Deep Alternate Kernel Fused Self-Attention Model-Based Lung Nodule Classification

Rani Saritha R. * and V. Sangeetha

Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India Email: ranisaritha3090@gmail.com (R.S.R.); Drsangeetha.v@kahedu.edu.in (V.S.) *Corresponding author

Abstract-Lung cancer causes death with delayed diagnosis and inadequate treatment. Hence there is a need for a computer-aided detection method that can identify the nodule category whether it is benign or malignant to avoid delays in diagnosis with the help of Computerized Tomography (CT) scans. This study proposed a novel architecture Deep Alternate Kernel Fused Self-Attention Model (DAKFSAM) which utilizes the characteristics of the residual network in different forms as well as incorporates the efficiency of the attention model. This model fuses the features extracted from different alternate kernel models in three levels of process with three kinds of alternate kernel models. The self-attention model takes multiple kernel flows' visual attention features and merges them into a form to improve nodule classification efficiency. The performance assessment utilizes the Lung Image Database Consortium-Image Database Resource Initiative (LIDC-IDRI) dataset, and the DAKFSAM mode, as proposed, achieves an F1-Score of 94.85%.

Keywords—pulmonary nodules, nodule detection, nodule classification, deep learning, convolutional neural networks, computer-aided diagnosis, medical imaging

I. INTRODUCTION

Lung cancer persists as the most widespread form of cancer, resulting in approximately 1.80 million annual deaths and standing as the leading cause of global fatalities [1, 2]. In 2020, the World Health Organization (WHO) projected nearly 2.21 million new cases of lung cancer, anticipating over 1.80 million fatalities [2]. Looking ahead to 2023, the American Cancer Society estimated that 238,340 individuals would receive a lung cancer diagnosis. Furthermore, WHO forecasts that by 2030, approximately 17 million people worldwide will be diagnosed with various forms of cancer. This estimation for the United States is based on considerations of factors such as population growth, demographic aging, lifestyle changes, and advancements in cancer detection and treatment [3]. Ciggy smoking is the key contributor to lung cancer and the major cause of death (80%). The second leading factor in lung cancer is radon gas exposure [4]. Conversely, a patient's life

quality and chance of survival can be considerably increased by an early identification and precise treatment of lung carcinoma. Specifically, the detection of pulmonary tumors in the initial phases of an automated system is significant and indispensable.

Lung cancer is a serious condition caused by abnormal or unnatural cells that multiply and eventually form a lung tumor, and are classified as benign and malignant. Malignant tissues grow quickly and can develop throughout the body, creating a health danger, whereas benign tissues do not grow quickly and are noncancerous. There are several techniques for diagnosing lung cancer, including Computerized Tomography (CT), isotope, X-ray, and the non-invasive Magnetic Resonance Imaging (MRI) [4]. Two commonly used modalities for anatomical imaging in the identification of diverse lung diseases are chest radiography, X-ray, and Computerized Tomography (CT) [4]. Computed tomography is recognized as one of the best techniques to identify lung cancer early among them [5]. With the advancement in CT tools high-resolution CT scans have become the imaging method of choice for the detection and analysis of lung illnesses [6]. Even though it suggests examining lung images with the increasingly better-automated resolution, manually assessing such images may be inefficient and extremely reliant on the skill and expertise of radiologists. Therefore, there is a great demand for Computer-Aided detection (CAD) systems for lung cancer classification [6].

Recently deep learning has made the tremendous capacity to enhance the performance of medical image analysis like segmentation, feature extraction, and classification. It may assist physicians in diagnosing the disease at the early or initial stage. So many researchers have developed automated lung cancer classification algorithms using deep learning. This present study employed a deep learning scheme to automatically classify images of lung nodules.

In this work, the performance of the deep Convolutional Neural Network (CNN) model is enchanted via the deep residual concept. It solves the degradation problem as well as improves the efficiency of the model. Normally, malignant nodules vary in this dimension. So, we need to extract features from the small-scale nodules as well as large-size nodules. There is a contradiction in the size of nodules and filters. This

Manuscript received December 23, 2023; revised March 1, 2024; accepted May 13, 2024; published November 8, 2024.

problem is solved with the fusion of different scale residual self-attention layers [6].

The article is divided into several parts. Section I outlines the background and significance of this research study. Section II works section crucially examines the existing research on the topic, highlighting key theories, methodologies, and findings relevant to the current study. Section III defines the methodology used for lung-nodule segmentation and distinguishes whether the nodules are malignant or benign. It also describes the dataset used for analyzing the nodules. Section IV defines the performance metrics used for analyzing the results and performance of the developed model. Section V is the conclusion part, in which the findings and significance of this research are briefed.

II. RELATED WORKS

The Lung Nodule classification has to be done by segmentation and classification stage; hence the related works are given in that order.

