IEEE *Access*
Multidisciplinary : Rapid Review : Open Access Journal

SPECIAL SECTION ON TRENDS, PERSPECTIVES AND PROSPECTS OF MACHINE LEARNING
APPLIED TO BIOMEDICAL SYSTEMS IN INTERNET OF MEDICAL THINGS

Health of Things Algorithms for Malignancy Level Classification of Lung Nodules

**MURILLO B. RODRIGUES[1], RAUL VICTOR M. DA NÓBREGA[1], SHARA SHAMI A. ALVES[1],
PEDRO PEDROSA REBOUÇAS FILHO[1], JOÃO BATISTA F. DUARTE[2], ARUN K. SANGAIAH[3],
AND VICTOR HUGO C. DE ALBUQUERQUE [2], (Member, IEEE)**

[1]Laboratório de Processamento de Imagens e Simulação Computacional, Instituto Federal de Educação, Ciencia e Tecnologia do Ceará, Maracanau 61939-140, Brazil
[2]Programa de Pós-Graduação em Informática Aplicada, Universidade de Fortaleza, Fortaleza 60811-905, Brazil
[3]School of Computing Science and Engineering, Vellore Institute of Technology (VIT), Vellore 632014, India

Corresponding author: Victor Hugo C. de Albuquerque (victor.albuquerque@unifor.br)

ABSTRACT Lung cancer is one of the leading causes of death world wide. Several computer-aided diagnosis systems have been developed to help reduce lung cancer mortality rates. This paper presents a novel structural co-occurrence matrix (SCM)-based approach to classify nodules into malignant or benign nodules and also into their malignancy levels. The SCM technique was applied to extract features from images of nodules and classifying them into malignant or benign nodules and also into their malignancy levels. The computed tomography exams from the lung image database consortium and image database resource initiative datasets provide information concerning nodule positions and their malignancy levels. The SCM was applied on both grayscale and Hounsfield unit images with four filters, to wit, mean, Laplace, Gaussian, and Sobel filters creating eight different configurations. The classification stage used three well-known classifiers: multilayer perceptron, support vector machine, and k-nearest neighbors algorithm and applied them to two tasks: (*i*) to classify the nodule images into malignant or benign nodules and (*ii*) to classify the lung nodules into malignancy levels (1 to 5). The results of this approach were compared to four other feature extraction methods: gray-level co-occurrence matrix, local binary patterns, central moments, and statistical moments. Moreover, the results here were also compared to the results reported in the literature. Our approach outperformed the other methods in both tasks; it achieved 96.7% for both accuracy and F-Score metrics in the first task, and 74.5% accuracy and 53.2% F-Score in the second. These experimental results reveal that the SCM successfully extracted features of the nodules from the images and, therefore may be considered as a promising tool to support medical specialist to make a more precise diagnosis concerning the malignancy of lung nodules.

INDEX TERMS Computer-aided diagnosis, pulmonary nodules, lung cancer, textural features, structural co-occurrence matrix, malignancy classification.

I. INTRODUCTION

The most common cancers that cause death are lung, prostate, breast and colon cancer. They represent 46% of all deaths due to cancer and lung cancer is responsible for more than a quarter (27%) of all cancers [1]. In developed countries, lung cancer patients have a 10 up to 16% chance of having a five-year survival rate [2]. Nevertheless, early detection of lung cancer, through computed tomography (CT), tends to improve survival in patients [1], insomuch that the five-year survival rate increases to 70% [3].

The medical specialist first identifies the pulmonary nodules from a CT scan, and then makes a possible prognosis based on the nodule morphology assessment including the clinical context. However, he often has to analyze a large number of nodules and make a prognosis quickly, and such tasks become burdensome under these circumstances [2], [4], [5]. Thus, computer-aided diagnosis (CAD) systems have arisen to overcome such situations. CAD systems are categorized into two groups, to wit, (*i*) detection systems (CADe) and, (*ii*) diagnostic systems (CADx) [2].

CADx systems perform an automatic diagnosis based on features extracted from the system input images. The automatic classification of the nodules into malignant or benign using CT images supports the medical specialist when assessing nodules [6] and at the same time the CADx system provides a second opinion to help in decision-making [5], [7], [8].

A combination of texture and shape features was used for the classification of pulmonary nodules into malignant or benign ones using several classifiers in [9]. The author used 33 cases, of which 14 were malignant and 19 were benign. The highest accuracy achieved was 90.91%.

A deep learning model of the Multi-Crop Convolutional Neural Network (MCCNN) to classify the lung nodules was introduced in [10]. The authors used 880 benign nodules and 495 malignant nodules from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI) dataset and obtained an accuracy of 87.14%.

A Multi-Scale Convolutional Neural Network (MCNN) for nodule diagnostic classification was proposed in [6]. This approach is robust against noisy inputs and it achieved 86.84% for nodule classification, also it outperformed the benchmark textural descriptors. Overall, the authors used 880 benign nodules and 495 malignant nodules from the LIDC/IDRI dataset.

