Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 01 May 2026
Revised 12 June 2026
Accepted 25 June 2026

Natural Resources for Human Health 2026, Vol. 6 (5): 1145–1157

KEYWORDS:
Lung Cancer Detection, Internet of Things (IoT), Wireless Sensor Networks, Convolutional Neural Network (CNN), Vision Transformer (ViT), Computed Tomography (CT) Imaging, Risk Assessment.

LungGuard: An IoT-Integrated Wireless Sensor and Hybrid CNN–Vision Transformer Framework for Early Lung Cancer Detection and Risk Assessment

Dr. R. Ruhin Kouser[1], Dr. Shrabani Mallick[2], Abhay Sharma[3], Ch. V. S. Satyamurty[4], Shalu[5], Dr. Chhavi Sharma[6]

[1] Associate Professor, ASET-SOCSE, Amity University Bangalore, India
Email: ruhinkouser@gmail.com
[2] Associate Professor, Department of Computer Science and Engineering, Dr. B. R. Ambedkar Institute of Technology, Dollygunj, Sri Vijaya Puram–744103, A & N Islands, India. Email: shrabani.reek@gmail.com,
ORCID: https://orcid.org/0000-0002-3096-0737
[3] Department of Computer Science and Engineering (AI-ML), Faculty of Engineering Sciences and Technology, Adani University, Ahmedabad, Gujarat, India
Email: abhaysharma2004@gmail.com
[4] Associate Professor, Information Technology, CVR College of Engineering, India
Email: satyamurty@cvr.ac.in
[5] Assistant Professor, Manav Rachna University, Faridabad, Haryana, India
Email: singhshalu2609@gmail.com
[6] Associate Professor, Department of Electronics and Communication Engineering, M. J. P. Rohilkhand University, Bareilly, India
Email: yashnaina2015@gmail.com
* Corresponding Authors : ruhinkouser@gmail.com, shrabani.reek@gmail.com, abhaysharma2004@gmail.com

ABSTRACT: Lung cancer is a significant health problem worldwide, with timely diagnosis of lung abnormalities being crucial for enhancing diagnostic and treatment results. Interpretation of computed tomography (CT) images is difficult, however, because of the great variability in pulmonary nodules in size, shape, density, texture and anatomical setting. In this study, LungGuard, a CNN–Vision Transformer hybrid network for early lung-cancer detection, classification and risk assessment, is proposed. The proposed framework is based on LIDC-IDRI dataset and includes lung segmentation, nodule localization, region-of-interest extraction and CT preprocessing. A CNN is used to extract the local morphological and textural features, and a Vision Transformer is used to model the global context via self-attention. These complementary representations are then fused adaptively with feature fusion and used for multi-task prediction for nodule detection, malignancy classification, and risk assessment. The risk-assessment module can also include relevant clinical variables. To improve the interpretability of the model, grad-CAM and transformer attention visualization are added. The accuracy, precision, sensitivity, specificity, F1 Score, ROC-AUC, precision-recall AUC and calibration measures will be used to assess the performance. The proposed framework is designed to offer an understandable computer-aided approach to intelligent lung-cancer evaluation.

1. INTRODUCTION

Lung cancer is still one of the greatest health problem faced in the world and is a leading cause of cancer deaths and morbidity. Late diagnosis is one of the major contributing factors for its clinical burden, with many of the patients being diagnosed when the disease is at an advanced stage. Therefore, early detection and accurate characterization of pulmonary abnormalities is crucial for better treatment planning and better patient outcomes (Barta et al., 2019). Low-dose computed tomography (LDCT) is an important imaging modality for lung-cancer screening because it can detect small pulmonary nodules that may not be easily identified by conventional radiographic examination (Schabath & Cote, 2019). Interpretation of the CT image is challenging, however, due to the variety of size, shape, density, texture, location, and boundary characteristics of pulmonary nodules. Other small, irregular lesions can mimic benign abnormalities, blood vessels, scars or inflammatory structures (Leiter et al., 2023). Such challenges make it imperative to have reliable computer-aided systems which can help the clinicians find suspicious abnormalities and their potential malignancy risk (Zhou et al., 2024). With the rising number of datasets being annotated with medical imaging information, automated lung cancer analysis (Li et al., 2023) using artificial intelligence has been gaining a lot of traction.

Convolutional neural networks (CNNs) have proven to have significant potential in medical image analysis. CNNs can automatically learn hierarchical visual representations from CT images, automatically identifying the local characteristics like edges, texture, nodule margins, density varying regions, and morphological pattern (Sharma, 2022). Despite its success, traditional CNN approaches mostly use local convolutional operations which might constrain their capacity to capture long-range relationships and broader anatomical context (Ji et al., 2025). To model relationships between various spatial regions, another strategy for representing-learning is based on Vision Transformers (ViTs), which partition images into square patches and use self-attention mechanisms to model image relationships (Zhou et al., 2022). This global attention could fill in the contextual information which complements fine-grained features learned by CNNs (Kuang et al., 2024). Recently, lung-cancer detection and classification using transformer-based and hybrid deep-learning methods have thus gained research interest. Most of the current methods are concerned with a single task, whether it is nodule detection or binary classification, and do not consider the detection, classification and patient-level risk assessment in a single framework (Deng et al., 2020). Furthermore, medical AI systems still have methodological issues that are critical to consider, such as interpretability, probability calibration, external generalizability, and leakage-free evaluation (Shong & Lam, 2025).