A. Lung Nodules Segmentation

The accurate and precise segmentation of lung tumors from CT pictures is a crucial step in the identification of lung nodules. Numerous approaches for effective lung segmentation have been developed for this purpose. Rakesh and Shanthi [7] developed the cuckoo search optimization approach for optimizing the segmentation of the lung, and subsequently the Active Contour (AC) technique for segmenting the lumps from the segmented lung picture. After performing the nodule segmentation operation, the post-processing was fine-tuned using the Markov Random Field (MRF) method. A regiongrowing method is used by Cascio *et al.* [8] to segment the lung nodule. They utilize the slice's central pixel as a seed.

Usman *et al.* [9] suggested a unique technique for the segmentation of lung nodules. Initially, a deep residual U-Net architecture-based adaptive Region of Interest (ROI) method is used to change the ROI's shape and position. Following that, two residual U-Nets are used to execute further patch-wise segmentations through the sagittal and coronal axes. Lastly, the axial, sagittal, and coronal axes' segmentation outputs are combined to create the absolute segmentation mask. To achieve lung nodule segmentation, Dehmeshki *et al.* [10] introduced a framework that utilized two-region growing strategies like contrast-based region growing and fuzzy connectivity.

Gong *et al.* [11] enhanced the U-Net design to segment lung nodules by using residual blocks with a squeeze-and-excitation module. To enhance the capability of network representation, SE-ResNet has incorporated both the encoding and decoding paths of the U-Net method. Savic *et al.* [12] developed a rapid marching method-based segmentation technique, which divides the picture into sections of related characteristics, and then fused by joining sections that expand with kmeans. A resource-conserving design called U-Det was proposed by Keetha *et al.* [13] utilized a Bi-FPN between the encoder and the decoder (bidirectional feature network) as a solution for segmentation. To increase segmentation effectiveness, it additionally uses the Mish activation function and mask class weights. To increase diagnostic performance, Kaviarasu and Sakthivel [14] suggested a novel strategy for lung segmentation that combined the Fuzzy C-means and K-means Clustering Technique (K-CT).

B. Lung Nodules Classification

The classification of lung nodules proposed by Mousa and Khan [15] used histogram equalization with homomorphic filtering for image enhancement using Specific Domain High-Frequency Emphasis Filtering (SPDHFEF). And then, several features are extracted like circularity, mean, and variance. After normalizing the results, the variances and means of a threshold nodule and non-nodule regions were selected as features for classification. A Support Vector Machine (SVM) classifier was utilized for lung-nodule classification. Narayanan *et al.* [16] introduced a new cluster-based classifier architecture for computer-aided identification of lung lumps. They created a unique optimal feature choice strategy for both classifier and cluster modules.

Orozco *et al.* [17] utilized GLCM features for lung nodule detection. The wavelet transform technique is used for segmentation. Support Vector Machine (SVM) classifier with Radial Basis Kernel is applied for lung nodule detection. The structures in the lung were segmented using Gaussian mixture models by Santos *et al.* [18]. Then, using texture descriptors such as Shannon's and Tsallis's Q entropy, distinguish nodules from non-nodules. In addition, the Hessian matrix was utilized to distinguish between blood and bronchi arteries, and SVM was employed to decrease false positives.

Zhou et al. [19] extracted 42-dimensional features from a CT image. And then the extracted features are reduced five times by using feature-level fusion based on rough set theory. Finally, optimize the kernel function of SVM by grid optimization approach. SVM is utilized for classifying lung nodules. To identify lung nodules, Asuntha and Srinivasan [20] suggested a novel deeplearning technique. Scale Invariant Feature Transform (SIFT), Wavelet-based features, Histogram of Oriented Gradients (HoG), Local Binary Pattern (LBP), and Zernike Moment are some of the feature descriptors that are used to extract features. To identify the optimal feature, the Fuzzy Particle Swarm Optimization (FPSO) method is applied. Deep learning algorithms are used to classify the data. The network's computational complexity is reduced using a unique FPSOCNN. Zhou et al. [21] developed a 3D faster R-CNN with a Unet-like structure to effectively learn the characteristics of nodules. It is suggested to classify nodules using a Gradient Boosting Machine (GBM) and 3D Dual Path Network (DPN) characteristics.

George *et al.* [22] suggested a YOLO-based architecture in which a single CNN forecasts several bounding boxes and their class probabilities from specified grids concurrently. To predict lung nodule malignancy, Baldwin *et al.* [23] introduced a Lung

cancer prediction CNN (LCP-CNN) model, which is used to assess and compare regions under the curve of the receiver operating features with the ground truth. Ding et al. [24] suggested a lung nodule identification approach that included 2D candidate generation as well as 3D False Positive Reduction (FPR). A faster RCNN and deconvolution structure is used to find candidates on axial slices, and more slices from the adjacent slices are included in the input. A technique for identifying pulmonary nodules based on multi-group patches derived from lung images was proposed by Jiang et al. [25]. Initially, Frangi-filters were employed to remove vessellike structures, also a slope analysis approach was developed to repair the juxta pleural nodules and eradicate the nodule outside the lung. Finally, using original CT images and binarization, a CNN model with a Multi-Crop (MC) pooling operation was built to acquire the expertise of physicians.