A lung nodule classification system was proposed in [11]. This system uses texture and morphological features from regions of interest containing lung nodules. The features were obtained by the GLCM method and discrete wavelet transform from 321 instances of lung nodules from the LIDC dataset. Deep features extracted from an autoencoder were used to classify lung nodules as either malignant or benign, using 4303 instances containing 4323 nodules from the LIDC dataset [5].

This work aims to classify the malignancy of lung nodules using the feature extractor Structural Co-occurrence Matrix (SCM) [12] and, a well-known classifier, in order to support the medical specialist when assessing nodules and also help him in decision-making. The performance of the multilayer perceptron (MLP), the support vector machine (SVM) and the k-nearest neighbors (k-NN) algorithms were compared to decide which classifier works better with the proposed SCM approach. Also, other experiments were conducted to assess the SCM approach compared to other extraction techniques, such as, the gray-level co-occurrence matrix (GLCM), local binary patterns (LBP), central moments (CM) and statistical moment (SM) [13]–[15]. Moreover, the results were compared to the pulmonary nodule classification approaches in [5], [6], [9]–[11]. All the experiments applied the extraction methods and classifiers on the same CT images. The results revealed that the effectiveness of the approach presented here and showed how it might help the medical specialist make a more precise diagnosis concerning the malignancy of lung nodules.

The remaining part of this work is organized as follows. In Section II we present the feature extractors and classification methods, and also detail the proposed approach and the experiments setup. Then, we present the results and its discussion in Section III. Finally, some conclusion remarks are presented in Section IV.

II. METHODOLOGY

An overview of the proposed approach here is depicted in Fig. 1. The first step consists of extracting ROIs (pulmonary nodules) from the LIDC/IDRI dataset. In the next step, different configurations of the SCM extractor are applied to the aforementioned ROIs in order to build datasets with the attributes of each node and its malignancy. Finally, in the classification step, the datasets, which were built in the previous step, go through a pre-processing; after which the classifiers are trained and tested. These steps are detailed in the next section.
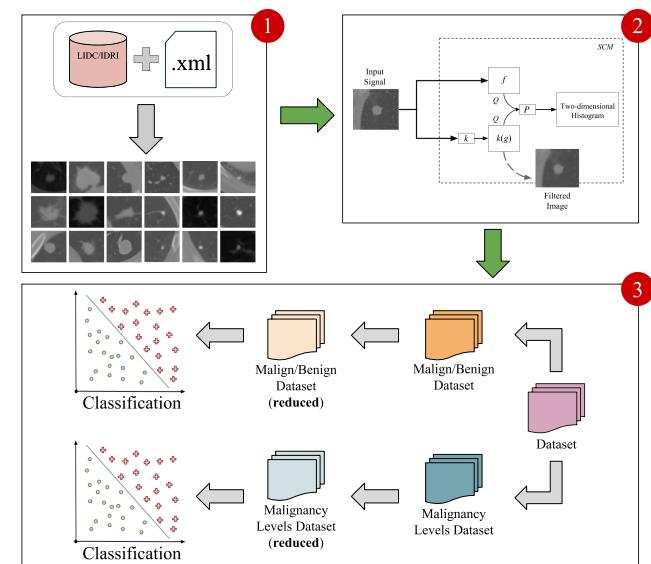
FIGURE 1. Flowchart of proposed approach: (1) ROIs Extraction; (2) Feature Extraction; (3) Malignancy Classification.

A. EXTRACTION OF THE ROIS

The LIDC/IDRI dataset [16] is composed of 1018 cases. Each case contains CT images of the chest and an associated XML file containing analytical data from four thoracic radiologists. In this dataset, 7371 injuries were labeled as "nodule" by at least one of the radiologists and 2669 of these injuries were defined as "a 3 mm nodule." Of these, 928 were marked 3mm by all four radiologists.

All data related to the nodules such as their coordinates within the exams and their malignancies are specified in the XML files. In order to obtain the ROI required in the feature extraction step, the centroid of each nodule is computed based on its coordinates. The centroid consists of three values x, y, and z, where the ordered pair (x, y) corresponds to the central position of the nodule in the slice z of the examination.
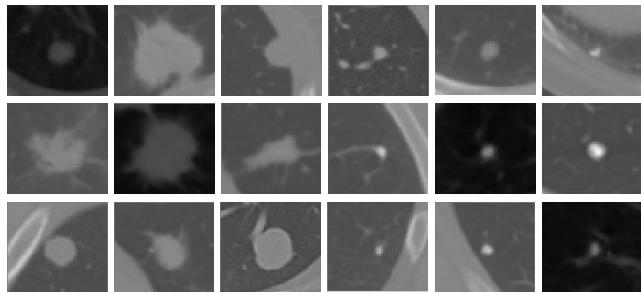
FIGURE 2. ROIs of malignant (first three columns) and benign (last three ones) nodules.

