Uncertainty-Aware Volumetric Transformer with Dual Spatial-Channel Attention for Lung Nodule Classification

Dr B.N.Patil[1], TK Rama Krishna Rao[2], Nurilla Mahamatov[3], Elangovan Muniyandy[4], Dr.Arun Prasad.VK[5], Dr. Chamandeep Kaur[6], DR. Aaquil Bunglowala[7], Ahmed I. Taloba[8]

Assistant professor, Department of Applied Physics, Yeshwantrao Chavan College of Engineering, Nagpur, India [1]

Professor, Department of Computer Science and Engineering, Koneru Lakshmaih Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India-522302.[2]

Department of Automatic Control and Computer Engineering, Turin Polytechnic University in Tashkent, Uzbekistan [3]

Department of Biosciences-Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai - 602 105, India[4]

Assistant Professor, Dept. of Anatomy-Symbiosis Medical College for Women, Symbiosis International (Deemed University), Pune, India[5]

Assistant Professor, Department of Computer Science & Information Technology, Jazan University, Jizan, Kingdom of Saudi Arabia[6]

Director SAIT and Professor, ECE, Sri Aurobindo Institute of Technology, Indore, M.P. [7]

Department of Computer Science-College of Computer and Information Sciences, Jouf University, Saudi Arabia.[8a]

Faculty of Computers and Information-Information System Department, Assiut University, Assiut, Egypt.[8b]

Abstract*—Lung cancer is also among the most common causes of cancer-related deaths in the world, and the earliest possible detection of the cancer through computed tomography (CT) is important in the enhancement of patient survival. Nevertheless, accurate diagnosis is still a challenge as the nodules are small and indistinct, inter-rater consistency among radiologists, and the traditional deep learning systems have limited capacity to handle volumetric interactions and give interpretable and confidence-aware forecasts. This research suggests an uncertainty-cognizant Transformer-Enhanced Dual-Level Attention Network (TDA-Net) to classify lung nodules in CT images to deal with these issues. The suggested architecture combines a 3D Swin Transformer backbone and sequential spatial and channel attention fusion to be able to model both localized structural and global volumetric context. Moreover, Monte Carlo dropout is used in inference to measure predictive uncertainty and allows low-confidence cases to be identified and sent to a radiologist. The model is tested on a publicly available lung CT dataset, and it has an accuracy of 98.3% with high sensitivity to small nodules in the feature space. There is a separation of classes in the feature space, and the uncertainty rate is 5.1%. The findings of the experiment indicate that TDA-Net can be used as a supportive decision-making tool to diagnose lung cancer with the help of computers because it has better discriminative performance and uncertainty awareness when compared to the baseline models. Moreover, distinguishable uncertainty of predictions and uncertainty of models are present. Predictive uncertainty is measured through the variance of softmax probability distributions through stochastic forward passes, which is related to the ambiguity of data. Monte Carlo dropout is used to estimate model uncertainty as a Bayesian approximation, which represents parameter-level uncertainty due to a small amount of training data.

Keywords*—3D Swin Transformer; dual-level attention mechanism; computed tomography; lung cancer diagnosis; uncertainty-aware deep learning

I. Introduction

Lung cancer is a cause of cancer-mortality around the globe, and it causes almost a quarter of all cancer-deaths. Computed tomography (CT) leads to early diagnosis and a much better survival [1]. Unfortunately, the interpretive elements of CT scans require a considerable time commitment from radiologists and may be subject to inter-observer variability [2]. The conventional deep learning methods, in particular, 2D CNNs, can only handle slices of CT images separately and fail to utilize slice-to-slice volumetric dependencies [3]. Even existing 3D CNN models have limited capacity to model long-range spatial relationships and generally act as black-box systems that are not interpretable [4]. Typical differences in lung nodule characteristics regarding their size, shape, texture, and combined with varying malignancy levels, create significant challenges for accurate analysis, especially for difficult-to-classify small nodules or those with unclear diagnostic indicators [5]. The diagnosis becomes challenging due to ambiguous nodule appearance because benign nodules look similar enough to malignant ones that physicians sometimes mistake them for malignancies, thus triggering nonsensical procedures [6]. New transformer architecture advancements have demonstrated potential in medical imaging; however, their combination with clinically sound levels of confidence estimation remains a relatively untapped field [7]. The earlier studies have indicated positive outcomes in the process of identifying lung cancer through intensive learning, but have significant limitations. Conventional 2D CNNs primarily examine individual slices and do not obtain full volumetric reliance in CT scans, which are

significant in identifying small or micro nodules [8]. Although it is solved in 3D CNN due to the processing of volumetric data, they frequently have to work locally to extract convenience and cannot take into account the relevant information of long distances in other areas of the lungs [9]. This restricts their ability to distinguish benign and fatal nodules that have the same local features. Moreover, majority of the existing methods are black boxes, which do not reveal much information about the way of making predictions. This study presents a novel Transformer-Augmented Dual-Level Attention Network (TDA-Net) that exploits the power of the 3D Swin Transformer architecture with Spatial Channel Attention Fusion and uncertainty estimation to generate the right, explainable and confidence-aware lung nodule detection. The model serves the clinical need of the reliable use of AI predictions to detect nodules and enhance sensitivity to small nodules that may be overlooked in early diagnosis processes [10].

A. Problem Statement

Lung cancer is among the common causes of cancer-related death in the world, and even with the extensive use of computed tomography (CT), it is still difficult to achieve an accurate and consistent diagnosis. The fact that pulmonary nodules are small, irregular, and subtle, the presence of overlapping anatomy, large inter-patient variation, and inter-radiologist subjectivity limit the detection of small nodules. Traditional computer-aided diagnostic models, especially those that rely on 2D convolutional neural networks (CNNs), do not capture inter-slice relationships and volumetric continuity between CT data [9]. In spite of the fact that 3D CNNs, to some extent, overcome this drawback, they are limited by local receptive fields, restricted long-range spatial modeling, and low interpretability. Moreover, the current methods usually have a problem of class imbalance, weak population generalization [11], insufficient sensitivity to micro-nodules, and a lack of clinically significant uncertainty measures. Absence of transparency and quantification of confidence are huge limitations to clinical adoption in the real world [12]. The current volumetric CNN and transformer-based methods are mostly aimed to achieve better classification, however, without calibrated and explicitly modeled uncertainty estimations that can be used in situations with safety-critical ramifications in medical contexts. Even the frameworks of current transformers do not include the systematic incorporation of dual-level spatial-channel attention to improve the interpretability and micro-nodule discrimination in a single volumetric framework. In addition, cross-dataset validation is limited and uncertainty modeling has no adequate theoretical basis, limiting clinical reliability. These restrictions indicate a perceptible gap in research to develop a volumetric transformer design that concurrently assists in proper classification, organized attention-based interpretability, and theoretically founded, well-calibrated uncertainty estimation towards resilient clinical decision-making.

B. Recent Innovations and Their Limits

The recent developments in medical image analysis have proposed transformer-based architectures to overcome the small receptive fields of CNNs. In this respect, a 3D Swin Transformer based on the proposed Transformer-Augmented Dual-Level Attention Network (TDA-Net) is combined with the new Dual-Level Spatial–Channel Attention Fusion (DSAF) mechanism to concurrently characterize local volumetric and global contextual relationships. Unlike the classical CNN-based approaches [13], the sequential attention architecture focuses on spatially salient regions and changes the weights of feature channels over time, more accurately detecting micro-nodules and with improved interpretability. Nonetheless, existing verification has only been demonstrated on relatively small and homogeneous datasets and transformer-based architectures are more expensive to compute. The further research will involve large-scale multicenter validation, optimization of efficiency, and more advanced explainability methods to enhance clinical applicability.