Zhang *et al.* [26] suggested lung nodule recognition based on several patches clipped from input images and enhanced the images by Frangifilter. Another key preprocessing step used in this study was vascular removal, which can emphasize nodules while weakening arteries. Four types of CNN models based on four lump levels are designed to precisely and rapidly find the sites of nodules. Furthermore, the CNN model was fed two sets of Lump candidates that designed pairs of pictures (Group 1 and 2).

To utilize segmentation as well as their uncertainty, Ozdemir et al. [27] suggested a two-stage Bayesian CNN model. Initially, Segmentation networks were trained with 2D axial CT slices. The original picture is then combined with segmentation-predicted mean and standard deviation maps to generate a 3-channel composite picture that is used as an input for 3D Bayesian CNN to identify nodules. Schultheiss et al. [28] developed RetinaNet architecture based on CNN for lung nodule detection. U-Net algorithm is adopted to segment the input image. The prospect of foreign bodies leading to incorrect choices in CNN-based nodule identification schemes was a significant component of this work. Zhou et al. [29] introduced a multi-crop CNN which can extract salient module features automatically by collecting different areas by using feature maps, as well as max-pooling operations, which are done at different periods.

III. EXPERIMENT AND METHOD

A. Lung Nodule Segmentation with Attention Gate- U-Net

In this work, lung nodule segmentation is processed with the help of U-Net with an attention gate. On publicly available benchmark datasets, Convolutional Neural Networks (CNNs) beat conventional methods like multi-atlas and graph-cut segmentation techniques [30] for medical image processing. The fundamental reason for this is that (i) Stochastic Gradient Descent (SGD) optimization is used to discover domain-specific picture characteristics, (ii) Kernels that have been learned are shared by every pixel, and (iii) Convolution techniques effectively use the structural data in medicinal pictures.

In particular, it has been shown that Fully Convolutional Networks (FCN), such as DeepMedic, U-Net, and holistically nested networks, exhibit strong and accurate performance in several tasks, including image segmentation, brain tumors, and cardiac Cardiac Magnetic Resonance Imaging [31–39].

Convolutional layers analyze local information one layer at a time to gradually extract higher dimensional picture representations. Finally, pixels in a highdimensional space are separated based on their semantics. Model predictions are therefore dependent on data gathered from a wide receptive field through this sequential procedure and a non-linear activation function. As a result, the feature map is created using the output of Layer 1 and is initially generated by utilizing a linear transformation. It is frequently selected as a rectified linear unit where c and I represent channel and spatial dimensions correspondingly. Feature activations can be expressed as follows: where the spatial subscript (i) is eliminated in the formulation for notational clarity, and the asterisk (*) stands for the convolutional process. The trainable kernel parameters define the characteristics of the function used in convolution Layer l.

Fig. 1 represents the U-Net with the Attention Gate (AG) architecture of the segmentation process. Fig. 2 demonstrates the use of attention coefficients to identify salient visual areas and trim feature responses to preserve just the activations relevant to the current task. The output of AGs is obtained by multiplying the attention coefficients and input features one at a time.



Fig. 1. Architecture of U-Net with Attention Gate (AG).

By default, a single scalar attention value is assigned to each pixel vector, which represents the number of feature maps in Layer I. This work suggests learning multi-dimensional attention coefficients when there are several semantic classes. This was motivated by Shen *et al.* [40], in which phrase embeddings were learned using multidimensional attention coefficients. Hence, every AG acquires the knowledge to concentrate on a portion of the target structures. According to Fig. 2, focus zones are determined for every pixel i using a gating vector. When using AGs for natural picture categorization as recommended in [41], the gating vector incorporates contextual information to filter lower-level feature responses.



Fig. 2. Attention gate workflow [41].

The gating coefficient is calculated using additive attention [42]. Empirical evidence supports the assertion that, despite demanding increased computational resources, this approach consistently yields more precise results compared to multiplicative attention [43]. The following formula describes additive attention:

$$q_{att}^{l} = \psi^{T} \left(\sigma_1 \left(W_x^T x_i^l + W_g^T g_i + b_g \right) \right) + b_{\psi} \quad (1)$$

$$\alpha_i^l = \sigma_2 \left(q_{att}^l \left(x_i^l, g_i; \Theta_{att} \right) \right) \tag{2}$$

where the sigmoid activation function is represented by $\sigma_2(x_{i,c}) = \frac{1}{1 + \exp(-x_{i,c})}$ and AG is defined by a number of factors Θ_{att} comprising: bias terms $b_{\psi} \in \mathbb{R}$, $b_g \in \mathbb{R}^{F_{int}}$ and linear transformations, $W_x \in \mathbb{R}^{F_1 x Fint}$, $W_g \in \mathbb{R}^{F_g x Fint}$, $\psi \in \mathbb{R}^{F_{int} x 1}$, Using channel-wise $1 \times 1 \times 1$ convolutions the linear transformations are calculated for input tensors. When the incorporated features g and are linearly translated to an intermediate space with a dimension of $\mathbb{R}^{F_{int}}$, this is denoted as vector concatenation-based attention, on other occasions [44].

