IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 January 2023, accepted 3 February 2023, date of publication 16 February 2023, date of current version 23 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3243104

RESEARCH ARTICLE

Multi-Orientation Local Texture Features for Guided Attention-Based Fusion in Lung Nodule Classification

AHMED SAIHOOD[1,2], HOSSEIN KARSHENAS[1], AND AHMAD REZA NAGHSH-NILCHI[1]

[1]Artificial Intelligence Department, Faculty of Computer Engineering, University of Isfahan, Isfahan 81746-73441, Iran
[2]Faculty of Computer Science and Mathematics, University of Thi-Qar, Nasiriyah, Thi-Qar 64001, Iraq

Corresponding author: Hossein Karshenas (h.karshenas@eng.ui.ac.ir)

ABSTRACT Computerized tomography (CT) scan images are widely used in automatic lung cancer detection and classification. The lung nodules' texture distribution throughout the CT scan volume can vary significantly, and accurate identification and consideration of discriminative information in this volume can greatly help the classification process. Deep stacks of recurrent and convolutional operations cannot entirely represent such variations, especially in the size and location of the nodules. To model this complex pattern of inter/intra dependencies in the CT slices of each nodule, a multi-orientation-based guided-attention module (MOGAM) is proposed in this paper, which provides high flexibility in concentrating on the relevant information extracted from different regions of the nodule in a non-local manner. Moreover, to provide the model with finer-grained discriminative information from the nodule volume, specifically-designed local texture feature descriptors (TFDs) are extracted from the nodule slices in multiple orientations. These TFDs not only represent the distribution of textural information across multiple slices of a nodule but also encode and approximate this distribution within each slice. The extended experimentation has shown the effectiveness of the non-local combination of these local TFDs through the proposed guided attention mechanism. According to the classification results obtained on the standard LIDC-IDRI dataset, the proposed approach has outperformed other counterparts in terms of accuracy and AUC evaluation metrics. Also, a detailed explainability analysis of the results is provided, demonstrating the correct functioning of the proposed attention-based fusion approach, which is required by medical experts.

INDEX TERMS Lung cancer classification, non-local guided attention, co-occurrence pattern, texture feature descriptor, long-range dependency.

I. INTRODUCTION

Lung nodule CT scan is formed from many slices, with sequential and long-range dependencies among the regions captured in these slices. Considering the orientation in volumetric data is essential in interpreting the 3D data structure geometry. The cross-sectional heterogeneity or relevance in the 3D volume of the lung nodule requires considering the importance of each slice relative to the other slices of the nodule. Thus, extracting high-level features, particularly the texture features, is subject to local and global depth analysis.

In recent years the deep learning approaches, the convolutional neural networks (CNN)s in particular [1], [2], [3], [4], [5], are applied for lung nodule classification with promising results. However, the repeated application of the local processes within the CNN layers on the texture features are ineffective for representing the long-range dependencies within a lung nodule and their complicated construction. Many deep learning-based computer-aided diagnosis (CAD) systems have sought to mitigate this issue. Different biomarker features in lung nodules are applied to give importance to each slice within the nodule. Applying biomarker features is not able to capture the long-range dependencies in different nodule orientations. The lung nodule does not have a uniform distribution in size and location throughout the CT slices [25]

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

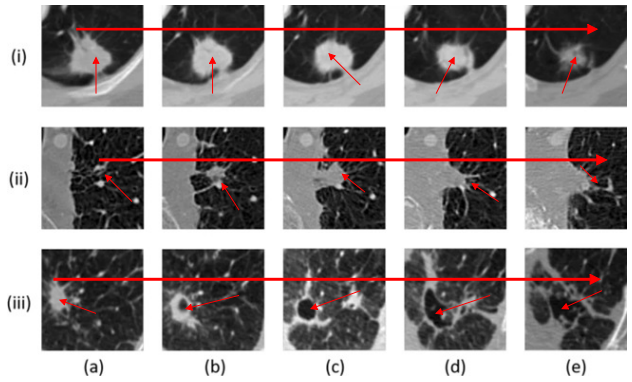


FIGURE 1. The lung nodule size and location variation through CT slices: The rows (i),(ii) and (iii) show different samples of lung nodule's slices and columns (a) to (e) show the nodule's size and location change through the slices (indicated by the red arrows).

(see Fig. 1), and single path CNNs cannot handle lung nodule size variability. Dual-path CNNs-based fusion methods have been applied to alleviate this shortcoming where each path uses different kernel sizes to extract multi-scale patterns [6], [7], [8]. However, they cannot capture long-range dependencies within nodule slices and are unable to provide a global understanding of the nodule.

Analysis of multi-modal medical images (such as PET and CT) is applied to provide complementary attention for both spatial and contextual information [9], [10]. Also, attention-based multiple instance learning is built based on pooling layers and long short-term memories [11]. These models are sensitive to the nature of features extracted from both modalities; hence, the static fusion process is not applicable, and they need to be fused commensurately. Employing deeper convolutional networks leads to ineffective information extraction, intricate training process, and disappearance of the gradient [37].

Recurrent neural networks (RNN), on the other hand, can capture the importance of inter-slice dependencies in CT scans [13], [14], [15]. Similar to convolutional operations, recurrent processes are applied progressively to sequentially ordered elements without encoding their position and orientation [16]; consequently, when repeated over relevant nodule slices, they cannot properly consider the unevenly distributed nodule sections through the slices.

Extending the scope of operations from local neighborhoods, non-local operations are flexible building blocks that enclose the convolutional/recurrent operations to generate an attention map aggregated from a specific orientation of the values in the lung nodule. It can be added to the initial or middle layers of the neural networks as a generic component allowing to handle both global and local information, in contrast to the fully connected layers, which are often used at the end. Researchers have proposed non-local operations to capture long-range relationships with deep neural networks [17]. The first non-local attention procedures applied to sequential three-dimensional data were computationally very expensive [18]. Ho et al. [19] sought to reduce the non-local operations' computational cost, but their method disregarded attention to the depth elements in 3D data. As a solution, 3D axial attention is proposed [20] where attention to the depth elements is considered, though the variation in nodule size and location within slices is disregarded. That is, the attention mechanism concentrates on some inconsequential features after screening out the extracted features, leading to performance degradation of the model. In other words, elements out of the nodule contours will be considered for attention, causing the model to attend to out of the nodule low-level features. On the other hand, as shown in Fig. 1, considering only column, row, and depth-wise attentions may oversight the spatial information located outside the contours of the stacked slices.

In medical image classification, focusing on the image's texture features is highly essential. Image textures lead to enhanced tissue analysis [22], [23], [24]. The grey-level co-occurrence matrix (GLCM) is one of the most prevalent lung nodule structure analysis methods [22]. It is a second-order statistical sequence texture feature extraction method that can describe the lung nodule structure. GLCM is expressed with the comparative frequency $\nu(i, j|\Delta x, \Delta y)$ for the pair of pixels appearing within a given neighborhood of $(\Delta x, \Delta y)$ distance, one with intensity i and the other with intensity j. The matrix $C_{i,j|\Delta x,\Delta y,\theta}$ includes the second-order statistical likelihood values to differentiate i and j grey levels at a certain distance and a particular angle θ. Whole-image GLCMs do not always adequately represent volumetric lung nodules. The random distribution of the global co-occurrences extracted from the images leads to misidentified high-level regions of interest through nodule slices. Proper texture analysis of the lung nodules should contribute to assessing the tissue heterogeneity.

To overcome these drawbacks, in this study, texture feature descriptors (TFDs) are first computed from the locally extracted co-occurrence patterns, preserving their spatial information distributions. Then, a multi-orientation TFD-based guided-attention module (MOGAM) is proposed for fusion in deep neural networks. The proposed fusion structure has the ability to find semantic relationships between each of the orientations' features and the nodule embedding space in a non-local manner, resulting in a high-level 3D representation of the nodule, which is used for effective lung cancer classification. The local co-occurrences are calculated considering binary patterns over the co-occurring quantized grey levels, and different texture-related features are extracted locally through windows imposed on these patterns in crosswise and lengthwise orientations. In crosswise orientation, we consider the nodule's slices to apply the co-occurrence patterns, while in the lengthwise orientation, the co-occurrence patterns are applied to the longitudinal nodule cuts. These features are then passed through non-local operations to devise attention maps for the input nodule. This would allow each element in one slice within a nodule to literally have spatial attention to all its related column, row, depth, and diagonal elements, as demonstrated in Fig.2. Thus

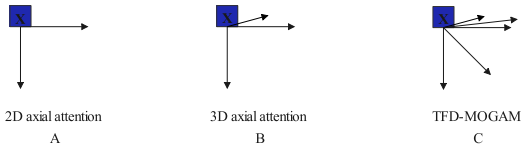


FIGURE 2. Element *X* attention capability: (A) An element *X* attending only to its column and row elements when conducting the 2D axial attention, where, as observed, non-consideration of volume depth in lung nodules leads to spatial information loss between slices, (B) Element *X* attends to its column, row, and depth when conducting the 3D axial attention, and (C) Applying the TFD-MOGAM, allows element *X* to attend to axial and non-axial directions, and thus have a higher flexibility in considering only nodule regions of the volume.

the proposed approach can determine the lung nodules' local and global texture weights in the intermediate layers of the network without local filters' intervention.

