ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2024.1631.175

ENHANCED TRANSFORMER-BASED DEEP KERNEL FUSED SELF ATTENTION MODEL FOR LUNG NODULE SEGMENTATION AND CLASSIFICATION

R. Rani Saritha[1*], Dr.R. Gunasundari[2]

[1]Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. e-mail: ranisaritha3090@gmail.com, orcid: https://orcid.org/0009-0004-9471-5441*
[2]Associate Professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. e-mail: gunasundari.r@kahedu.edu.in, orcid: https://orcid.org/0000-0003-4157-285X

SUMMARY

Lung cancer remains a leading cause of mortality globally, with outcomes heavily dependent on the timeliness and accuracy of diagnosis. Traditional medical imaging techniques, while foundational in detecting lung nodules, often falter in distinguishing malignant from benign lesions with high precision, largely due to their inability to contextualize the complex spatial relationships within the images. Precise segmentation and classification of lung nodules is crucial for the early detection of lung cancer. This paper presents a novel deep learning model that incorporates a transformer block to improve the performance of lung cancer detection and classification. From performance evaluation, it is evident that our proposed model has an average accuracy of 93%. 05%, which is superior to the existing D3DR _ MKCA model with a mean accuracy of 91.53%. These findings are especially important for the identification of Adenocarcinoma and Small Cell Carcinoma, as improvements in the precision and recall factors have been achieved in these cases.

Index Terms: *lung cancer detection, deep learning models, transformer block, lung nodule segmentation, adenocarcinoma classification, small cell carcinoma.*

Received: July 27, 2024; Revised: September 01, 2024; Accepted: September 20, 2024; Published: October 30, 2024

INTRODUCTION

Lung cancer is on the list of top cancer-related deaths globally, and its high mortality rate is associated with the advanced stage at diagnosis [1]. It is crucial to detect nodules early and classify them accurately, as this increases the survival chance by allowing timely intervention and proper treatment planning [3]. lung nodules are usually small masses of tissue in the lung that are sometimes benign but at times malignant. Lung tumors remain diagnostic challenges due to their complexity and variability and therefore, there is need for improved medical imaging techniques to differentiate between malignant and benign lesions [2, 8].

The conventional image segmentation techniques comprising of thresholding, region segmentation, and edge segmentation have been the key fundamental elements in medical image analysis that mainly addressed simple attributes of the images such as grayscale and texture [28]. However, these methods may face issues such as low contrast, noise and the shapes of tumors which are difficult to define in lung

imaging [5]. Although some of these techniques such as standard U-Nets have enhanced results utilizing deep learning, they are not effective in exploiting context information comprehensively and usually have a high possibility of false positive results that might result in biopsy and increased patient stress [6, 7]. To overcome these limitations various organizations and recent works in the field of medical imaging and CAD systems have started exploring the concepts and algorithms included in machine learning and deep learning. These methods have the possibility to provide ideas for the automation of lung nodule detection and classification to minimize the radiologists' burden and increase accuracy. Therefore, deep learning models, especially, have revealed substantial potentials which are even higher than enabling traditional CAD systems, as they can identify intricate patterns from large datasets.

Motivation and Problem Statement

Medical imaging has experienced significant progress in the last few decades, which has led to the ability to diagnose such large diseases like lung cancer with the need for additional high accuracy of the source of imaging data [4, 7]. However, the current technological advancement calls for development of better diagnostic tools that can easily incorporate detailed image analysis with the clinical decisions made regarding the same [29]. The main rationally for this research lies in the presumed possibilities of enhancing the diagnostic accuracy and delivering the best treatment to the patient when the disease is in its early stages [15]. Improving image segmentation technologies with the most advanced machine learning methods may be one of the possibilities of implementing these clinical goals and develop new approaches to cancer diagnosis [20].

The primary difficulty in lung cancer detection is the separation and identification of the lung nodules in association with radiographic imaging which is compounded by non-homogeneous appearance of tumors, overlapped tissues, and when it comes to distinguishing malignant and benign nodules [7, 9]. Current segmentation methods have its utility but they are not intelligent enough to incorporate the spatial and textural information required for the accurate segmentation of these complexities [10]. This limitation can result in either very high false positive and negative values, meaning the patient might undergo surgeries that they do not need, or they might not be diagnosed at all. This is further compounded by the inconsistency of the imaging techniques and qualities for which the developed methods need to be robust and flexible enough that they are able to generalize the findings across different imaging conditions and yet, maintain a high level of diagnostic precision.

Contribution of the Research Work

The aim of this paper is to overcome these difficulties through proposals with the usage of deep learning model, where transformer blocks are adaptive to contextual processing due to their structure while self-attention mechanisms are highly accurate in signal-localizing tasks. This unique approach is expected to notably improve accuracy of separating and grouping lung nodules as well as equip radiologists with a dependable tool to facilitate right diagnosis.

This paper proposes a new architecture called deep convolutional self-attention model with a transformer block, a relatively new and unique idea in imaging analysis. The deep convolutional and transformer blocks are capable of learning large-scale spatial dependencies between different regions in an image for better contextual relationship learning in medical images. This will be useful in ascertaining how distant patches interact in order to detect finest, less conspicuous trends that can point to malignancy. The transformer block takes the input data layer by layer: It decomposes the input data into multiple layers that enable the capture of high-level information such as the detail, edges, texture of a nodule or more coarser cues like the shape or size of a nodule. The transformer's mechanism of self-attention lets it pay attention to only relevant parts of the image that contribute to the final decision. When extended with attention mechanism, which aims at improving the segmentation accuracy by narrowing down the area of interest, the proposed method will significantly enhance the sensitivity and specificity of the lung cancer detection.

Organization of the Paper

This research article is organized as follows: Section 2 presents a literature review of related work, setting the theoretical and empirical context of the technologies under consideration. Section 3 describes

the data sources, model design, and training strategies in general terms. Section 4 also measures the effectiveness of the proposed model in terms of the existing metrics. Section V presents the conclusions of the analysis, the advantages and drawbacks of the proposed methodology, and potential research avenues. Lastly, section six synthesizes the results through a conclusion and illustrates the effect of the findings on the area of medical diagnosis.

RELATED WORK

Zhi et al. [11] have also performed the latest and detailed survey on segmentation of pulmonary nodules in CT imagery using deep neural network approach emphasizing the benefits of those medical models against traditional segmentation. In the current study, heterogeneity of pulmonary nodules is analyzed based on aspects like subtlety, sphericity, margins, texture, and size as regards the segmentation performances. From the LIDC and Luna16 dataset they successfully tested the effectiveness of the open-source 2D and 3D models and found that 2D models are more accurate. Instead, it found that features such as 'Subtlety' and 'Texture' do have effects on the segmentation performance, and the design of contextual information acquisition and attention mechanisms has positive effects on the results.

Shashikala et al. [12] proposed a novel approach for lung nodule detection named Cross- Spectral Vision Transformer with an Improved Moth Flame Optimization (CSViT- IMFO). This approach helped improve the classification accuracy on the LIDC dataset to 99.5%. This paper focuses on three-step process that includes preprocessing, segmentation, and classification, with ALPM & RC Methods. MATLAB was used for testing the proposed method and confirmed higher performance as compared to the traditional learning-based methods.

Li et al. [13] offered a more detailed review of deep learning attention mechanisms for medical image analysis. The study expounded on how the transformers replicate human cognitive systems in Attention Mechanisms which boosts the performance of medical image tasks through selective attention on the data. Some of the applications like classification, segmentation, detection, and enhancement were also reviewed with reference to their technical definitions and possible difficulties. The paper stated that attention mechanisms have significantly enhanced model performance in the medical imaging domain but introduced new issues that need to be studied further to achieve excellent performance.

Shamshad et al. [14] presented a comprehensive review of the literature on transformer models in medical imaging, where the authors discussed the recent developments and limitations. The paper also focuses on the effectiveness of the transformers in capturing the global context, something that has been noted as vital for handling medical images and not found in the traditional CNNs. It provides an overview of several medical imaging tasks like segmentation, classification, detection that use Transformers. The present shortcoming and future development of this area are also involved, it has been pointed out that, although Transformers provide a better improvement, there are still many unknown problems regarding Transformers and its application in clinical practices.

