Enhanced Edge-Deployable Deep Learning System for Lung Cancer Type and Stage Classification Using CT Imagery

C. Venkatesh [1,2], Shashi Kant Gupta[3]

[1]Post-Doctoral Researcher, Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia. ***(Student ID: LUCPDF202604060091) (Application Number: LUC/CPGS/PDF/2025291002).***

[2]Associate Professor, Department of Electronics and Communication Engineering, Annamacharya University, Rajampet, Andhra Pradesh, India

[3]Adjunct Professor, Lincoln University College, Malaysia

[1,2]venky.cc@gmail.com, [1,2]pdf.venkatesh@lincoln.edu.my, [3]raj2008enator@gmail.com

[1,2]0000-0001-9323-177X, [3]0000-0001-6587-5607

Abstract: One of the most fatal types of cancer worldwide is lung cancer and therefore an accurate and efficient diagnostic system is required for early detection and the stage of the disease. While deep learning approaches have made substantial strides in the diagnosis from Computed Tomography (CT) scans, several current methods are either computationally heavy or depend on ensemble models or transformer-based networks, which are not suitable for healthcare settings with limited computing resources. Furthermore, some procedures only classify cancers and do not include detailed staging using the TNM (Tumor–Node–Metastasis) classification. To solve these problems, this paper presents LiteSOS-Net, a lightweight deep learning framework to classify lung cancer type and predict its TNM stage simultaneously from CT images. The study starts with a thorough review of the literature to define the important gaps in the recent research that would impact clinical usage of the existing methods.

The proposed framework consists of bilinear image resizing, guided-filter denoising, Multi-Peak Generalized Histogram Equalization for enhancing the image contrast, Attention U-Net for segmentation, and a classification network based on ShuffleNet with channel-shuffle gating. The Osprey Optimization algorithm is employed to further optimize the network parameters to increase the accuracy of the network predictions whilst maintaining computational efficiency. Several publicly available benchmark CT image datasets are proposed to evaluate the proposed framework. Its light weight design allows for edge deployment in limited resource healthcare applications. The research gaps identified and methodology proposed is presented in this paper, and the experimental validation and performance analysis in the next phase of the study.

Keywords: Lung Cancer; Computed Tomography; Lightweight Neural Networks; ShuffleNet; Edge Computing; Osprey Optimization.

Introduction

One of the greatest hazards to health is still lung cancer. Diagnostic accuracy (early detection and TNM based staging) is very important to ensure optimal selection of treatment modalities, better prognosis and greater chances of survival. Computed Tomography (CT) is now the preferred imaging technique for lung cancer screening because it allows visualization of small pulmonary nodules and subtle structural abnormalities in the lungs. Nevertheless, the manual interpretation of computed tomography images is laborious and can be subjective due to inter-observer differences, and thus the need for strong computer-aided diagnostic systems that can assist clinical decision making [1].

In the medical field, deep learning has become a strong solution for medical image analysis, making significant strides in automated lung nodule detection, segmentation and classification. Survey studies in recent years have shown that most current methods use either CNN-based or transformer-based models or a combination of the two to enhance diagnostic accuracy. While these approaches have shown promising results, they suffer from high computational complexity, significant memory usage, and slow inference speeds, limiting their application especially in limited healthcare settings. Furthermore,

most reported studies concentrate on either the nodule detection or the lung cancer classification problem, with only relatively few studies considering the combined problem of lung cancer classification and classification of the tumour stage TNM in a single framework [2-3].

Encouraged by these constraints, this paper introduces LiteSOS-Net, a lightweight model for cancer type classification and TNM-based staging with CT images. This proposed framework incorporates two bilinear image resizing, two guided-filter denoising, two Multi-Peak Generalized Histogram Equalization for contrast enhancement, Attention U-Net for segmentation, and a classification backbone using ShuffleNet with channel-shuffle gating. To further optimize the model parameters for predictive performance and remain computationally efficient, the Osprey Optimization Algorithm is used. Given its lightweight nature, the proposed framework suits well for working on edge devices, and can be easily implemented in remote and resource-constrained healthcare environments [4].

The rest of the sections follow. The related literature and the gaps in the research are reviewed in Section II, proposed LiteSOS-Net framework presented in Section III, and conclusion & future research directions are planned in Section IV.

Related Works

In 2025, Zhu et al. [5] presented a lung nodule detection system using CT imaging. The proposed model utilized the model architecture YOLOv8, which can be transplanted to other domains and is very lightweight to realize rapid inference, and can be used to accurately locate nodules in real time. The framework mainly emphasizes detection but lacks support for subclassification of lung cancer and prediction of TNM stage.