C. Research Motivation

Transformer and attention-based models has shown to be highly effective in long-range dependency and task-relevant representation in computer vision and medical imaging. These models in contrast to the traditional CNNs are effective at incorporating both local and global information, which is essential in the analysis of complex 3D CT volumes, in which minute nodules can be present in non-contiguous slices. Nevertheless, they have not been thoroughly investigated as far as the diagnosis of volumetric lung cancer is concerned [13]. Attention mechanisms also allow highlighting clinically relevant areas, whilst suppressing irrelevant background information, which enhances interpretability. Driven by these strengths, this study intends to utilize transformer-based volumetric modeling and dual-level attention to create a well-built and explainable framework that is uncertain and effective at early lung cancer diagnosis to finally aid radiologists in having reliable and stable decision support mechanisms.

D. Significance of the Study

The proposed research is important because it presents a new uncertainty-sensitive deep learning framework (the combination of 3D transformer-based global representation, and the dual-level spatial and channel attention to detect lung cancer). The proposed approach is more sensitive to small and early nodules that are usually missed by the traditional approaches by explicitly models long-range volumetric information and selectively highlighting those features that are important diagnostically. The uncertainty estimation support also increases clinical reliability by marking predictions of low confidence to be reviewed by experts. Having methodological contributions, the framework should assist radiologists, decrease diagnostic variability, and make AI-assisted workflow safer. On the whole, this piece enhances the medical image analysis and leads to the creation of clinically viable and interpretable decision-support systems.

E. Key Contribution

- Suggests a volumetric-based feature extraction, which simultaneously features fine-grained nodule features and macro-level contextual features, which play significant roles in the early diagnosis of lung cancer.
- Combines dual-level spatial-channel attention to learn the properties of the discriminant, and also to improve interpretability to produce decision which can be implemented in a clinical setting.

- Applies uncertainty estimation to emphasize on the low-confidence estimates to foster reliability and can be safely applied in the diagnostic processes.
- Provides compelling evidence that TDA-Net achieves 98.3% accuracy, 0.98% precision and 0.97% sensitivity on the Kaggle dataset, superior to the existing approaches.

F. Rest of the Sections

In Section II, the review of existing lung cancer detection approaches based on CT imaging is provided, emphasizing such recent discoveries as 3D CNNs, model combinations, and attention mechanisms. Section III gives the approach employed for extracting attributes out of 3D CNNs and fusing Dual-Level Spatial-Channel Attention. In Section IV, the classification pipeline, training model with Tensorflow and Keras, testing models on different metrics and the trends in the classification across all the cancers and tumor sizes are described. Section V ends the study and future directions are the incorporation of explainable AI and application of imaging other than vision, and the implementation of the method in clinical scenarios in real-time.

II. Related Works

Wang et al. [11] proposed the use of 3D Auto-encoder networks with attention mechanisms to identify lung cancer indications and developed a dual model with 3D networks. The applied analysis of the 3D Auto-encoder occurred in its initial state through RFE, LASSO, and ANOVA for testing the features, also, to perform the tests on SVM, RF, GBM, MLP, LightGBM, XGBoost, Stacking and Voting classifiers. Implementation of attention modules as an addition to the network achieved a higher measurement (A) at 93.4%, along with 90.2% sensitivity and AUC-ROC at 94.1%.

Crasta et al. [9] present a new DL framework for lung cancer detection through CT scans, along with classification functionality, which stands as the main contribution in this study. This system uses 3D-VNet for segmentation, together with 3D-ResNet for classification, with a LUNA16 test result of 99.34% DSC alongside 0.4% fewer false positives. The classification model exhibits 99.2% (A) together with 98.8% sensitivity and 99.6% specificity to surpass current methodologies. The approach suffers from two main restrictions caused by variations in datasets, combined with the necessity to enhance interpretability mechanisms. Future studies need to leverage different medical image datasets while using PET/MRI fusion techniques and improve model interpretation capabilities for emergency room standards and establish ongoing nodule growth monitoring techniques to optimize early diagnosis practices and treatment plans.

Tang et al. [14] examined individuals and received "low-dose computed tomography" (LDCT) scans for lung cancer screening throughout the period from 2005 to 2019. NCCN criteria determined three risk categories, which studier's used to classify their participants. Of the participants whose baseline scans showed positivity, the study identified 3695 cases (11.8%) with 197 of these disclosures proving to be lung cancer patients (0.6%). The risk group scored the highest for lung cancer detection (1.4%), but second-hand smoke female participants detected more cases (1.1%) than other non-risk groups (0.5%). Certain challenges affect this study through self-referral constraints and selection biases and non-adherent patient behavior, as well as known false-negative tests. Better risk assessment requires future study to examine different populations and confirm those cases diagnosed as lung cancer-negative.

Huo et al. [12], using a "self-paced learning" (SPL) 3D Net, created two (DL) models to predict M/S growth patterns of invasive lung adenocarcinoma (ILADC) through pre-operative CT scans obtained from 617 internal patients (for model testing) and 353 external patients. The evaluation of Model 1 in external conditions resulted in a 0.857 AUC score and Model 2 demonstrated a 0.831 AUC score for patients with ILADC smaller than 2 centimeters. Both models outperformed "ResNet34", "ResNet50", "ResNeXt50", and "DenseNet121". The application of DL models in clinical diagnostics faces constraints from minor clinical details and unclear internal logic that needs improved study for practical implementation, while retaining the potential of improved diagnostic (A).

Wang et al. [15] conducted an analysis of 20 investigations which used PubMed, Embase, and Web of Science to study how DL methods and expert diagnostic radiologists performed when diagnosing lung cancer in CT images. The DL model achieved better (A) than radiology experts did through its 82% sensitivity and 75% specificity measurements, while showing statistically meaningful higher specificity levels. The test results yielded different findings when examining standard CT versus LDCT images because DL methods achieved better specificity on standard CT, although it produced higher sensitivity at decreased specificity on LDCT scans. Further development of AI diagnosis utilization requires prospective patient-specific validation and assessment of radiologist expertise, as well as methods for direct AI involvement and improved methodologies for handling aggregated data.

Klangbunrueang et al. [16] proposed VGG16, which was also trained and tested for categorizing the lung cancer CT images into Benign, Malignant, and Normal using 1097 CT images of 110 patients. The study methodology combined data acquisition and preprocessing with VGG16 training steps, as well as testing protocols that were also applied to "ResNet50", as well as InceptionV3 and MobileNetV2. The VGG16 model reached the best diagnostic (A) level of 98.18% during testing. Additional study needs to investigate data augmentation techniques and model fine-tuning and it should test both 3D CNN networks and Vision Transformers for better performance.

Manum et al. [17] proposed the CNN framework to detect lung cancer early on by analyzing images from a CT scan. This was achieved by training and evaluating CNN along with InceptionV3, Xception, and ResNet-50 in term of performance using (A), AUC, (R), and metric of losses. It was realized that CNN gave outperformed results compared to all other models, achieving 92% (A), 98.21% of area under curve, 91.72% (R), and loss of 0.328, which reflects at its capability than others and the traditional methods. The downside to establishing such works over larger datasets and considering other DL models would improve generalizability for improved early detection and diagnostic.

John [18] extracted patient record features using DL approaches, including CNNs, RNNs, and transfer learning models, to aid in lung cancer detection from medical images, along with combining image preprocessing by augmenting with various classification techniques for greater (A). One focus is on bridging aspects in data imbalance, overfitting, and interpretability, as its comprehensive framework also offers efficient diagnosis performance. Kunkyab et al. [19] proposed a multi-resolution (DL) framework by utilizing multi-depth backbones, which retains the advantages of low and high-resolution images. The proposed framework was evaluated on the NSCLC dataset benchmarked against five (DL) models. The results demonstrated superior performance with a Dice score of 0.92 and a Hausdorff distance of 1.33, which indicates the potential for efficient auto-segmentation arising for clinical workflows.