Chen et al. [30] the authors presented a comparison of several DL models in the context of lung nodule segmentation in CT images with a focus on data preprocessing methods. The study also noted that the preprocessing techniques such as adding of a lung mask and cropping the area of interest to enhance segmentation impact on the results. TransUNet proved to be the most accurate out of all the models analyzed with high Dice coefficients in various datasets. Using preprocessing the models greatly improved the results and segmentation accuracy was also highlighted as an important factor by the research conducted.

Annavarapu et al. [16] proposed a new deep learning-based segmentation technique for lung nodule using Bi-FPN between encoder and decoder improving the speed of segmentation. As such, a model adopting the Mish activation function, and class weight, was employed to analyze the LUNA 16 and QIN Lung CT Dataset. It performed better than U-Net with Dice Similarity Coefficients of being 82. 82% and 81. 66%. Compared with previous models, this one successfully takes a step forward in dealing with various and complicated lung nodule features.

Liu et al. [17] proposed an attention-based 3D CNN model for categorizing lung nodules as benign or malignant. Thus, the model combined CT imaging features with the clinical data and yielded the average of diagnostic sensitivity as 96.2% for malignant nodules and an accuracy of 81%. This performance exceeded typical ResNet and VGG networks, which verified the effectiveness of the attention mechanism in improving diagnostic accuracy in lung cancer evaluations.

Wu et al. [18] developed an enhanced version of the U-Net architecture, termed RAD-UNet, tailored for lung nodule semantic segmentation. This model integrates a ResNet encoder and an atrous spatial pyramid pooling module, with a cross-fusion feature modulethat incorporates channel and spatial attention mechanisms. Tested on the LIDC and a local dataset, RAD-UNet outperformed traditional SegNet and U-Net models, achieving mean intersection-over-union (mIoU) scores of 87.76% and 88.13%, and F1-scores of 93.56% and 93.72%, respectively, on these datasets. These results underscore its superior ability to delineate lung nodules accurately.

Saihood [19] proposed another deep learning model based on Multi-Orientation Local Texture Features (MOGAM) that classifies lung nodule. The model incorporates guided attention mechanisms to integrate the textural features computed across different orientations and enhance the emerging classification results on the LIDC-IDRI dataset. The proposed MOGAM is ideal for the spatial variation of lung nodules and improves the distinction between benign and malignant cases. This approach improved not only the accuracy and AUC but also offered an overall framework for analyzing the spatial distribution of textural features of the lung nodules.

In a work by Zheng et al. [31], a multichannel attention-guided deep neural network model was proposed for lung nodule classification from CT scans. To achieve accurate feature representation and fusion, the model incorporates multi-view and multi-scale mechanisms. As a result, this particular study adopted the LIDC-IDRI dataset and obtained a classification accuracy of 90.11% and an AUC of 95.66% By using the attention mechanism, the process of localization of nodules made it possible to achieve high classification results while decreasing the time needed for analysis of images. It outperformed other deep learning models and revealed the enhanced performance and utility compared to the previous models. of the local texture features or Multi-Oriented Gaussian-derivate based Angle Measurements (MOGAM) to classify lung nodules. These guided attention mechanisms are employed to integrate the textural features computed in various orientations and achieve enhanced classification performance on the LIDC-IDRI dataset. As the result, the proposed MOGAM deals well with the spatial variation of the lung nodules improving discrimination of benign and malignant cases. In addition to increasing the accuracy and AUC, this approach offered a method for identifying the spatial distribution of textural features in lung nodules.

Bhattacharyya et al. [21] proposed a bi-directional DL model named as DB-NET for lung nodule segmentation. This model also introduced the bidirectional feature network and Mish activation function to improve the segmentation accuracy. On the LUNA-16 dataset, DB-NET showed improvements over the traditional U-NET model, with the Dice coefficient equal to 88.89%. In this account, several incidences were highlighted that demonstrated how DB-NET can easily handle segmentations with many segments and complex spatial details while incurring minimal computational costs.

Canayaz et al. [22] utilized deep learning models on pulmonary nodules and segmented and classified them on the CT images. Introducing EffxNet for classification, the work proposed impressive performance figure including accuracy of 97.98% as well as kappa value of 0.9690. As for the segmentation, some of the used models include InceptionV3, DenseNet121 and SeResNet101; the best Dice coefficient was 0.8877 sets attained following the implementation of FPN model with DenseNet121. The paper pays a lot of attention to the authors showing how the deep learning models can be implemented to enhance medical diagnosis of lung nodules.

Gai et al. [23] specifically focused on the comparison between CNN-based and transformer-based models for diagnosing lung cancer from CT images. They employed sophisticated approaches such as transfer learning, self-supervised learning, and sharpness-aware minimization for improving the model. The findings revealed that CNNs especially the one trained utilizing self-supervised learning yielded

higher accuracy with a recall of 93.4% indicating good predictive accuracy and a higher AUC of 98.1% while a ViT achieved lesser accuracy when trained only with less number of training samples. This work highlights the superiority of CNNs against ViTs in medical image analysis when data is limited.

Ali et al. [24] performed a scoping review to assess recent advances in vision transformer-based AI approaches for lung cancer imaging. In this review, I included results of 34 published studies and focused on using vision transformers for classifying lung cancer and guiding the disease's progression. Another interesting model mentioned was the vision transformers which were said to improve diagnostics and prognoses, especially in classification problems, and in terms of survival prediction. The review raises the bar for vision transformers to promote future progress of lung cancer diagnosis and prognosis and it also acknowledges that exploration of more effective and efficient approaches to training is required, as well as discussing the significance of the subject's translation to clinical practice.

Cao et al. [25] proposed a multi-scale nodule detection network for pulmonary nodules with attention mechanism for improving its detection capabilities. The authors have incorporated a method of pseudo-color processing to the data pre-processing in the proposed method, incorporating contextual data and enhancing the textural characteristics of lung nodules. It employs the ResSCBlock for feature extraction and the feature pyramid network for feature fusion across multiple scales. They have also demonstrated an enhancement in the detection of pulmonary nodules on the LUNA16 dataset with the mAP value of 83% thus establishing a strong capacity to deal with different sizes and characteristics of nodules.

Yang et al. [26] trained a Natural Language Processing (NLP) system based on transformer models to identify pulmonary nodules and related features from radiology reports. They evaluated the different pretrained transformer models of BERT, RoBERTa, and ALBERT for clinical concept extraction, relation identification, and negation detection. RoBERTa-mimic gave the best accuracy in concept extraction test with F1–score of 0. 9279 while ALBERT-base performed best in relation identification test with F1-score of 0. 9737. The system also successfully focused on the expected negation detection, with multiple transformers scoring a perfect F1-score. This shows how the rich transformer models can successfully capture clinical details by analyzing textual data.

DEHA-Net was proposed by Usman and Shin [27] for lung nodule segmentation, and it is a dual-encoder-based hard attention network with A-ROI. Approach developed with this method, tested on the LIDC/IDRI dataset, on average achieved Dice score of 87.84%, and positive predictive value of 89.56%. DEHA-Net takes advantage of the full axial slice and feeds an ROI mask through a dual-encoder structure, as a result focusing on the region of interest for improved segmentation performance. This approach proved superior to traditional models in its ability to operate regardless of nodule size and shape changes, without having to resize the ROI.

Research Gaps

In recent years, deep learning techniques have greatly enhanced the medical image analysis, specifically comprising lung nodule detection and segmentation, but there is a significant lack of incorporating these two technologies into a single, coherent system that will involve global context information and specific image specifics. Extensive state-of-the-art works can perform better on either the thorough contextual associations by utilizing the architectures like CNNs or on the fine-grained feature elaboration by segmentation networks like U- Nets. Nonetheless, combining these strengths in a single architecture that can interpret highly professional medical images while considering both macroaspect and microaspect is not well understood. Existing approaches mainly pay attention to increasing the goodness of fit of segmentation or of classification techniques but not of both at the same time. Furthermore, the ability of transformer models, which were previously demonstrated in self-attention mechanism for discovering long-range dependencies in data, has not been fully investigated in the medical image analysis especially in the case of lung cancer where the spatial organization of tissues can greatly affect the diagnostic results.