The main issues of concern in this article consist of: 1) lung nodules are usually arbitrary in their size (the nodule diameter can vary from 3 to 30 mm among the nodules [21]), requiring to learn how the extracted feature representations adapt to various spatial scales. CNN models are sensitive to nodules' scale variations due to their feature map size limitation. In this article the semantic relationships between each orientation and nodule embedding space is found to consider multi-scale nodule variations; 2) loss of high-level features extracted from global GLCMs in specific locations. In the proposed approach, the TFDs are estimated locally within windows imposed on co-occurrence masks, for each pair of grey level intensities (i, j); 3) vital information loss regarding the global 3D shape of the nodule, i.e. variation of lung nodule size and location through CT slices (Fig. 1), as the result of resorting to progressive local convolutional operations. The focus of the proposed method is on capturing long-range dependencies by computing non-local attention maps from the TFDs computed in crosswise and lengthwise orientations, encoding the important regions of the volumetric nodule; 4) redundancy and high correlation of the queries and keys in computing the non-local attention maps may lead to overfitting. The queries in the proposed MOGAM are obtained from two different set of TFDs (in different orientations), allowing the module to freely query all axial and even diagonal elements for each input element; 5) although recurrent neural networks could consider the long-range dependencies of the lung nodules for classification [13], [14], [15], repeated application of recurrent operations over relevant nodule's slices lead to loss of interpretability of the lesions' appearance, considered very essential for the radiologists. As demonstrated in this paper, the shape and location of the nodule's regions are properly projected through the computed non-local attention maps of the proposed method.

The main contributions of this paper are:

1) A new method for extracting TFDs locally, using the co-occurrence patterns applied on the nodule's volume in crosswise and lengthwise orientations to handle inter/intra dependencies of the CT scan slices.
2) A new MOGAM for effective information fusion guided by multi-orientations TFDs.
3) A deep neural network architecture based on the proposed MOGAM that can simultaneously detect and classify lung nodules with high accuracy (no nodule's region pre-detection is required), outperforming its latest counterparts in the LIDC-IDRI dataset realm.

The rest of this article is structured as follows: in section II, we introduce related works on classification of lung nodules based on attention mechanisms, section III describes the procedure for TFD extraction and the proposed MOGAM approach. In section IV, we present the results obtained through different experiments, and their discussion is provided in section IV-F. Finally, section V concludes the paper and provides an outlook of future work.

II. RELATED STUDIES

The use of non-local operations to capture long-range relevant data with deep neural networks, especially for medical image processing, is gaining momentum. Wang et al. [18] proposed a non-local procedure that computes the return of non-local operations applied to each element in the data as a weighted sum of the features to all other elements in the input feature maps. Concatenating all the elements in one vector to generate a single attention map is of concern. This attention map is very expensive in terms of computation. Ho et al. [19] sought to reduce the non-local operations' computational cost by considering two attention maps without concatenating the elements: one allows each element to attend to its column elements and the other to its row elements. Al-Shabi et al. [20] proposed the non-local operations for 3D axial attention by devising three attention maps to allow each element to attend to its height, width, and depth. Bera and Biswas [26] proposed a non-local procedure for denoising the CT images by applying the self-similarity attention in the neighborhood to compute the classical non-local means approaches. Rundo et al. [27] combined the 3D densely connected convolutional layers based on spatio-temporal non-local attention methods. Al-Shabi et al. [28] proposed a progressive growing channel attentive non-local network for lung nodules classification by adding a channel attention mechanism (ProCAN) to the non-local network proposed in [20] to enhance the attentive ability. Rundo et al. [29] applied a self-augmentation method to yield 3D X-ray images from augmented images through reinforcement learning. The non-local operations are used to process the volumetric images.

Recently, non-local attention modules (NLAM) have been adopted for different applications like capturing spatial and temporal dependencies among video frames [31], foreground objects detection enhancement in the YOLOv4 structure [32], single-image rain streak removal [33] and exploiting the global context information for stereo matching [34]. However, redundancy and high correlation of the queries and keys in computing the non-local attention maps may lead to overfitting. Li et al. [35] used the binary nodules mask to

guide the model to automatically consider both nodule and the whole-lung information.

Non-local operations are also used to enhance information representation in the segmentation process. Wang et al. [30] proposed non-local U-Nets equipped with flexible global accumulation blocks for medical image segmentation. These blocks can be installed into the U-Net as size-preserving approaches. Qu et al. [6] proposed a model containing multi-scale and multi-view details for multi-phase pancreas segmentation. They devised two non-local attention processes to enrich the high-level feature presentation: 1) a location attention process that yields cross-phase dedicated feature correlations to overcome the misalignment regions, and 2) the depth-wise attention process that runs to determine the channel reliances.

The available attention mechanisms are based on spatial or channel attention subject to multiple parallel paths in the networks,also attention-based pooling [57] has been emphasized recently. Wen et al. [36] proposed two parallel non-local operations based-attention module named NVCF: 1) the one that provides the global representation of the lung nodules based on ResNet18 and ShuffleNetV2 and 2) the one that extracts non-local features by running non-local operations [17]. Xia et al. [37] used a residual attention network based on a squeeze-and-excitation network to extract spatial and contextual features. They built a multi-scale attention network that pays attention to high-level features. Zhang and Yang [38] proposed the shuffle attention model, where shuffling operations are run to combine channel and spatial features. They grouped the channels to process each feature group in parallel, then computed the channel attention by applying global average pooling and spatial attention within each group. Both of the feature maps are concatenated with the exact channel count. Ranjbarzadeh et al. [39] proposed cascaded CNNs where both the local and global features are separated into two paths: the first path detects the global part by extracting pixels located in the border of the volumetric nodule, and the second label the local feature for each slice. They proposed distancewise attention to consider the nodules' variation in both location and size within the volumes. Xu et al. [40] proposed a hybrid attention procedure that includes two parallel paths for extracting spatial attention and channel attention maps. They applied the channel coefficients in different feature maps to determine the similarity between channels and spatial coefficients with other feature maps to make the model attend to the similarity within feature maps. Shan and Yan [41] presented a spatial and channel attention network consisting of two attention blocks to handle spatial and channel-wise associations. The information extracted from the blocks is integrated via a decoder. Chen et al. [9] built their attention module by fusing features extracted from dual-path neural networks considering (PET/CT) multimodality.

The available models can be divided into two main categories: 1) non-local attention that seeks to give immediate importance to the crucial parts of the data rather than ministering all data equally, and 2) attention-based non-local operations where attention is applied merely to the high-level spatial features extracted from nodule slices or global features extracted from the channels of the extracted feature maps, as presented in Table 1.

TABLE 1. Overview of corresponding works on spatial and channel attention.

Researcher	Spatial attention	Channel attention	Mechanism
Bera et al. [22]	✓	×	Non-local attention
Rundo et al. [23]	✓	✓	Non-local attention
Al-Shabi et al. [24]	✓	✓	Non-local attention
Wen et al. [28]	✓	×	Dual paths network
Zhang et al. [29]	✓	✓	Shuffling operations
Ranjbarzadeh et al. [30]	✓	×	Cascaded convolutional neural network
Xu et al. [36]	✓	✓	Hybrid attention mechanism
Wang et al. [9]	✓	×	Non-local attention
Al-Shabi et al. [13]	✓	×	Non-local attention
Li et al. [35]	✓	×	guided-attention
This study	✓	✓	MOGAM attention

III. METHOD

This section explains the proposed procedure for texture feature descriptors computations, and then the design of MOGAM is elaborated in the following subsection.

A. TEXTURE FEATURE DESCRIPTORS

To calculate local texture features from a 3D nodule $X \in R^{h\times w\times d}$, the intensities in each nodule's slice are quantized to q levels, followed by revolving the slice in all possible directions (Fig.3 – left column), to detect the co-occurrence of patterns (i, j) by comparing the slice and its revolved version in a specific direction. As a result, for each of the directions, $q \times q$ binary co-occurrence patterns (BCOPs) are extracted, showing the location of a specific directed co-occurrence pattern in the slice.

A $k \times k$ count filter is applied to the extracted BCOPs to obtain the frequency of each directed co-occurrence pattern in a local neighborhood, where k is the size of the 2D filter. The resulting local co-occurrence counts (LCOCs) provide fine-grained second order statistics from the locally attainable regions of each slice, and can be used to calculate different discriminative texture descriptors - the TFDs. To keep the complexity of feature extraction low, the following descriptor functions [42] are considered for computation in this study using LCOCs:

$$mean = \frac{1}{q^2}\sum_{i=1}^{q}\sum_{j=1}^{q} i\, CO_{i,j} \tag{1}$$

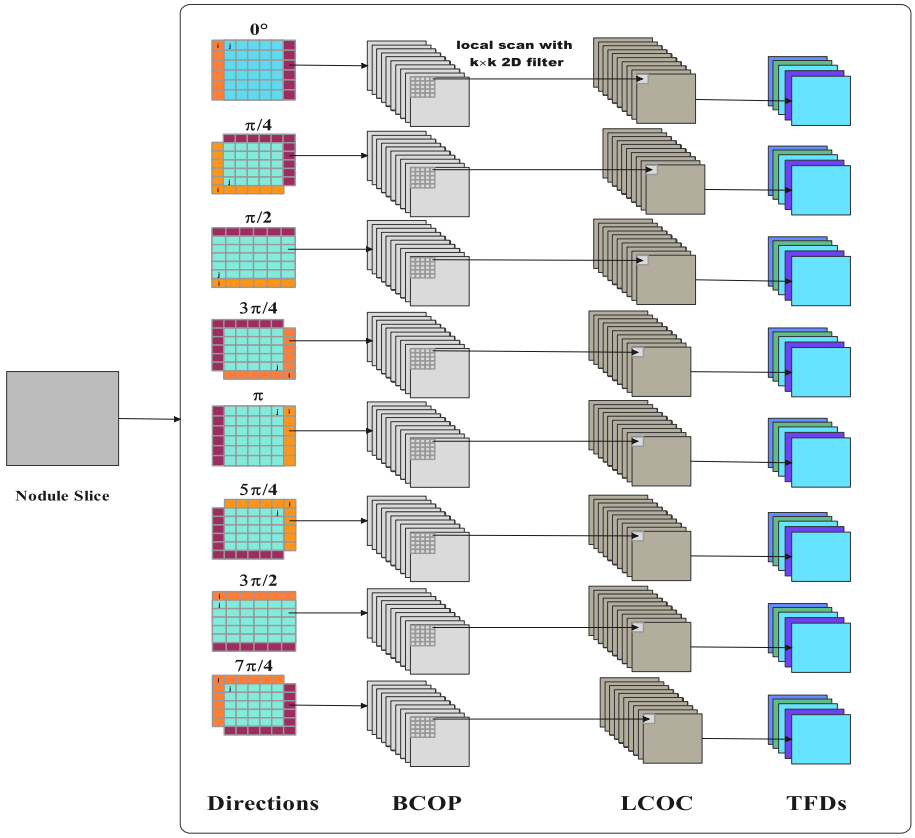


FIGURE 3. The main steps in the TFDs extraction procedure for one slice input.