In 2024, Majumder et al. [6] established an ensemble CNN based framework for lung cancer classification from computed tomography (CT) scans. The model was enhanced by using an ensemble strategy that combines multiple CNN architectures, improving the classification accuracy. However, the design of the ensemble resulted in significant increase in computational complexity, memory, and inference time, making it impractical for use in resource-limited healthcare settings.

In 2023, Jafari et al. [7] presented a framework comprising an improved U-Net structure and a Swin Transformer. The transformer-based feature extraction enhanced the contextual representation and detection performance. The attention mechanism added significant computational complexity and model complexity, however, which made it difficult to deploy to the edge in real time.

In 2023, Hendrix et al. [8] create a nodule detection system using CT images. The proposed approach showed good diagnostic capability on larger datasets. The use of a computationally intensive backbone network resulted in slow inference time and limited its applicability to lightweight embedded platforms.

Research Gaps

It is seen that the existing methods are mainly depends on transformer-based techniques or ensemble learning or complex multi-stage networks, leading to higher computational complexity and lower feasibility for real-time deployment. Moreover, there has been limited research on edge-assisted diagnosis for resource-limited and rural healthcare settings. Current works focus mainly on either lung cancer detection or lung cancer classification, and few studies report on the development of lightweight frameworks that can simultaneously classify lung cancer subtypes and predict its TNM stage. These restrictions demonstrate the need for a common framework based on efficient computation and performance that could be readily implemented in the clinic, motivating the proposed LiteSOS-Net.

Proposed Methodology

The proposed LiteSOS-Net is an automated lung cancer detection system based on lightweight deep-learning framework and Computed Tomography (CT) image as shown in Figure 1. The framework is able to complete two clinical tasks concurrently, namely lung cancer subtype classification and TNM-based stage prediction. The entire process consists of six sequential steps. The design of the architecture aims to ensure high diagnostic accuracy while also being computationally efficient, enabling it to be utilized in resource-limited healthcare settings.

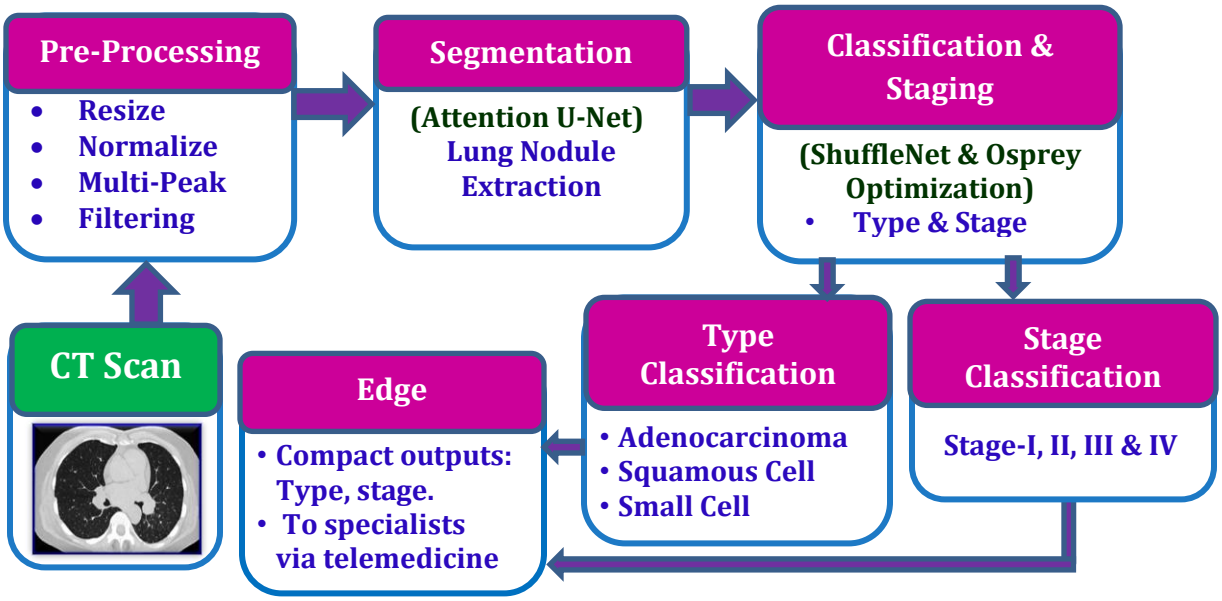

Figure 1. Block diagram of Proposed Methodology.

