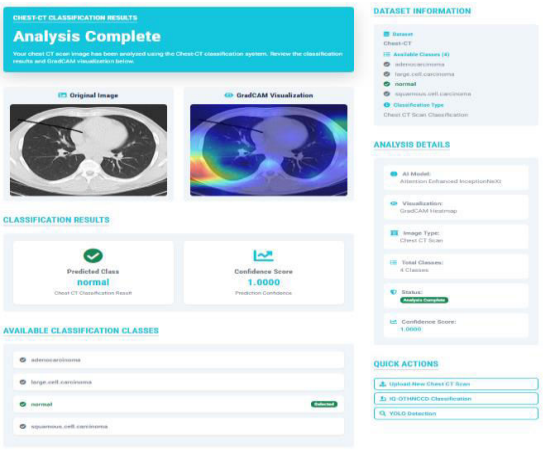


Fig.5 Predicted Result

Figure 5 displays the chest CT classification result labeled "Normal," accompanied by a confidence score of 1.0000, signifying absolute prediction certainty, further substantiated by Grad-CAM imagery emphasizing pertinent lung areas.

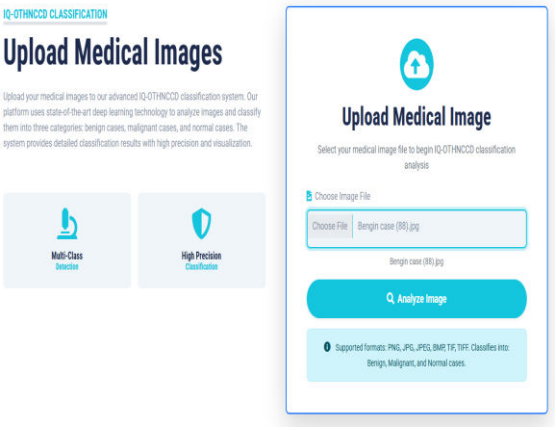


Fig.6 Choose File

Figure 6 depicts the IQ-OTHNCCD medical image upload interface, allowing users to input medical images for multi-class classification. The system accommodates several formats and conducts high-precision analyses categorizing samples as benign, malignant, or normal.

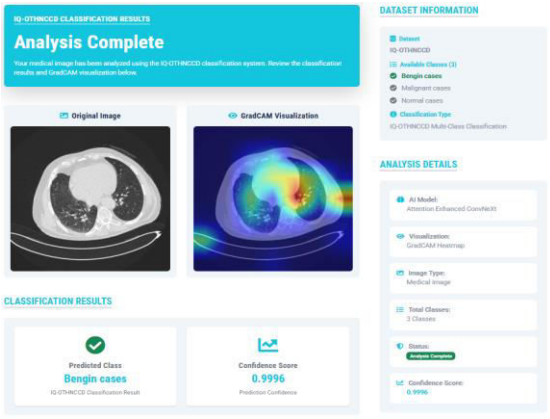


Fig.5 Predicted Result

Figure 5 displays the IQ-OTHNCCD classification results categorized as “Benign cases,” accompanied by a confidence score of 0.9996, signifying exceptional predictive accuracy, corroborated by Grad-CAM imagery that emphasizes critical diagnostic areas.

5. CONCLUSION

Lung cancer continues to be the primary cause of cancer-related deaths, requiring precise and prompt detection to enhance patient outcomes and inform treatment decisions. The system employs an attention-augmented hybrid CNN–Vision Transformer (ViT) framework for the classification and diagnosis of lung cancer, leveraging two publically accessible datasets: the IQ-OTH/NCCD Lung Cancer Dataset and the Chest CT-Scan Images Dataset. The preprocessing processes encompass image resizing to 299×299 pixels, data augmentation, tensor normalization, and stratified splitting to preserve class equilibrium. Classification is executed via various baseline architectures, such as ResNet50, DenseNet169, EfficientNetV2-Medium, ConvNeXt-Base, InceptionNeXt-Base, MobileViT-Small, ConViT-Base, Swin-Base, MaxViT-Base, and DeiT3-Base, whilst detection is managed via YOLOv5, YOLOv8, YOLOv9, and YOLOv11. Improvements comprise hybrid CNN–

ViT architecture with InceptionNeXt blocks, grid and block attention methods, GradCAM for interpretable visuals, and a Flask-based interface for efficient deployment and user engagement. ConvNeXt-Base achieves a classification accuracy of 99.09% on the IQ-OTH/NCCD dataset, InceptionNeXt-Base attains 99.01% accuracy on the chest CT-scan dataset, and YOLOv5 records a mean average precision of 72.8% for detection. The system exhibits exceptional dependability, balanced performance, and interpretability, offering automated, explainable, and clinically pertinent decision support for lung cancer diagnosis, facilitating efficient and resilient real-world implementation.