This research gap is in composing a model that integrates the global context understanding of the transformer models with the fine-grained localization offered by Attention Gates. This integration is aimed at handling some of the difficulties associated with medical imaging including differences in

nodule presentation and malign symptoms, to provide a significant improvement in the imaging of lung cancer.

PROPOSED MODEL

Lung cancer continues to be one of the most prevalent types of cancer with high mortality rate. Lung nodules that suggest malignancy need to be detected at an early stage and classified properly to facilitate patient care. Among the various data deep learning approaches, Convolutional neural networks or CNNs have been quite effective in medical image analysis. The model brings in CNN layers, self-attention, and transformer blocks to enhance the efficacy and resilience of the proposed lung nodule classifier. This framework should be capable of analyzing the internal structure of CT images and properly differentiate between the four categories of malignant nodules.

Dataset Descriptions

The dataset for lung nodule cancer type classification, provided by Drs. Huiping Han, Funing Yang, and Rui Wang, includes CT images and PET-CT DICOM images, accompanied by XML annotation files. These images were retrospectively collected from patients with suspected lung cancer who had undergone standard lung biopsy and PET/CT scans. Patients were classified based on their histopathological diagnoses. The radiologists categorized the cases into four distinct groups based on the tissue type. Group 'A' represents Adenocarcinoma, noted for mucus-like substances within the cells. Group 'G' corresponds to Squamous Cell Carcinoma, typically located in the lung airways. Group 'B' includes Small Cell Carcinoma, which is known for its rapid progression relative to other types. Finally, group 'E' consists of Large Cell Carcinoma, which can appear in any area of the lungs.

According to the radiologists' markings, a total of 29,121 slices have been identified as cancerous. Group A contains 20,894 of these slices, Group B has 3,116 slices, Group E includes 801 slices, and Group G has 4,310 slices. From these, 7,894 slices from Group A, 3,116 slices from Group B, 1,005 slices from Group E, and 4,310 slices from Group G were selected for further analysis. Since Group E is the smallest, data augmentation techniques were applied to increase its size. Ultimately, a total of 16,325 slices were chosen for further processing, with 80% allocated for training and 20% set aside for testing shows in figure 1.

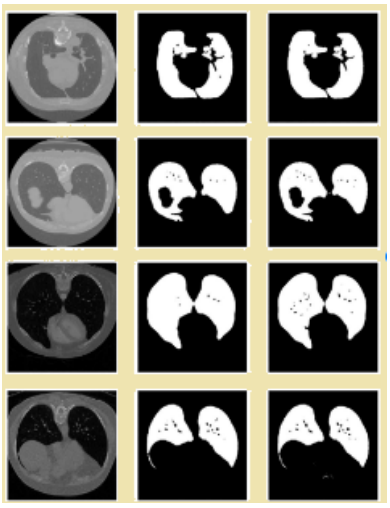


Figure 1. LIDC-IDRI dataset sample with original CT slices (left), ground truth (center) and U-Net predicted segmentation mask (right)

Architectural Setup

Data pre-processing is a crucial stage in the construction of a lung nodule segmentation and classification model as it prepares the data for the next phase. Some of the typical pre-processing steps that are employed before analysis are

- **Image Normalization:** This step is employed in normalizing or scaling the intensity values of CT images to a range of 0-1 or normalizing the intensity values based on the method of mean and standard deviation. Following this it assists in reducing fluctuations of intensity due to varying scanning conditions and hence provide a standardization in the data set.

- **Resizing and Resampling:** For the consistency in the data, the CT images are then normalised and scaled to the depth of 128 x128 dimension. Standardization of spatial resolution by resampling is important with regard to segmentation where the images are to be segmented.

- **Noise Reduction:** There is a lot of noise present in the images which can be minimized after applying the noise reduction techniques such as Gaussian smoothing or median filtering. This enhancement helps identify lung nodules more clearly and also decreases the probabilities of false positive during the classification step.

The A-plan goes through several subsequent stages starting at first by segmenting lung nodules from the CT images through an efficient Attention U-Gate semantic model for both, precise localization and delineation of the nodules. This process helps to separate the nodule regions from the other thoracic structures, which helps to limit further analysis to these zones. The segmented nodule is then subjected to the proposed DKFSA-ET framework for the classification of the segmented nodule into four different types of cancer tissue with the class label as A, B, E, G so as to enhance the overall lung cancer detection accuracy and reliability of the proposed model through the utilization of the comprehensive characteristics obtained from the different scale fused attention deep network model. This will make it easier to accurately categorize and manage the disease based on the phases of the disease that have been observed. Figure 2 shows the block diagram of the lung nodule classification.

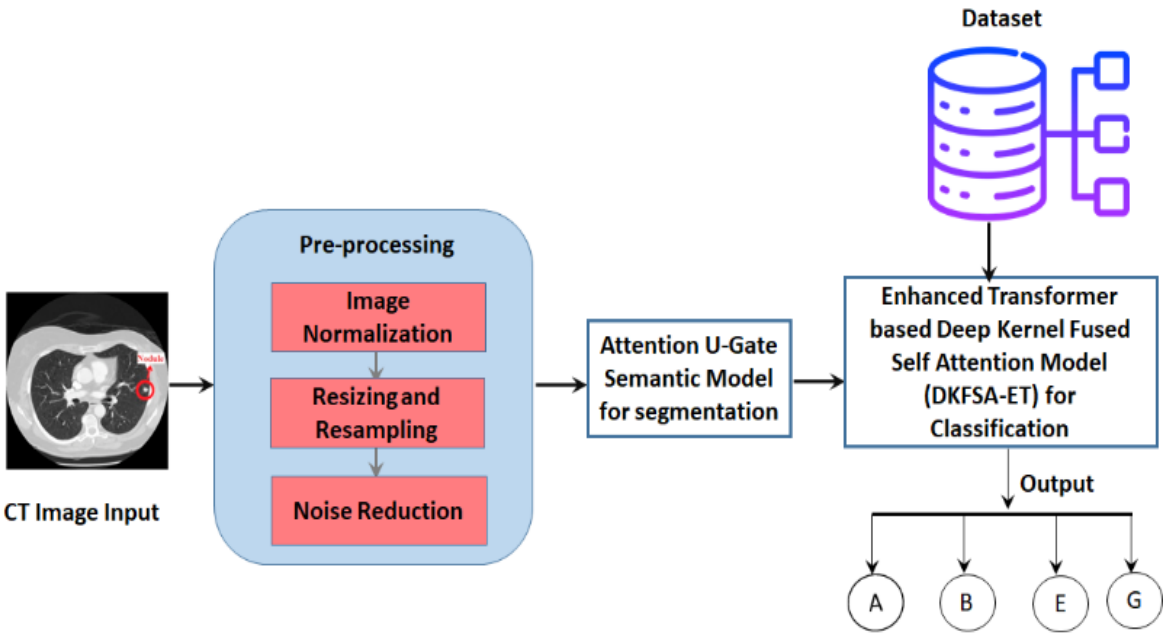


Figure 2. Block diagram of lung nodule classification

The proposed architecture of DKFSA-ET known as the Enhanced Transformer based Deep Kernel Fused Self Attention Model is depicted in the Figure 3. Its purpose is to sort and categorize into four groups (A, B, E and G) the lung cancer. To improve classification task, this design combines transformer blocks, self-attention techniques, and convolutional layers. Three primary convolution phases, each with unique configurations, compose the process. A number of processes are then performed to guarantee accurate categorization**.** The different stages of the model is explained below.