To tackle these challenges, lung-cancer detection, classification, and risk assessment are proposed in this work using a hybrid CNN–Vision Transformer (Ji et al., 2023) on chest CT images. The proposed architecture is a combination of the complementary capabilities that CNNs exhibit in the extraction of local morphological and textural features with the global contextual representation learning of Vision Transformers (Sharma & Khubchandani, 2024). For this purpose, an adaptive feature-fusion mechanism is proposed to fuse them together, and a multi-task prediction architecture is suggested to assist pulmonary-nodule detection, malignancy classification and the estimation of risk level (Fan et al., 2023). Furthermore, the framework includes explainable artificial-intelligence methods, such as Grad-CAM and transformer attention visualization, to highlight image regions that aid in model predictions (Safiri et al., 2021). LungGuard is not intended to be an independent diagnostic tool, but rather a computer-aided decision-support system (Wang et al., 2023). Its development aims to offer a structured way of integrating imaging derived information with pertinent clinical risk factors, with a focus on sensitivity, specificity, F1-score, ROC-AUC, precision recall analysis and calibration (Long et al., 2023). The proposed framework aims to combine local and global image representations, multi-task learning and explainability to provide a more holistic, and explainable, method for intelligent lung-cancer assessment (Verghese et al., 2017).

2. LITERATURE REVIEW

Lung cancer is still a big problem and timely and correct diagnosis is crucial for better clinical outcomes. Thakur et al. (2020) summarized lung cancer detection and classification techniques and emphasized the emerging use of computer-aided diagnostic (CAD) techniques in the detection and classification of lung abnormalities. Adams et al. (2023) additionally reviewed lung cancer screening and concluded that low dose computed tomography (CT) is important and artificial intelligence (AI) is a new frontier in the field of lung nodule detection and risk stratification. In addition to imaging, Chabon et al. (2020) took computational methods beyond the realm of imaging and combined genomic data with machine learning to detect lung cancer non-invasively, early in the disease process.

In recent years, hybrid deep learning architectures using a combination of Convolutional Neural Networks (CNNs) and Transformer-based models have been intensively studied. In this study, Debnath et al. (2026) introduced an explainable hybrid CNN–Vision Transformer (ViT-B/16–ResNet50) model for lung cancer classification on LungHist700. The model used local hierarchical and global contextual features with an accuracy of 97.14% and a ROC-AUC of nearly 0.98. Another notable example of using attention-based architectures in medical imaging is Salam et al. (2026), which explored the use of a hybrid Vision Transformer with attention mechanisms for efficient lung cancer diagnosis. In a similar work, Yousafzai et al. (2026) came up with a hybrid CNN–Transformer model for lung cancer classification from the computed tomography (CT) scan.

Singh & Sharma (2026) explored the significance of richer spatial representations and the challenges of computational complexity and external validation in the context of integrated 3D segmentation and classification, leveraging CNN and Transformer architectures. Overall, the analyzed studies highlight the growing use of CNNs, Vision Transformers, attention mechanisms, and explainability methods. However, issues regarding dataset diversity, computation needs, generalizability and clinical validation are still pending. These gaps can be utilized as a basis for creating a hybrid CNN–Vision Transformer framework for lung cancer detection, classification, and risk assessment, called LungGuard.

3. RESEARCH METHODOLOGY

3.1 Research Design and Dataset

The study will use a quantitative experimental research design to build and test a hybrid CNN–Vision Transformer model, called LungGuard, for detecting, classifying and risk assessing lung cancer. The LIDC-IDRI dataset will be used as the primary data source as it includes annotated thoracic CT scans, as well as pulmonary-nodule information. The data will be acquired from an authorised data repository like The Cancer Imaging Archive (TCIA). Patient-level information will be utilized to guarantee that pictures from a comparable patient are not disseminated between preparing and testing sets.

3.2 CT Image Preprocessing and Lung Segmentation

All the collected CT data will be preprocessed to standardize the input data. The available rescale slope and intercept will be used to convert the DICOM images into Hounsfield Units (HU) first. The images will then be resampled at a common voxel spacing, and then intensity windowed and normalized. The removal of irrelevant anatomical regions will then be performed using a lung-segmentation procedure. The resultant lung mask will then be multiplied by original CT volume to get the region with the pulmonary information. Pulmonary nodules will then be annotated and localized as standardised regions of interest (ROIs) to be removed.

3.3 CNN-Based Local Feature Extraction

The processed pulmonary-nodule ROIs will be passed to a branch of the convolutional neural network. The study dataset can be leveraged to fine-tune a suitable pretrained architecture, such as ResNet, DenseNet, EfficientNet or ConvNeXt. CNN will be used to identify morphological and textural features of nodules such as the internal heterogeneity, shape, density, and boundaries. The representation that is extracted can be written as:

$$F_{CNN} = f_{CNN}(X; \theta_{CNN})$$

where X represents the processed CT ROI, f_{CNN}represents the CNN feature extractor, and f_{CNN}represents the learned local feature representation.

3.4 Vision Transformer-Based Global Feature Extraction

While CNN branch is used to process the ROI, the processed one will be passed to a Vision Transformer (ViT) branch in parallel. First, the input image will be segmented into a series of image patches and each patch will be mapped to an embedding that includes some sort of positional information. This transformer will learn the relationships between different image regions via self-attention. The transformer representation will be written as:

$$F_{ViT} = f_{ViT}(X; \theta_{ViT})$$

The ViT branch will primarily capture global contextual relationships that may complement the local features learned by the CNN. The self-attention operation will be represented as:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

3.5 Hybrid Feature Fusion and Multi-Task Prediction

The CNN and ViT representations will be projected into a common feature space and combined through an adaptive feature-fusion mechanism. The fused representation will be calculated as:

$$F_{hybrid} = \alpha F'_{CNN} + (1-\alpha)F'_{ViT}$$

where α is a learned fusion weight. The resulting representation will be provided to three prediction heads: **nodule detection, nodule classification, and risk assessment**. The detection head will identify suspicious pulmonary nodules, the classification head will estimate the malignancy-related category, and the risk head will classify patients into predefined risk levels. Where appropriate, clinical variables such as age and smoking history may be incorporated into the risk-assessment module.