A. Acquisition of CT images.

The suggested framework starts by collecting chest CT images from the benchmark datasets that are already publicly available. These datasets are obtained from different CT scanners, different imaging protocols, so that the images acquired will have differences in the spatial resolution, image contrast and noise characteristics. Thus, any acquired images have to undergo a pre-processing stage to create standardized inputs for subsequent analysis. The obtained image is then passed to the pre-processing stage where improvement of the obtained image is done before segmentation.

B. Image Pre-processing

In the pre-processing stage, the aim is to enhance the quality and uniformity of CT images, remove unwanted variations from the images while retaining diagnostically relevant structures. This stage involves the following four consecutive operations: bilinear image resizing, intensity normalization, guided-filter denoising and Multi-Peak Generalized Histogram Equalization (MPGHE). All CT images are first transformed to a common spatial resolution by bilinear interpolation to have a uniform size for the input of the network. After resizing, the images are normalized to eliminate changes in intensity that may have occurred due to various imaging devices, which is done by the Z-score method. Next, a guided filter is used to reduce the effects of image noise, by retaining the edges of the nodules. Lastly, MPGHE increases local contrast resulting in more visibility of pulmonary nodules and other clinically relevant structures. The whole pre-processing process can be represented as shown in eq. (1).

$$F_p = H(G(N(R(I)))) \quad \text{..... (1)}$$

Let $R(\cdot)$, $N(\cdot)$, $G(\cdot)$ and $H(\cdot)$denote resizing, normalization, guided filtering and histogram equalization, respectively. The output of this stage is a noise cancelled and contrast enhanced CT image that is suitable for the input of the segmentation network.

C. Segmentation with Attention U-Net

The optimized CT image is then fed into an Attention U-Net which is able to accurately detect the area of interest (lung) and pulmonary nodules. The encoder captures the multi-level semantic information while the decoder gradually recovers the segmentation map, using skip connections [9]. The attention gates filter out extraneous background information and accentuate informative areas so that the network can focus on areas of interest such as suspicious nodules. The segmented lung region is attained by using Eq. (2).

$$F_s = F_p \odot M \text{..... (2)}$$

In this, M is the predicted segmentation mask and $\odot$ is an element-wise multiplication. The result of the segmentation stage consists of only the lung parts of interest that are segmented and the background tissues have been removed. This segmented image is then passed to the feature extraction to limit unnecessary computations and to enhance the classification performance.

D. Lightweight Feature Extraction and Lung Cancer Classification

Through the proposed LiteSOS-Net backbone, lung segmentation is performed on the segmented lung image. The network is based on grouped convolution, channel shuffle, and depth-wise separable convolution, which allow the network to efficiently extract discriminative features with a low computational cost. These operations drastically decrease trainable parameters and floating-point operations, while maintaining the representation of features, in comparison with conventional convolutional networks.

Global Average Pooling is used to transform feature maps into a feature vector of small storage size. The Osprey Optimization Algorithm (OOA) is used during training to optimize the network parameters and improve the convergence [10]. Finally, a Softmax classifier makes a prediction for the probability of each lung cancer type.The classification probability is determined by using Eq. (3).

$$P(T_i) = \frac{e^{z_i}}{\sum_{j=1}^{c} e^{z_j}} \text{..... (3)}$$

where c denotes the number of all of the cancer classes.

The lung cancer subtype predicted using the eq. (4).

$$T = \arg\max P(T_i) \text{..... (4)}$$

Deep features are extracted and at the same time are fed into the TNM-guided staging module to assess the clinical stage of the disease.

E. TNM – Guided Stage Prediction.

The proposed system after classification of cancer subtypes, assigns the disease stage based on the Tumor Node Metastasis (TNM) staging system. The segmented tumor is then used to extract morphological characteristics which, together with deep features learned by LiteSOS-Net, are used to determine the most probable clinical stage. This holistic approach allows for the diagnosis and staging to be done together in a single approach.

The stage predicted is calculated by eq. (5).

$$S = \arg\max P(S_i) \text{..... (5)}$$

Each stage of the TNM is associated with a probability $P(S_i)$.

This step yields the predicted clinical stage, which can be helpful as part of a cancer classification as a subtype and supplement this classification for treatment planning.

F. Edge Deployment

The trained LiteSOS-Net model is then deployed to lightweight formats to be used in the clinical environment in real time. The proposed architecture is compact and has fewer computational resources, less memory usage and lower inference time as compared to the conventional deep learning models. Therefore, it fits very well in resource-restricted clinical settings, such as rural hospitals, mobile health systems and other clinical settings.