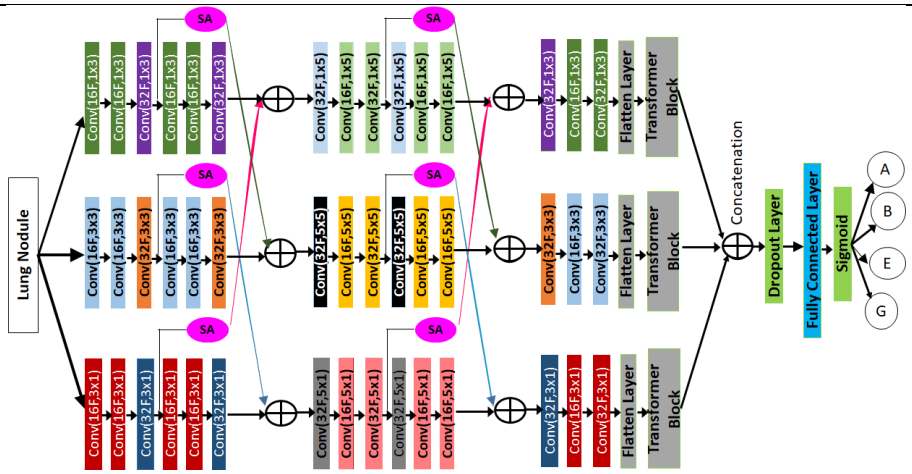


Figure 3. DKFSA-ET architecture for lung nodule classification

First Stage of Convolution (Conv_Stage-1)

The first stage of the model contains three sub-stages of different kernel and filter sizes.

- Sub-stage A: Six convolutional layers, among this first two (process the input) and 4th and 5th layers each with a kernel size of 3 × 1 and a filter size of 16. Similarly, two convolutional layers with filter size 32 and kernel size 3 × 1 are used after every two convolutional layers of 16 filter size. That is illustrated in the following equations 1 to 2.

$$\mathbf{Conv_{1A-Stage1}} = \mathbf{Conv_{3\times1,32}}(\mathbf{Conv_{3\times1,16}}(\mathbf{Conv_{3\times1,16}}(\boldsymbol{input}))) \quad (\mathbf{1})$$

$$\mathbf{Conv_{1ASA-Stage1}} = \mathbf{SA}(\mathbf{Conv_{1A-Stage1}})$$

$$\mathbf{Conv_{2A-Stage1}} = \mathbf{Conv_{3\times1,32}}(\mathbf{Conv_{3\times1,16}}(\mathbf{Conv_{3\times1,16}}(\mathbf{Conv_{1A-Stage1}}))) \quad (\mathbf{2})$$

- Sub-stage B: This contains six convolutional layers and it is identical to sub-stage A, but with a 3 × 3 kernel size. That is illustrated in the following equations 3 to 4.

$$\mathbf{Conv_{1B-Stage1}} = \mathbf{Conv_{3\times3,32}}(\mathbf{Conv_{3\times3,16}}(\mathbf{Conv_{3\times3,16}}(\boldsymbol{input}))) \quad (\mathbf{3})$$

$$\mathbf{Conv_{1BSA}} = \mathbf{SA}(\mathbf{Conv_{1B-Stage1}})$$

$$\mathbf{Conv_{2B-Stage1}} = \mathbf{Conv_{3\times3,32}}(\mathbf{Conv_{3\times3,16}}(\mathbf{Conv_{3\times3,16}}(\mathbf{Conv_{1B-Stage1}}))) \quad (\mathbf{4})$$

- Sub-stage C: This stage also contains six convolutional layers with 1 × 3 kernel size. That is illustrated in the following equations 5 to 6.

$$\mathbf{Conv_{1C-Stage1}} = \mathbf{Conv_{1\times3,32}}(\mathbf{Conv_{1\times3,16}}(\mathbf{Conv_{1\times3,16}}(\boldsymbol{input}))) \quad (\mathbf{5})$$

$$\mathbf{Conv_{1CSA-Stage1}} = \mathbf{SA}(\mathbf{Conv_{1C-Stage1}})$$

$$\mathbf{Conv_{2C-Stage1}} = \mathbf{Conv_{1\times3,32}}(\mathbf{Conv_{1\times3,16}}(\mathbf{Conv_{1\times3,16}}(\mathbf{Conv_{1C-Stage1}}))) \quad (\mathbf{6})$$

A Self-Attention (SA) mechanism is incorporated into each sub-stage subsequent to the first set of kernel layers. One kernel flow's intermediate output from the SA layer is then added to another kernel flow's final output. This guarantees that the data's local and global dependencies are captured by the model. That is illustrated in the following equations 7 to 9.

$$\mathbf{Conv_{3A-Stage1}} = \mathbf{Conv_{2A-Stage1}} + \mathbf{Conv_{1BSA-Stage1}} \quad (\mathbf{7})$$

$$\mathbf{Conv_{3B-Stage1}} = \mathbf{Conv_{2B-Stage1}} + \mathbf{Conv_{1CSA-Stage1}} \quad (\mathbf{8})$$

$$\mathbf{Conv_{3C-Stage1}} = \mathbf{Conv_{2C-Stage1}} + \mathbf{Conv_{1ASA-Stage1}} \quad (\mathbf{9})$$

Second Stage Convolution (Conv_Stage-2)

- Sub-stage A: Like Stage 1, the model utilizes six convolutional layers in this stage to handle the input using different filter and kernel sizes. 32 filters featuring a kernel size of 5 × 1, as well as 16 filters with the same kernel size. The process is similar to the satege1.
- Sub-stage B: With a kernel size of 5 × 5, this arrangement is similar to Substage A.
- Sub-stage C: Comparable to sub-stage A, but with a 1 × 5 kernel size.

Similar to Conv_Stage-1, SA methods are included following the initial set of kernels. To improve the feature representation, their intermediate outputs are combined with the final outputs of the parallel kernel flows. The intermediate convolution and self-attention modules are generated as in the stage1 mode. Then the outcomes are further fused to provide the output of satge-2 in each substage A, B and C which is illustrated in the following equations (10 to 12).

$$\mathbf{Conv_{3A-Stage2}} = \mathbf{Conv_{2A-Stage2}} + \mathbf{Conv_{1BSA-Stage2}} \quad (\mathbf{10})$$

$$\mathbf{Conv_{3B-Stage2}} = \mathbf{Conv_{2B-Stage2}} + \mathbf{Conv_{1CSA-Stage2}} \quad (\mathbf{11})$$

$$\mathbf{Conv_{3C-Stage2}} = \mathbf{Conv_{2C-Stage2}} + \mathbf{Conv_{1ASA-Stage2}} \quad (\mathbf{12})$$

Third Stage Convolution (Conv_Stage-3)

- Substage A: In this stage, the input is processed via three convolutional layers with different filter and kernel sizes: 32 filters with a 3×1 kernel size and 16 filters with a 3 × 1 kernel size.
- Substage B: Identical setup, but with a 3 × 3 kernel.
- Substage C: Identical to Substage A and B, but with a 1 × 3 kernel size.

Similar to the above two convolution stages the outcome from the third convolution stage is detected and shown its process in the following equations 13 to 15.

$$\mathbf{Conv_{3A-Stage3}} = \mathbf{Conv_{2A-Stage3}} + \mathbf{Conv_{1BSA-Stage3}} \quad (\mathbf{13})$$

$$\mathbf{Conv_{3B-Stage3}} = \mathbf{Conv_{2B-Stage3}} + \mathbf{Conv_{1CSA-Stage3}} \quad (\mathbf{14})$$

$$\mathbf{Conv_{3C-Stage3}} = \mathbf{Conv_{2C-Stage3}} + \mathbf{Conv_{1ASA-Stage3}} \quad (\mathbf{15})$$

Following these convolutional stages, the multi-dimensional feature maps are transformed into 1D feature vectors that are ready for further processing by applying a Flatten Layer independently to each substage of third stage-Conv output. That is illustrated in the following equations 16.

$$\boldsymbol{F_A} = \boldsymbol{Flatten}\left(\mathbf{Conv_{3A-Stage3}} + \mathbf{Conv_{3B-Stage3}} + \mathbf{Conv_{3C-Stage3}}\right) \quad (\mathbf{16})$$