The ROI obtained from such a centroid consists of an image sized 49×49 centered in the ordered pair (x, y) (nodule center). Fig. 2 shows some ROIs of malignant and benign nodules.

B. FEATURE EXTRACTION TECHNIQUES

Feature extraction techniques aim to obtain features from an image or region of interest (ROI) and fully represent such images or ROI. In this work, feature extractors were applied to the ROIs that contain nodules and then the malignancy of the nodules is defined during the classification step by applying machine learning techniques. In this section, we briefly present the four feature extractors that were used to validate our proposal.

The gray-level co-occurrence matrix (GLCM) proposed by [17] analyzes the co-occurrence of gray intensities between related pairs of pixels by means of distance and direction. The GLCM results allow a set of attributes including texture, energy, contrast, correlation, among others to be computed.

The Local binary pattern (LBP) technique was initially proposed by [18] based on the work of [19].In LBP, a label (a binary number) is assigned to each pixel based on its neighborhood. The proposal in [18] is invariant to grayscale and variant to rotation. Nevertheless, an LBP version that is invariant to both grayscale and rotation can be found in [20].

Statistical moments (SM) are scalar magnitudes that describe the spatial distribution of the pixels of an image or a ROI. SM is one of the most commonly applied methods for image feature extraction. Central moments (CM) consist in moments centered on regions, thus [21] proposed the so called Hus Moments (HM) that are moments combining invariant to rotation, translation and scale.

1) STRUTUCAL CO-OCCURENCE MATRIX

The Structural Co-occurrence Matrix (SCM) technique is a method to analyze the relationship among low-level structures of two discrete signals in a n-dimensional space by mapping the co-occurrences among input signal structures in a two-dimensional histogram [12].

In the generic SCM model, consider two discrete signals $f \in D_f \subset \mathbb{C}^n$ and $g \in D_g \subset \mathbb{C}^n$ which have the same

number of samples in each dimension. In addition, consider a $k: D_g \rightarrow D_f$ that modifies the signal structures g. Such function intends to increase the structural differences between f and g. In the case of image applications, both f and g signs are grayscale images defined in $D_f \subset \mathbb{Z}^2$ with dimensions $m \times n$ where L is the number of levels.

The SCM is a matrix $M = \{m_{i,j}\}$ that stores the co-occurrence of the structures of the signals $Q(f)$ and $Q(k(g))$ as shown in Equation 1 [12]

$$m_{i,j} = \#\{(i,j)|P(i,j), i = Q(f_p), j = Q(k(g))_{p+d}\}, \quad (1)$$

where Q is a clustering algorithm of N classes or a quantization function of L to N levels. The goal of $Q$is to define how the low-level structures of the signals are organized into scales. The property P indicates whether the f and $k(g)$ signal intensities will be considered based on the similarity of their quantized signal intensities from f and $k(g)$. $\#\{\cdot\}$ represents the cardinality of the subset of pairs $(i,j) \in I \times I$, $I = \{0, \cdots, N-1\}$, that satisfies the property P; f_p means the sample value in the p position and $k(g)_{p+d}$ represents the sample value $k(g)$ given an offset d.

A set of scalar attributes computed from the SCM represent the structural information stored in M in a numerical format. The attributes are categorized into three groups: the statistical group, the information group and the divergence group [12]. The attributes of the statistical group focus on the structural level similarity stored in M. The information group comprehends an attribute that classifies the order of the data in M by measuring the randomness of the structural information. Finally, the divergence group describes the mutual information (dissimilarity) among the marginal distributions of M which take into account that its elements are independent variables.
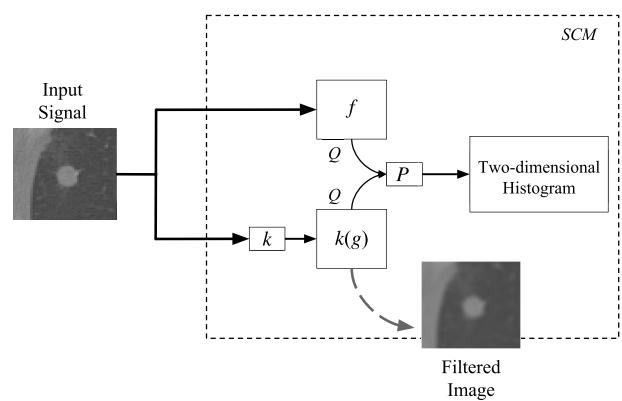
FIGURE 3. SCM model applied to extract image features.

C. NODULES FEATURE EXTRACTION

Since we already have the ROIs, the SCM feature extractor is applied and eight different datasets are generated. The SCM model used in our approach is depicted in Fig. 3. First, we have the nodules images represented as the input signal

and the signals f and g are two equivalent nodule images in grayscale and in Hounsfield Units (HU) defined in $D_f \subset \mathbb{Z}^2$ with $m \times n$ dimensions and L levels.