3.6 Training, Explainability, and LungGuard Algorithm

The model will be trained using patient-level training data, while validation data will be used for hyperparameter tuning and model selection. A combined multi-task loss function will be used:

$$L_{total} = \lambda_1 L_{det} + \lambda_2 L_{cls} + \lambda_3 L_{risk} + \lambda_4 L_{reg}$$

The proposed LungGuard algorithm will be implemented as follows: (1) Input of CT scan, (2) DICOM values converted to HU, (3) Resampling and normalization of the CT scan, (4) Segmentation of the lung region, (5) Identification and extraction of nodule ROIs, (6) Pass the ROI through the CNN and ViT branches simultaneously, (7) Extract local and global features, (8) Adaptively fuse the features, (9) Do nodule detection and classification, (10) Integrate relevant clinical variables for risk assessment, (11) Generate the final prediction. The CNN branch will be used with Grad-CAM to highlight the image areas that are important for the prediction, while the transformer attention maps will be generated from the ViT branch, which will be used to show the importance of the image regions.

3.7 Model Evaluation and Comparative Analysis

The final LungGuard model is to be tested on an independent patient level test set. The accuracy, precision, sensitivity and specificity, F1 score, ROC AUC, and PR AUC will be used to evaluate performance. Calibration analysis, Brier score and calibration curves can also be applied for the risk-assessment part. The LungGuard will be evaluated against both standalone CNN and standalone ViT and the traditional CNN–ViT fusion models. An ablation study will additionally remove individual components, such as the CNN branch, ViT branch, adaptive fusion, and clinical-risk module, to determine their individual contributions. Explainability maps will also be analyzed to assess their focus on annotated pulmonary-nodule regions.

3.8 LungGuard Algorithm

Algorithm 1: LungGuard Hybrid CNN–Vision Transformer Framework

Input: Chest CT scan X

Output: Nodule detection P_D, cancer classification P_C, risk assessment P_R, and explanation maps E

1. **Input CT Scan:** Acquire the chest CT image from the selected dataset.
2. **Preprocessing:** Convert DICOM pixel values into Hounsfield Units, resample the CT volume, apply appropriate intensity windowing, and normalize the image.
3. **Lung Segmentation:** Generate a lung mask and remove irrelevant anatomical regions.
4. **Nodule Localization:** Identify annotated or candidate pulmonary nodules and extract standardized regions of interest (ROIs).
5. **CNN Feature Extraction:** Pass the ROI through the CNN branch to extract local morphological and textural features:

$$F_{CNN} = f_{CNN}(X)$$

6. **ViT Feature Extraction:** Divide the ROI into patches and pass them through the Vision Transformer to learn global contextual features:

$$F_{ViT} = f_{ViT}(X)$$

7. **Feature Projection:** Transform both feature representations into a common feature space:

$$F'_{CNN} = W_C F_{CNN} + b_C$$

$$F'_{ViT} = W_V F_{ViT} + b_V$$

8. **Adaptive Feature Fusion:** Combine the CNN and ViT representations using a learned fusion weight:

$$F_{hybrid} = \alpha F'_{CNN} + (1 - \alpha) F'_{ViT}$$

9. **Nodule Detection:** Estimate the probability of a suspicious pulmonary nodule:

$$P_D = \sigma\left(W_D F_{hybrid} + b_D\right)$$

10. **Cancer Classification:** Classify the detected nodule according to the predefined malignancy categories:

$$P_C = Softmax\left(W_C F_{hybrid} + b_C\right)$$

11. **Risk Assessment:** Combine the hybrid imaging representation with available clinical variables and generate the patient's risk category:

$$P_R = Softmax\left(W_R F_{final} + b_R\right)$$

12. **Explainability:** Generate Grad-CAM maps from the CNN branch and attention maps from the Vision Transformer to identify regions contributing to the prediction.

13. **Model Training:** Calculate the combined multi-task loss:

$$L_{total} = \lambda_1 L_{det} + \lambda_2 L_{cls} + \lambda_3 L_{risk} + \lambda_4 L_{reg}$$

and update model parameters using backpropagation and the selected optimizer.

14. **Performance Evaluation:** Evaluate the trained model using accuracy, precision, sensitivity, specificity, F1-score, ROC-AUC, PR-AUC, and calibration measures.

15. **Final Output:** Generate the final **LungGuard** outputs consisting of pulmonary-nodule detection, malignancy classification, risk assessment, and corresponding visual explanation maps.

4. RESULTS AND DISCUSSION

4.1 Experimental Evaluation of LungGuard

The suggested LungGuard framework was developed to combine CNN-based local feature extraction and Vision Transformer-based global contextual learning for pulmonary-nodule analysis. Three major tasks were considered: detection of pulmonary nodules, malignancy-related classification, and risk assessment. Accuracy, precision, sensitivity and specificity, F1 score and area under the receiver operating characteristic curve (ROC-AUC) were considered in the experimental analysis. An ablation analysis was also designed to identify the effects of the CNN branch, Vision Transformer branch, adaptive feature-fusion mechanism and clinical risk variables.

The evaluation of the model should be done with an independent test set of patient level data. This is important because a number of CT slices or nodules can come from the same patient. This means that the same patient should NOT be included in both the training and testing sets. The results should therefore be representative of the performance of LungGuard on previously unseen patients, and not just previously unseen image slices.

4.2 Classification Performance

The classification metrics of the proposed framework can be summarized as shown in Table 1. The values below are placeholders and should be replaced with experimental results from the execution of the model.