B. Deep Alternate Kernel Fused Self-Self-Attention Model (DAKFSAM) for Lung Nodule Classification

This work designed a novel architecture Deep Alternate Kernel Fused Self-Attention Model (DAKFSAM) for lung nodule classification with integration of different scales of residual network with self-attention module.

The workflow of the proposed model is shown in Fig. 3. Fig. 4 shows the overall framework of the proposed deep architecture. The attention was used to mix every input sequence in a weighted manner, with the best suitable vectors acquiring the largest weights, to provide the decoder flexibility to employ the input vectors' most pertinent components [45–48]. Since, which completely replaced recurrence with self-attention, deep learning has adopted using attention as the major method for representation learning [49].

Focusing on one scenario as opposed to numerous settings is the definition of self-attention. Using the same

context for the keys, query, and values provides an additional choice. Modern models have been developed for a variety of activities as a result of the ability to directly replicate long-distance interactions, which makes use of the advantages of contemporary technology [50–56]. Adding self-attention to convolution models has resulted in improvements in a variety of applications.



Fig. 3. Workflow of Deep Alternate Kernel Fused Self-Attention Model (DAKFSAM).

A complete attention model may be created by using the independent self-attention layer in place of spatial convolutions. Similar to convolution, self-attention independent models initially remove a small area for a given pixel $x_i j \in R^{(d_i n)}$ in locations $ab \in N_k(i, j)$ with spatial size k cantered around $x_i j$, known as the memory block. As opposed to past studies on attention in vision, which concentrated on global attention among all pixels, this sort of local attention is different [57, 58]. A complete attention model cannot apply global attention at all layers; instead, it can only be used once the input has undergone considerable spatial downsampling.



Then, employing single-headed attention, the pixel result $y_i j \in R^{(d_{(out)})}$ is determined as given in the Eq. (3).

$$y_{ij} = \sum_{a,b \in \mathcal{N}_k (i,j)} softmax_{ab} \left(q_{ij}^{T_k} b_{ab} \right) v_{ab}$$
(3)

where the queries $q_{ij} = W_Q x_{ij}$, keys $k_{ab} = W_K x_{ab}$, and values $v_{ab} = W_V x_{ab}$ are linear-transformations of both pixels in location [*ij*] and its surrounding pixels. The term *softmax*_{ab} refers to a softmax that was applied to each logit that was estimated close to *ij*. w_Q , W_K , and $W_V \in \mathbb{R}^{d_{out} X d_{in}}$ are all learned transforms. Similar to convolutions (Eq. (1), local self-attention collects spatial data over neighbourhoods, but it does so by combining weights from different value vectors into a convex combination (*softmax*_{ab}(.)) parameterized by content interactions. This calculation is carried out once for each pixel *ij*.

The concept of residual block is done in a different form than the existing residual model. Here the output from the Self-attention model of different scales is added to the output feature map of different scales. The input image undergoes convolution through multiple layers (1 to 6) to extract hierarchical features with varying resolutions. Employing 2D convolutional operations, these layers use different kernel sizes and strides, as depicted in Fig. 4. In the initial stage, the segmented lung nodule undergoes processing with three types of kernels $(1 \times 3, 3 \times 3, 3 \times 1)$ in the first level. Each kernel, applied twice with 16 filters, is succeeded by an additional convolution with the same kernel size but 32 filters. This is followed by a self-attention layer, and the output is subjected to two more convolutions with the same kernel in its flow with a filter size of 16 and one more with a filter size of 32. The output from the intermediate selfattention of one kernel flow is then added to the final output of another kernel flow.

The process initiates with the first convolutional layer, where a series of 2D convolutional operations with a 3×1 kernel size is applied, followed by subsequent layers. The second layer is constructed based on features extracted by the first layer. Layer 3, designed to process the input image directly, utilizes 3×3 kernel sizes. The spatial attention mechanism (SA Layer 3) enhances discriminative features. Building upon these Layers, layer 4 continues feature extraction using 3×3 convolutional operations on the output of Layer 3. Additionally, Layers 5 and 6 employ 1×3 kernel sizes for further feature refinement.

Spatial attention mechanisms (SA Layer 1, SA Layer 3, SA Layer 5) are strategically integrated after specific convolutional layers to emphasize crucial regions in the feature maps. Fig. 4 visually represents the connectivity and flow of information through the convolutional layers and spatial attention mechanisms, showcasing the intricate architecture designed for effective lung nodule detection and classification. Alternate kernel integration involves adding the attention feature map from one kernel flow to the output from the first-level process of another kernel flow. Addition operations occur between the output feature maps of specific convolutional layers (2, 4, and 6 layers) and their corresponding spatial attention layers (SA Layer 3, SA Layer 5, SA Layer 1). These added features are fed into the next level of processing with three types of kernels $(5 \times 1, 5 \times 5, 1 \times 5)$ for the second level. The same process repeats for one more level with kernel sizes $(1\times3, 3\times3, \text{ and } 3\times1)$. Ultimately, features from the three flows are concatenated, redundant features are removed with a dropout layer and adaptive average pooling is performed to represent the final feature vector. This vector is then classified with a SoftMax layer to determine whether the given nodule is benign or malignant (see Fig. 5).