Then, consider a $k : D_g \rightarrow D_f$ that modifies the signal structures g. The Gaussian, Laplacian, Mean and Sobel filters were applied as the k function in order to find which provided the SCM extractor with the best performance. These low-pass (Gaussian and Mean) and high-pass (Laplacian and Sobel) filters were chosen since low pass filters soften the image details while high pass filters highlight the details. Thus, the SCM extractor behavior can be evaluate according to filter type.

D. MACHINE LEARNING TECHNIQUES
The aim of classifiers is to divide the feature space into regions that represent samples from a category [22]. In this work, the classifiers analyzed the features extracted from images of lung nodules and then identified weather a nodule was malignant or benign as well as its malignancy level. We conducted various experiments with well-known classification methods in order to select one to work with our proposal.

The multilayer perceptron (MLP) is a combination of single-layer perceptrons towards solving non-linear problems [22]. MLP is described by having at least one hidden layer between the input and output layer. The input layer is responsible for receiving the information to be analyzed, then the responses of one layer are passed on to the next layer as impulses that drive among linked neurons which are boosted by their corresponding weights. To conclude, an activation function in the output layer returns the results related to the input layer values [23].

The main goal of support vector machines (SVM) is to find an optimal separating hyperplane in which all samples from one class are set apart from the other class samples. SVM is based on the theory of statistical learning [24] that provides principles to improve the generalization ability of models [25].

K-nearest neighbors (k-NN) is a non-parametric classifier since it makes no assumption about the data distribution [22]. In the classification process, k-NN considers each training set element as a tuple with the respective label indicating its class. Each tuple is given as a point in an n-dimensional space, thus given a new tuple X with no label, k-NN computes the distances from X to all tuples in the training set and then assigns to X the most frequent class in the k closest tuples [22].

E. MALIGNANCY CLASSIFICATION
The LIDC/IDRI dataset [16] contains all the information concerning the nodules such as their coordinates within the exams and their malignancies specified in the XML files. The malignancy information consists of a value from 1 to 5 that indicates an increasing degree of malignancy. In this work, we conducted two experiments (*i*) a two-class problem (malignant and benign) in which the nodules with a malignancy value of ''3'' were disregarded and the nodules with a value of ''1'' or ''2 '' were labeled as benign and nodules valued ''4'' or ''5'' were labeled as malignant; (*ii*) we considered a 5-class problem in which the nodules were labeled according to their given malignancy values, that is, nodules having malignancy values equal to 1 were labeled as class 1, nodules having malignancy values equal to 2 were labeled as class 2, and so on.

Before training and testing the classifiers with the datasets generated by the SCM extractor, the number of samples was reduced in order to have a balanced dataset. This reduction was conducted by selecting the 500 samples closest to the mean value of each class. Thus, the reduced dataset in the binary classification problem consisted of 1000 samples divided into 500 samples labeled as malignant and 500 as benign. The reduced 5-class classification problem totaled 2500 samples in which each class consisted of 500 samples.

The reduced datasets underwent experiments with the following three classifiers: (*i*) MLP configured with an 8/20/2 neurons architecture (8 neurons in the input layer, 20 neurons in the single hidden layer and 2 neurons in the output layer), (*ii*) SVM with radial base function kernel (RBF) and, (*iii*) k-NN with k = 5. The MLP and SVM hyperparameters were chosen through a grid search technique with 10-folds cross-validation which finds the best parameters through an analysis of the results obtained from a combination of parameters. The grid search for SVM was conducted on $C \in [2^{-5}, 2^{-4}, 2^{-3}, \ldots, 2^{15}]$ and $\gamma \in [2^{-15}, 2^{-14}, \ldots, 2^3]$ hyperparameters [26]. The grid search for the MLP classifier architecture was used to find the number of hidden layers (HL) and neurons in each classifier. Moreover, we used the back-propagation algorithm as training with a learning rate of $\eta = 0.01$ and the momentum $m = 0.01$, and we defined the maximum number of iterations = 1000 or accuracy = 0.000001 as the stop criterion.

Each classifier was trained and tested 10 times. The results were given as the averages (10 iterations) for the following metrics: accuracy (ACC), sensitivity (True Positive Rate, TPR), specificity (SPC), positive predictive value (PPV), and F-Score. The ACC represents the ratio of samples correctly classified. TPR represents the proportion of positive samples correctly classified divided by the total number of positive samples. Similar to TPR, SPC is the proportion of negative samples correctly classified by the total number of negative samples. The PPV is the ratio of positive samples correctly classified and all samples classified as positive. The F-Score measure is the harmonic mean between PPV and TPR.

All experiments were conducted on an Intel Core i5 processor, 8GB RAM and MacOS X El Capitan 10.11.2 operating system. The extractors (Section II-B) and classifiers (Section II-D) were implemented using the OpenCV 3.0 library.