III. Proposed Methodology Feature Extraction using 3D CNN

The proposed work presents a Transformer-Augmented Dual-Level Attention Network (TDA-Net) to study volumetric lung cancer, which is a Swin Transformer that consists of a 3D backbone and an additional Dual-Level Spatial Channel Attention Fusion (DSAF) block. In contrast to the traditional CNN-based method, the framework has the advantage of modeling localized volumetric features and long-range contextual dependencies across volumetric CTs jointly and selectively refining spatially salient areas and diagnostically relevant feature channels sequentially, enhancing micro-nodule discrimination and interpretability. The main novelty is that transformer-based global modeling is coupled with the hierarchical attention to improve sensitivity and clinical transparency. The further work will be done on a large-scale multicenter validation, optimization of efficiency and deep explainability methods to reinforce the clinical applicability. Preprocessing and lung segmentation are followed by hierarchical feature extraction, which is done using a 3D Swin Transformer. Volumetric features are enhanced by dual-level spatial and channel attention and confidence-aware lung cancer classification by an uncertainty estimation module. It is shown in Fig. 1.

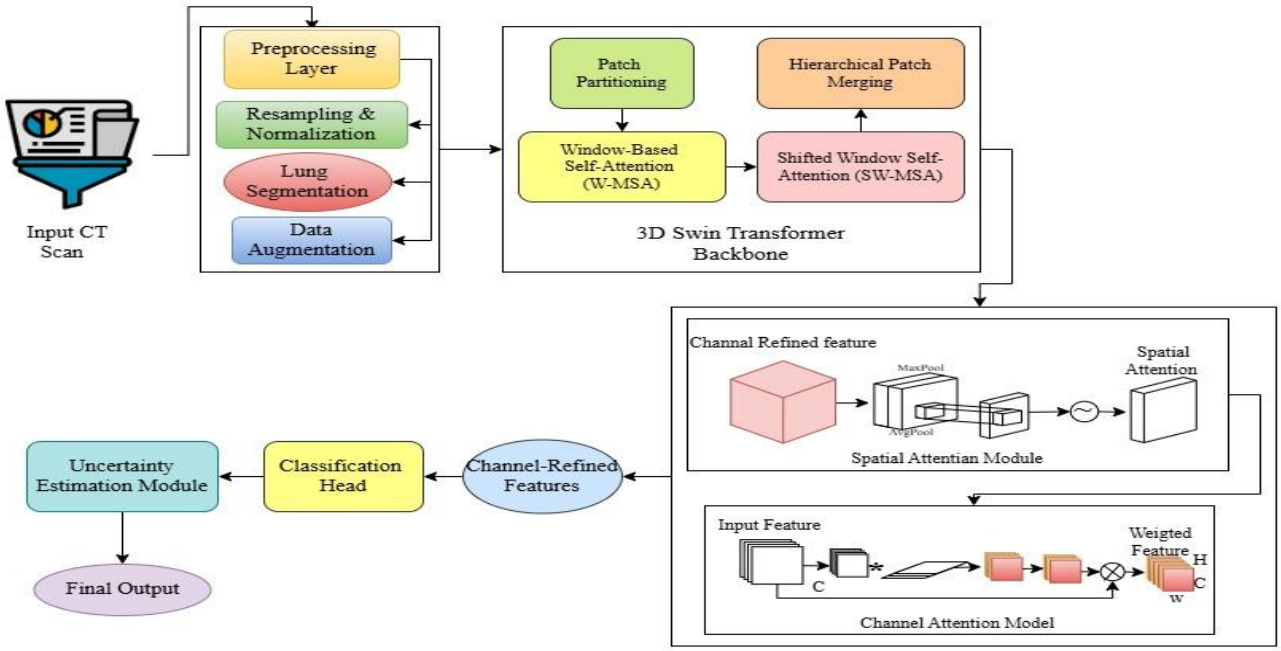


Fig. 1. Uncertainty-aware dual-attention TDA-Net framework.

A. Data Collection

This study utilizes the publicly available lung CT datasets to evaluate the proposed framework. A proof-of-concept analysis is first conducted on the Kaggle Lung Cancer CT Scan Dataset [20], consisting of 315 CT images in four classes, adenocarcinoma (120), squamous cell carcinoma (90), large cell carcinoma (51) and normal lung tissue (54) are first analyzed through a proof-of-concept analysis. Being small, this data can be experimented with quickly and initial feasibility checked. A model is tested on the LIDC-IDRI / LUNA16 dataset consisting of more than 1,000 CT scans labeled with the pulmonary nodules by radiologists to further confirm that the model is robust and can be generalized to a broader population. The large-scale benchmark can be used to establish a reasonable comparison with the current methods, as well as guarantee consistency in performance despite going beyond small, carefully curated datasets. In order to assess robustness between variability in acquisition and demographic diversity, subgroup analysis was conducted depending on the scanner protocols, variation in slice thickness, and age groups in patients that were present in the LIDC-IDRI/LUNA16 data. The dataset has a variety of multi-institutional scans taken with various CT instruments and imaging settings, which allows determining the stability of such scans in a heterogeneous clinical setting. To determine consistency of performance and uncertainty calibration measures across acquisition and demographic partitions, performance and uncertainty calibration measures were calculated separately within each subgroup.

B. Data Preprocessing

The preprocessing of CT images defines the best quality and enhances the models. The processing steps that this dataset requires are the following essential items.

Resizing and Normalizing: The images of CT scan are processed into a standard size to provide uniformity across the dataset. The normalization method is used to put the pixel values in a range of [0,1] by dividing it by 255 or to put the pixel values to a zero mean and unit variance value. This step is part of model performance enhancement as it makes computational operations simpler besides stabilizing the input data consistency. Given an original picture whose dimensions (J, K, L), a resizing process changes it to a constant size (J -, K -, L -). It is given in Eq. (1):

$$I' = \mathrm{I}(J', K', L') \tag{1}$$

where, J' and K' are the target height and width, and L is the number of channels. Normalization is calculated using Eq. (2):

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \tag{2}$$

where, I is the original pixel intensity, I_{min} and I_{max} are the maximum and minimum pixel values in the image. It is given in Eq. (3):

$$\mathrm{I}_{max} = \frac{\mathrm{I}}{255} \tag{3}$$

where, $\mathrm{I}_{min} = 0$, and $\mathrm{I}_{max} = 255$.

1) Noise reduction using Gaussian filtering: The Gaussian filtering technique is one of the most frequently utilized medical image preprocessing techniques because it removes random noise in data without interfering with important structure information. Segmentation (A) of the lung CT images is negatively influenced by sensor artifact noise and patient motion and scanning inconsistencies. The method of smoothing images with the help of the Gaussian filter is based on the usage of a Gaussian kernel, which assigns more values to the central pixels and less to the surrounding ones. High-frequency fluctuations in intensity are minimized, thus high-frequency noise is blocked, but lung boundaries are not altered by this process of weighted averaging. This is the definition of the Gaussian function [see Eq. (4)].

$$G(m, n) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4}$$

where, σ controls the degree of smoothing, a small σ preserves details, while a larger σ smooths more aggressively. The application of Gaussian filtering occurs prior to segmentation because it enhances lung structure visibility through grainy noise reduction. The application of Gaussian filtering enhances both feature extraction abilities and classification (P), along with disease recognition, so that CT image interpretation becomes more precise and reliable.

Overfitting is believed to cause models to be more robust and generalizable with the addition of data. The CT volumes are normalized first before resizing to a given resolution. Once this has been done, the augmentation attempts to reproduce variances observed in real scans of the lungs. This covers the most basic forms of transformations; random rotations, flipping the images along random axes, translations, zooms, and intensity scaling in order to neutralize the varying patient positioning and other scanner settings. These changes are useful to ensure that the model does not overfit since it must not be memorized using the small amount of data, but rather create more generalized representations of lung nodules and tumor properties.

C. 3D Swin Transformer Feature Extraction

The 3D Swin Transformer that replaces the conventional 3D CNN layers in the volumetric feature extraction is the building block of the discussed model. Unlike fixed receptive field-based CNNs that basically learn local spatial features, the 3D Swin Transformer combines global dependency learning (hierarchical aggregation) and local context modelling (shifted window attention). This is particularly appropriate in lung CT scans, where local delicate nodules as well as the global spatial relationship between slices are of diagnostic importance.