Fig. 5. Local attention layer with spatial extent k = 3 [59].

C. Dataset Descriptions

The dataset descriptions section thoroughly depicts the dataset. In the proposed study, the segmentation and classification processing are carried out on a database

from Lung Image Database Consortium-Image Database Resource Initiative (LIDC-IDRI) that is freely accessible. Computer-aided methods the identification, in classification, and assessment of pulmonary knobs are allowed in the biggest archive of CT images from LIDC-IDRI. It consists of 1,018 thoracic CT scan pictures of cancer cases with properly annotated lesions of 1,010 patients. The data were gathered from various medical imaging companies across the world. Four radiologists from LIDC completed the annotations. A total of 7,371 annotated lesions are present in the database, of which 2,669 have a diameter greater than or equal to 3 mm and 4,702 have a diameter less than 3 mm.

The experiment is also carried out on the nodules directly without the segmentation; the necessary dataset is accessible at (https://github.com/mundher/local-global). The authors have made available a dataset from LIDC-IDRI that contains a total 848 nodules of which 442 benign nodules and 406 malignant nodules, processed in 17 different methods. Furthermore, by splitting the original dataset into training and testing sets, cross-validation is used to assess predictions. The original dataset is split into K-fold cross-validation with k as 10 and with 14,416 enhanced nodules each for the tenfold cross-validation.

IV. RESULTS AND DISCUSSION

A. Performance Matrics

In evaluating the performance of the lung image classification models, various performance metrics are employed to provide a comprehensive understanding of their effectiveness. One of the primary metrics is accuracy, which offers an overall assessment of the classified images. It is expressed as the ratio of accurate predictions to the total number of instances. Precision, another vital metric, focuses on the proportion of truly positive patterns labeled as positive among all correctly or incorrectly classified positive samples. Recall, a complementary metric, gauges the model's ability to correctly classify positive patterns. F1-Score, a harmonic mean between recall and precision [60-62]. The Dice coefficient, an alternative measure, assesses the agreement between predicted and actual positive instances. Sensitivity evaluates the model's ability to predict true positives for each available category. On the other hand, specificity assesses the model's capacity to predict true negatives for each category. Lastly, the Intersection over Union (IoU) is employed in object detection algorithms to differentiate between correct and incorrect detections [63, 64].

Accuracy_m =
$$\frac{T_P + T_N}{T_P + T_N + F_P + F_N}$$
 (4)

$$Precision_{\rm m} = \frac{T_P}{T_P + F_P}$$
(5)

$$\operatorname{Recall}_{\mathrm{m}} = \frac{T_{-}P}{T_{-}P + F_{-}N} \tag{6}$$

$$F1-Score = \frac{2 \times Precision_m \times Recall_m}{Precision_m + Recall_m}$$
(7)

Dice_m =
$$\frac{2 \times T_P}{(T_P + F_P) + (T_P + F_N)}$$
 (8)

$$Sensitivity_{m} = \frac{T_P}{(T_P + F_N)}$$
(9)

$$Specificity_{m} = \frac{T_{N}}{(T_{N} + F_{P})}$$
(10)

$$IoU_{m} = \frac{Area of Overlap}{Area of Union}$$
(11)

B. Performance Analysis on Segmentation

Fig. 6 shows the sample CT slices of the dataset and Fig. 7 shows the segmentation of nodules with the help of U-net with attention gate model.



Fig. 6. Sample CT images.



Fig. 7. Segmented nodule images sample.

Fig. 7 visually represents the identification of the knob presence in both the ground truth and segmented nodule

images. Performance metrics for various algorithms in lung-nodule segmentation are presented in Table I.

Models	Accuracy	Sensitivity	Specificity	Dicem	IoUm
PSPNet [1]	0.931	0.851	0.962	0.868	0.778
PAN [2]	0.927	0.837	0.960	0.858	0.765
2D U-Net [3]	-	0.890	(-)	0.830	_
2D Seg U-Det [4]	-	0.850	(-)	0.820	_
3D FCN [5]	-	-	(-)	0.690	_
3D Nodule R-CNN [6]	-	-	(-)	0.640	_
2D CNN [7]	-	-	(-)	0.610	_
Attention U-Net	0.943	0.905	0.973	0.876	0.783

TABLE I. PERFORMANCE METRIC OF SEGMENTATION ALGORITHMS

Table I illustrates the algorithmic performance in the segmentation of lung nodules. The table distinctly indicates that the Attention U-Net algorithm yields superior accuracy, sensitivity, specificity, dice, and IoU compared to alternative techniques.