III. RESULTS AND DISCUSSION
This section presents and analyzes the experimental results obtained by applying the MLP, SVM and k-NN classifiers on

the two datasets generated by the SCM extractor: (*i*) binary classification problem with malignant and benign labels, and (*ii*) the 5-class classification problem with the 5 levels of malignancy. The results are analyzed in two stages, the first one makes a comparison among SCM configured with the Media, Gaussian, Laplacian and Sobel filters combined with the three aforementioned classifiers. In addition, the SCM is applied to the ROIs in grayscale and Hounsfield Units (HU). The goal in the first stage is to obtain the best combination of the input image, filter and classifier. In the second stage we compared SCM results with the extractors presented in Section II-B. The results of the classifiers and their runtimes were both taken into account.

A. BINARY CLASSIFICATION - BENIGN AND MALIGNANT
In this section we present the results and validate the SCM performance for the binary classification problem.

1) OPTIMAL FILTER AND CLASSIFIER TO USE WITH SCM IN PULMONARY NODULE MALIGNANCY CLASSIFICATION
Several tests with the SCM extractor were performed in order to find the best combination of input information, filter and classifier. The tests were performed in grayscale and HU images using four different filters in the SCM: Mean, Gaussian, Laplacian and Sobel filters.

The average results (10 iterations) of all combinations are presented in Table 1 on the basis of accuracy, sensitivity, specificity, positive predictive value, and F-Score.

TABLE 1. Average rate (10 iterations) of Accuracy (ACC), sensitivity (TPR), specificity (SPC), positive predictive value (PPV) and F-Score of SCM. The firsts and second higher values are in bold.

Extractor	Classifier	TPR(%)	SPC(%)	PPV (%)	F-Score(%)	ACC (%)
SCM Mean HU	MLP	**98,200**	92.600	92.992	**95.525**	**95.400**
	SVM	**98.200**	**95.200**	**96.748**	**96.700**	**96.700**
	k-NN	96.400	94.200	94.324	95.351	95.300
SCM Gaussian HU	MLP	96.200	79.800	82.646	88.909	88.000
	SVM	**98.800**	91.800	92.336	95.458	95.300
	k-NN	94.200	92.800	92.899	93.545	93.500
SCM Laplacian HU	MLP	20.000	80.000	50.000	28.571	50.000
	SVM	57.600	71.000	66.512	61.736	64.300
	k-NN	60.800	69.000	66.230	63.399	64.900
SCM Sobel HU	MLP	52.400	72.000	65.174	58.093	62.200
	SVM	77.200	82.000	81.092	79.098	79.600
	k-NN	78.400	74.200	75.239	76.787	76.300
SCM Mean Grayscale	MLP	90.000	10.000	50.000	64.285	50.000
	SVM	64.200	56.800	59.776	61.909	60.500
	k-NN	42.400	70.200	58.725	49.245	56.300
SCM Gaussian Grayscale	MLP	60.000	40.000	50.000	54.545	50.000
	SVM	61.200	59.200	60.000	60.594	60.200
	k-NN	48.000	66.800	59.113	52.980	57.400
SCM Laplacian Grayscale	MLP	75.800	59.400	65.120	70.055	67.600
	SVM	54.600	87.400	81.250	65.311	71.000
	k-NN	64.000	74.200	71.269	67.439	69.100
SCM Sobel Grayscale	MLP	83.600	65.800	70.967	76.767	74.700
	SVM	83.800	79.800	80.576	82.156	81.800
	k-NN	78.400	80.200	79.837	79.112	79.300

Table 1 shows that the SCM with the mean filter and SCM with Gaussian filter applied on the ROIs in HU were the only methods with ACC greater than 90%.

When the SCM was applied on the grayscale images, using the Sobel filter, it obtained an ACC value of ≈ 80% whereas when using the other filters the results were not satisfactory (≤ 71%). In addition, the results obtained using ROIs in HU were below 65% with the Laplace filter and below 80% with the Sobel filter.

The best results in Table 1 are for the SCM extractors with mean filter and Gaussian filter both applied to HU images. The SVM classifier, with the SCM Mean HU, achieved the highest TPR, PPV, F-Score and ACC values.

The experiments conducted with the low-pass filters (Gaussian and Mean) obtained better results when compared to the high-pass filters (Laplace and Sobel). Thus, the SCM extractor is able to obtain more relevant characteristics from nodule images when low-pass filters are applied, since these filters attenuate the details and remove noise from the original images. Therefore, an image with less noise and with less details guarantees a better SCM extractor performance.

The average runtime for training and testing of each classifier for each extractor and also the average extraction runtime of a sample are given in Table 2.

TABLE 2. Runtime average of training, test and extraction of SCM and classifiers. The firsts and second higher values are in bold.