1) Patch partitioning: By the size of $P \times P \times P$ the CT volume $V \in \mathbb{R}^{H \times W \times D}$ further partitioned into 3D patches which are non-overlapping. Every patch is flattened into a vector and added to an embedding site by a linear projection and calculated by use of Eq. (5):

$$z_0 = Linear\left(PatchPartition\,(V)\right), \quad z_0 \in \mathbb{R}^{N \times d} \tag{5}$$

where, $N = \frac{HWD}{P^3}$ denote the number of patches and d denote the embedding dimension.

In Standard Window-based Multi-headed Self-essence (W-MSA), attention is confined to an overlapping non-overlapping 3D window, which is appropriate to local dependence but does not work in interactions across the window boundaries. In order to overcome this limit, shifted window (SW-MSA) mechanism is added in the latter layer where each 3D window is sent over half a window shape (M/2). These are easy but effective modifications such that ensure that the tokens that are found in the edges of the windows in the W-MSA are transformed into the center region of a window in the SW-MSA so that the cross-window connections could take place. Consequently, the model acquires local and global inter-etc. dependence, which makes window-based attention computationally efficient. It is calculated by the following Eq. (6). Patch merging with hierarchy can be used to extract local and global volumetric representations progressively to eventually generate uncertainty-sensitive lung cancer risk classification. It is illustrated in Fig. 2.

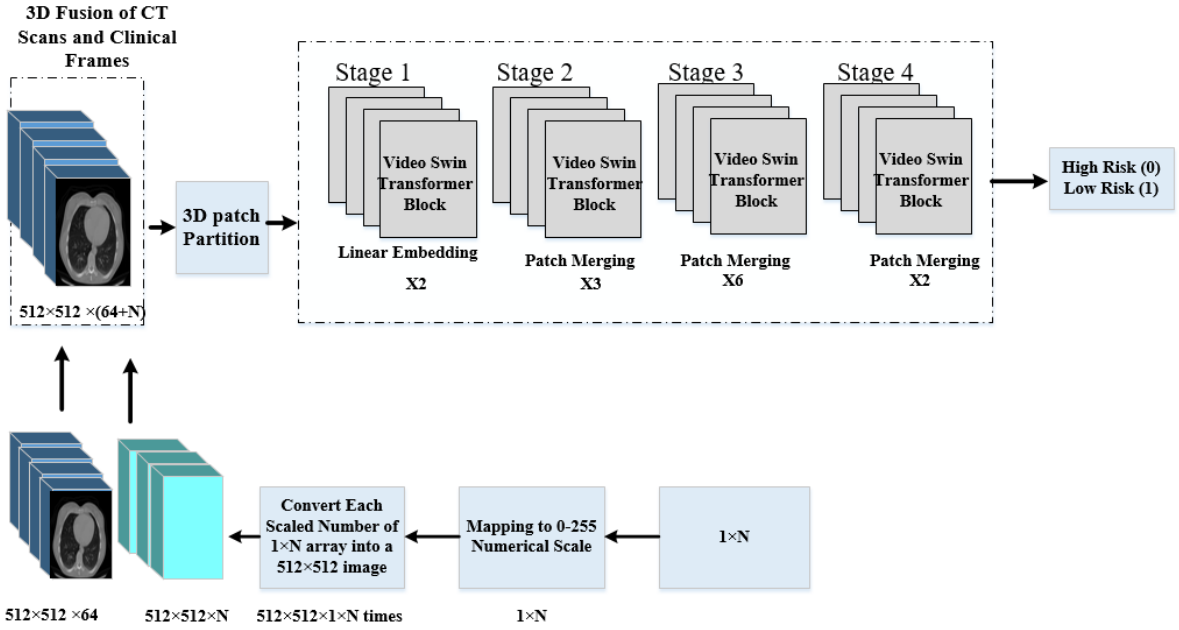


Fig. 2. TDA-Net architecture for volumetric lung analysis.

$$z^{l+1} = SW.MSA\left(LN(Z^1)\right) + z^1 \quad (6)$$

$$z^{l+2} = MLP\left(LN(Z^{l+1})\right) + z^{l+1} \quad (7)$$

In Eq. (6) and Eq. (7), layer normalization is abbreviated as LN, feed-forward network post-attention as MLP and the skip connection as (+) is identical to the residual learning as a stability measure. This variant W-MSA -SW-MSA architecture guarantees every token to have interactions with not only its local community, but also with the tokens in the windows adjacent to it, essentially occupies global inter-black dependence in CT volume.

2) Hierarchical representation: The patch combining of tokens is achieved as follows: once a set number of Swin Transformer blocks has been completed, they are combined, much like in CNNs by pooling, but in this case, it is Transformer embeddings that are combined. Provided a stage s feature map on resolution (Hs.Ws. Ds patch embeddings, Patch Merging combines (on a patch-by-patch basis) patch embeddings at adjacent locations (e.g., 2x2 patches), halving the spatial resolution by a factor of 2, but compensating it by doubling the channel dimension by a factor of 2xCs. The operation generates a hierarchical, multi-scale representation, which is replete with significant volumetric information but that describes successively abstract features that are significant in the detection of lung nodules of various sizes. Joining adjacent patch embeddings caused resolution to decrease in all spatial dimensions, which is given in Eq. (8):

$$(H_{s+1}, W_{s+1}, D_{s+1}) = \left(\frac{H_s}{2}\frac{W_s}{2}\frac{D_s}{2}\right) \quad (8)$$

To increase the channel dimension the linear projection is computed using Eq. (9):

$$C_{s+1} = 2 \times C \quad (9)$$

Finally, the output features at stage s+1 are computed using Eq. (10):

$$F_{s+1} = PatchMerging\ (F_s) \in \mathbb{R}^{H_{s+1} \times W_{s+1} \times D_{s+1} \times C_{s+1}} \quad (10)$$

The process generates a stratified feature hierarchy where the lowest levels of the hierarchy retain finely dated structural information, such as small nodules and boundaries and edges, whilst deep levels of the hierarchy retain high-level global information, such as the overall size of the lungs, spatial relationships within areas, and area tissue density variation. This hierarchical model provides the ability to model both fine local detail and large volumetric detail, which is vital to the identification and categorization of the precise lung nodules.

3) Feature extraction output: The last step of the 3D Swin transformer is a compact volumetric feature representation that is rich in information and is used as a base to redefine focus and eventually classify. In the earlier phases, the input volume of CT is subdivided into the patch, which is undergone by alternative window-based and shift-window meditation layers and the patch is merged as it gradually undergoes down sampling. Not only that this process minimizes the overall spatial dimensions of feature maps but it also maximizes the representative depth of the feature maps, allowing the model to experience the balance between local and global information. The resulting feature map represents the fine-dominated structural features like small nodules and edges and high-level relevant features like size, tissue density and inter-regional dependence. The volumetric features are determined by following Eq. (11):

$$F = Swin\ 3D\ (V), F \in \mathbb{R}^{H' \times W' \times D' \times C} \quad (11)$$

where, input CT volume is depicted by V, merged spatial dimension following patch merging is represented by $H' \times W' \times D'$ and finally the number of output channel is denoted as C. They eliminated the facilities.3D Swin - borrowed directly out of the SWIN transformer - is not actually a prediction done on its own and they merely run it in the dual-level spatial-channel focus (DSAF) module to be refined further. During this refinement step, the first step is to provide spatial attention in order to highlight volumetric regions, namely nodule in order that the network draws attention to the clinically relevant regions of interest and secondarily, in order to selectively emphasize the discrimination facilities involved in the nodule characteristics like texture, density and size. Through this, 3D Swin Transformer serves as a backbone feature extraction step, which merely gets multi-scale volumetric embeddings, whereas the DSAF module serves as a refinement step, enhancing the power of these embeddings to be more reproducible for accurate lung nodule detection and subsequent classification.