The Attention U-Net Segmentation method demonstrates a superior Dice coefficient, surpassing the 2D CNN [7] approach by 0.2659 and outperforming the 3D Nodule R-CNN [6] method by 0.2359. Additionally, it exhibits a 0.1859 improvement over the 3D FCN [5] method and a 0.0559 advantage over the 2D Seg U-Det [4] method. Furthermore, it outshines the 2D U-Net [3] method by 0.0459 in terms of the dice coefficient. In Fig. 8, the Attention U-Net method exhibits an improvement of 0.0182 over Pan [2] methods and is 0.0075 superior to PSPNet [1]. Fig. 8 underscores the high Dice Similarity Coefficient (DSC) achieved by Attention U-Net segmentation in comparison to other segmentation approaches. Table II provides a comprehensive outline of the DAKFSAM method performance compared to existing methods for lung nodule classification using the dataset-1 nodule images with a visual representation presented in Fig. 9.



Fig. 8. Analysis of segmentation algorithms based on dice.

TABLE II. CLASSIFICATION PERFORMANCES OF PROPOSED METHOD VS EXISTING METHOD (USING DATASET 1)

	Methodology	AUC	Accuracy	Precision	Sensitivity/Recall	
	ResNet50 [2]	86.82%	77.62%	0.80	0.71	
	ResNet18 [2]		78.21%	0.79	0.75	
	Densenet121 [1]		84.57%	0.87	0.80	
	Deep residual network [3]	88.36%	80.26%	0.81	0.78	
	DAKFSAM (Proposed Method)		93.26% 85.14%		0.8488	
)	Performance Comparisor	n Across Differer	nt Methodologies			
					Methodology ResNet50 [2] ResNet18 [2] DenseNet121 [1] Deep Residual Netwo	
					DAKFSAM (Proposed 1	
-						
0.0	Precision Recall	Acc	uracy	AUC		

Fig. 9. Analyses of precision, recall, accuracy and F1-Score using Dataset-1.

The DAKFSAM method demonstrates a precision improvement of 0.0463 over the Deep residual network [3] method, surpassing ResNet18 [2] by 0.07 and Resnet50 [2] method by 0.05. Fig. 9 visually highlights the heightened precision achieved by the DAKFSAM method compared to alternative approaches. In terms of recall, the DAKFSAM method outperforms Deep residual network [3] by 0.0688, DenseNet121 [1] method by 0.05, ResNet18 [2] by 0.10, and ResNet50 [2] by 0.14, all percentage values. Furthermore, the DAKFSAM method attains a higher F1-Score exceeding the Deep residual network [3] method by 0.05782%. Similarly, it outperforms DenseNet121 [1] by 0.019%, ResNet18 [2] by 0.08305%, and ResNet50 [2] method by 0.10022%. Table III presents a detailed outline of the classification performance achieved by employing the attention gate U-Net segmentation. The analysis is specifically focused on dataset-2 CT images.

The DAKFSAM method achieves 1.98% higher accuracy compared to the Deep Residual Network [3] method. It also outperforms the DenseNet121 [1] method by 8% and the ResNet18 [2] method by 11.67%. Additionally, it is 14% more accurate than the ResNet50 [2] method.

According to Fig. 10, the DAKFSAM method achieves higher accuracy compared to other methods. Specifically, Fig. 11 offers a 2.71% higher F1-Score than the Deep Residual Network [3] method. The proposed method also outperforms the DenseNet121 [1], ResNet18 [2] and ResNet50 [2] methods by 8.29%, 12.49%, and 14.29%, respectively.

Methods

ResNet18 [2] DenseNet121 [1] Deep Residual Network [3] DAKFSAM (Proposed M

TABLE III. CLASSIFICATION PERFORMANCES OF PROPOSED METHOD VS EXISTING METHOD [USING DATASET 2]

Methods	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
ResNet50 [2]	80.23	81.26	81.26	80.23	80.56
ResNet18 [2]	82.56	73.69	83.56	84.56	82.36
DenseNet121 [1]	86.23	85.78	86.23	86.58	86.56
Deep residual network [3]	92.25	92.66	91.25	92.47	92.14
DAKFSAM (Proposed Method)	94.23	94.57	93.56	94.78	94.85



Fig. 10. Analysis of accuracy.



Fig. 11. Analysis of precision, recall, and F1-Score (using dataset 2).

V. CONCLUSION

This paper introduces an effective deep residual network incorporating various forms of kernel fusion for the classification of lung nodules. To segment the nodules from the CT-slices, the U-Net with an attention gate model is employed in this work. The segmented nodules undergo further classification using the proposed DAKFSAM. The utilization of alternate kernel fused model and self-attention concept enhances the classification performance. The three-level process of deep fused model extracts all discriminative features to accurately represent whether a nodule is noncancerous or cancerous. Experimental results on the LIDC-IDRI dataset demonstrate the efficiency of the DAKFSAM design. The proposed DAKFSAM method achieves a 2% increase in accuracy compared to the Deep Residual network, and the F1-Score of the D3DR MKCA method surpasses the Deep Residual Network model by 2.71% higher than the Deep residual network model. This study underscores the superior performance of the suggested method in lung nodule classification, offering potential benefits for clinical diagnosis.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Rani Saritha R. conceptualized and designed the study, performed the data analysis, and drafted the manuscript. She also coordinated the research project and contributed to the interpretation of the results. V. Sangeetha assisted with the design of the methodology, conducted the experimental validations, and substantially revised the manuscript; all authors had approved the final version.