Extractor	Classifier	Trainig Time (s)	Test Time (s)	Extraction Time (s)
SCM Mean HU	MLP	**1.757766±0.946278**	0.000001±0.000001	0.0480 ± 0.0041
	SVM	**2.029692±0.017843**	0.000136±0.000135	
	k-NN	0.000018±0.000002	0.003278±0.003246	
SCM Gaussian HU	MLP	2.154019±0.738326	0.000001±0.000001	**0.0440 ± 0.0046**
	SVM	**1.994519±0.035163**	0.000360±0.000356	
	k-NN	0.000028±0.000018	0.003151±0.003119	
SCM Laplacian HU	MLP	2.701015±0.170395	0.000001±0.000001	**0.0448 ± 0.0040**
	SVM	3.365795±0.493034	0.001086±0.001075	
	k-NN	0.000020±0.000002	0.003069±0.003039	
SCM Sobel HU	MLP	2.695329±0.726748	0.000001±0.000001	0.0457 ± 0.0051
	SVM	2.608690±0.068234	0.000589±0.000584	
	k-NN	0.000020±0.000003	0.003257±0.003224	
SCM Mean Grayscale	MLP	2.225363±0.732661	0.000001±0.000001	0.0488 ± 0.0082
	SVM	2.739410±0.168567	0.000938±0.000929	
	k-NN	**0.000017±0.000002**	**0.002090±0.002069**	
SCM Gaussian Grayscale	MLP	3.009984±0.419738	0.000001±0.000001	0.0460 ± 0.0084
	SVM	2.800269±0.257210	0.000973±0.000964	
	k-NN	0.000020±0.000002	0.002575±0.002549	
SCM Laplacian Grayscale	MLP	2.486071±0.062094	0.000001±0.000001	0.0469 ± 0.0084
	SVM	2.797863±0.022985	0.000851±0.000842	
	k-NN	**0.000018±0.000002**	0.002146±0.002125	
SCM Sobel Grayscale	MLP	2.548915±0.113428	0.000001±0.000001	0.0477 ± 0.0087
	SVM	2.515292±0.024407	0.000597±0.000591	
	k-NN	0.000019±0.000004	**0.002133±0.002112**	

The average runtimes of the training and testing steps are very similar among the extractors. Among the classifiers, the *k*-NN achieved the fastest average runtime for training and the slowest average runtime for testing. The MLP and SVM classifiers achieved the shortest training runtime values. Among the classifiers which obtained the best metrics values, the SCM Sobel Grayscale extractor with MLP classifier achieved the slowest training average runtime = 2.55 seconds whereas the SCM Mean HU extractor with *k*-NN classifier achieved the slowest test average runtime = 0.003 seconds.

Related to the extraction runtimes, all the SCM variations reached similar average runtime values which were less than 0.05 seconds. seconds. Therefore, when the SCM is combined with the right filter, it is able to extract relevant characteristics from the pulmonary nodule images and the experimental classification results proved that such extracted features produced high accuracy rates as well as fast runtimes.

2) COMPARISON WITH FEATURE EXTRACTORS REPORTED IN THE LITERATURE
The analysis performed in Section III-A.1, pointed out that the SCM Mean HU was the best approach to extract nodule features among the SCM extractor alternatives. In this section, this approach is compared with the GLCM, LBP,

Central Moments and Statistical Moments extractors. They were applied using the same images and their extracted features were validated by the same classifiers.

The goals here are to show that the SCM Mean HU extractor outperforms the other extractors commonly used in the literature. The experimental results are presented in Table 3 and cover the ACC, SPC, PPV and F-Score mean values of 10 iterations of each combined extractor to each classifier.

TABLE 3. Average rate (10 iterations) of Accuracy (ACC), sensitivity (TPR), specificity (SPC), positive predictive value (PPV) and F-Score of SCM, GLCM, LBP, CM and SM. The firsts and second higher values are in bold.

Extractor	Classifier	SPC(%)	TPR(%)	PPV (%)	F-Score(%)	ACC (%)
SCM Mean HU	MLP	**98.200**	92.600	92.992	**95.525**	95.400
	SVM	**98.200**	**95.200**	95.339	**96.748**	**96.700**
	k-NN	**96.400**	**94.200**	94.324	95.351	95.300
GLCM	MLP	51.000	62.000	57.303	53.968	56.500
	SVM	64.600	72.800	70.370	67.361	68.700
	k-NN	63.800	70.060	68.400	66.045	67.200
LBP	MLP	68.554	79.918	78.173	73.048	74.100
	SVM	80.800	80.600	80.638	80.719	80.700
	k-NN	82.200	65.600	70.497	75.900	73.900
CM	MLP	80.761	51.297	62.287	70.331	66.000
	SVM	84.600	64.000	70.149	76.699	74.300
	k-NN	78.200	67.200	70.450	74.123	72.700
SM	MLP	83.000	55.800	65.251	73.063	69.400
	SVM	79.200	75.600	76.447	77.799	77.400
	k-NN	67.800	76.400	74.179	70.846	72.100

Observing the results presented in Table 3 the LBP was the only one which obtained the closest ACC value to the SCM, with ≈ 80%. As one can see, the SCM extractor outperformed all the other results. These superior results prove the efficiency of our approach when applied to pulmonary nodule malignancy classification.