$$\sigma = \sqrt{\sum_{i=1}^{q}\sum_{j=1}^{q}(i - mean)^2 \, CO_{i,j}} \tag{2}$$

$$contrast = \sum_{i=1}^{q}\sum_{j=1}^{q}(i-j)^2 \, CO_{i,j} \tag{3}$$

$$dissimilarity = \sum_{i=1}^{q}\sum_{j=1}^{q}|i-j| \; CO_{i,j} \tag{4}$$

$$homogeneity = \sum_{i=1}^{q}\sum_{j=1}^{q}\frac{CO_{i,j}}{1+(i-j)^2} \tag{5}$$

$$ASM = \sum_{i=1}^{q}\sum_{j=1}^{q}(CO_{i,j})^2 \tag{6}$$

$$energy = \sqrt{ASM} \tag{7}$$

$$max. = max(CO_{i,j}) \tag{8}$$

$$entropy = \sum_{i=1}^{q}\sum_{j=1}^{q} CO_{i,j} \; \log_p CO_{i,j} \tag{9}$$

where $CO_{i,j}$ is the local frequency of the co-occurrence pattern (i, j). Figs.4 gives a visual demonstration of these TFDs for a sample lung CT slice. As demonstrated in this figure, a feature map of size equal to the input slice is obtained for each descriptor function, with σ showing the standard deviation, and ASM standing for the angular second moment. An important advantage of the proposed TFD extraction procedure is the simplicity of its incorporation into deep neural networks as layer-wise operations (shown as columns in Fig.3). This paves the way for an end-to-end training of the whole classification model.

The same procedure is applied to each of the d slices in the nodule's volume, yielding $d \times 8 \times p$ (represented by $\hat{d}$ for brevity) TFD maps, each of size $h \times w$, considering all 8 possible directions in 2D and representing the number of descriptor functions by p. This allows the relationships within each slice (i.e. intra-slice relationships) to be encoded by TFDs, and thus considers the TFDs extracted from different slices to be independent. To account for the dependencies existing between different slices of a nodule (i.e. inter-slice dependencies), a separate set of TFDs are extracted from a lengthwise slicing of the nodule's volume (depicted in Fig.6), using the same TFD extraction procedure. Fig.5 shows the extracted TFDs in this orientation for a sample lung CT slice. It is important to note that in the lengthwise orientation of the nodule, the input to the TFD extraction procedure is formed from concatenating the spatially analogous regions of the nodule's slices. For this orientation, $w \times 8 \times p$ (represented by $\hat{w}$ for brevity) maps of size $h \times d$ are obtained, called TFD_2 to distinguish it from the earlier crosswise TFDs which are called TFD_1 in this paper.

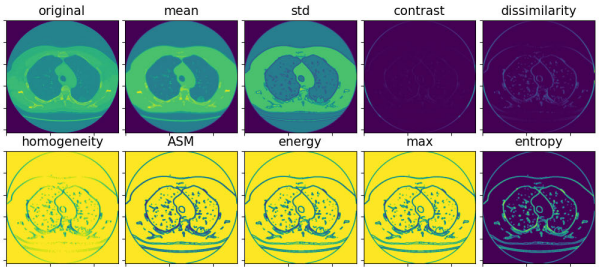


FIGURE 4. The nine statistical texture descriptors computed for one slice in crosswise orientation.

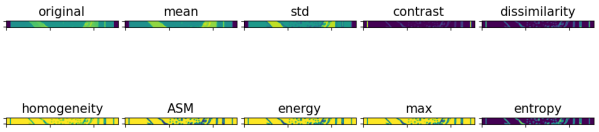


FIGURE 5. The nine statistical texture descriptors computed for one slice in lengthwise orientation.

Such a multi-orientation processing of the nodule provides complementary information, extracted simultaneously, that can be very useful for decision making if properly fused, as explained in the next subsection.

B. MOGAM

Various high-level representations of the condensed regions are obtained when screening the nodule with local windows in different orientations for extracting TFDs. For example, the homogeneity of the regions can be detected using the *mean* descriptor. Transforming these multi-orientation TFDs to a 3D understanding of the regions distribution in the volume is required for effective nodule classification. This transformation can be obtained by a fusion structure, combining the influential information provided by TFDs based on their impact on the classification result.

CNNs have been widely used for information fusion using their stack of convolutional and pooling layers. While these operations allow for automatic extraction of high-level representations, they are usually limited to local neighborhoods. On the other hand, in sequential processing of TFDs obtained from the nodule's slices using recurrent structures such as long short-term memories [43], [56], only specific orders of attention to the discriminative information within TFDs is possible. Attention mechanism [44] has been proposed to overcome these limitations. In this mechanism adaptive importance weights are computed for different pieces of information being fused using the training data. For an image, the attention weights of each position x_i based on all other x_j positions can be computed as follows [18]:

$$a_i = \frac{1}{C(x_i)} \sum_{\forall j} h(x_i, x_j) f(x_j) \tag{10}$$

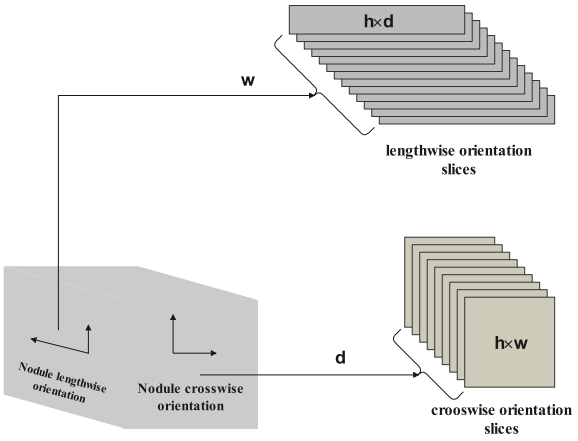


FIGURE 6. Nodule's crosswise and lengthwise orientations.

where a_i shows the computed weight for x_i, f is a transformation function of input data obtained at all x_j positions, h finds the relationship between two elements x_i and x_j, and C is the normalization function. For example, f can be computed by applying convolutions with different kernels on the input data and h can be simply the dot product between the two inputs [17]. An important property of this approach is that the information considered for combination is no longer restricted to being local or sequential, and can be in an arbitrary non-local range, adding high flexibility to the fusion model. This approach is utilized in the method proposed in this paper.

The main input information to the proposed fusion structure are TFD_1 and TFD_2, the two sets of TFDs extracted from the nodule (in different orientations), along with the input features from the nodule (see Fig. 7). A CNN is used in the proposed approach to learn the transformation function for each set of TFDs extracted from the nodule. This allows the model to learn arbitrarily sophisticated texture representations from the TFDs in addition to resizing and aligning them before fusion. For the sake of simplicity, these transformations are denoted using the following two functions:

$$G_1 = \Phi(TFD_1) \tag{11}$$

$$G_2 = \Theta(TFD_2) \tag{12}$$

The input features to the fusion module are also spatially embedded by applying c different convolutional kernels of size $1 \times 1 \times 1$ to the input:

$$F_i = W_i^F \circledast X, \forall i \in \{1, \cdots, c\} \tag{13}$$

where W_i^F denotes the parameters of the ith kernel, $\circledast$ represents the convolution operator and X is the input to the module. The inter/intra slice dependencies encoded in the textural feature maps G_1 and G_2, obtained from the multi-orientation TFDs, are then used to determine the importance of the volumetric regions constituting the nodule's input feature maps, according to Eq. 10:

$$A_1 = SoftMax(G_1 F) \tag{14}$$

$$A_2 = SoftMax(G_2F) \quad (15)$$
$$Z_1 = A_1F \quad (16)$$
$$Z_2 = A_2F \quad (17)$$

Four different attention maps are considered for input weighting, related to height (h), width (w), depth (d) and channels (c) dimensions, each providing a different perspective to the importance of regions in the input feature map. Multiple reshaping of the data is considered in the non-local operations for computing the attention maps, A_{1i} and A_{2i}, as well as in computing the resulting weighted feature maps, represented by Z_{1i} and Z_{2i} in Fig.7, where $i \in \{1, \cdots, 4\}$ corresponds to different dimensions. Summation pooling is applied in the last layer of the module to merge these feature maps, and obtain the fused feature map in the output.