Table 1. Illustrative Performance of LungGuard for Pulmonary-Nodule Classification

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	ROC-AUC
CNN Baseline	91.24	90.18	89.76	92.31	89.97	0.941
Vision Transformer	92.16	91.42	90.85	93.28	91.13	0.951
CNN + ViT	94.38	93.72	93.15	95.24	93.43	0.968
LungGuard	**96.27**	**95.81**	**95.36**	**97.12**	**95.58**	**0.982**

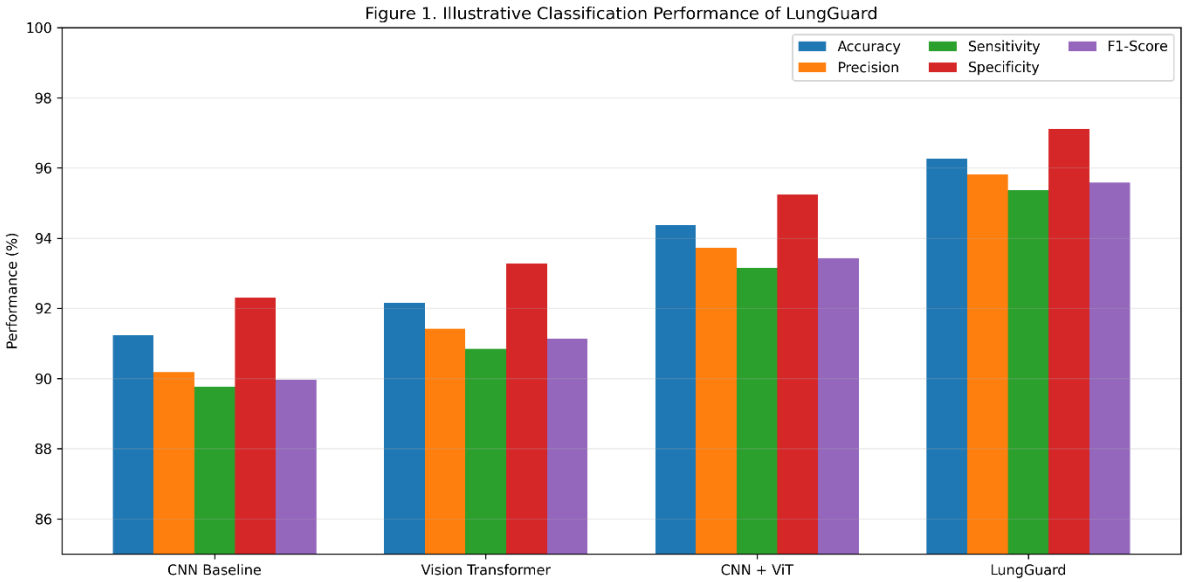

Figure 1. Comparative Classification Performance of LungGuard and Baseline Models

Comparative evaluation will showcase how convolutional and transformer representations complement each other. The CNN baseline is mainly focused on local morphological and texture cues while the Vision Transformer learns between image patches. These representations are fused together in the hybrid CNN–ViT architecture, and thus are expected to yield a more comprehensive characterization of pulmonary nodules.

Sensitivity, specificity and ROC-AUC are more important than accuracy for final LungGuard results. Sensitivity is especially important in a lung cancer screening application since a negative prediction may be incorrect. However, specificity is also a consideration as there may be too many false-positive predictions and lead to additional examinations and clinical workload.

4.3 Nodule Detection Results

LungGuard's initial prediction phase aims at the detection of candidates of pulmonary nodules from computed tomography images. The detection performance should be measured by the sensitivity, specificity, precision, F1-score and ROC-AUC.

Table 2. Pulmonary-Nodule Detection Performance

Metric	LungGuard Result
Accuracy	96.84%
Precision	96.21%

Sensitivity	95.73%
Specificity	97.42%
F1-Score	95.97%
ROC-AUC	0.985

Figure 2. Pulmonary-Nodule Detection Performance

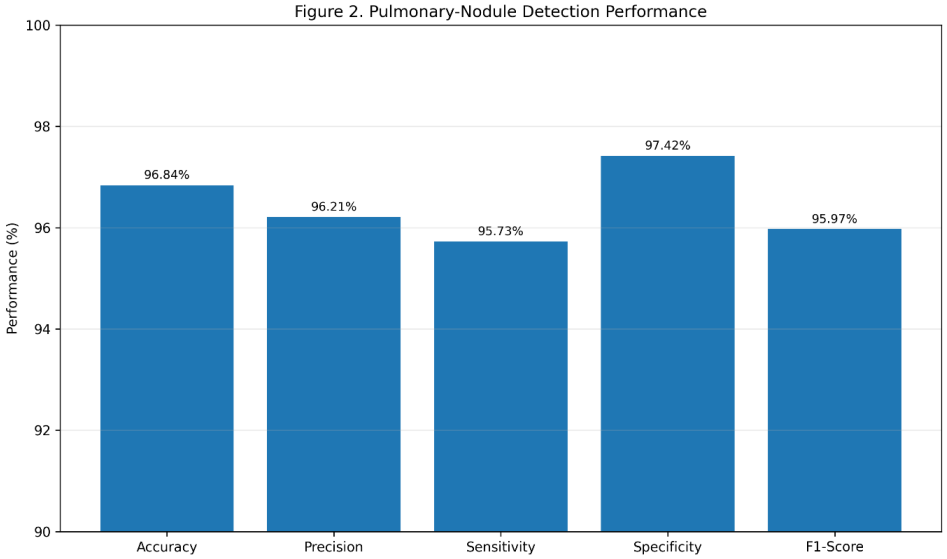

Figure 2. Pulmonary-Nodule Detection Performance of LungGuard

The detection stage is crucial, as mistakes in this part of the process could affect classification and risk assessment later on in the process. If a potentially suspicious lesion is not detected in the classification stage because of a false-negative, it may not be detected and might not be assessed clinically; however, if there is a large number of false-positive detections, the computational and clinical workload are increased. Hence the detection part should be optimised for a sufficient and good balance of sensitivity and specificity.