TABLE 4. Time of training, testing and extraction of SCM, GLCM, LBP, CM, SM and classifiers. The firsts and second higher values are in bold.

Extractor	Classifier	Training Time (s)	Test Time (s)	Extraction Time (s)
SCM Mean HU	MLP	**1.757766±0.946278**	0.000001±0.000001	
	SVM	2.029692±0.017843	**0.000136±0.000135**	0.0480 ± 0.0041
	k-NN	0.000018±0.000002	0.003278±0.003246	
GLCM	MLP	**2.270910±0.867463**	0.000001±0.000001	
	SVM	2.967113±0.024136	0.000868±0.000859	6.0877 ± 1.0365
	k-NN	0.000023±0.000004	0.003399±0.003365	
LBP	MLP	44.048176±1.032490	0.000001±0.000001	
	SVM	16.411175±0.137494	0.005731±0.005673	0.1747 ± 0.0427
	k-NN	0.000116±0.000070	0.010815±0.010707	
CM	MLP	2.420326±0.023988	0.000001±0.000001	
	SVM	**2.731912±0.025732**	**0.000733±0.000726**	0.0053 ± 0.0022
	k-NN	**0.000016±0.000001**	0.002059±0.002038	
SM	MLP	2.621801±0.024420	0.000001±0.000001	
	SVM	3.152975±0.028960	0.000804±0.000795	0.0054 ± 0.0019
	k-NN	0.000018±0.000001	**0.002328±0.002304**	

The Table 4 presents the average runtimes of extraction, training and testing. The lowest average training runtimes of the MLP and SVM classifiers were obtained with the information from the SCM Mean HU extractor. The lowest average test runtime using the SVM classifier was also obtained by our approach. However, the lowest average runtime with the *k*-NN classifier was achieved by the CM extractor, followed by SM. CM and SM extractors reached the lowest average extraction runtime per sample of ≈ 0.0054 seconds, followed by SCM Mean HU.

Despite the SM and CM extractors having the lowest extraction runtimes, the SCM Mean HU still represents a better approach, since the average accuracy rate obtained reached 96.7%.

Therefore, our approach, the SCM Mean HU extractor, achieved not only good metrics but it also outperformed the runtime of the GLCM, LBP, CM and SM extractors, which are considered standard approaches in literature.

The average confusion matrix obtained from the 10 iterations of the SCM Mean HU, GLCM, LBP, CM and SM with the classifiers is presented in Table 5. This average confusion matrix shows the low number of false positives and false negatives as well as the high number of samples classified correctly of SCM.

Observing the Table 5, we highlight the low number of misclassification with the SCM extractor. The low number of false positives and false negatives is justified by the structural analysis from SCM that is capable of highlighting relevant information of the images, so the classification stage turns more efficient. In addition, it can be highlighted that the SCM extractor achieved higher true positive and true negative rates among analyzed approaches.

3) COMPARISON WITH OTHER APPROACHES FOR PULMONARY NODULE CLASSIFICATION

In Section III-A.1, the SCM with mean filter applied to images in Hounsfield Units, called SCM Mean HU, was selected as the best approach among the alternative extractors, especially when its extracted features were applied as input to an SVM classifier. In Section III-A.2, the SCM Mean HU was compared to other extractors commonly used in the literature (GLCM, LBP, CM and SM). This comparison was conducted by applying all extractors to the same input images, in order to prove the efficiency of the SCM Mean HU.

In this section, the results of the SCM Mean HU are compared to the results of the other proposed approaches for pulmonary nodule malignancy classification. It should be mentioned that the information concerning the nodules is from the same dataset LIDC/IDRI, but the images provided to the extractors are different, since the method of selecting the ROIs that contain the nodules varies among the authors.

The Table 6 presents the accuracy values of both our approach and the related works. In some of these studies, comparisons were also made with other methods.

The accuracy values presented in Table 6 show that the SCM Mean HU combined with the SVM classifier is an efficient approach to classify pulmonary nodules. This approach obtained an accuracy of 96.7% which is higher than the values reported in the related works.

In addition to presenting the highest accuracy value, the proposed approach used a balanced database with a large number of images in which 500 samples were malignant nodules and 500 samples were benign nodules.

In the literature, some works used a low number of nodule images, which may hamper the method generalization ability. In other studies, which used a considerable number of nodule images, there are an unbalanced base with a larger number of benign nodules.

TABLE 5. Numerical average confusion matrix for SCM Mean HU, GLCM, LBP, CM and SM.

True Class	Classified As	SCM Mean HU			GLCM			LBP			CM			SM		
		MLP	K-NN	SVM	MLP	K-NN	SVM	MLP	K-NN	SVM	MLP	K-NN	SVM	MLP	K-NN	SVM
Benign	Benign	49.10	48.20	49.10	25.50	31.90	32.30	35.10	41.10	40.40	40.30	39.10	42.30	41.50	33.90	39.60
	Malignant	0.90	1.80	0.90	24.50	18.10	17.70	16.10	8.90	9.60	9.60	10.90	7.70	8.50	16.10	10.40
Maligant	Benign	3.70	2.90	2.40	19.00	14.70	13.60	9.80	17.20	9.70	24.40	16.40	18.00	22.10	11.80	12.20
	Malignant	46.30	47.10	47.60	31.00	35.30	36.40	39.00	32.80	40.30	25.70	33.60	32.00	27.90	38.20	37.80

TABLE 6. Related works with the best result in bold.