The feature maps obtained from TFDs are used as queries in the proposed MOGAM to find highly correlated regions of the nodule's volume, passed as keys to the module. It allows various related, possibly long-range (i.e., non-local), features of the nodule to be queried based on the textual information encoded in TFDs. This gives the proposed fusion structure an advantage over other similar works [14], [17] since TFDs are extracted not only from axially correlated regions of the nodule but also from those with diagonal dependencies. Moreover, Employing two different TFD orientations for obtaining the queries provides additional flexibility to the elements being attended. It can also decrease the chance of attention weights overfitting [18] as the elements of nodule's volume and their dependencies are considered from multiple orientations.

Depending on the number of kernels used for transforming the input feature maps (in spatial embedding) and the TFDs (in Φ and Θ functions), the size of the feature maps obtained in the proposed MOGAM varies. Thus, a stack of these MOGAM-based layers can create arbitrarily complex patterns of information fusion, using a hierarchical view to the 3D encoding of the nodule textural features provided by TFDs. Fig 8 shows a typical deep network architecture involving three MOGAM fusion layers with various filter sizes. In this architecture which is used for the experiments in this study, the feature maps obtained from the last MOGAM layer are passed through a 3D global average pooling layer before being fed to a fully-connected layer with *SoftMax* activation function for classification of the nodule into benign or malignant.

IV. EXPERIMENTS AND RESULTS

A detailed experimental analysis of the proposed approach is presented in this section. First, the data used for experiments is introduced. Then, the experimental environment and implementation details are provided. Next, the results of extensive experiments are presented and discussed. Moreover, the interpretability of proposed model is also examined.

A. DATASET

The LIDC-IDRI dataset [41], consisting of 1018 CT scan images, annotated by four expert radiologists, is used for the experiments in this study. Only the CTs annotated by at least three experts are used. Cases with more than three malignancy scores are averaged and categorized as malignant, while those with less than three are averaged and categorized as benign. The instances with malignancy scores of precisely 3 (i.e., have equal like to be benign or malignant) are scrapped, ending in 1670 nodules (935 benign and 735 malignant). In each CT scan the slices surrounding a nodule are isolated from the rest. Upon observing the maximum nodule depth size being 33 slices, the maximum volume size surrounding each nodule is considered to be $512 \times 512 \times 33$.

B. EXPERIMENTAL DESIGN

The nine descriptor functions mentioned in Eq.1 are considered for experiments (i.e. $p = 9$). The architecture of the CNNs considered for implementing the transformation functions Φ and Θ, applied on TFD_1 and TFD_2, are specified in Tables 2 and 3, respectively. In all convolutional layers of Φ (Table 2) the size of the 3D kernel is set to three. The layers are organized in pairs, where the first layer of each pair provides an input embedding, using a stride of (1, 1, 1), and the second layer reduces the input size by combining the adjacent elements in the specified dimension with a proper stride. The first three pairs operate on the eight TFD directions with the stride set to (1, 1, 2) and the last two pairs are responsible for the nine statistical functions computed for TFDs with a stride of (1, 1, 3). This architecture is almost mirrored for Θ function (Table 3) with an additional reshaping at the end which is required for aligning the outputs of the two functions. The progressive transformation of TFDs computed for different directions and descriptor functions through convolutional layers gradually captures the strong relationships among these TFDs and avoids information dilution.

The transformations demonstrated in Tables 2 and 3 using different layers is duplicated for each kernel. Different number of MOGAMs are considered in the implementation of the deep fusion model and tested in the experiments. The channel count (i.e., number of kernels) is equally set for both Φ and Θ functions in each MOGAM. The best outcome is obtained when three consecutive MOGAMs are used for fusion, where the number of filters is set to 16, 32 and 64 for the first, second and third modules, respectively. The TFD inputs to these layers are pre-calculated (before feeding to the proposed deep neural network) with a window size of five ($k = 5$) for computing LCOCs and the intensities of slices are quantized into eight levels ($q = 8$) when determining the masks for BCOPs.

The experimental results are reported for an implementation based on the TensorFlow version 2.4.0 framework, on a Windows server 64 OS system with a GEFORCE RTX 2080 GPU and 64 GB of RAM. The *Adam* optimizer with the cross-entropy loss function is used for training

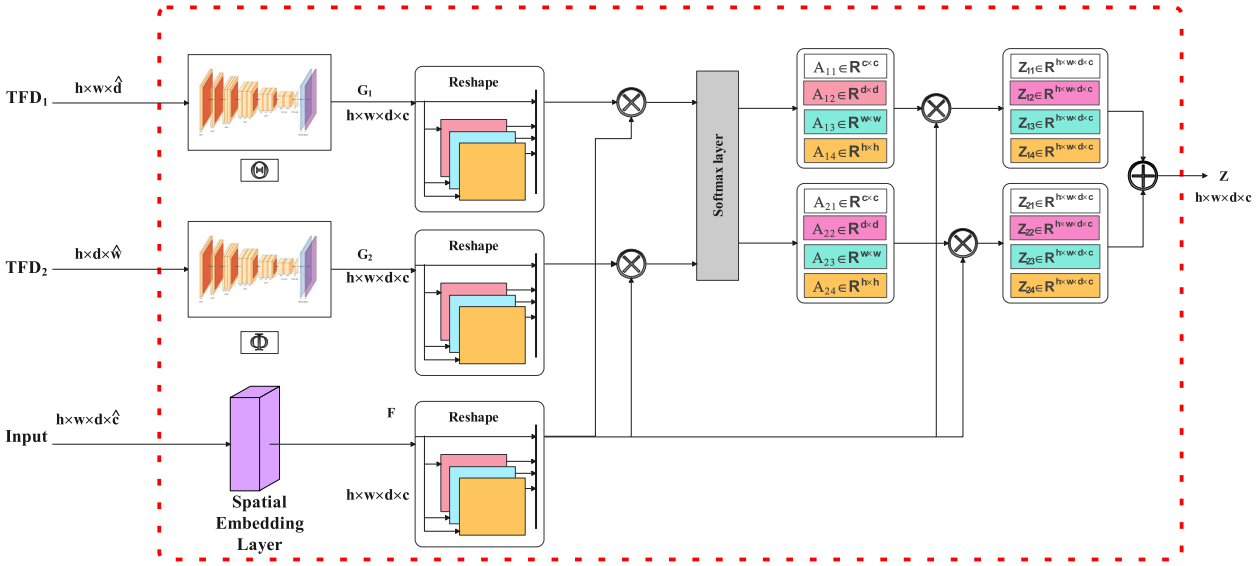


FIGURE 7. The architecture of the proposed MOGAM.

TABLE 2. The operations in Φ applied to obtain the feature map G_1.

Layers	Input shape	Kernel size	Stri -des	Output shape
Conv3	$h \times w \times \hat{d}$	3×3×3	1,1,1	$h \times w \times \hat{d}$
Conv3	$h \times w \times \hat{d}$	3×3×3	1,1,2	$h \times w \times (\hat{d}/2)$
Conv3	$h \times w \times (\hat{d}/2)$	3×3×3	1,1,1	$h \times w \times (\hat{d}/2)$
Conv3	$h \times w \times (\hat{d}/2)$	3×3×3	1,1,2	$h \times w \times (\hat{d}/4)$
Conv3	$h \times w \times (\hat{d}/4)$	3×3×3	1,1,1	$h \times w \times (\hat{d}/4)$
Conv3	$h \times w \times (\hat{d}/4)$	3×3×3	1,1,2	$h \times w \times (\hat{d}/8)$
Conv3	$h \times w \times (\hat{d}/8)$	3×3×3	1,1,1	$h \times w \times \left(\hat{d}/8\right)$
Conv3	$h \times w \times \left(\hat{d}/8\right)$	3×3×3	1,1,3	$h \times w \times \left(\hat{d}/24\right)$
Conv3	$h \times w \times \left(\hat{d}/24\right)$	3×3×3	1,1,1	$h \times w \times \left(\hat{d}/24\right)$
Conv3	$h \times w \times \left(\hat{d}/24\right)$	3×3×3	1,1,3	$h \times w \times d$

TABLE 3. The operations in Θ applied to obtain the feature map G_2.

Layers	Input shape	Kernel size	Stri -ders	Output shape
Conv3	$h \times d \times \hat{w}$	3×3×3	1,1,1	$h \times d \times \hat{w}$
Conv3	$h \times d \times \hat{w}$	3×3×3	1,1,2	$h \times d \times (\hat{w}/2)$
Conv3	$h \times d \times (\hat{w}/2)$	3×3×3	1,1,1	$h \times d \times (\hat{w}/2)$
Conv3	$h \times d \times (\hat{w}/2)$	3×3×3	1,1,2	$h \times d \times (\hat{w}/4)$
Conv3	$h \times d \times (\hat{w}/2)$	3×3×3	1,1,1	$h \times d \times (\hat{w}/4)$
Conv3	$h \times d \times (\hat{w}/4)$	3×3×3	1,1,2	$h \times d \times (\hat{w}/8)$
Conv3	$h \times d \times (\hat{w}/8)$	3×3×3	1,1,1	$h \times d \times (\hat{w}/8)$
Conv3	$h \times d \times (\hat{w}/8)$	3×3×3	1,1,3	$h \times d \times (\hat{w}/24)$
Conv3	$h \times d \times (\hat{w}/24)$	3×3×3	1,1,1	$h \times d \times (\hat{w}/24)$
Conv3	$h \times d \times (\hat{w}/24)$	3×3×3	1,1,3	$h \times d \times w$
Reshape	$h \times d \times w$	×	×	$h \times w \times d$

the deep neural models, where the learning rate is set to 0.0001 and a batch size of 175 nodules is considered, specified with trial and error. Different metrics are used for lung nodule classification evaluation in the literature. This study applies accuracy, sensitivity and specificity rate on the separate test data set in addition to ROC-AUC. The 10-fold cross-validation method is adopted for model evaluation in all experiments.