REFERENCES

- D. D. S. Goncalves, M. N. Yukuyama, M. Y. S. Miyagi, T. J. V. Silva, C. Lameu, N. A. B. Chacra, and G. L. B. D. Araujo, "Revisiting flubendazole through nanocrystal technology: Statistical design, characterization and its potential inhibitory effect on xenografted lung tumor progression in mice," *Journal of Cluster Science*, vol. 34, pp. 261–272, 2023.
- [2] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, pp. 209–249, 2021.
- [3] B. U. Dhaware and A. C. Pise, "Cancer detection using bayasein classifier and FCM segmentation," in *Proc. IEEE, International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, 2016, pp. 170–174.
- [4] Cancer Facts and Figures 2022. [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2022.html#:~:text=Estimated% 20numbers%20of%20new%20cancer,factors%2C%20early%20d etection%2C%20and%20treatment
- [5] R. Manickavasagam and S. Selvan, "Automatic detection and classification of lung nodules in CT image using optimized neuro fuzzy classifier with cuckoo search algorithm," *Journal of Medical Systems*, vol. 43, no. 9, 2019.
- [6] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep

convolutional neural networks," *Pattern Recognition*, vol. 85, pp. 109–119, 2019.

- [7] S. Rakesh and M. Shanthi, "Nodule segmentation of lung CT image for medical applications," in Proc. Global Transitions Proceedings, 2021, vol. 2, no. 1.
- [8] D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, "Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models," *Comput. Biol. Med.*, vol. 42, pp. 1098–1109, 2012.
- [9] M. Usman et al. (2020). Volumetric lung nodule segmentation using adaptive ROI with multi-view residual learning. [Online]. Available: https://www.nature.com/articles/s41598-020-69817-y
- [10] J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach," *IEEE Trans. Med. Imaging*, vol. 27, pp. 467–480, 2020.
- [11] L. Gong, S. Jiang, Z. Yang, G. Zhang, and L. Wang, "Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 11, pp. 1969–1979, 2019.
- [12] M. Savic, Y. Ma, G. Ramponi, W. Du, and Y. Peng, "Lung nodule segmentation with a region-based fast marching method," *Sensors*, vol. 21, 1908.
- [13] N. Keetha, P. Babu, A. Samson, and C. Annavarapu, "U-Det: A Modified U-Net architecture with bidirectional feature network for lung nodule segmentation," arXiv preprint, arXiv:2003.09293, 2020.
- [14] K. Kaviarasu and V. Sakthivel, "K-Means clustering using fuzzy C-Means based image segmentation for lung cancer," *South Asian Journal of Engineering and Technology*, vol. 2, no. 17, pp. 10–17, 2022.
- [15] W. A. H. Mousa and M. A. Khan, "Lung nodule classification utilizing support vector machines," in *Proc. International Conference on Image Processing*, 2022, vol. 9.
- [16] B. N. Narayanan *et al.*, "Optimized feature selection-based clustering approach for computer-aided detection of lung nodules in different modalities," *Pattern Anal. Applic.*, vol. 22, pp. 559– 571, 2019.
- [17] M. H. Orozco *et al.*, "Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine," *Biomed. Eng. Online*, vol. 14, no. 9, 2015.
- [18] A. M. Santos, A. O. C. Filho, A. C. P. Silva, R. A. Nunes, and M. Gattass, "Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 27–39, 2014.
- [19] T. Zhou, H. Lu, J. Zhang, and H. Shi, "Pulmonary nodule detection model based on SVM and CT image feature-level fusion with rough sets," *Biomed. Res. Int.*, 8052436, 2016.
- [20] A. Asuntha and A. Srinivasan, "Deep learning for lung cancer detection and classification. multimedia tools and applications," *Multimedia Tools and Applications*, vol. 79, pp. 7731–7762, 2020.
- [21] W. Zhu et al., "Deep lung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," in Proc. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 673–681.
- [22] J. George, S. Skaria, and V. V. Varun, "Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans," *Medical Imaging 2018: Computer-Aided Diagnosis*, pp. 347–355, 2018.
- [23] D. Baldwin *et al.*, "External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules," *Thorax*, vol. 75. 2020.
- [24] J. Ding *et al.*, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, 2019.
- [25] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, "An automatic detection system of lung nodule based on multigroup patch-based deep learning network," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, 2018.
- [26] G. Zhang, Z. Yang, L. Gong, S. Jiang, L. Wang, and H. Zhang, "Classification of lung nodules based on CT images using squeeze-and-excitation network and aggregated residual transformations," *Radiol. Med. Apr.*, vol. 125, no. 4, 2020.
- [27] O. Ozdemir, B. Woodward, and A. A. Berlin, "Propagating uncertainty in multi-stage Bayesian convolutional neural

networks with application to pulmonary nodule detection," arXiv preprint, arXiv:1712.00497, 2017.