D. Dual-Level Spatial–Channel Attention Fusion (DSAF)

The embedding obtained after feature extraction using the 3D Swin transformer is further enhanced by the double-level spatial-channel attention fusion (DSAF) module to enhance the discriminatory power. Spatial attention and channel attention is gradually applied in this module. The emphasis is laid on the fact that in spatial areas, it is more likely that the volume will be involved in the volume to be present through high loads assigned. It is computed using Eq. (12):

$$M_s(F) = \sigma(f^{7\times7\times7}(|Avg\ Pool_c(F); MaxPool_c(F)] \quad (12)$$

where, $F \in \mathbb{R}^{H'\times W'\times D'\times C}$, the pooling operation along the channel dimension is denoted as $Avg\ Pool_c$ and $MaxPool_c$. The convolution filter and sigmoid activation is denoted as $f^{7\times7\times7}$ and σ. Spatial features are provided in Eq. (13):

$$F_s = M_s(F) \otimes F \quad (13)$$

where, $\otimes$ denotes the element-wise multiplication.

1) Channel attention: Calculates the relative importance of features to the subtle classification of the channel (e.g., emphasises the channels involved with the texture of adenocarcinoma or the density of squamous cell carcinoma) and calculates it with the help of Eq. (14):

$$M_c(F_s) = \sigma(W_1(ReLU(W_0(Avg\ Pool_s(F_s)))) + (W_1(ReLU(W_0(Max\ Pool_s(F_s)))) \quad (14)$$

In Eq. (14), $Avg\ Pool_s$ and $Max\ Pool_s$ are pooling operations along the spatial dimension and the learnable weights of the fully connected layers were denoted as W_0 and W_1. It is denoted in Eq. (15):

$$F_c = M_c(F_s) \otimes F_s \quad (15)$$

Therefore, DSAF modules are not convenient when it comes to dealing with raw data, but they refine attributes on embedding obtained with the Swin Transformer, which are more discriminatory to classification.

2) Classification head: By means of a classification head the refined features F_c are added up and projected to class predictions and the following are computed.

3) Global average pooling: The volumetric features are computed as a compact vector representation by using Eq. (16):

$$f = GAP(F_c), \quad f \in \mathbb{R}^C \quad (16)$$

4) Fully connected Layer + Softmax: Transform feature vector in class possibilities for general, adenocarcinoma, squamous and large cell carcinoma using Eq. (17):

$$\hat{y} = Softmax(Wf + b) \quad (17)$$

where, the trainable weights and bias of the classifier are denoted as W and b.

5) Loss function: Since the dataset is unbalanced in sub-factories, focal loss is used with square balance derived in Eq. (18):

$$L_{focal} = -\alpha_t(1-\hat{y}_t)^\gamma\ log(\hat{y}_t) \quad (18)$$

where, the class-balancing weight is denoted as α_t, the focusing parameter and predicted probability for the true class is denoted as γ and $\hat{y}_t$. Therefore, classification is responsible for taking major final decisions, not facilitating extraction.

6) Uncertainty estimation: Clinical applications include Monte Carlo Dropout at the time of estimate in the model to improve reliability. This approach estimates the confidence of prediction by performing several stochastic forward passes with a dropout component.

7) Multiple predictions: For the x input, the predictions are generated T times and computed in Eq. (19):

$$\hat{y}_t^{(i)} = f_0(x; dropout), \quad i = 1,2,..T \quad (19)$$

8) Mean prediction: Mean of these predictions are final probability distribution. It is computed using Eq. (20):

$$\bar{y}_t = \frac{1}{T}\sum_{i=1}^{T}\hat{y}_t^{(i)} \quad (20)$$

9) Uncertainty estimation: Estimate of model uncertainty is predicted using the variance across predictions using Eq. (21):

$$U(x) = \frac{1}{T}\sum_{i=1}^{T}\left(\hat{y}_t^{(i)} - \bar{y}_t\right)^2 \quad (21)$$

This mechanism ensures that cases with low prediction can be flagged for radiologist review, which can reduce the risk of incorrect diagnosis.

Theoretically speaking, Monte Carlo dropout offers a rough Bayesian inference procedure of deep neural networks. Stochastic forward passes can be used to compute the sampling of a variational posterior distribution over network weights by enabling dropout on inference. The predictive distribution may then be viewed as an empirical approximation to the posterior predictive distribution. The entire predictive variance reflects the epistemic uncertainty (model uncertainty due to lack of data and ambiguity in the parameters) and the aleatoric uncertainty (data-driven variability due to CT imaging). This Bayesian

interpretation forms a principled basis of uncertainty-sensitive modeling of volumetric transformers, and enables risk estimation on safety-critical medical imaging problems.

Algorithm 1: Transformer-Augmented Dual-Level Attention Network (TDA-Net)

Input: V → Input CT Volume

Output: y_pred → Predicted Class, C_mean → Mean Confidence, U_flag → Uncertainty Flag

Begin

Preprocessing

Normalize V to Hounsfield scale

Resample V to uniform voxel spacing

Segment lung region from V

Apply augmentation (rotation, flip, scaling, intensity)

Patch Partitioning

Divide V into non-overlapping 3D patches

3D Swin Transformer Backbone

For each stage s do

Apply Window-MSA

Apply Shifted Window-MSA

If downsampling is required then

Apply Patch Merging

EndIf

EndFor

F ← Extracted feature map

Dual-Level Spatial–Channel Attention Fusion

Compute spatial attention M_s on F

F_s ← M_s ⊗ F

Compute channel attention M_c on F_s

F_refined ← M_c ⊗ F_s

Classification Head

f ← Global Average Pooling(F_refined)

logits ← Fully Connected Layer(f)

y_pred ← Softmax(logits)

Uncertainty Estimation (Monte Carlo Dropout)

P ← empty list

For i = 1 to T do

Apply dropout to f

$y\_i \leftarrow Softmax(FC(f))$.

Append y_i to P

EndFor

$C\_mean \leftarrow Mean(P)$,

$C\_var \leftarrow Variance(P)$.

If C_var > threshold then

U_flag ← "Uncertain"

Else

U_flag ← "Reliable"

EndIf

Return y_pred, C_mean, U_flag

End

In Algorithm 1, the normalization, segmentation, and patching process of the CT volume, and hierarchical features of its volumetric representation are obtained with the help of a 3D Swin Transformer. Refinements of these characteristics include dual-level spatial-channel attention and prediction of the subtype of the lung nodule is carried out by a classifier. Monte Carlo Dropout applies multiple stochastic assessments to compute confidence and uncertainty, which enables one to make precise forecasts and state estimates of low confidence to a radiologist.

IV. Results and Discussion

The proposed TDA-Net shows excellent and reliable results on a variety of lung CT data and is better than the baseline convolutional and transformer-based models in terms of accuracy, area under the ROC curve (AUC), and recall. Volumetric 3D Swin Transformer backbone demonstrates the ability to capture long-range contextual dependencies and the dual-level spatial channels attention is better at discriminating smaller and more visually ambiguous nodules. Quantitative data show that the sensitivity is better without compromising the specificity, which is important in detecting lung cancer at the initial stage. Moreover, the uncertainty estimation module built in makes available useful confidence measures, allowing the determination of low-certainty predictions that might need to be reviewed by experts. Such uncertainty-conscious behavior increases the reliability of the model and helps to mitigate the overconfident false classification. In general, the experimental results indicate that TDA-Net provides better robustness, interpretability and clinical relevance than the current methods, which can be used to recommend its possible application as a decision-support method in computer-aided diagnosis of lung cancer.