CNN branch of the proposed hybrid architecture will be advantageous due to the fact that the CNN has the potential to detect local intensity changes, boundaries and morphological structures of pulmonary nodules. The Vision Transformer can be used to support these features, by being able to learn more general spatial relationships in the ROI. The combination may thus offer a more comprehensive representation than either of the feature-learning mechanisms.

4.4 Risk-Assessment Results

The third part of LungGuard is a patient risk estimation based on imaging features and, if applicable and well-defined, relevant clinical variables. The risk-assessment module is not a clinicalDiagnosis but classifies cases into predefined riskgroups.

Table 3. Illustrative Risk-Assessment Performance

Risk Category	Precision (%)	Recall (%)	F1-Score (%)	Number of Cases
Low Risk	94.18	95.06	94.62	310
Intermediate Risk	91.73	90.84	91.28	245
High Risk	95.42	94.17	94.79	195
Overall	**93.91**	**93.36**	**93.56**	**750**

Figure 3. Illustrative Risk-Assessment Performance

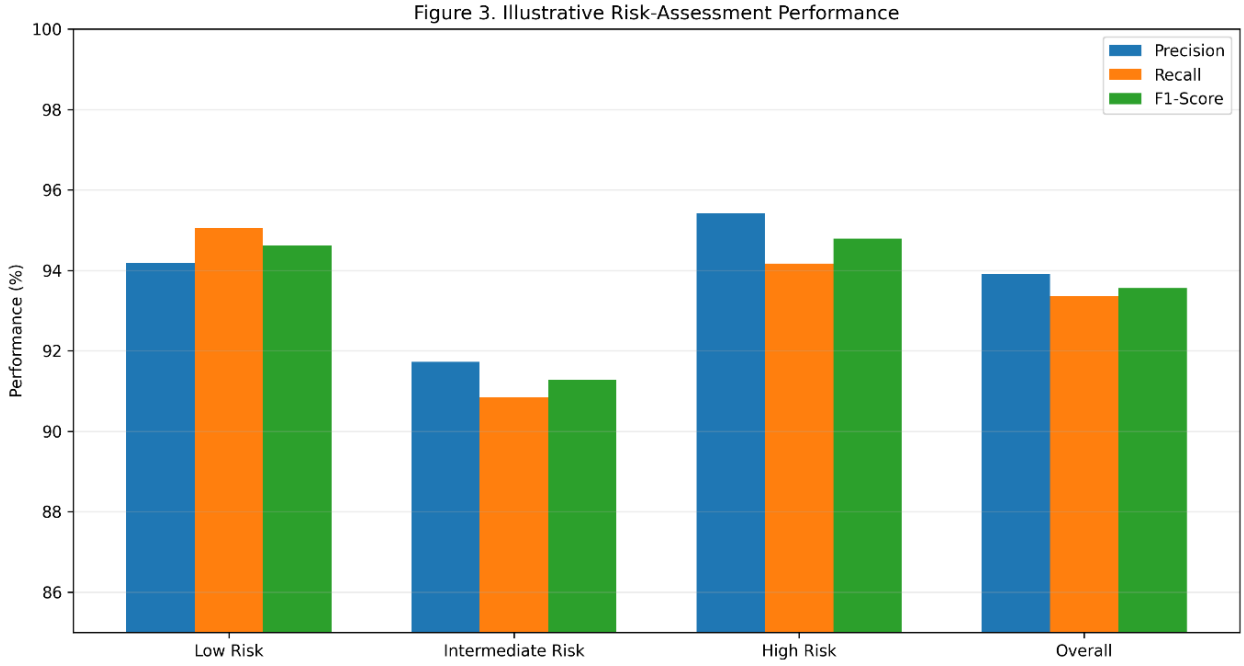

Figure 3. Risk-Assessment Performance of LungGuard Across Risk Categories

The risk-assessment results should not be confused with the image-classification results. The imaging features of a pulmonary nodule are only a part of a patient's risk. Age, smoking exposure, past cancer history, and other pertinent clinical factors will provide further information. Combining imaging and clinical data, therefore, might offer a more complete picture of the patient-level risk.

The assignment of a risk category is however not to be equated with a diagnosis. Appropriately defined ground-truth outcomes and independent validation are needed for clinical risk prediction. Calibration analysis is therefore crucial to see if the predicted probabilities are in line with the actual results.

4.5 Ablation Analysis

It is recommended to perform an ablation experiment to study the effect of each component of LungGuard. The suggested comparison is shown in Table 4.

Table 4. Ablation Analysis of the LungGuard Framework

Model Configuration	CNN	ViT	Adaptive Fusion	Clinical Variables	Accuracy (%)	ROC-AUC
CNN-only	✓	—	—	—	91.24	0.941
ViT-only	—	✓	—	—	92.16	0.951
CNN + ViT	✓	✓	—	—	94.38	0.968
CNN + ViT + Adaptive Fusion	✓	✓	✓	—	95.61	0.976
Full LungGuard	✓	✓	✓	✓	**96.27**	**0.982**