- [28] M. Schultheiss *et al.*, "A robust convolutional neural network for lung nodule detection in the presence of foreign bodies," *Sci. Rep.*, vol. 10, 2020.
- [29] S. W. Zhou *et al.*, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognition*, vol. 61, pp. 663–673, 2017.
- [30] R. Wolz *et al.*, "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE TMI*, vol. 32, no. 9, 2013.
- [31] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conference* on Computer Vision and Pattern Recognition, 2015, pp. 3431– 3440.
- [32] K. Kamnitsas et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.
- [33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [34] C. Y. Lee et al., "Deeply-supervised nets," Artificial Intelligence and Statistics, pp. 562–570, 2015.
- [35] S. Xie and Z. Tu, "Holistically-nested edge detection," in Proc. IEEE International Conference on Computer Vision, 2015, pp. 1395–1403.
- [36] H. R. Roth *et al.*, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Medical Image Analysis*, vol. 45, pp. 94–107, 2018.
- [37] H. R. Roth *et al.*, "Hierarchical 3D fully convolutional networks for multi-organ segmentation," arXiv preprint, arXiv:1704.06382, 2017.
- [38] K. Kamnitsas *et al.*, "Ensembles of multiple models and architectures for robust brain tumour segmentation," *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 450–462, 2018.
- [39] W. Bai et al., "Human-level CMR image analysis with deep fully convolutional networks," arXiv preprint, arXiv:1710.09289, 2017.
- [40] T. Shen *et al.*, "Disan: Directional self-attention network for rnn/cnn-free language understanding," arXiv preprint, arXiv:1709.04696, 2017.
- [41] F. Wang et al., "Residual attention network for image classification," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3156–3164.
- [42] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint, arXiv:1409.0473, 2017.
- [43] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," arXiv preprint, arXiv:1508.04025, 2015.
- [44] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.
- [45] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," *Advances in Neural Information Processing Systems*, pp. 577–585, 2005.
- [46] Y. Wu *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv preprint, arXiv:1609.08144, 2016.
- [47] J. K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. International Conference on Machine Learning*, 2015, pp. 2048–2057.

- [48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, pp. 5998– 6008, 2017.
- [49] C. Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, C. Hawthorne, A. M. Dai, M. D. Hoffman, and D. Eck, "Music transformer," *Advances in Neural Processing Systems*, vol. 4, 2018.
- [50] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, 8, 2019.
- [51] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pretraining of deep bidirectional transformers for language understanding," arXiv preprint, arXiv:1810.04805, 2018.
- [52] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, and D. Tran, "Image transformer," in *Proc. 35th International Conference on Machine Learning*, 2018, vol. 80.
- [53] N. Shazeer, Y. Cheng, N. Parmar, D. Tran, A. Vaswani, P. Koanantakool, P. Hawkins, H. Lee, M. Hong, C. Young, R. Sepassi, and B. A. Hechtman, "Mesh-tensorflow: Deep learning for supercomputers," *Advances in Neural Information Processing Systems*, vol. 31, 2018.
- [54] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," arXiv preprint, arXiv:1803.02155, 2018.
- [55] A. S. Forruque, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. B. M. S. Ali, and A. H. Gandomi, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13521–13617, 2023.
- [56] I. Bello, B. Zoph, A. Vaswani, J. Shlens, and Q. V. Le, "Attention augmented convolutional networks," in *Proc. IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3286–3295.
- [57] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, "Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers," arXiv preprint, arXiv:1801.05173, 2018.
- [58] Z. R. Shen, M. Y. Zhang, H. Y. Zhao, S. Yi, and H. S. Li. "Efficient attention: Attention with linear complexities," in *Proc. IEEE/CVF Winter Conference on Applications of computer Vision*, 2021, pp. 3531–3539.
- [59] X. Yang, "An overview of the attention mechanisms in computer vision," *Journal of Physics: Conference Series*, vol. 1693, no. 1, 012173, 2020.
- [60] F. Lia *et al.*, "Evaluate the malignancy of pulmonary nodules using the 3D deep leaky noisy-or network," arXiv preprint arXiv:1711.08324, 2017.
- [61] R. L. Siegel and K. D. Miller, "Cancer statistics," *Cancer J. Clin.*, pp. 1–43, 2023.
- [62] T. Wang, "The lung image database consortium image collection (LIDC-IDRI)," *The Cancer Imaging Archive*, vol, 32, no. 2, 2020.
- [63] N. Florez *et al.*, "Lung cancer in women: The past, present, and future, clinical lung cancer," *Clinical Lung Cancer*, vol. 25, no. 1, January 2024.
- [64] J. J. Lin and J. F. Gainor, "Current opportunities and challenges in ALK-positive lung cancer," *Translational Lung Cancer Research*, vol. 13, no. 1, 2024.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License (<u>CC BY-NC-ND 4.0</u>), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.