C. RESULTS

The performance of the proposed nodule classification approach based on TFD extraction and MOGAM fusion, represented with TFD-MOGAM, is evaluated and compared with available baseline non-local neural networks and the most recent model-based attention methods for lung nodule classification in Table 4. These methods include NVCF [36], shuffle attention [38], ProCAN [28], hybrid attention mechanism [40], multi-scale attention [37] and multi-modality attention [9], which are implemented and tested on the LIDC-IDRI dataset. As observed, the proposed model improved AUC by 1.48%, accuracy by 1.2%, sensitivity by 5.3% and specificity by 0.35% compared with the best results reported in the literature.

A closer investigation of the proposed TFD-MOGAM method performance is shown in Fig.9, where its ROC curve is compared with those of the baseline attention-based models including 2D axial attention [19] and 3D axial attention [17].

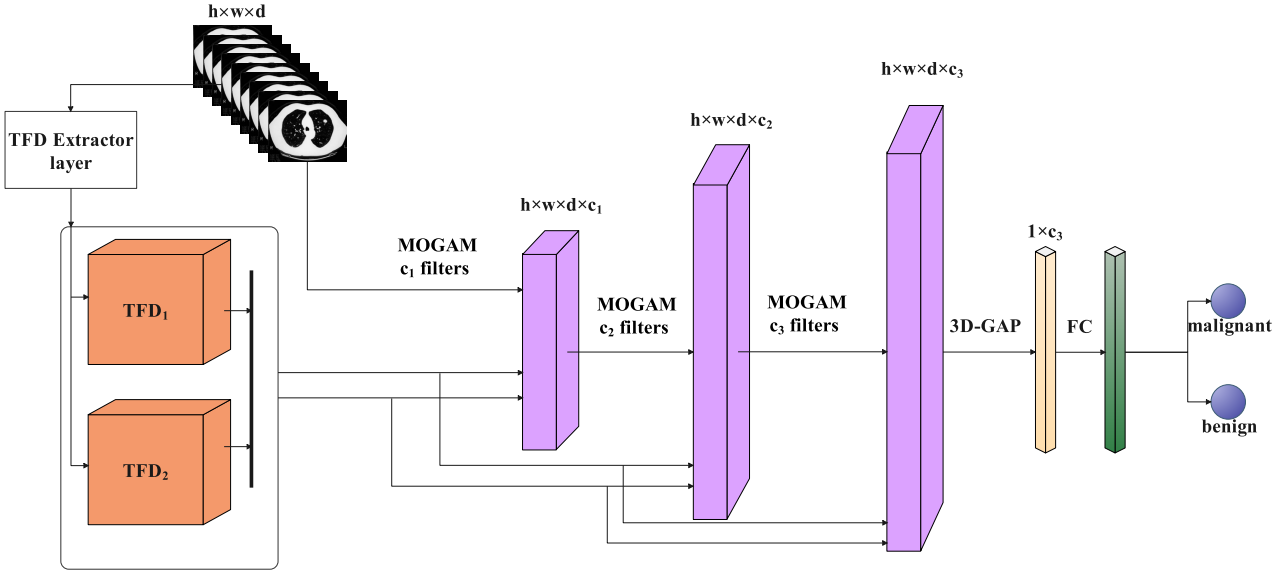


FIGURE 8. A deep neural architecture for lung nodule classification employing the proposed TFD-MOGAM-based in three layers for information fusion.

TABLE 4. Comparison of the proposed model with the most recent approaches for lung cancer classification.

Methods	AUC (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
2D axial attention [14]*	92.8	93.5	91.2	89.2
3D axial attention [17]*	94.1	95.8	93.8	92.59
NVCF [29]	93.3	94.4	94.58	93.8
Multi-Modality Attention [12]	94.3	95.35	97.3	93.8
Multi-Scale Attention [37]	92.7	96.2	94.33	90.32
Shuffle Attention [30]	93.41	94.6	91.3	95.1
Hybrid attention mechanism [38]	91.54	93.55	89.2	94.15
C-ConvNet/C-CNN [36]	92.21	94.99	95.3	87.3
ProCAN [26]	94.32	96	92.3	95.35
TFD-MOGAM	95.8	97.2	97.6	95.7

* *Baseline methods.*

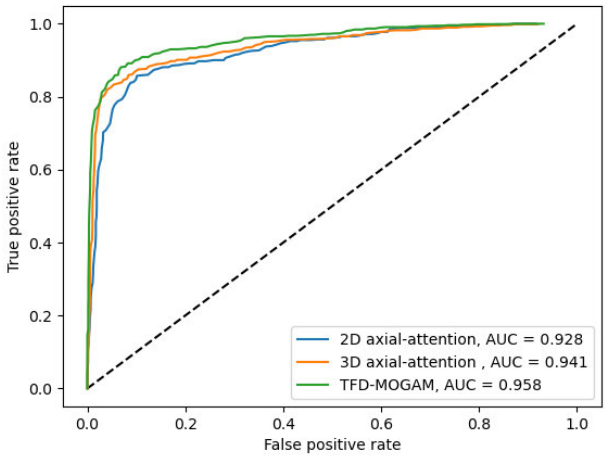


FIGURE 9. ROC curves and the average AUCs for the baseline models and the proposed model.

With the help of the proposed MOGAM for feature fusion, TFD-MOGAM exhibits a higher AUC of 3% compared with 2D axial attention and 1.7% compared with 3D axial attention in nodule classification.

D. ABLATION EXPERIMENTS

In this section the impact of the constituting components on the overall performance of the proposed TFD-MOGAM is evaluated. One of the main advantages of TFD-MOGAM is in using the feature descriptors capturing various textural information of the regions across nodule's volume. The results of an ablation study on the choice of these descriptors are tabulated in Table 5, obtained over one hundred nodules randomly sampled in a validation set in order to perform optimum model selection. It is evident from the results that while certain subset of descriptors may decrease the overall performance, the best accuracy, sensitivity, specificity and AUC results are obtained when all of the nine feature descriptors introduced in Eqs. 1-9 are utilized; the second-best results are obtained with homogeneity, ASM and energy descriptors.

Locality of the computed descriptors is also very important for correct representation of textural information captured from volumetric nodule. To illustrate the superiority of local TFDs in comparison with those computed from the global GLCM, in the second ablation study the nine feature descrip-

TABLE 5. Results of the ablation experiments for different texture feature descriptors in TFD-MOGAM.

Texture Descriptor	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
Mean	74.1	71.52	69.3	72.2
Mean + STD + Contrast + Dissimilarity	83.8	88.32	75.6	81.1
Homogeneity	71.1	79.9	81.3	74.54
Homogeneity + ASM + Energy	92.3	91.02	87.9	89.2
Max + Entropy	79.9	91.5	88.3	89.9
All	97.2	97.6	95.7	95.8

TABLE 6. Results of the ablation experiment for comparing the use of global GLCM and local co-occurrences in TFD-MOGAM.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
global TFD-MOGAM	88.2	91.6	87.65	86.9
TFD-MOGAM	97.2	97.6	95.7	95.8

TABLE 7. Results of the ablation experiment on diffrent number of MOGAM layers for fusion in deep neural network.

#MOGAM blocks	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
1-MOGAM	93.2	94.1	91	91.8
2-MOGAMs	93.5	94	92	92.3
3-MOGAMs	97.2	97.6	95.7	95.8
4-MOGAMs	97	96.6	95.2	95.1

tors are computed from global GLCMs obtained in eight different directions, instead of utilizing sliding local windows. The experimental results shown in Table 6 confirm the initial hypothesis that the spatial placement of the textural information is required for correct lesion detection and classification, which is not provided in global GLCMs.

Furthermore, we conduct an ablation experiment to specify the number of MOGAMs required for information fusion in the proposed model. The results of this experiment is presented in Table 7. It can be seen that increasing the number of MOGAMs improves the classification performance in terms of all evaluation metrics. However, with more than three consecutive MOGAMs in the network architecture, the performance starts to degrade which can be attributed to excess complexification of the model resulting in overfitting. Therefore, three MOGAMs (as specified in Section IV-B) are employed in the proposed deep fusion model for all of the experiments conducted in this study.

E. COMPUTATIONAL COMPLEXITIES

One of the important aspects of analyzing classification models performance is the computational time required to construct these models (training complexity), and also the time complexity of deploying them for actual classification (test time requirements). Fig. 10 shows the average test time of the proposed TFD-MOGAM compared with other nodule classification methods, obtained on the same computing environment and the LIDC-IDRI dataset. The results indicate that the proposed model requires more time than the other classification models, except the Shuffle attention model which has the highest time complexity. This increment in the time requirements of the proposed model is expected due to a more complex multi-stage fusion structure which is necessary for obtaining better classification performance.

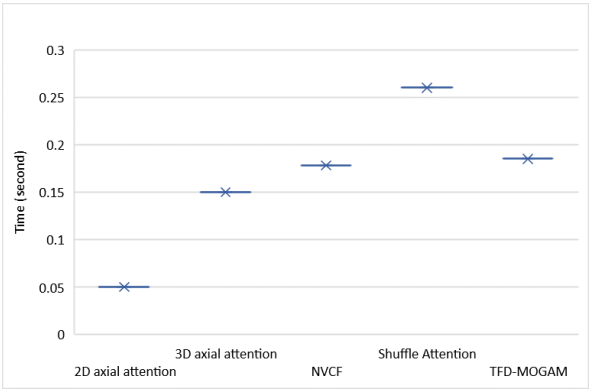


FIGURE 10. The test time requirements of different lung nodule classification models.