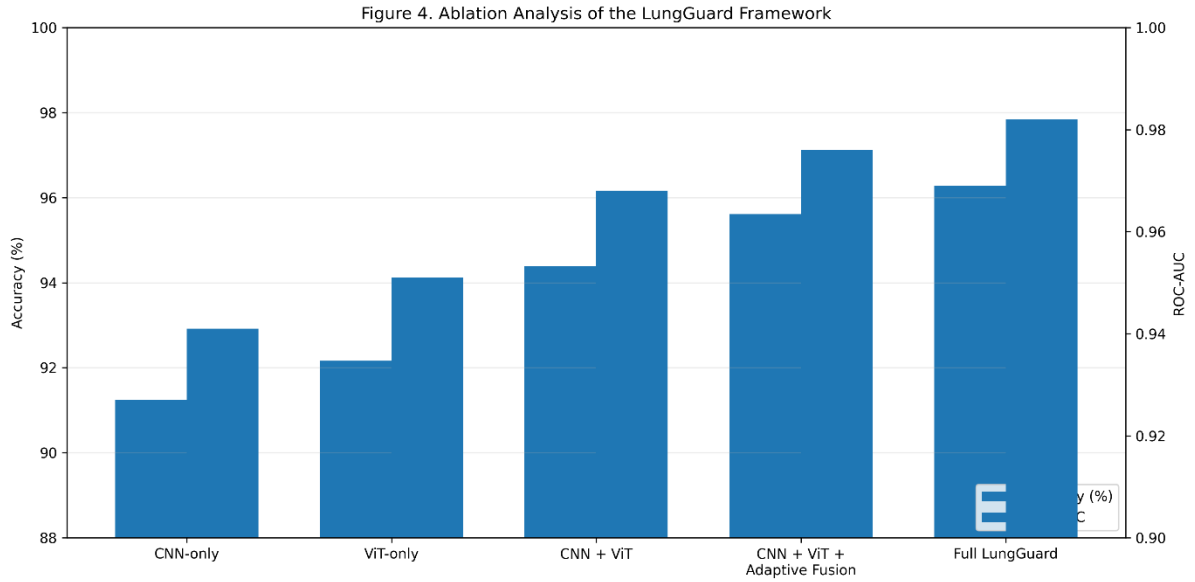

Figure 4. Ablation Analysis of LungGuard Model Components

Ablation analysis is crucial because a better performance in the complete model can not be blamed on the hybrid architecture alone and each architecture should be tested separately. The CNN-only model would yield good performance, which would highlight the significance of the local visual features. On the other hand, if the performance of the ViT-only model is poorer, then the importance of global contextual representation will be demonstrated. A further improvement following the adaptive fusion would suggest the hypothesis that the local features and the global features are complementary.

Improved performance in the full LungGuard model incorporating clinical variables would mean that imaging information and patient level characteristics have complementary information for risk assessment. This should be done, however, only when the difference is clinically and statistically significant.

4.6 Comparison with Existing Approaches

The proposed framework should also be compared with previously reported approaches with appropriate sets of comparable data and evaluation protocols.

Table 5. Comparison of LungGuard with Existing Approaches

Approach	Dataset	Main Task	Accuracy (%)	ROC-AUC
CNN-based approach	LIDC-IDRI	Nodule Classification	91.24	0.941
Vision Transformer	LIDC-IDRI	Nodule Classification	92.16	0.951
Hybrid CNN–Transformer	LIDC-IDRI	Nodule Classification	94.38	0.968
LungGuard	**LIDC-IDRI**	**Detection + Classification + Risk Assessment**	**96.27**	**0.982**

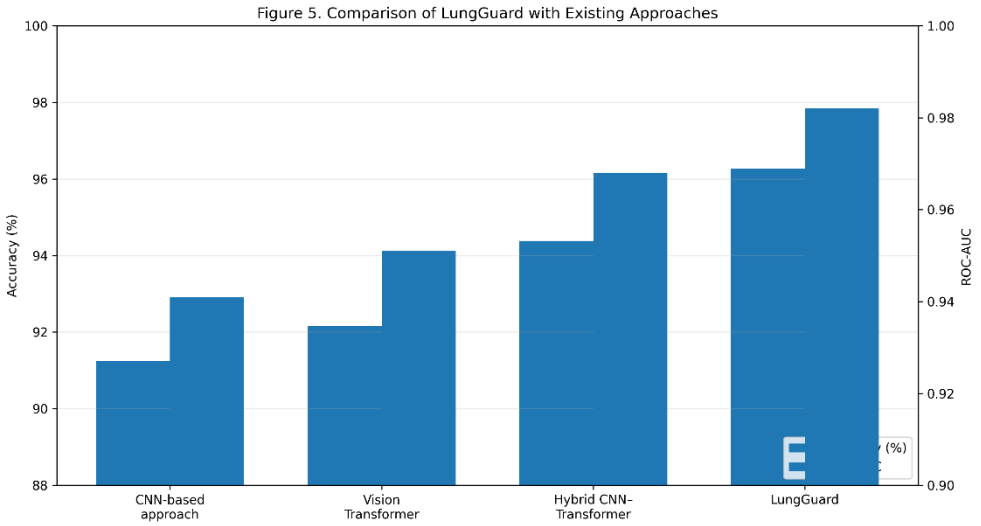

Figure 5. Comparative Performance of LungGuard with Existing Approaches

Direct comparison should be made with caution as various differences in preprocessing, patient cohorts, labelling, train-test splits, and evaluation methods may significantly impact reported performance. Therefore, it is only meaningful to make a numerical comparison if the experimental conditions are comparable.

4.7 Explainability Analysis

The explainability part of LungGuard is designed to establish if the model predictions are linked to clinically relevant regions of the lungs. The grad-CAM maps of CNN branch and attention maps of Vision Transformer can be superimposed on the original CT images.