The system can be improved by incorporating larger, multi-institutional datasets to boost generalization across varied demographics and imaging modalities. Integrating multi-class categorization for diverse lung cancer subtypes and stages can enhance clinical relevance. Advanced attention mechanisms and transformer topologies may be utilized to enhance the detection of nuanced patterns in CT images. Real-time deployment on cloud platforms or edge devices can improve accessibility and scalability in clinical environments. Furthermore, the incorporation of automated reporting, risk assessment, and decision-support dashboards might aid radiologists in optimizing workflow, hence facilitating swifter, more precise, and interpretable lung cancer diagnosis and therapy.

REFERENCES

[1] Revathi, T. (2025). Hybrid Transformer-Based Framework for Multi-Type Lung Cancer Detection Using Attention-Guided Feature Fusion. Procedia Computer Science, 270, 4533-4542.

[2] Shariff, V., Paritala, C., & Ankala, K. M. (2025). Optimizing non small cell lung cancer detection with convolutional neural networks and differential augmentation. Scientific Reports, 15(1), 15640.

[3] Golkarieh, A., Kiashemshaki, K., Boroujeni, S. R., & Isakan, N. A. (2025). Advanced U-Net architectures with CNN backbones for automated lung cancer detection and segmentation in chest CT images. arXiv preprint arXiv:2507.09898.

[4] Kumar, V., Prabha, C., Sharma, P., Mittal, N., Askar, S. S., & Abouhawwash, M. (2024). Unified deep learning models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images. BMC medical imaging, 24(1), 63.

[5] Zhang, C., Aamir, M., Guan, Y., Al-Razgan, M., Awwad, E. M., Ullah, R., ... & Ghadi, Y. Y. (2024). Enhancing lung cancer diagnosis with data fusion and mobile edge computing using DenseNet and CNN. Journal of Cloud Computing, 13(1), 91.

[6] S. H. Hosseini, R. Monsefi, and S. Shadroo, “Deep learning applications for lung cancer diagnosis: A systematic review,” *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 14305–14335, Feb. 2024, doi: 10.1007/s11042-023-16046-w.

[7] R. L. Siegel, A. N. Giaquinto, and A. Jemal, “Cancer statistics, 2024,” *CA, A Cancer J. Clinicians*, vol. 74, no. 1, pp. 12–49, Jan. 2024, doi: 10.3322/caac.21820.

[8] A D Venkatesh, K Bhaskar, G Swapna, & G Viswanath. (2025). Advanced Hybrid Learning Architecture for Precision Cardiovascular Risk Assessment. In International Journal of Health Sciences and Pharmacy (IJHSP) (Vol. 9, Number 1, pp. 50–61). Zenodo. https://doi.org/10.5281/zenodo.15448632

[9] D. Riquelme and M. Akhloufi, "Deep learning for lung cancer nodules detection and classification in CT scans," *AI*, vol. 1, no. 1, pp. 28–67, Jan. 2020, doi: 10.3390/ai1010003.

[10] A. Atmakuru *et al*., "Deep learning in radiology for lung cancer diagnostics: A systematic review of classification, segmentation, and predictive modeling techniques," *Expert Syst. Appl.*, vol. 255, Dec. 2024, Art. no. 124665, doi: 10.1016/j.eswa.2024.124665.

[11] R. Javed *et al*., "Deep learning for lungs cancer detection: A review," *Artif. Intell. Rev.*, vol. 57, no. 8, p. 197, Aug. 2024, doi: 10.1007/s10462-024-10807-1.

[12] L. Wang, "Deep learning techniques to diagnose lung cancer," *Cancers*, vol. 14, no. 22, p. 5569, Nov. 2022, doi: 10.3390/cancers14225569.