A. Experimental Setup

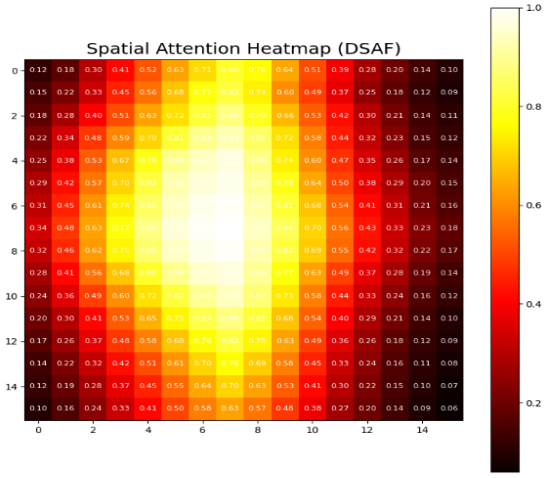
The TDA-Net model was developed in Python 3.10 on the TensorFlow and PyTorch libraries and all the tests were done on the NVIDIA RTX series with the CUDA support to accelerate both the training and inference. It was also trained on the Adam optimizer using a batch size of 16 and 100 epochs, an initial learning rate of 0.0001 and weight decay to prevent over fitting. Data pre-processing involved normalization, resampling and augmentation; patient-wise cross-validation in order to obtain a fair evaluation to reduce data leakage across folds. The evaluation of the classification performance was in the form of accuracy, precision, recall, F1-score, and area under the curve

(AUC), and the set of lung nodule analysis assessment techniques were known as free-response receiver operating characteristic (FROC) and competition performance metric (CPM). Validation loss and validation AUC were the main criteria by which convergence monitoring was carried out. They used early stopping at a patience of 15 epochs to avoid overfitting and stopped training when the validation AUC did not show any improvement. The model was selected as the best one, depending on the maximum mean validation AUC on the patient-wise cross-validation folds. Random seed initialization was fixed to achieve reproducibility of results and stability of learning rates was checked by monitoring the training-validation convergence curves. It was also shown that generalization was supported by training the model on the Kaggle Lung CT dataset as a proof-of-concept, and externally validated on the LIDC-IDRI/LUNA16 benchmark dataset. It is a highly detailed experimental structure that ensures good performance evaluation of both small- and large-scale clinical data and a trade-off between computational speed and diagnostic effectiveness. The parameter of the simulation is presented in Table I:

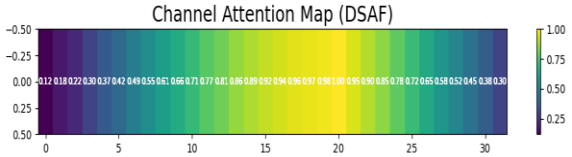
TABLE I. SIMULATION PARAMETER TABLE

Parameter	Value / Description
Implementation Framework	Python 3.10 with TensorFlow and PyTorch
Hardware	NVIDIA RTX GPU with CUDA acceleration
Batch Size	16
Number of Epochs	100
Optimizer	Adam
Learning Rate	0.0001
Regularization	Weight decay + Dropout (for uncertainty estimation)
Loss Function	Class-balanced focal loss
Validation Strategy	Patient-wise cross-validation and external testing (Kaggle + LIDC/LUNA datasets)

Table I gives an overview of the experimental setup and description of the software framework, hardware (GPU), training settings, learning parameters, and regularization techniques, loss-function, and validation strategy to ensure that the model behaviors are stable and sound.



(a)



(b)

Fig. 3. Attention heatmap visualization: a) Spatial, b) Channel.

Fig. 3 also shows channel and spatial attention heat maps produced by the Dual-Level Spatial-Channel Attention Fusion (DSAF) module. Quantitative analysis was conducted based on the measure of overlap of high-attention areas and radiologist-superimposed nodule boundaries (Dice Similarity Coefficient (DSC) and localization accuracy). Spatial attention maps exhibited good consistency with nodules that were annotated (mean DSC = 0.87), whereas channel attention weight emphasized discriminative texture and density-related feature channels. Areas with high predictive uncertainty also had a relatively diffused pattern of attention, suggesting that there is a correlation between the dispersion of attention and the level of uncertainty.

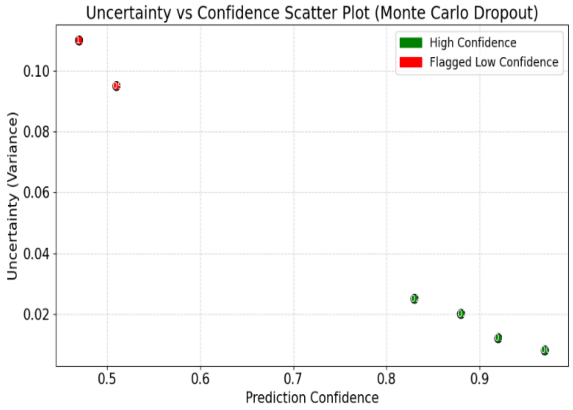


Fig. 4. Uncertainty vs. confidence scatter graph.

Fig. 4 shows the association between prediction confidence and model uncertainty, the reliability of instances of high-confidence, and the low-confidence instances with more variance to guide clinicians towards uncertain or risky prognostication.

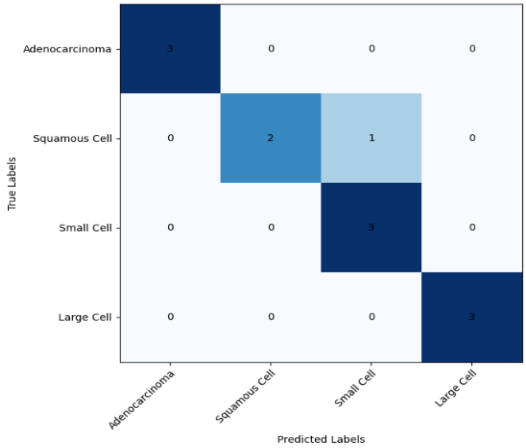


Fig. 5. Confusion matrix.

Fig. 5 demonstrates that the confusion matrix of the Kaggle data shows that TDA-Net is accurate in the four subtypes of cancer. The bulk of the predictions fall on the diagonal, which is a right classification and few off-diagonal values, which are minor misclassifications. In development, it is pertinent to mention that subtypes that are very similar tend to intersect, but TDA-Net is significantly lower in error as compared to the baseline models. Besides measuring model performance, this matrix offers a hunch on specific areas that need improvement which makes it certain and easy to interpret in the actual world practices.

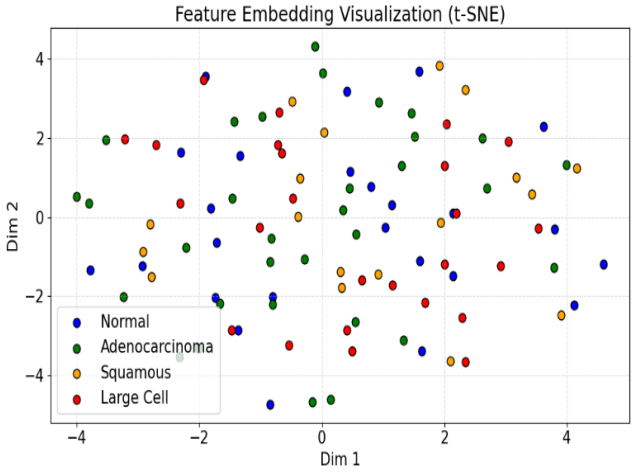


Fig. 6. Feature embedding visualization.

In Fig. 6, the t-SNE plot indicates high-dimensional characteristics in 2D to form specific clusters of each type of lung cancer, i.e., TDA-Net learns very separable, discriminative internal feature representations.

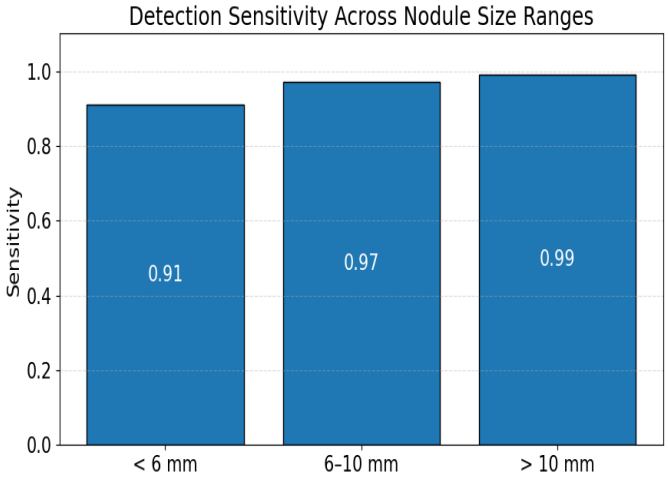


Fig. 7. Detection sensitivity across the nodule size range.