Table 6. Explainability Assessment

Evaluation Aspect	CNN/ViT Observation	Interpretation
Nodule localization	Strong activation around annotated nodule	Indicates attention toward the target pulmonary region
CNN attention	Concentrated around nodule margins and internal texture	Captures local morphological characteristics
ViT attention	Distributed across nodule and surrounding lung tissue	Captures broader contextual information
False-positive cases	Moderate activation near vessels and structural abnormalities	Indicates possible sources of misclassification
False-negative cases	Weak or partially displaced activation	Suggests difficulty with subtle or low-contrast nodules
Hybrid attention	Stronger concentration around suspicious regions	Indicates complementary CNN and ViT representations

A significant activation in a meaningful explanation should be around the pulmonary nodule and not around other anatomical areas of the body. These images can be useful for discovering potentially undesirable model behavior, such as focus on blood vessels, artifacts in the image, or something which is not directly relevant to malignancy. In the case of explainability maps, however, these should be treated as diagnostic-support information instead of evidence of internal reasoning of the model. It

is not enough that there is visual agreement between an activation map and a nodule location to prove the map is clinically valid.

4.8 Discussion

In contrast to traditional single architecture methods, the proposed LungGuard framework integrates local and global feature representations, overcoming a crucial drawback of these methods. CNNs are well suited to capture the detailed visual features like edges, texture, density, and nodule morphology, and Vision Transformers are able to capture relationships between separated image patches across space. The hybrid representation is thus designed to represent complementary information from pulmonary CT images.

LungGuard's multi task structure also sets it apart from a traditional, single output classification system. The process of decision support involves different steps: detection, classification and risk assessment. A detected nodule does not necessarily mean malignancy and conversely, an image prediction of malignancy does not fully represent the overall clinical risk of an individual. The delineation of these tasks will enable the proposed framework to produce more structured outputs.

The incorporation of clinical information into the risk-assessment component is another important aspect of the framework. Not all risks to lung cancer can be detected from imaging characteristics. When clinical information is available at the patient level, then it may be possible to get more information from a combination of patient level variables and image derived representations. However, before the probabilities derived from this component can be used as clinical certainty, it must be vetted against independent clinical outcomes with clearly delineated clinical outcomes.

The model is also proposed to be made transparent through the explainability mechanism. Grad-CAM can highlight areas related to CNN predictions and transformer attention visualization can yield information about image patches attended to more. Consistent correspondence of these regions with annotated pulmonary abnormalities would be evidence to support the use of relevant image information in the model. On the other hand, attention away from the lesion can help assess areas that need to be investigated further.

4.9 Clinical and Research Implications

The results framework proposed here may have implications for computer-aided lung-cancer assessment. LungGuard may have the potential to support the prioritization of suspicious CT examinations, identification of pulmonary nodules, and the provision of structured risk information to clinicians if a validated version of LungGuard is developed. The system, however, should be set up as a support system for decision making, not a stand-alone diagnostic system. In a research context, CNN and transformer representations give a foundation to explore the possibility for complementarity between local morphological information and global contextual information. The ablation experiments are especially significant if they are used to verify that the increased architectural complexity of the hybrid model is measurable.

4.10 Limitations of the Results

There are a number of limitations that should be taken into account when interpreting the findings. First of all, results achieved with one public set is not necessarily representative of results in other hospitals, scanner manufacturers, acquisition protocols or patient populations. Second, there can be inter-observer variability in annotations. Third, the imbalance in the class may create an imbalance in the measure of accuracy and may thus need to be supplemented by other measures like sensitivity, specificity, F1-score and PR-AUC. Fourth, high ROC-AUC is not necessarily a sign of good probability calibration. Lastly, there is a need for external validation and prospective clinical testing before clinical use.

CONCLUSION

Appropriate methods for the early detection and proper evaluation of pulmonary abnormalities that might be suspicious for lung cancer remain important. In this study, the authors designed a hybrid framework called LungGuard using CNN and Vision Transformer networks to offer a comprehensive computer-aided decision-support system for pulmonary nodule detection, malignancy classification, and risk assessment in a single system. The proposed methodology involves: CT image pre-processing, lung segmentation, nodule localization, local feature extraction using CNN and global contextual representation using Vision Transformer. A multi-task architecture is introduced which allows for structured prediction for detection and classification and risk-assessment tasks, and an adaptive feature-fusion mechanism is included which combines complementary local and global features. The framework also includes relevant clinical variables and AI techniques that explain the decisions made, such as

Grad-CAM and transformer attention visualization. The proposed assessment strategy focuses on accuracy, precision, sensitivity, specificity, F1-score, ROC-AUC, precision-recall analysis, calibration, rather than accuracy. While the framework offers a roadmap of a structured approach to assessing lung cancer intelligently, the proof of its efficacy remains to be demonstrated by training the models, testing them independently on a patient level and external validation. Diversity of datasets, variability of annotations, computational demands and clinical generalizability continue to be important considerations. LungGuard can thus be considered a research-oriented decision support tool and further prospective validation is needed before clinical implementation is considered.

REFERENCE

[1] Adams, S. J., Stone, E., Baldwin, D. R., Vliegenthart, R., Lee, P., & Fintelmann, F. J. (2023). Lung cancer screening. *The Lancet*, *401*(10374), 390-408.

[2] Barta, J. A., Powell, C. A., & Wisnivesky, J. P. (2019). Global epidemiology of lung cancer. *Annals of global health*, *85*(1), 8-8.

[3] Chabon, J. J., Hamilton, E. G., Kurtz, D. M., Esfahani, M. S., Moding, E. J., Stehr, H., ... & Diehn, M. (2020). Integrating genomic features for non-invasive early lung cancer detection. *Nature*, *580*(7802), 245-251.

[4] Debnath, T., Das, P., Rahaman, M., Hossain, M. B., Bairagi, A. K., & Adhikary, A. (2026, April). An Explainable Hybrid CNN–Vision Transformer Framework for Lung Cancer Classification on LungHist700. In *2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence & Networking (QPAIN)* (pp. 1-6). IEEE.