The quadric complexity makes the non-local based elements flattening method [17] very high expensive. Through the 2D axial attention [19] and 3D axial attention [20] the complexity of the computation has been reduced to $O(n\sqrt{n})$. In the proposed TFD-MOGAM method, the complexity of the computation each of the crosswise and the lengthwise attention map multiplication is $O(n\sqrt{n})$. But, since these operations are not repeated the order does not change. Thus, the complexity of the proposed method is less than non-local based elements flattening method [17] but a bit more than those of 2D axial attention [19] and 3D axial attention [20] which is negligible.

The average accumulative training time of the previous five deep classification models per the first five epochs is depicted in Fig. 11. Again, it can be seen that the growth in computational time of the Shuffle attention model is very fast, and the other models based on 3D processing of nodules perform almost similarly, while 2D axial attention has the smallest slope in time requirement.

F. DISCUSSION

The salient regions greatly impacting the nodule classification results may have small sizes, be located outside of the nodules perceived contours, or be spatially distributed in a local or non-local proximity of each other. The multi-orientation computation of local TFDs allows decomposing the textural information contained in different regions of the nodule's volume according to small windows slid over the nodule in different orientations, captured in

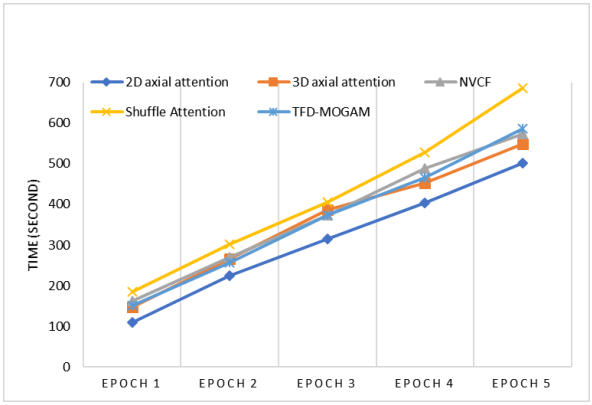


FIGURE 11. The training time requirements of different lung nodule classification models.

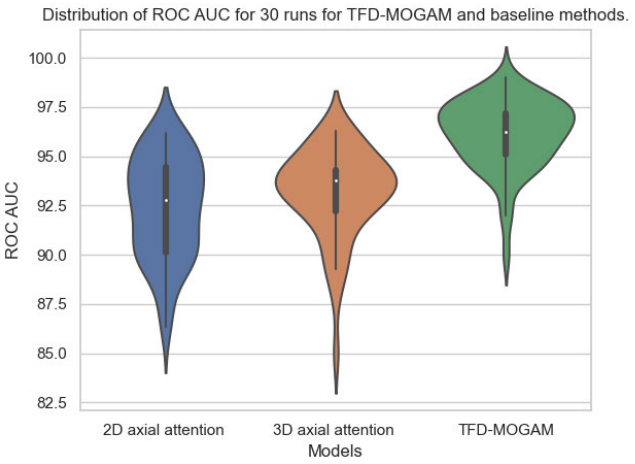


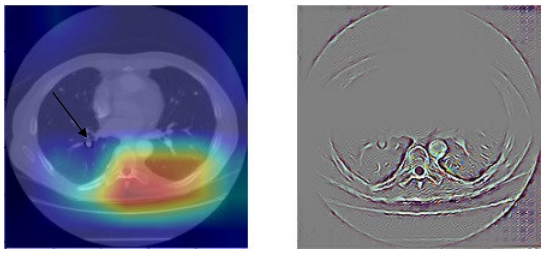
FIGURE 12. Distribution of ROC-AUC for 30 runs of the proposed TFD-MOGAM-based model and baseline methods.

various directions and from different descriptor functions perspective. Such fine-grained information can then be used for identifying those regions with arbitrary spatial resolution in the classification process, even for small lesions. Attentive weights provide a flexible mechanism for this purpose and can capture the local and non-local correlation of the regions for better classification. In the proposed MOGAM the detailed information encoded in TFDs are utilized for such attendance to important regions of the to salient regions in the nodule's volume, both spatially and channel-wise.

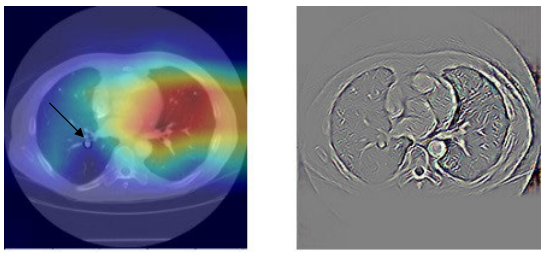
The results presented in Table 4 confirm the advantage of the proposed approach for overcoming the challenges in analyzing nodule information compared with other counterparts in classification. Notably, direct application of 3D-CNN structures on the 3D volume of lung nodule, which uses local 3D convolution operations, imposes limitations on the models' ability to handle the long-range dependencies in a volumetric nodule [55], yielding less impact on the classification performance.

The ablation experiments' outcomes indicate that fusion of different feature descriptors through the non-local attention is effective for lung nodule classification, which can be due to the sensitivity of the model to the TFDs extracted from the local co-occurrence masks. Different interpretations of textures through different feature descriptors can significantly extend the available information.

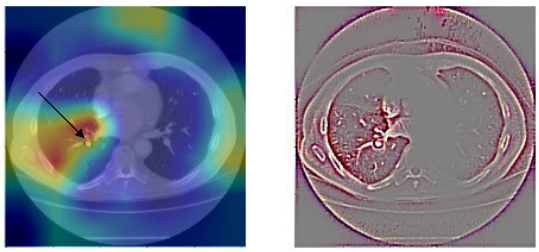
Through sequences of runs performed with the proposed model and the baseline models considering the LIDC-IDRI dataset, the distribution of the AUCs for these models are depicted in Fig. 12. The violin plot of TFD-MOGAM is fatter, has a smaller tail and a larger average AUC, showing a more robust performance compared with the other two baseline models. To test the statistical significance of the dissimilarities in the performance of the proposed and baseline models, a statistical analysis is conducted using the t-test. The obtained p-value when comparing TFD-MOGAM with 2D axial attention and 3D axial attention is respectively less than 0.0003 and less than 0.00001, revealing the significant advantage of the proposed TFD-MOGAM in lung nodule classification in terms of the AUC measure.



(a) Grad-CAM and Guided-GradCAM for 2D axial attention



(b) Grad-CAM and Guided-GradCAM for 3D axial attention

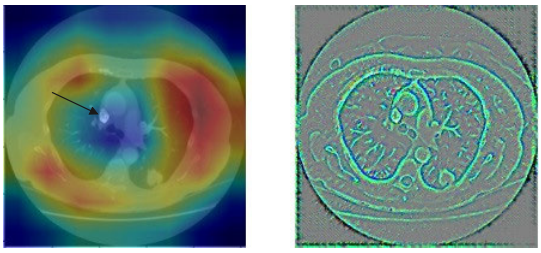


(c) Grad-CAM and Guided-GradCAM for TFD-MOGA

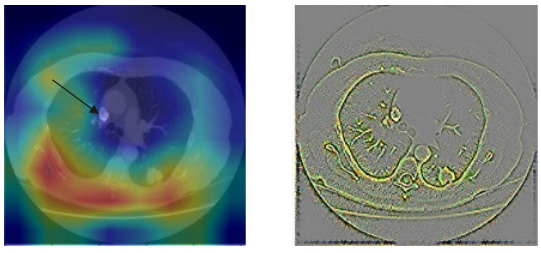
FIGURE 13. The visualization results by grad-CAM (left) and Guided-GradCAM (right) from TFD-MOGAM and other baseline methods for a benign nodule.

G. RESULT INTERPRETABILITY

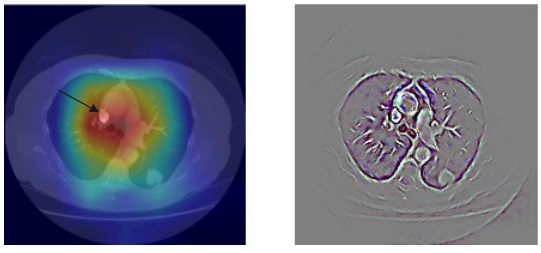
To show how the TFD-MOGAM module classifies lung nodules considering attention to spatial and channel information semantically in the regions of interest, we selected a sample of benign and malignant nodules to gain insight in interpreting



(a) Grad-CAM and Guided-GradCAM for 2D axial attention



(b) Grad-CAM and Guided-GradCAM for 3D axial attention



(c) Grad-CAM and Guided-GradCAM for TFD-MOGAM

FIGURE 14. The visualization results by grad-CAM (left) and Guided-GradCAM (right) from TFD-MOGAM and other baseline methods for a malignant nodule.

the decision-making based on the proposed TFD-MOGAM by adopting Grad-CAM [51].

Regarding the benign nodule sample, Fig.13 (left) shows the Grad-CAM on the predicted important coarse regions for decision making. Fig.13 (right) displays the output of guided-Grad-CAM, where the best-localized relevant regions obtained through discriminative classes by Grad-CAM are combined i.e., fine-grained regions visualized through the proposed TFD-MOGAM. As observed in Fig. 13c, for TFD-MOGAM, the salient regions are purified, and more accurate attentional signals on the lung nodules are used for classification than Figs. 13a and 13b. This can be attributed to the model's ability to capture long-range dependencies in CT scans and concentrate more even on small lesions, thus, an improvement in classification efficacy.