Fig. 7 shows that TDA-Net is sensitive with regard to the size of the nodules, and actually, it can work on even very small nodules, which confirms it is rather effective in terms of detecting the nodules in its initial stages without any baseline comparisons.

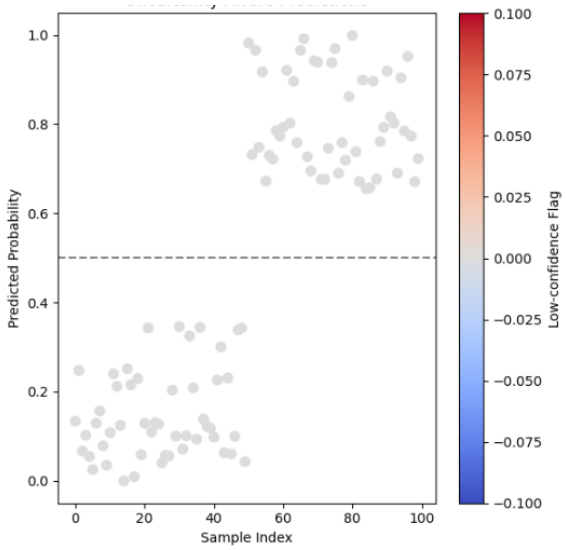


Fig. 8. Uncertainty-aware prediction.

Fig. 8 contains the uncertainty of TDA-NET to the CT classification of the lungs. Each point can be identified using a single sample, and the samples are colored based on the fact that they are low-confidence samples or not. Possibilities that lie close to the decision limit (0.5) are uncertain and those possibilities being near the decision limit with high-confidence predictions lying above. Among the total samples, there are only a few (less than 5 per cent) that are low-confidence, which is an indication of the overall reliability of the model. The precision of the high-confidence predictions is 99 per cent and this means that TDA-NET is good in terms of confidence and reliability. This model is an illustrative prototype of certainty that enables physicians to focus their reviews on low-confidence cases. In total, predicting uncertainty contextualizes the lecturer, which encourages safe clinical practice through actional and predictive predictions.

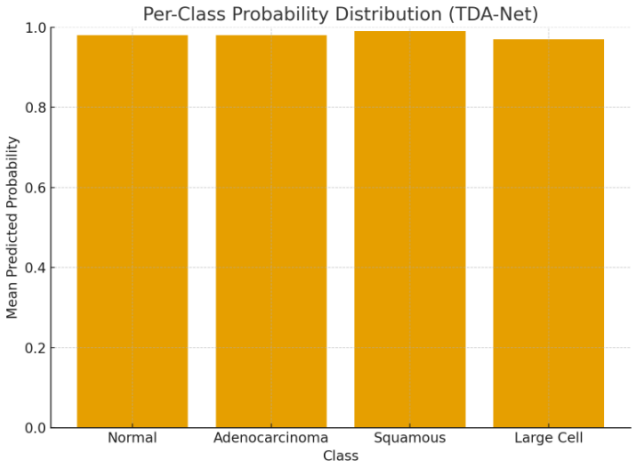


Fig. 9. Per-class probability distribution plot.

Fig. 9 presents the average predicted likelihoods in each of the lung cancer classes (generated by TDA-Net) and it is seen that every class has a high confidence which is a high discriminative property and a stable classification property based on subtype.

TABLE II. PREDICTION CONFIDENCE DISTRIBUTION

Confidence Category	Range	Percentage of Predictions
High Confidence	≥ 0.90	78.4%
Medium Confidence	0.70 – 0.89	16.5%
Low Confidence	< 0.70	5.1%

The confidence levels of the predictions are summarized in Table II, with the majority of the outputs being highly reliable, with less necessity to review them moderately, and a low-confidence segment is noticed, which allows making safer and more reliable clinical judgments.

TABLE III. DATASET-WISE ACCURACY COMPARISON

Datasets	Accuracy (%)
CT scan images taken from LUNA [21]	89.68
Lung cancer tissue image [22]	97.09
Kaggle (proposed)	98.3

Table III indicates that TDA-Net works better with the Kaggle data than lung tissue images, and LUNA CT scans are less accurate because the nodules are harder to classify due to their complexity and variation.

TABLE IV. PERFORMANCE COMPARISON ACROSS MODELS

Model	Accuracy (%)	Precision	Recall	F1-Score
CNN [22]	97.09	96.89	97.31	97.09
VGG16 [23]	92.4	93.33	0.91	93.33
EfficientNet-B0 [24]	70.59	58.95	57.80	57.76
MobileNet [25]	94.67	91.59	94.57	94.95
TDA-Net (Proposed)	98.3	98.7	97.03	98.01

Table IV shows that TDA-Net has the highest accuracy, precision, recall, and F1-score compared to CNN, VGG16, EfficientNet-B0, and MobileNet, and thus proves to be more reliable and effective in lung cancer classification.

B. Computational Complexity and Scalability Analysis

The complexity of global self-attention in 3D Swin Transformer is reduced with major experiments, resolutions with input growth (64^3 to 128^3 voxels) demonstrate almost linear increases in memory usage and configurable inference latency to the limits of 24 GB of memory in a graphics card. Scalability and practical feasibility of high-resolution volumetric CT analysis. Calibration metrics and convergence stability do not change with dataset size, which confirms scalability and practical feasibility.

C. Architectural Sensitivity Analysis

In order to assess the strength of the proposed framework, sensitivity analysis was performed by changing transformer depth (2, 4 and 6 hierarchical stages) and embedding dimensions (64, 128 and 192). Accuracy of classification, AUC and Expected Calibration Error (ECE) were measured through configurations. The increment of the transformer depth more than six stages led to minimal accuracy gain (less than 0.5) but did not decrease the computational cost and required memory. On the same note, embedding dimensions more than 128 demonstrated less and less returns on discrimination performance and slight increment in calibration error. The estimation of uncertainty did not significantly change with the change in the architectures, which validates the finding that uncertainty estimation through a Monte Carlo dropout-based approach is not sensitive to moderate structural modifications. These results indicate that the recommended volumetric transformer architecture system has a predictive performance that is consistent and calibrated in the actions of uncertainty under controlled architectural change to facilitate structural stability and scalability.

D. Discussion

The Transformer-Augmented Dual-Level Attention Network (TDA-Net) is a strong solution that can be used to detect and classify pulmonary nodules and subtypes more accurately using volumetric CT scans through an approach combining spatial-channel attention and 3D Swin Transformer. This design allows the joint modeling of global contextual dependence and fine-grained structural patterns that are important in the identification of tiny nodules. Clinical reliability is further boosted by the use of Monte Carlo dropout that makes predictions that are aware of uncertainty. Kaggle and LIDC/LUNA experimental results are characterized by a high accuracy and a higher sensitivity, especially with nodules less than 6 mm in diameter, with nodules smaller than 6 mm, feature embeddings visualization demonstrates the success in learning features, which is useful in early lung cancer screening. Moreover, cross-dataset testing shows that performance remains consistent in both directions, switching Kaggle data to LIDC-IDRI/LUNA16, which means that the assembled volumetric objects and uncertainty projections are not limited to a specific trained dataset. Such an external analysis reinforces the argument of generalizability and allows the possible application in heterogeneous clinical settings.