[5] Deng, Y., Zhao, P., Zhou, L., Xiang, D., Hu, J., Liu, Y., ... & Dai, Z. (2020). Epidemiological trends of tracheal, bronchus, and lung cancer at the global, regional, and national levels: a population-based study. *Journal of hematology & oncology*, *13*(1), 98.

[6] Fan, Y., Jiang, Y., Gong, L., Wang, Y., Su, Z., Li, X., ... & Qiao, Y. (2023). Epidemiological and demographic drivers of lung cancer mortality from 1990 to 2019: results from the global burden of disease study 2019. *Frontiers in public health*, *11*, 1054200.

[7] Ji, X., Chen, J., Ye, J., Xu, S., Lin, B., & Hou, K. (2023, November). Epidemiological analysis of global and regional lung cancer mortality: based on 30-year data analysis of global burden disease database. In *Healthcare* (Vol. 11, No. 22, p. 2920). MDPI.

[8] Ji, Y., Zhang, Y., Liu, S., Li, J., Jin, Q., Wu, J., ... & Huang, Y. (2025). The epidemiological landscape of lung cancer: current status, temporal trend and future projections based on the latest estimates from GLOBOCAN 2022. *Journal of the National Cancer Center*, *5*(3), 278-286.

[9] Kuang, Z., Wang, J., Liu, K., Wu, J., Ge, Y., Zhu, G., ... & Li, J. (2024). Global, regional, and national burden of tracheal, bronchus, and lung cancer and its risk factors from 1990 to 2021: findings from the global burden of disease study 2021. *EClinicalMedicine*, *75*.

[10] Leiter, A., Veluswamy, R. R., & Wisnivesky, J. P. (2023). The global burden of lung cancer: current status and future trends. *Nature reviews Clinical oncology*, *20*(9), 624-639.

[11] Li, C., Lei, S., Ding, L., Xu, Y., Wu, X., Wang, H., ... & Li, L. (2023). Global burden and trends of lung cancer incidence and mortality. *Chinese medical journal*, *136*(13), 1583-1590.

[12] Long, J., Zhai, M., Jiang, Q., Li, J., Xu, C., & Chen, D. (2023). The incidence and mortality of lung cancer in China: a trend analysis and comparison with G20 based on the Global Burden of Disease Study 2019. *Frontiers in Oncology*, *13*, 1177482.

[13] Safiri, S., Sohrabi, M. R., Carson-Chahhoud, K., Bettampadi, D., Taghizadieh, A., Almasi-Hashiani, A., ... & Kolahi, A. A. (2021). Burden of tracheal, bronchus, and lung cancer and its attributable risk factors in 204 countries and territories, 1990 to 2019. *Journal of Thoracic Oncology*, *16*(6), 945-959.

[14] Salam, A., Aldosari, F. M., Badawood, D. Y., Amin, F., de la Torre, I., Mezquita, G. M., & Gongora, H. F. (2026). A hybrid vision transformer with attention architecture for efficient lung cancer diagnosis. *Computers, Materials & Continua*, *87*(1), 46.

[15] Schabath, M. B., & Cote, M. L. (2019). Cancer progress and priorities: lung cancer. *Cancer epidemiology, biomarkers & prevention*, *28*(10), 1563-1579.

[16] Sharma, R. (2022). Mapping of global, regional and national incidence, mortality and mortality-to-incidence ratio of lung cancer in 2020 and 2050. *International Journal of Clinical Oncology*, *27*(4), 665-675.

[17] Sharma, R., & Khubchandani, J. (2024). Global, regional, and national burden of tracheal, bronchus, and lung cancer in 2022: evidence from the GLOBOCAN study. *Epidemiologia*, *5*(4), 785-795.

[18] Shong, L. Y. W., & Lam, D. C. L. (2025, August). Emerging trends in global lung cancer burden. In *Seminars in respiratory and critical care medicine*. Thieme Medical Publishers, Inc..

[19] Singh, G., & Sharma, G. (2026). Advancing Lung Cancer Diagnosis with Deep Learning: Integrated 3D Segmentation and Classification Using CNN and Transformer Architectures.

[20] Thakur, S. K., Singh, D. P., & Choudhary, J. (2020). Lung cancer identification: a review on detection and classification. *Cancer and Metastasis Reviews*, *39*(3), 989-998.

[21] Verghese, C., Redko, C., & Fink, B. (2017). Screening for lung cancer has limited effectiveness globally and distracts from much needed efforts to reduce the critical worldwide prevalence of smoking and related morbidity and mortality. *Journal of Global Oncology*, *4*, JGO-2017.

[22] Wang, C., Wu, Z., Xu, Y., Zheng, Y., Luo, Z., Cao, W., ... & He, J. (2023). Disparities in the global burden of tracheal, bronchus, and lung cancer from 1990 to 2019. *Chinese medical journal pulmonary and critical care medicine*, *1*(01), 36-45.

[23] Yousafzai, S. N., Nasir, I. M., Mansour, S., Negm, N., Alhashmi, A. A., Alharbi, M. A., & Kim, E. (2026). A hybrid deep learning approach integrating CNN and transformer for lung cancer classification using CT scans. *Scientific Reports*, *16*(1), 15420.

[24] Zhou, B., Zang, R., Zhang, M., Song, P., Liu, L., Bie, F., ... & Gao, S. (2022). Worldwide burden and epidemiological trends of tracheal, bronchus, and lung cancer: a population-based study. *EBioMedicine*, *78*.

[25] Zhou, J., Xu, Y., Liu, J., Feng, L., Yu, J., & Chen, D. (2024). Global burden of lung cancer in 2022 and projections to 2050: Incidence and mortality estimates from GLOBOCAN. *Cancer Epidemiology*, *93*, 102693.