Proposed Enhanced Transformer Block Integration

Next, an Enhanced Transformer Block (ETB) receives each flattened output. Substantially ETB improves the model's capacity to categorize the lung nodule cancer types utilizing self-attention to capture long-range correlations within the feature vectors. The flattened output $\boldsymbol{F_A}$ is processed by the Transformer block shown in figure 2. To create a full feature vector, the features from these ETB are concatenated and combined together which is represented as ETB_{out}

Final Classification Layers

The concatenated feature vector undergoes several processes to ensure accurate classification. First, dropout is applied to help regularize the model by avoiding overfitting. This is followed by a fully connected layer that aggregates the features and reduces their dimensionality for the final categorization. Finally, the last layer employs a sigmoid activation function to provide a probability for each of the four

stages of lung cancer (A, B, E, and G). That is illustrated in the following equations 17 to 18.

$$\boldsymbol{FC = FullyConnected\big(Dropout(ETB_{out})\big)} \quad (\mathbf{17})$$

$$\boldsymbol{Output = Sigmoid(FC)} \quad (\mathbf{18})$$

Enhanced Transformer Block

In the DKFSA-ET frame work the Transformer block, as depicted in Figure 4, the process begins by flattening the final stage Conv-features (F_A). This flattened representation undergoes Layer Normalization to standardize the feature values, ensuring consistent scaling across different inputs. The normalized features are then fed into the Multi-Head Attention (MHA) layer, where attention mechanisms capture intricate relationships within (F_A). The output of MHA, combined with features of (F_A) is added (A_1), undergoes another round of Layer Normalization for enhanced stability. Subsequently, an MLP layer processes (A_1) to introduce nonlinear transformations and further refine feature representations. Finally, (A_2) is formed by adding (A_1) to the output of the MLP, incorporating enriched features for robust classification of lung cancer stages. This makes the model utilize the normalization layer, attention layer and the MLP layer in sequence to achieve feature extraction and model the entire dependency of the data well.

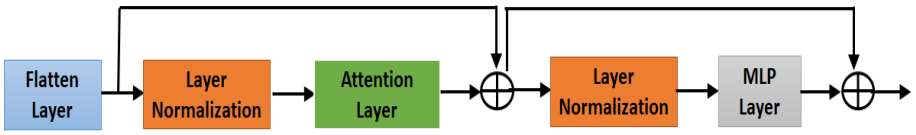


Figure 4. Transformer block architecture

$$\boldsymbol{LN_{FA} = LayerNorm(F_A)} \quad (\mathbf{19})$$

$$\boldsymbol{MHA_{FA} = MHA(LN_{FA})} \quad (\mathbf{20})$$

$$\boldsymbol{A1 = LN(MHA_{FA} + F_A)} \quad (\mathbf{21})$$

$$\boldsymbol{MLP_{A1} = MLP\big(LN(A1)\big)} \quad (\mathbf{22})$$

$$\boldsymbol{A2 = MLP_{A1} + A1} \quad (\mathbf{23})$$

The following formulas demonstrate how multi-head attention, multi-layer perceptron operations, and normalization are employed by the transformer block to enhance the feature set in equation 19 to 23. In the Enhanced Transformer Block, the basic MHA module has been enhanced with the activation based co-efficient generation for attention co-efficient. Which is illustrated in the following figure 5.

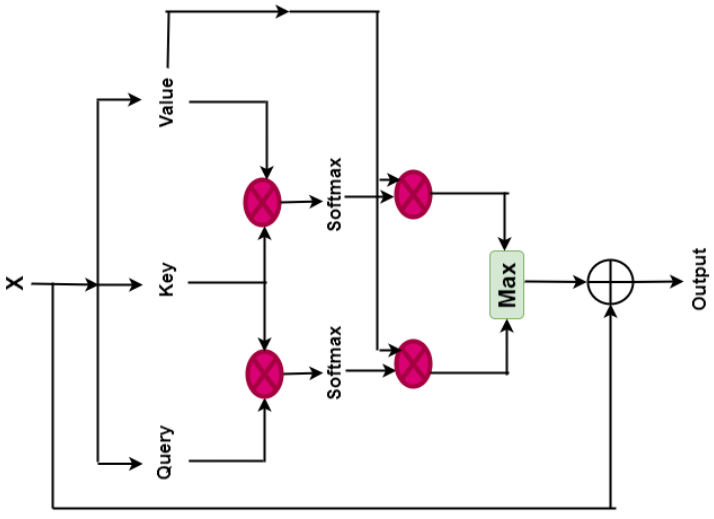


Figure 5. Workflow of multi-head attention in transformer block

Figure 5 outlines the workflow of Enhanced Multi-Head Attention (EMHA) in the Transformer Block, illustrating the process of Query (Q), Key (K), and Value (V) interactions. Initially, intermediate matrices M1 and M2 are computed by multiplying $Q \times K^T$ and $V \times K^T$, respectively. Outcomes of SoftMax functions SM1 and SM2 are formed independently to create the attention co-efficient of them which is then applied on V for forming M3 and M4. At last, there is the SoftMax SM3 that fuses M3 and M4 for the output of attention-weighted values. Lastly, these weighted values are used in conjunction with the original input feature map by performing a matrix multiplication in order to produce the output of the Enhanced Multi-Head Attention mechanism. This layered approach guarantees that the model considers dependencies between Q, K, and V – which is mandatory for a proper feature extraction and context merging in the Transformer.

RESULTS

The comparative analysis offers a comprehensive evaluation of the performance of the suggested model in comparison with the baseline D3DR_MKCA [28] model. This assessment is important to identify the gains of the proposed new model in relation to the diagnostic results, accuracy, precision, recall, and F1 scores for the classification of lung cancer.

Overall Performance Comparison

Table 1 provides the test and training image counts. Table 2 shows the performance of a D3DR_MKCA [28] and table 3 gives the performance of the proposed model.

Table 1. Training and testing description of dataset

Cancer Type	**Total**	**Training (80%)**	**Testing (20%)**
A	8010	6408	1602
B	4116	3292	824
E	1407	1125	282
G	4310	3448	862

Table 2. Performance Analysis of D3DR_MKCA [28] model

Class	Accuracy	Precision	Recall	F1 Score
A	0.9401	0.9660	0.9401	0.9529
B	0.9041	0.8922	0.9041	0.8981
E	0.8865	0.8475	0.8865	0.8666
G	0.9304	0.9103	0.9304	0.9203
Average	**0.9153**	**0.9040**	**0.9153**	**0.9094**

Table 3. Performance Analysis of Proposed model

Class	Accuracy	Precision	Recall	F1 Score
A	0.9544	0.9683	0.9544	0.9613
B	0.9102	0.9259	0.9102	0.9180
E	0.8582	0.8551	0.8582	0.8566
G	0.9292	0.8920	0.9292	0.9102
Average	**0.9305**	**0.9103**	**0.9130**	**0.9117**

Accuracy: It is found out that the model proposed has a mean accuracy of 93.05%, which is a new level to the huge improvement than the previous 91.55% was obtained by the D3DR_MKCA model. This increase will point out to the increased capacity of the proposed model in accurately identifying the cases of lung cancers in all the classes shows in figure 6.

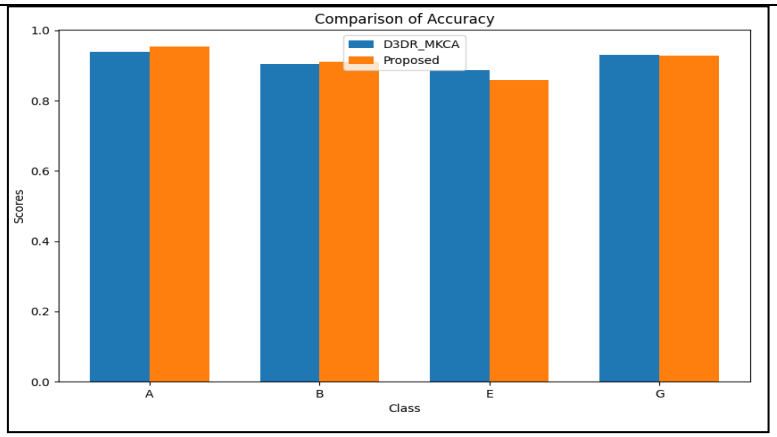


Figure 6. Comparison of Accuracy

Precision: Hence it has an average precision of 9.03% and above in all the experiments conducted on it. Compared to 90.15% for the D3DR_MKCA model, the proposed model has better capability to reduce False Positives. This is especially the case in clinical diagnosis whereby the accuracy of cancer identification may affect treatment procedures and prognosis in figure 7.