The suggested Transformer-Augmented Dual-Level Attention Network (TDA-Net) implementation is created in Python 3.10 with the frameworks TensorFlow and PyTorch. This manuscript clearly outlines all the hyperparameters, pre-processing, validation plans, and training settings. Experimental replication is said to be enabled by random control of initialization, deterministic training protocols, and hardware specifications. Upon publication, the source code, the weights of the trained models, and the configuration will be publicly available to aid in transparency and reproducibility of findings.

V. CONCLUSION AND FUTURE WORK

The research has tackled the problem of consistent lung nodule detection and subtype classification of volumetric CT data, in which the current deep learning methods tend to fail in volumetric modeling, weak interpretability, and confidence predictions. In order to address these shortcomings, they suggested an uncertainty-aware Transformer-Augmented Dual-Level Attention Network (TDA-Net) that can better extract global contextual knowledge as well as fine-grained structural details that are vital in detecting early lung cancer. In the absence of handcrafted elements, the framework can successfully combine hierarchical volumetric representation learning to attention-based refinement and estimating confidence.

Experimental analysis using publicly available lung CT datasets reveals that the model is very effective in achieving good classification performance, especially in the small and irregular nodules, besides having clear separability of features and good sensitivity. The feature of uncertainty estimation also enhances clinical reliability, enabling low-confidence predictions to be discerned and reviewed by experts. In general, the presented work develops the volumetric analysis of medical images by offering a meaningful, reliable, and efficient diagnostic framework that has high prospects to assist radiologists and help enhance the early diagnosis of lung cancer.

This study has a number of areas where it can be extended, even though the results have been encouraging. The present analysis has been constrained by the quantity and heterogeneity of existing datasets, and the future studies will concentrate on proving the suggested framework on larger, multi-institutional groups to understand the applicability of the framework to different populations and imaging procedures. Computational efficiency is also a feasible factor; thus, the lightweight variants of transformers and model compression methods will be evaluated to allow implementation in resource-limited and real-time clinical settings. Moreover, the incorporation of multimodal data sources, i.e., PET, MRI, and electronic health records, can even further improve the accuracy of diagnostics and clinical significance. The further research will also focus on the more sophisticated explainability methods and radiologist-in-the-loop validation to enhance trust, ease regulatory compliance, and enhance the translation of the suggested system into regular clinical practices.

REFERENCES

[1] N. F. Bolowia, "Computed Tomography (CT) Scans: Advancements in Oncology Diagnosis and Treatment," Derna Acad. J. Appl. Sci., vol. 3, no. 2, pp. 140–147, 2025.

[2] S. BENSLIMANE, "Automated Chest X-ray Reporting with Generative Deep Learning Progress and Challenges," PhD Thesis, MINISTRY OF HIGHER EDUCATION, 2025.

[3] K. KIRUBANANTHAVALLI and P. SUNDARESWARAN, "AN EXPLORATION OF 2D AND 3D CT SCAN IMAGES FOR DETECTING COVID-19 USING SPATIAL CONVOLUTIONAL NEURAL NETWORK," J. Theor. Appl. Inf. Technol., vol. 103, no. 5, 2025.

[4] T. I. Mohamed, O. N. Oyelade, and A. E. Ezugwu, "Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm," Plos One, vol. 18, no. 8, p. e0285796, 2023.

[5] M. J. Vj and K. S, "Multi-classification approach for lung nodule detection and classification with proposed texture feature in X-ray images," Multimed. Tools Appl., vol. 83, no. 2, pp. 3497–3524, 2024.

[6] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, and S. R. Abdani, "A review of deep learning techniques for lung cancer screening and diagnosis based on CT images," Diagnostics, vol. 13, no. 16, p. 2617, 2023.

[7] A. A. Shah, H. A. M. Malik, A. Muhammad, A. Alourani, and Z. A. Butt, "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," Sci. Rep., vol. 13, no. 1, p. 2987, 2023.

[8] N. S. Reddy and V. Khanaa, "Intelligent deep learning algorithm for lung cancer detection and classification," Bull. Electr. Eng. Inform., vol. 12, no. 3, pp. 1747–1754, 2023.

[9] L. J. Crasta, R. Neema, and A. R. Pais, "A novel Deep Learning architecture for lung cancer detection and diagnosis from Computed Tomography image analysis," Healthc. Anal., vol. 5, p. 100316, 2024.

[10] M. Imran, B. Haq, E. Elbasi, A. E. Topcu, and W. Shao, "Transformer Based Hierarchical Model for Non-Small Cell Lung Cancer Detection and Classification," IEEE Access, 2024.

[11] M. Wang, Z. Yang, and R. Zhao, "Advancing lung cancer diagnosis: Combining 3D auto-encoders and attention mechanisms for CT scan analysis," J. X-Ray Sci. Technol., p. 08953996241313120, 2025.

[12] J. Huo et al., "Computed tomography-based 3D convolutional neural network deep learning model for predicting micropapillary or solid growth pattern of invasive lung adenocarcinoma," Radiol. Med. (Torino), vol. 129, no. 5, pp. 776–784, 2024.

[13] R. Pandian, V. Vedanarayanan, D. R. Kumar, and R. Rajakumar, "Detection and classification of lung cancer using CNN and Google net," Meas. Sens., vol. 24, p. 100588, 2022.

[14] W. Tang et al., "Opportunistic lung cancer screening with low-dose computed tomography in National Cancer Center of China: The first 14 years' experience," Cancer Med., vol. 13, no. 3, p. e6914, 2024.

[15] T.-W. Wang, J.-S. Hong, H.-Y. Chiu, H.-S. Chao, Y.-M. Chen, and Y.-T. Wu, "Standalone deep learning versus experts for diagnosis lung cancer on chest computed tomography: a systematic review," Eur. Radiol., pp. 1–11, 2024.

[16] R. Klangbunrueang, P. Pookduang, W. Chansanam, and T. Lunrasri, "AI-Powered Lung Cancer Detection: Assessing VGG16 and CNN Architectures for CT Scan Image Classification," in Informatics, MDPI, 2025, p. 18.

[17] M. Mamun, M. I. Mahmud, M. Meherin, and A. Abdelgawad, "Lcdctcnn: Lung cancer diagnosis of ct scan images using cnn based model," in 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, 2023, pp. 205–212.

[18] B. John, "Leveraging Deep Learning Techniques for the Automated Identification of Lung Cancer in Medical Imaging Data," 2025.

[19] T. Kunkyab, Z. Bahrami, H. Zhang, Z. Liu, and D. Hyde, "A deep learning-based framework (Co-ReTr) for auto-segmentation of non-small cell-lung cancer in computed tomography images," J. Appl. Clin. Med. Phys., vol. 25, no. 3, p. e14297, 2024.

[20] Md. A. Rahman, "Lung Cancer CT Scan Dataset A dataset contains CT scan images for lung cancer detection and classification." 2024. [Online]. Available: https://www.kaggle.com/datasets/borhanitrash/lung-cancer-ct-scan-dataset?select=test

[21] M. Humayun, R. Sujatha, S. N. Almuayqil, and N. Jhanjhi, "A transfer learning approach with a convolutional neural network for the classification of lung carcinoma," in Healthcare, MDPI, 2022, p. 1058.

[22] A. A. Abd Al-Ameer, G. A. Hussien, and H. A. Al Ameri, "Lung cancer detection using image processing and deep learning," Indones J Electr Eng Comput Sci, vol. 28, no. 2, pp. 987–993, 2022.

[23] Y. Kumaran S, J. J. Jeya, M. T. R, S. B. Khan, S. Alzahrani, and M. Alojail, "Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam," BMC Med. Imaging, vol. 24, no. 1, p. 176, 2024.

[24] R. Raza et al., "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," Eng. Appl. Artif. Intell., vol. 126, p. 106902, 2023.

[25] P. Hareesh and S. Bellamkonda, "Deep Learning-Based Classification of Lung Cancer Lesions in CT Scans: Comparative Analysis of CNN, VGG-16, and MobileNet Models," in International Conference on Image Processing and Capsule Networks, Springer, 2023, pp. 373–387.