By observing the malignant nodule sample, we can see our TFD-MOGAM network has more attention to coarse (Fig.14 (left)) and fine-grained (Fig.14(right)) features of the malignant nodules. It is notable from Fig. 14 (left) that our model could identify and concentrate on the regions of interest where the Grad-CAM is applied to CT images containing malignant nodules. Also, in Fig. 14 (right), the guided-Grad-CAM is displayed. Consequently, our model is considered interpretable with increased adoption rates.

V. CONCLUSION

A new and practical non-local attention module is proposed where the Co-occurrence masks are applied on quantized intensities. The texture features descriptors are extracted locally from these masks and are estimated from the lung nodules in their crosswise and lengthwise orientations. These features are extracted for devising the attention maps to be applied on the nodule to salience the high-level features.

Ablation experiments indicate that the applied co-occurrence masks could significantly improve the sensitivity of the extracted texture feature descriptors compared to applying the global GLCM. The co-occurrence masks determine the textural and locational structures of the nodules in the CT images clearly, Fig.13.

The long-range dependencies in lung nodule cross-sectional slices are mitigated in [20] by applying three attention maps computed through the non-local operations, where the nodules are of equal size through their slices, causing attention to missing an essential portion of the consequent slices due to their size variation. This study could overcome this shortcoming by considering eight directions of the co-occurrence masks computed from the crosswise and lengthwise dimensions of the nodule.

Experimental results show that performing element-wise multiplication on a given feature map could highlight the texture weights, Figs.11 and 13. Although this model consumes slightly more time than its counterparts in training (Fig. 12) and testing (Fig. 14), its performance is enhanced significantly, Fig. (10).

This newly proposed module yields visible context fusion and has overcome the inadequate representation of the CT scans' construction, which have long-range dependencies on a crosswise slice that sequentially comprises the nodules within the CT scan. It allows the model to concentrate on small lesions by considering their attention to all related slices. This method is verified through the quantitative and qualitative studies run on the LIDC-IDRI dataset. Experimental outcomes indicate that this TFD-MOGAM can seriously enhance the interpretation and representation of texture features on medical images. The class activation mapping indicates that this model reveals the interpretable weights that focus on the region of interest in the high-level features. The importance of each cross-sectional slice of the nodule is proportionally increasing according to the nodule's overall length section cuts representation. The results indicate that this model outperforms its counterparts in terms of lung nodule classification.

In future studies, we would like to devise a trainable co-occurrence masks extractor module so that rather than extracting these masks as preprocessing, a learnable co-occurrence layer would be applied and propose attentional maps by applying different texture feature types (e.g., Gabor filters). Labeling the nodules is an expensive process. Moreover, indeterminate nodules (those with a median malignancy score of 3) constitute almost 44% of the LIDC dataset. Thus, one important future work is to take advantage of these

nodules in the training process using semi-supervised learning approach to improve the performance of nodule classification. Generative adversarial networks [57] are one of the possible methods for this purpose where the discriminator is used to not only to detect the fake samples but also to classify the input samples. Therefore, the indeterminate nodules similar to the labeled nodules can be similarly classified. Another possible method is to use the biomarker data as an extra modality by considering their similarity in indeterminate nodules to those for the benign or malignant nodules.

REFERENCES

[1] Y. Chen, Y. Wang, F. Hu, and D. Wang, "A lung dense deep convolution neural network for robust lung parenchyma segmentation," *IEEE Access*, vol. 8, pp. 93527–93547, 2020, doi: 10.1109/ACCESS.2020.2993953.

[2] A. Halder, S. Chatterjee, and D. Dey, "Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103347, doi: 10.1016/j.bspc.2021.103347.

[3] Z. Li, J. Zhang, T. Tan, X. Teng, X. Sun, H. Zhao, L. Liu, Y. Xiao, B. Lee, Y. Li, and Q. Q. Zhang, "Deep learning methods for lung cancer segmentation in whole-slide histopathology images—The ACDC@LungHP challenge 2019," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 429–440, Feb. 2021, doi: 10.1109/JBHI.2020.3039741.

[4] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, E. Tsougenis, Q. Huang, M. Cai, and P.-A. Heng, "Weakly supervised deep learning for whole slide lung cancer image analysis," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3950–3962, Sep. 2020, doi: 10.1109/TCYB.2019.2935141.

[5] A. Sreekumar, K. R. Nair, S. Sudheer, H. G. Nayar, and J. J. Nair, "Malignant lung nodule detection using deep learning," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Jul. 2020, pp. 209–212, doi: 10.1109/ICCSP48568.2020.9182258.

[6] T. Qu, X. Wang, C. Fang, L. Mao, J. Li, P. Li, J. Qu, X. Li, H. Xue, Y. Yu, and Z. Jin, "M3Net: A multi-scale multi-view framework for multi-phase pancreas segmentation based on cross-phase non-local attention," *Med. Image Anal.*, vol. 75, Jan. 2022, Art. no. 102232, doi: 10.1016/j.media.2021.102232.

[7] A. Hering, S. Häger, J. Moltz, N. Lessmann, S. Heldmann, and B. Van Ginneken, "CNN-based lung CT registration with multiple anatomical constraints," *Med. Image Anal.*, vol. 72, Aug. 2021, Art. no. 102139, doi: 10.1016/j.media.2021.102139.

[8] J. Lu, R. Jin, E. Song, G. Ma, and M. Wang, "Lung-CRNet: A convolutional recurrent neural network for lung 4DCT image registration," *Med. Phys.*, vol. 48, no. 12, pp. 7900–7912, Dec. 2021, doi: 10.1002/mp.15324.

[9] L. Chen, K. Liu, H. Shen, H. Ye, H. Liu, L. Yu, J. Li, K. Zhao, and W. Zhu, "Multimodality attention-guided 3-D detection of nonsmall cell lung cancer in 18F-FDG PET/CT images," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 4, pp. 421–432, Apr. 2022, doi: 10.1109/trpms.2021.3072064.

[10] C. E. Brookmeyer, S. Bhatt, E. K. Fishman, and S. Sheth, "Multimodality imaging after liver transplant: Top 10 important complications," *RadioGraphics*, vol. 42, no. 3, pp. 702–721, May 2022.

[11] S. Liang, T. Wang, C. Chen, H. Liu, C. Qin, and Y. Feng, "RSEA-Net: Residual squeeze and excitation attention network for medical image segmentation," *Res. Square*, 2022, doi: 10.21203/rs.3.rs-1419097/v1.

[12] Y. Wang, C. Zhou, H.-P. Chan, L. M. Hadjiiski, and A. Chughtai, "Fusion of multiple deep convolutional neural networks (DCNNs) for improved segmentation of lung nodules in CT images," in *Proc. SPIE*, vol. 12033, pp. 593–596, Apr. 2022, doi: 10.1117/12.2612360.

[13] M. M. Farhangi, N. Petrick, B. Sahiner, H. Frigui, A. A. Amini, and A. Pezeshk, "Recurrent attention network for false positive reduction in the detection of pulmonary nodules in thoracic CT scans," *Med. Phys.*, vol. 47, no. 5, pp. 2150–2160, May 2020, doi: 10.1002/mp.14076.

[14] S. Sasikumar, P. N. Renjith, K. Ramesh, and K. S. Sankaran, "Attention based recurrent neural network for lung cancer detection," in *Proc. 4th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC)*, Oct. 2020, pp. 720–724, doi: 10.1109/I-SMAC49090.2020.9243556.

[15] A. Saihood, H. Karshenas, and A. N. Nilchi, "Deep fusion of gray level co-occurrence matrices for lung nodule classification," *PLoS ONE*, vol. 17, no. 9, pp. 1–26, 2022.

[16] H. Lin, W. Zheng, and X. Peng, "Orientation-encoding CNN for point cloud classification and segmentation," *Mach. Learn. Knowl. Extraction*, vol. 3, no. 3, pp. 601–614, Aug. 2021, doi: 10.3390/make3030031.

[17] J. Wang, Y. Guo, Y. Ying, Y. Liu, and Q. Peng, "Fast non-local algorithm for image denoising," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 1429–1432, doi: 10.1109/ICIP.2006.312698.

[18] X. Wang, R. B. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7794–7803, doi: 10.1109/CVPR.2018.00813.

[19] J. Ho, N. Kalchbrenner, D. Weissenborn, and T. Salimans, "Axial attention in multidimensional transformers," 2019, *arXiv:1912.12180*.

[20] M. Al-Shabi, K. Shak, and M. Tan, "3D axial-attention for lung nodule classification," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 16, no. 8, pp. 1319–1324, Aug. 2021, doi: 10.1007/s11548-021-02415-z.

[21] S. G. Armato et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Feb. 2011, doi: 10.1118/1.3528204.

[22] E. Bebas, M. Borowska, M. Derlatka, E. Oczeretko, M. Hladunski, P. Szumowski, and M. Mojsak, "Machine-learning-based classification of the histological subtype of non-small-cell lung cancer using MRI texture analysis," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102446, doi: 10.1016/j.bspc.2021.102446.

[23] M. A. Khan, V. Rajinikanth, S. C. Satapathy, D. Taniar, J. R. Mohanty, U. Tariq, and R. Damasevicius, "VGG19 network assisted joint segmentation and classification of lung nodules in CT images," *Diagnostics*, vol. 11, no. 12, p. 2208, 2021.

[24] K. Prakash and S. Saradha, "Efficient prediction and classification for cirrhosis disease using LBP, GLCM and SVM from MRI images," *Proc. Mater. Today*, 2021, doi: 10.1016/j.matpr.2021.03.418.