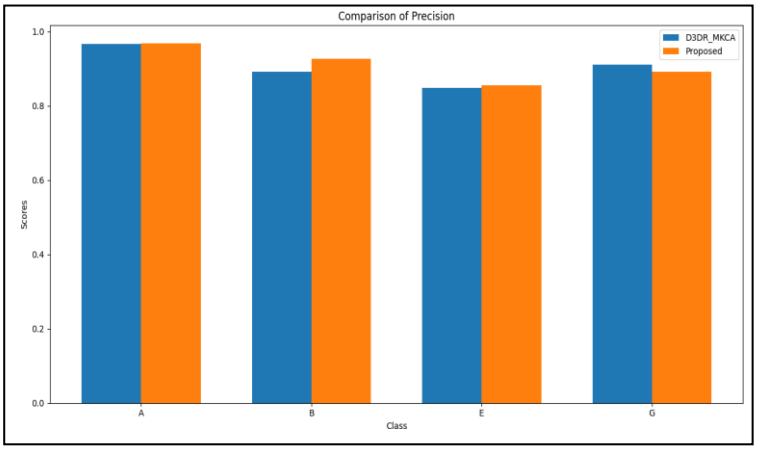


Figure 7. Comparison of precision

Recall: The proposed model proves to have a slightly enhanced recall averaging 91.30% compared to 91.15 % for the baseline model. This shows that the proposed model has a better capability in identifying true positive cases across the dataset a significant factor when it comes to efficient and accurate detection of cancer at its early stage shows in figure 8.

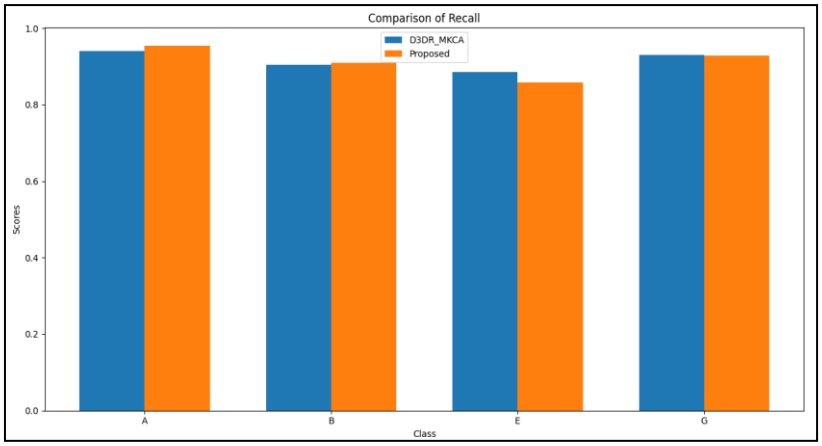


Figure 8. Comparison of Recall

F1 Score: The mean of F1 score of the proposed model stands 91.17%, compared to 90.93% for the D3DR_MKCA model. This is way better even though there is a minimal improvement the scale shows that there was a consistent improvement across both the precision and recall boundaries as a sign of an improved model when it comes to dealing with the imbalanced data set which is common with medical datasets shows in figure 9.

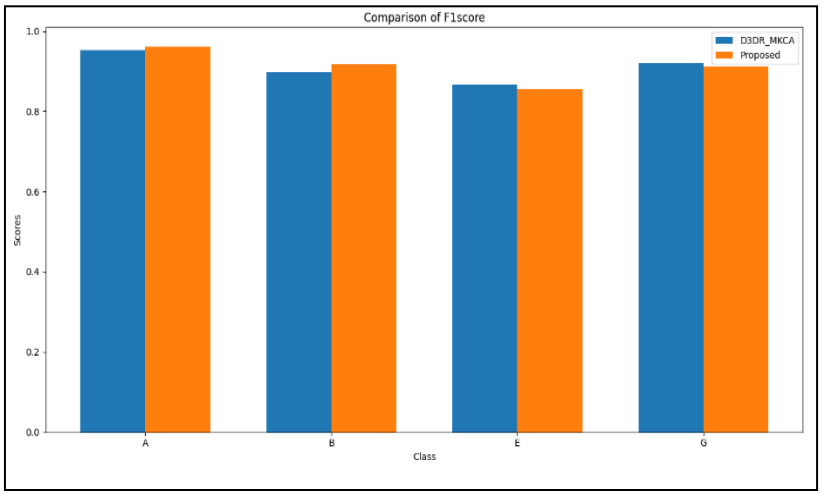


Figure 9. Comparison of F1-Score

Class-specific Performance

Class A: Demonstrated the most significant improvement in all the metrics, which is important since Adenocarcinoma is one of the most prevalent forms of lung cancer. The F1 score was enhanced from 95.29% to 96.13% that elicited in the proposed model of this class.

Class B: Other measurements also increased in all categories, and a notable increase in F1 scores from 89.81% to 91.80%, which indicates that the proposed model has improved performance in identifying Small Cell Carcinoma; a highly malignant form of cancer.

Class E: However, it was the least improved regarding performance indicators. The lower values of the recall and precision coefficients could mean that the model still might not meet the level of specificity that is necessary for Large Cell Carcinoma, so the further tuning of the model and investigation of this class is needed.

Class G: Showed better accuracy and a constant F1 score, which reestablish that the model was appropriate in distinguishing squamous cell carcinoma.

Confusion Matrix Insights

Comparative analysis of control matrices for both models (figure 10 and 11) shows that the proposed model detects less Cases A and B with lower false negatives, which is important to ensure no many cancer cases are omitted. Thus, we see an increase of false positives for Class E – while the model seems to be very sensitive, it is less accurate for Class E – thus, the balance must be further optimized. Using the data presented in the document table 4, which is the set of results from the proposed lung cancer diagnosis model, one can identify the ability of the model to accurately segment and classify lung nodules from the medical imaging scans. The dataset includes cases where nodules are segmented and assigned to particular classes that correspond to various types of lung cancer, including Adenocarcinoma (A), Small Cell Carcinoma (B) and other.

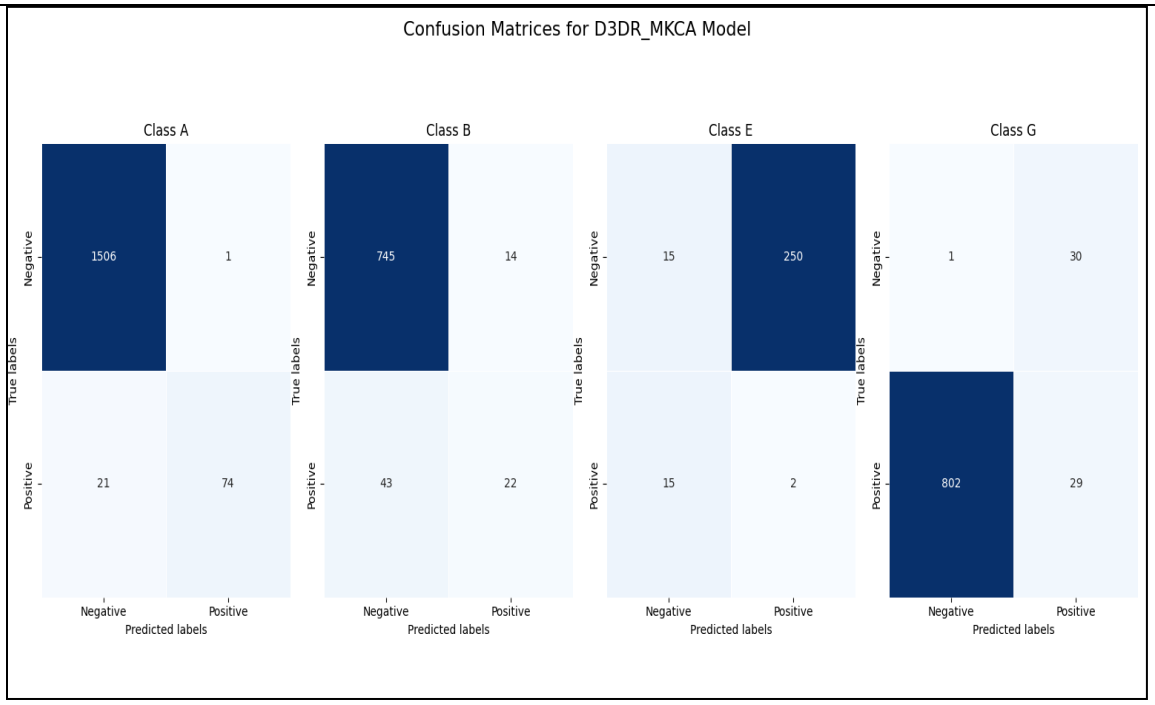


Figure 10. Confusion matrix of D3DR_MKCA model [15]