The overall prediction confidence is computed as the average of the individual prediction confidence scores and is expressed in eq. (6).

$$C = \max\ (P(T_i) \times P(S_i)) \text{..... (6)}$$

where $P(T_i)$, $P(S_i)$ are the likelihoods of the cancer subtype prediction and TNM stage prediction respectively. The proposed framework can be used for computer-aided diagnosis (CAD) of lung cancer for practical healthcare applications as the result of this framework will be the predicted lung cancer subtype, the TNM based clinical stage and also the confidence score of the predicted stage.

Conclusion

In this paper, a lightweight model, LiteSOS-Net, is developed to perform lung cancer subtype classification and staging based on the TNM classification system simultaneously from computed tomography (CT) images. The proposed framework combines the following modules: (1) Attention U-Net for accurate lung nodule segmentation, (2) a lightweight backbone with channel shuffle operations, (3) the Osprey Optimization Algorithm (OOA) for improved model optimization, and (4) a TNM-guided staging module for comprehensive disease assessment. The proposed framework overcomes high computation complexity, absence of integrated classification and staging and limited applicability for the edge deployment real-time. Moreover, the ease of use of LiteSOS-Net enables its implementation in resource-limited healthcare settings and its ability to deliver an efficient diagnostic performance. Further studies will be carried out in the future with the aim of also fully experimentally validating the proposed framework using benchmark CT image sets. The Quantitative metrics and TNM stage prediction accuracy will be used to evaluate the performance of LiteSOS-Net. Comparative analysis with the existing techniques and deployment experiments on the edge will also be conducted to showcase the effectiveness and applicability of the proposed framework.

References

1. S. L. Tan, G. Selvachandran, R. Paramesran, and W. Ding, "Lung cancer detection systems applied to medical images: A state-of-the-art survey," Archives of Computational Methods in Engineering, vol. 32, no. 1, pp. 343–380, Jan. 2025. https://doi.org/10.1007/s11831-024-10141-3
2. 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 of lung cancer on chest computed tomography: A systematic review," European Radiology, vol. 34, no. 11, pp. 7397–7407, 2024. https://doi.org/10.1007/s00330-024-10804-6
3. C. Gao, L. Wu, W. Wu, Y. Huang, X. Wang, Z. Sun, M. Xu, and C. Gao, "Deep learning in pulmonary nodule detection and segmentation: A systematic review," European Radiology, vol. 35, no. 1, pp. 255–266, Jan. 2025. https://doi.org/10.1007/s00330-024-10907-0
4. G. Cai, Y. Cai, Z. Zhang, Y. Cao, L. Wu, D. Ergu, Z. Liao, and Y. Zhao, "Medical artificial intelligence for early detection of lung cancer: A survey," Engineering Applications of Artificial Intelligence, vol. 159, Part B, Art. no. 111577, Nov. 2025. https://doi.org/10.1016/j.engappai.2025.111577
5. W. Zhu, X. Wang, J. Xing, X. S. Xu, and M. Yuan, "YOLOv8-BCD: A Real-Time Deep Learning Framework for Pulmonary Nodule Detection in Computed Tomography Imaging," Quantitative Imaging in Medicine and Surgery, vol. 15, no. 9, pp. 8189–8208, 2025. https://doi.org/10.21037/qims-24-975
6. S. Majumder, N. Gautam, A. Basu, A. Sau, Z. W. Geem, and R. Sarkar, "MENet: A Mitscherlich Function Based Ensemble of CNN Models to Classify Lung Cancer Using CT Scans," PLoS ONE, vol. 19, no. 3, e0299625, 2024. https://doi.org/10.1371/journal.pone.0299625
7. H. Jafari, K. Faez, and H. Amindavar, "Swin-Tempo: Temporal-Aware Lung Nodule Detection in CT Scans as Video Sequences Using Swin Transformer-Enhanced U-Net," arXiv:2310.02640, 2023.
8. W. Hendrix et al., "Deep Learning for the Detection of Benign and Malignant Pulmonary Nodules in Non-Screening Chest CT Scans," Communications Medicine, vol. 3, no. 1, 2023. https://doi.org/10.1038/s43856-023-00377-0
9. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Proc. MICCAI*, pp. 234–241, 2015, https://doi.org/10.1007/978-3-319-24574-4_28
10. B. Abdollahzadeh, S. Mirjalili, and H. Faris, "Osprey Optimization Algorithm: A Novel Nature-Inspired Metaheuristic Optimizer," *Artificial Intelligence Review*, vol. 57, 2024, https://doi.org/10.1007/s10462-024-10929-4