[25] M. Usman, B.-D. Lee, S.-S. Byon, S.-H. Kim, B.-I. Lee, and Y.-G. Shin, "Volumetric lung nodule segmentation using adaptive ROI with multi-view residual learning," *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, Jul. 2020, doi: 10.1038/s41598-020-69817-y.

[26] S. Bera and P. K. Biswas, "Noise conscious training of non local neural network powered by self attentive spectral normalized Markovian patch GAN for low dose CT denoising," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3663–3673, Dec. 2021, doi: 10.1109/TMI.2021.3094525.

[27] F. Rundo, G. L. Banna, L. Prezzavento, F. Trenta, S. Conoci, and S. Battiato, "3D non-local neural network: A non-invasive biomarker for immunotherapy treatment outcome prediction. Case-study: Metastatic urothelial carcinoma," *J. Imag.*, vol. 6, no. 12, p. 133, Dec. 2020, doi: 10.3390/jimaging6120133.

[28] M. Al-Shabi, K. Shak, and M. Tan, "ProCAN: Progressive growing channel attentive non-local network for lung nodule classification," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108309, doi: 10.1016/j.patcog.2021.108309.

[29] F. Rundo, A. Genovese, R. Leotta, F. Scotti, V. Piuri, and S. Battiato, "Advanced 3D deep non-local embedded system for self-augmented X-ray-based COVID-19 assessment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 423–432.

[30] Z. Wang, N. Zou, D. Shen, and S. Ji, "Non-local U-Nets for biomedical image segmentation," in *Proc. AAAI*, 2020, pp. 6315–6322.

[31] Z. Liu, F. Du, W. Li, X. Liu, and Q. Zou, "Non-local spatial and temporal attention network for video-based person re-identification," *Appl. Sci.*, vol. 10, no. 15, p. 5385, Aug. 2020, doi: 10.3390/APP10155385.

[32] M. Jiang, L. Song, Y. Wang, Z. Li, and H. Song, "Fusion of the YOLOv4 network model and visual attention mechanism to detect low-quality young apples in a complex environment," *Precis. Agricult.*, vol. 23, no. 2, pp. 559–577, Apr. 2022, doi: 10.1007/s11119-021-09849-0.

[33] L. Zhu, Z. Deng, X. Hu, H. Xie, X. Xu, J. Qin, and P. A. Heng, "Learning gated non-local residual for single-image rain streak removal," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 6, pp. 2147–2159, Jun. 2021, doi: 10.1109/TCSVT.2020.3022707.

[34] Z. Rao, M. He, Y. Dai, Z. Zhu, B. Li, and R. He, "NLCA-Net: A non-local context attention network for stereo matching," *APSIPA Trans. Signal Inf. Process.*, vol. 9, no. 1, pp. 1–13, 2020, doi: 10.1017/ATSIP.2020.16.

[35] Z. Li, S. Wang, H. Yu, Y. Zhu, Q. Wu, L. Wang, Z. Wu, Y. Gan, W. Li, B. Qiu, and J. Tian, "A novel deep learning framework based mask-guided attention mechanism for distant metastasis prediction of lung cancer," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, May 17, 2022, doi: 10.1109/TETCI.2022.3171311.

[36] Y. Wen, L. Chen, H. Chen, X. Tang, Y. Deng, Y. Chen, and C. Zhou, "Non-local attention learning for medical image classification," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2021, pp. 1–6, doi: 10.1109/ICME51207.2021.9428267.
[37] K. Xia, J. Chi, Y. Gao, Y. Jiang, and C. Wu, "Adaptive aggregated attention network for pulmonary nodule classification," *Appl. Sci.*, vol. 11, no. 2, pp. 1–15, 2021, doi: 10.3390/app11020610.
[38] Q.-L. Zhang and Y.-B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2235–2239, doi: 10.1109/ICASSP39728.2021.9414568.
[39] R. Ranjbarzadeh, A. B. Kasgari, S. J. Ghoushchi, S. Anari, M. Naseri, and M. Bendechache, "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," *Sci. Rep.*, vol. 11, no. 1, pp. 1–17, May 2021, doi: 10.1038/s41598-021-90428-8.
[40] L. Xu, L. Wang, S. Cheng, and Y. Li, "MHANet: A hybrid attention mechanism for retinal diseases classification," *PLoS ONE*, vol. 16, no. 12, 2021, Art. no. e0261285, doi: 10.1371/journal.pone.0261285.
[41] T. Shan and J. Yan, "SCA-Net: A spatial and channel attention network for medical image segmentation," *IEEE Access*, vol. 9, pp. 160926–160937, 2021, doi: 10.1109/ACCESS.2021.3132293.
[42] Y. Hu, Z. Liang, B. Song, H. Han, P. J. Pickhardt, W. Zhu, C. Duan, H. Zhang, M. A. Barish, and C. E. Lascarides, "Texture feature extraction and analysis for polyp differentiation via computed tomography colonography," *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1522–1531, Jun. 2016, doi: 10.1109/TMI.2016.2518958.
[43] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[44] M.-H. Guo, T. X. Xu, J. J. Liu, Z. N. Liu, P. T. Jiang, T. J. Mu, S. H. Zhang, R. R. Martin, M. M. Cheng, and S. M. Hu, "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, pp. 331–368, Mar. 2022, doi: 10.1007/s41095-022-0271-y.
[45] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 60–65, doi: 10.1109/CVPR.2005.38.
[46] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *IEEE Access*, vol. 7, pp. 8975–9000, 2019, doi: 10.1109/ACCESS.2018.2890743.
[47] T. Toizumi, S. Zini, K. Sagi, E. Kaneko, M. Tsukada, and R. Schettini, "Artifact-free thin cloud removal using gans," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 1, Sep. 2019, pp. 3596–3600.
[48] A. Porebskil, N. Vandenbrouckel, and L. Macaire, "Haralick feature extraction from LBP images for color texture classification," in *Proc. 1st Workshops Image Process. Theory, Tools Appl.*, 2008, pp. 1–8.
[49] K. Mei, A. Jiang, J. Li, and M. Wang, "Progressive feature fusion network for realistic image dehazing," in *Computer Vision—ACCV* (Lecture Notes in Computer Science), vol. 11361. Cham, Switzerland: Springer, 2019, pp. 203–215, doi: 10.1007/978-3-030-20887-5_13.
[50] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, and E. A. Kazerooni, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Feb. 2011, doi: 10.1118/1.3528204.
[51] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 2017, pp. 618–626, Dec. 2017, doi: 10.1109/ICCV.2017.74.
[52] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," 2013, *arXiv:1312.6034*.
[53] R. Fu, Q. Hu, X. Dong, Y. Guo, Y. Gao, and B. Li, "Axiom-based grad-CAM: Towards accurate visualization and explanation of CNNs," 2020, *arXiv:2008.02312*.
[54] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 24–25.
[55] Y. Han, H. Qi, L. Wang, C. Chen, J. Miao, H. Xu, Z. Wang, Z. Guo, Q. Xu, Q. Lin, H. Liu, J. Lu, F. Liang, W. Feng, H. Li, and Y. Liu, "Pulmonary nodules detection assistant platform: An effective computer aided system for early pulmonary nodules detection in physical examination," *Comput. Methods Programs Biomed.*, vol. 217, Apr. 2022, Art. no. 106680, doi: 10.1016/j.cmpb.2022.106680.
[56] N. I. Yangfan, Y. A. N. G. Yuanyuan, X. I. E. Zhe, Z. H. E. N. G. Dezhong, and W. A. N. G. Weidong, "Multi-feature extraction of pulmonary nodules based on LSTM and attention structure," *J. Shanghai Jiao Tong Univ.*, vol. 56, no. 8, pp. 1078–1088, 2022.
[57] A. Saihood, H. Karshenas, and A. R. N. Nilchi, "Spatial-channel attention-based stochastic neighboring embedding pooling and long-short-term memory for lung nodules classification," in *Proc. 12th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Nov. 2022, pp. 477–485, doi: 10.1109/ICCKE57176.2022.9960025.

AHMED SAIHOOD received the B.Sc. degree in computer engineering from Basra University, Iraq, in 2009, and the M.Tech. degree in information technology from Maharishi Markandeshwar University (MMU), India, in 2013. He is currently pursuing the Ph.D. degree with the Artificial Intelligence Department, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran. He is also a faculty of computer sciences and mathematics at the University of Thi-Qar, Iraq.

HOSSEIN KARSHENAS received the B.E. degree in computer engineering from Shahid Beheshti University, Tehran, Iran, in 2006, the M.E. degree in artificial intelligence and robotics from the Iran University of Science and Technology (IUST), Tehran, in 2009, and the Ph.D. degree in artificial intelligence from the Technical University of Madrid (UPM), Madrid, Spain, in 2013. He is currently an Assistant Professor at the Artificial Intelligence Department, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran. His main research interests include estimation of distribution algorithms, probabilistic graphical models, continuous and multi-objective optimization, computational intelligence, and machine learning, where he has published several articles in peer-reviewed journals.

AHMAD REZA NAGHSH-NILCHI received the B.S., M.S., and Ph.D. degrees from the University of Utah, Salt Lake City, all in electrical engineering. He is currently an Associate Professor with the University of Isfahan, Isfahan, Iran. He was the Chairperson of the Computer Engineering Department for three terms. He is the author or the coauthor of several journal articles and conference papers. He has written a section of a book. He has collaborated with internationally known institutions and peers. He was a Research Scholar with the National University of Ireland, Galway, Ireland, in 2011, and with the University of California, Irvine, in 2012. His current research interests include medical image, signal processing, and intensive computing. He is the Editor-in-Chief of the *Iranian Journal of Engineering Sciences*. He was listed in Who's Who in the World in 2011.