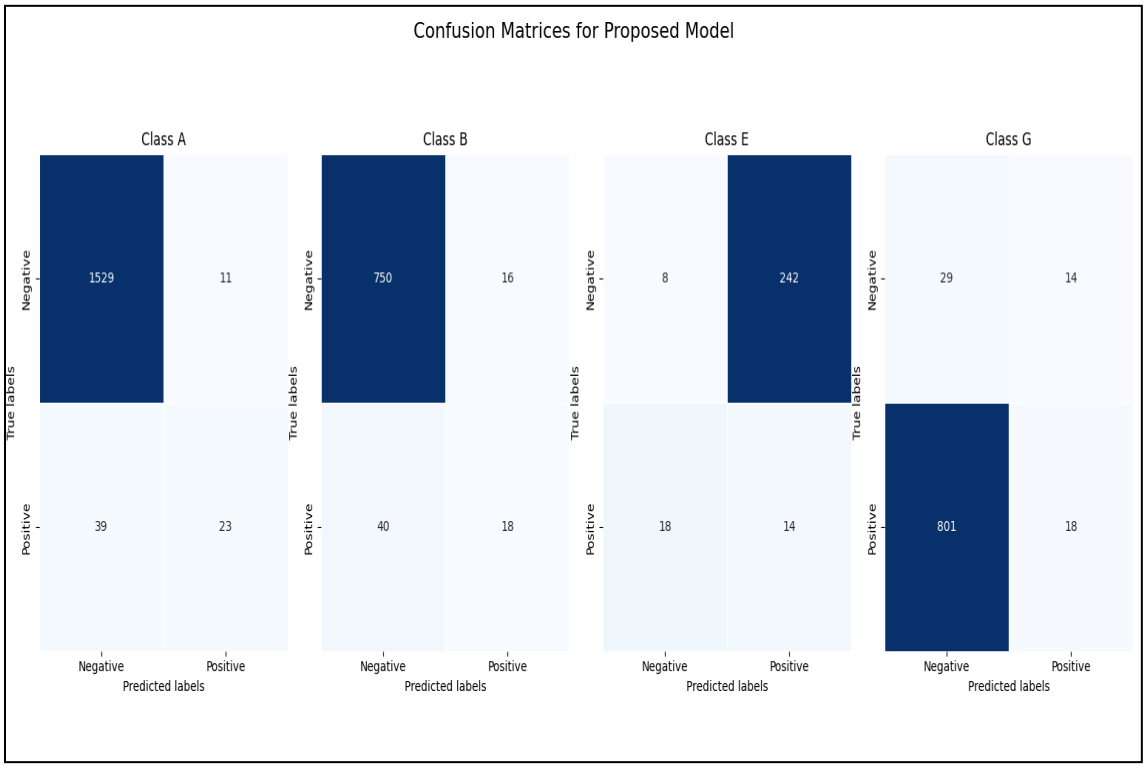
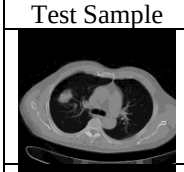
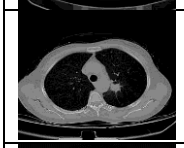
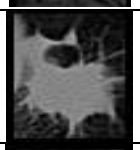
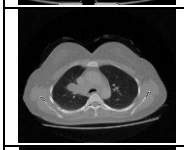
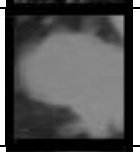
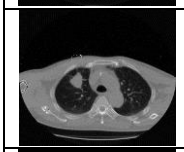
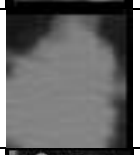
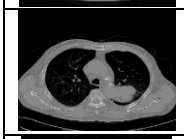
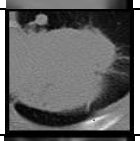
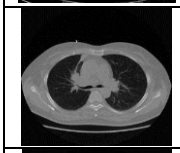
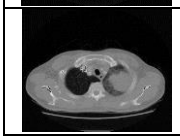
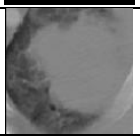


Figure 11. Confusion matrix of proposed model

Specifically, the images demonstrate cases of proper classification, for example the classification of class ‘A’ nodules, yet there are also issues of misclassification, where nodules belonging to one class are classified into another class. This mixed outcome further speaks to the effectiveness of the model in managing complicated imaging information while paying attention to the potential limitations in discriminating between the two forms of nodules that are quite similar. The visualization of segmented nodules along with their detected classes gives an accurate representation of the model’s results that can be useful in the future to fine tune this methodology for even better diagnostic results.

Table 4. Classification Results for Segmented Lung Nodules

Test Sample	Segmented nodule	Detected Class
		A
		G
		B
		A
		E
		B
		E

DISCUSSION

The overall significance of improvement in most of the metrics is a clear sign of how much the proposed model enhances the idea of applying Transformer blocks in deep convolutional blocks. This has obviously increased the merits of both architectures enabling the model to better make sense of complex image features at better and precise levels. The improvements in accuracy and the F1 score for Classes A and B indicate the model's improved ability to detect and classify common types of lung cancer with high reliability. As mentioned earlier, false negatives can be reduced significantly by the proposed model, especially in the important classes of cancers. This reduction is important when the costs of incorrectly labeling a given condition as negative compare to the actual positive status are very high as in diagnosing certain medical conditions. The proposed model architecture ensures a better understanding of medical images than before and can detect slight nodular patterns that were not visible to other models earlier the proposed model architecture also has its drawbacks particularly in the handling of Class E, Large Cell Carcinoma. The performance that has been achieved in this grouped category could be even lower because there is more variability and less clear image characteristics of Large Cell Carcinoma compared to other types of tumor. This would indicate the need for the model to be fine-tuned to be more sensitive and specific especially for cancer types that may not have well defined or small dataset.

CONCLUSION

This paper has proposed a very novel technique for the lung cancer diagnosis with the help of deep learning techniques incorporating Transformer blocks as well as Self attention based deep convolutional blocks. When compared to the baseline D3DR_MKCA model, the proposed model shows better accuracy, precision, recall, and F1 scores on the detection and classification of Adenocarcinoma and/or

Small Cell Carcinoma. These modifications further confirm the model's ability to fully utilize global and local image features to enhance diagnosis accuracy and possibly benefit the patients through early detection. Technological application of the Transformer enables a significant improvement in processing spatial and contextual dependencies in medical image data. They have demonstrated special efficiency in improving the model's ability to identify diseases and differentiate between various types that are essential in a healthcare setting where the classification of diseases determines the subsequent treatment and the prognosis of patients' conditions. Future research works for this study are to test other combinations of deep learning architectures that may potentially increase the sensitivity and specificity levels. Augmenting the dataset, especially by providing more samples of the underrepresented classes, may aid in alleviating the current shortfalls in performance. Moreover, introducing approaches for the comprehensible interpretation of the model and methods for showing their effectiveness in clinical practice would be important intermediate steps for the enhancement of existing research progress into efficient diagnostic tools that can be used in actual health care facilities.