[13] A. Halder and D. Dey, "MorphAttnNet: An attention-based morphology framework for lung cancer subtype classification," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105149, doi: 10.1016/j.bspc.2023.105149.

[14] G Loge, T Sunil Kumar Reddy, G Swapna, & G Viswanath. (2025). Interpretable AI for Precision Brain Tumor Prognosis: A Transparent Machine Learning Approach. In International Journal of Health Sciences and Pharmacy (IJHSP) (Vol. 9, Number 1, pp. 180–195). Zenodo. https://doi.org/10.5281/zenodo.15523628

[15] L. Ma, H. Wu, and P. Samundeeswari, "GoogLeNet-AL: A fully automated adaptive model for lung cancer detection," *Pattern Recognit.*, vol. 155, Nov. 2024, Art. no. 110657, doi: 10.1016/j.patcog.2024.110657.

[16] N. Gautam, A. Basu, and R. Sarkar, "Lung cancer detection from thoracic CT scans using an ensemble of deep learning models," *Neural Comput. Appl.*, vol. 36, no. 5, pp. 2459–2477, Feb. 2024, doi: 10.1007/s00521-023-09130-7.

[17] N. A. Wani, R. Kumar, and J. Bedi, "DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence," *Comput. Methods Programs Biomed.*, vol. 243, Jan. 2024, Art. no. 107879, doi: 10.1016/j.cmpb.2023.107879.

[18] A. Heidari *et al*., "A new lung cancer detection method based on the chest CT images using federated learning and blockchain systems," *Artif. Intell. Med.*, vol. 141, Jul. 2023, Art. no. 102572, doi: 10.1016/j.artmed.2023.102572.

[19] A. R. Bushara *et al*., "An ensemble method for the detection and classification of lung cancer using computed tomography images utilizing a capsule network with visual geometry group," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104930, doi: 10.1016/j.bspc.2023.104930.

[20] R. Raza *et al*., "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 106902, doi: 10.1016/j.engappai.2023.106902.

[21] J. Subash and S. Kalaivani, "Dual-stage classification for lung cancer detection and staging using hybrid deep learning techniques," *Neural Comput. Appl.*, vol. 36, no. 14, pp. 8141–8161, May 2024, doi: 10.1007/s00521-024-09425-3.

[22] M. Nahiduzzaman *et al*., "A novel framework for lung cancer classification using lightweight convolutional neural networks and ridge extreme learning machine model with Shapley additive

exPlanations (SHAP)," *Expert Syst. Appl.*, vol. 248, Aug. 2024, Art. no. 123392, doi: 10.1016/j.eswa.2024.123392.

[23] I. Naseer *et al*., "Lung cancer classification using modified U-Net based lobe segmentation and nodule detection," *IEEE Access*, vol. 11, pp. 60279–60291, 2023, doi: 10.1109/ACCESS.2023.3285821.

[24] Kumar, C. S., Sirisati, R. S., Gudditti, V., Rao, K. S., & Challa, R. K. (2022, December). A smart recommendation system for medicine using intelligent NLP techniques. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 1081-1084). IEEE. https://doi.org/10.1109/ICACRS55517.2022.10029078

[25] H. Xiao, Q. Liu, and L. Li, "MFMANet: Multi-feature multi-attention network for efficient subtype classification on non-small cell lung cancer CT images," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104768, doi: 10.1016/j.bspc.2023.104768.

[26] R. Mahum and A. S. Al-Salman, "Lung-RetinaNet: Lung cancer detection using a RetinaNet with multi-scale feature fusion and context module," *IEEE Access*, vol. 11, pp. 53850–53861, 2023, doi: 10.1109/ACCESS.2023.3281259.

[27] S. U. Atiya, N. V. K. Ramesh, and B. N. K. Reddy, "Classification of non-small cell lung cancers using deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 13261–13290, Jul. 2023, doi: 10.1007/s11042-023-16119-w.

[28] H. Alyasriy, "The IQ-OTHNCCD lung cancer dataset," vol. 1, doi: 10.17632/BHMDR45BH2.1.

[29] *Chest CT-Scan Images Dataset*. Accessed: Oct. 8, 2024. [Online]. Available: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images

[30] Swapna, G., Sreenivasulu, K., Deepika, M., Baseer, K. K., Neerugatti, V., & Viswanath, G. (2025). Brain tumour detection using MRI images in CNN.