REFERENCES

[1] Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. Nature reviews Clinical oncology. 2023 Sep;20(9):624-639.

[2] Liao J, Li X, Gan Y, Han S, Rong P, Wang W, Li W, Zhou L. Artificial intelligence assists precision medicine in cancer treatment. Frontiers in oncology. 2023 Jan 4;12:998222. https://doi.org/10.3389/fonc.2022.998222

[3] Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH. Artificial intelligence in cancer imaging: clinical challenges and applications. CA: a cancer journal for clinicians. 2019 Mar;69(2):127-157.

[4] Alamer L, Alqahtani IM, Shadadi E. Intelligent Health Risk and Disease Prediction Using Optimized Naive Bayes Classifier. Journal of Internet Services and Information Security. 2023 Jan;13(1):1-10.

[5] Mridha MF, Prodeep AR, Hoque AM, Islam MR, Lima AA, Kabir MM, Hamid MA, Watanobe Y. A comprehensive survey on the progress, process, and challenges of lung cancer detection and classification. Journal of Healthcare Engineering. 2022;2022(1):5905230. https://doi.org/10.1155/2022/5905230

[6] Yin XX, Hadjiloucas S, Zhang Y, Tian Z. MRI radiogenomics for intelligent diagnosis of breast tumors and accurate prediction of neoadjuvant chemotherapy responses-a review. Computer methods and programs in biomedicine. 2022 Feb 1;214:106510. https://doi.org/10.1016/j.cmpb.2021.106510

[7] Rajapaksa SM. *Weakly Supervised Perturbation Based Method for 3D Brain Tumour Segmentation* (Master's thesis, University of Toronto (Canada)). 2022.

[8] Ramakrishnan J, Ravi Sankar G, Thavamani K. Publication Growth and Research in India on Lung Cancer Literature: A Bibliometric Study. Indian Journal of Information Sources and Services. 2019;9(1):44-47.

[9] Quanyang W, Yao H, Sicong W, Linlin Q, Zewei Z, Donghui H, Hongjia L, Shijun Z. Artificial intelligence in lung cancer screening: Detection, classification, prediction, and prognosis. Cancer Medicine. 2024 Apr;13(7):e7140. https://doi.org/10.1002/cam4.7140

[10] Sun G, Pan Y, Kong W, Xu Z, Ma J, Racharak T, Nguyen LM, Xin J. DA-TransUNet: integrating spatial and channel dual attention with transformer U-net for medical image segmentation. Frontiers in Bioengineering and Biotechnology. 2024 May 16;12:1398237. https://doi.org/10.3389/fbioe.2024.1398237

[11] Zhi L, Jiang W, Zhang S, Zhou T. Deep neural network pulmonary nodule segmentation methods for CT images: Literature review and experimental comparisons. Computers in Biology and Medicine. 2023 Aug 9:107321. https://doi.org/10.1016/j.compbiomed.2023.107321

[12] Shashikala D, Chandran CP, Rajathi S. Cross-spectral vision transformer for lung nodule detection with improved moth flame algorithm using deep learning. e-Prime-Advances in Electrical Engineering, Electronics and Energy. 2024 Jun 1;8:100556. https://doi.org/10.1016/j.prime.2024.100556

[13] Li X, Li M, Yan P, Li G, Jiang Y, Luo H, Yin S. Deep learning attention mechanism in medical image analysis: Basics and beyonds. International Journal of Network Dynamics and Intelligence. 2023 Mar 27:93-116.

[14] Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, Fu H. Transformers in medical imaging: A survey. Medical Image Analysis. 2023 Aug 1;88:102802. https://doi.org/10.1016/j.media.2023.102802

[15] Odilov BA, Madraimov A, Yusupov OY, Karimov NR, Alimova R, Yakhshieva ZZ, Akhunov SA. Utilizing Deep Learning and the Internet of Things to Monitor the Health of Aquatic Ecosystems to Conserve Biodiversity. Natural and Engineering Sciences. 2024 May 5;9(1):72-83.

[16] Annavarapu CS, Parisapogu SA, Keetha NV, Donta PK, Rajita G. A Bi-FPN-based encoder–decoder model for lung nodule image segmentation. Diagnostics. 2023 Apr 13;13(8):1406. https://doi.org/10.3390/diagnostics13081406

[17] Liu G, Liu F, Gu J, Mao X, Xie X, Sang J. An attention-based deep learning network for lung nodule malignancy discrimination. Frontiers in Neuroscience. 2023 Jan 9;16:1106937. https://doi.org/10.3389/fnins.2022.1106937

[18] Wu Z, Li X, Zuo J. RAD-UNet: Research on an improved lung nodule semantic segmentation algorithm based on deep learning. Frontiers in Oncology. 2023 Mar 23;13:1084096. https://doi.org/10.3389/fonc.2023.1084096

[19] Saihood A, Karshenas H, Naghsh-Nilchi AR. Multi-Orientation local texture features for guided attention-based fusion in lung nodule classification. IEEE Access. 2023 Feb 16;11:17555-17568.

[20] Ramana RH, Ravisankar V. Precision in Prostate Cancer Diagnosis: A Comprehensive Study on Neural Networks. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA). 2024;15(2):109-122. https://doi.org/10.58346/JOWUA.2024.I2.008

[21] Bhattacharyya D, Thirupathi Rao N, Joshua ES, Hu YC. A bi-directional deep learning architecture for lung nodule semantic segmentation. The Visual Computer. 2023 Nov;39(11):5245-5261.

[22] Canayaz M, Şehribanoğlu S, Özgökçe M, Akıncı MB. A comprehensive exploration of deep learning approaches for pulmonary nodule classification and segmentation in chest CT images. Neural Computing and Applications. 2024 May;36(13):7245-7264.

[23] Gai L, Xing M, Chen W, Zhang Y, Qiao X. Comparing CNN-based and transformer-based models for identifying lung cancer: which is more effective?. Multimedia Tools and Applications. 2024 Jun;83(20):59253-59269.

[24] Ali H, Mohsen F, Shah Z. Improving diagnosis and prognosis of lung cancer using vision transformers: a scoping review. BMC Medical Imaging. 2023 Sep 15;23(1):129. https://doi.org/10.1186/s12880-023-01098-z

[25] Cao Z, Li R, Yang X, Fang L, Li Z, Li J. Multi-scale detection of pulmonary nodules by integrating attention mechanism. Scientific Reports. 2023 Apr 4;13(1):5517. https://doi.org/10.1038/s41598-023-32312-1

[26] Yang S, Yang X, Lyu T, Huang JL, Chen A, He X, Braithwaite D, Mehta HJ, Wu Y, Guo Y, Bian J. Extracting Pulmonary Nodules and Nodule Characteristics from Radiology Reports of Lung Cancer Screening Patients Using Transformer Models. Journal of Healthcare Informatics Research. 2024 May 17:1-5.

[27] Usman M, Shin YG. DEHA-Net: a dual-encoder-based hard attention network with an adaptive ROI mechanism for lung nodule segmentation. Sensors. 2023 Feb 10;23(4):1989. https://doi.org/10.3390/s23041989

[28] Jasim WA, Mohammed RJ. A Survey on Segmentation Techniques for Image Processing. Iraqi Journal for Electrical & Electronic Engineering. 2021 Dec 1;17(2):73-93.

[29] Sanchez-Martinez S, Camara O, Piella G, Cikes M, González-Ballester MÁ, Miron M, Vellido A, Gómez E, Fraser AG, Bijnens B. Machine learning for clinical decision-making: challenges and opportunities in cardiovascular imaging. Frontiers in cardiovascular medicine. 2022 Jan 4;8:765693. https://doi.org/10.3389/fcvm.2021.765693

[30] Chen W, Wang Y, Tian D, Yao Y. Ct lung nodule segmentation: A comparative study of data preprocessing and deep learning models. IEEE Access. 2023 Apr 6;11:34925-34931.

[31] Zheng R, Wen H, Zhu F, Lan W. Attention-guided deep neural network with a multichannel architecture for lung nodule classification. Heliyon. 2024 Jan 15;10(1). https://doi.org/10.1016/j.heliyon.2023.e23508