Approach	Dataset	#Nodules	Accuracy
SCM Mean HU	LIDC/IDRI	1000	**96.70%**
Krewer *et al.*[9]	LIDC/IDRI	33	90.91%
Shen *et al.* [10]	LIDC/IDRI	1375	87.14%
Shen *et al.* [6]	LIDC/IDRI	1375	86.84%
Sergeeva *et al.* [11]	LIDC	321	81.3%
Kumar *et al.* [5]	LIDC	4323	75.01%

B. MULTICLASS CLASSIFICATION - LEVEL OF MALIGNANCY

This section aims to present and validate the results obtained in applying the SCM Mean HU extractor on malignancy level classification of nodules, considering the 5 levels of malignancy according to the LIDC/IDRI dataset.

1) OPTIMAL FILTER AND CLASSIFIER TO USE WITH SCM IN PULMONARY NODULE MALIGNANCY CLASSIFICATION

In this section, we present the results obtained with the MLP, SVM and k-NN classifiers trained and tested on the datasets generated by the SCM extractor. Similarly to Section III-A.1, the SCM extractor is used with the Mean, Gaussian, Laplacian, and Sobel filters. In addition, it is also applied on the HU and grayscale images. Table 7 presents the ACC, TPR, SPC, PPV and F-Score average values of 10 iterations for each classifier with the SCM combined with a specific filter as well as the input image type. The best results are highlighted in bold.

TABLE 7. Average rate (10 iterations) of Accuracy (ACC), sensitivity (TPR), specificity (SPC), positive predictive value (PPV) and F-Score of SCM. The firsts and second higher values are in bold.

Extractor	Classifier	SPC(%)	TPR(%)	PPV (%)	F-Score(%)	ACC (%)
SCM Mean HU	MLP	66.08	38.91	31.72	30.72	55.64
	SVM	79.33	52.78	48.40	49.01	71.00
	k-NN	78.76	48.57	47.48	47.57	69.87
SCM Gaussian HU	MLP	64.93	37.69	31.04	29.45	54.51
	SVM	78.64	**55.18**	47.40	48.40	70.25
	k-NN	79.66	50.38	49.08	49.18	71.08
SCM Laplacian HU	MLP	65.22	31.73	31.39	30.40	54.01
	SVM	75.56	44.29	43.12	42.72	65.82
	k-NN	76.43	44.91	44.52	44.53	66.97
SCM Sobel HU	MLP	68.21	37.73	34.12	31.80	56.74
	SVM	79.54	48.99	48.40	47.94	70.63
	k-NN	76.39	46.27	44.87	44.88	67.19
SCM Average Grayscale	MLP	57.50	33.56	24.80	20.05	47.30
	SVM	**82.35**	54.27	52.68	52.66	**74.45**
	k-NN	**82.42**	53.76	**52.80**	52.82	74.43
SCM Gaussian Grayscale	MLP	73.30	43.55	39.95	40.58	63.79
	SVM	82.03	54.21	51.76	50.58	73.60
	k-NN	82.23	**54.78**	**52.96**	**53.20**	74.36
SCM Laplacian Grayscale	MLP	64.06	31.41	29.48	26.66	52.54
	SVM	75.51	41.41	42.68	41.16	65.30
	k-NN	72.17	38.72	39.00	38.69	61.71
SCM Sobel Grayscale	MLP	64.03	31.53	30.27	28.32	52.69
	SVM	79.81	49.38	48.28	46.75	70.53
	k-NN	79.17	49.03	48.44	48.60	70.46

For the 5-class (levels of malignancy) classification problem, the settings of filter and input image type that obtained the best results were the Mean and Gaussian filter applied on the grayscale images (SCM Mean Grayscale and SCM Gaussian Grayscale). The SCM Mean Grayscale achieved the highest accuracy and specificity when its results were used in the SVM classifier. On the other hand, the SCM Gaussian Grayscale reached the highest values for TPR, PPV and F-Score when used with the *k*-NN classifier.

In the 5-class classification problem, the F-Score values were all less than 50% in the SCM applications on the HU images unlike the binary classification results presented in Section III-A.1. While for the grayscale images, only combinations with the Mean and Gaussian filters obtained F-Score values greater than 50%, except when used with the MLP classifier.

The training, test and extraction information (average per sample) runtimes are presented in Table 8.

TABLE 8. Mean time of training, test and extraction of SCM and classifiers. The firsts and second higher values are in bold.

Extractor	Classifier	Trainig Time (s)	Test Time (s)	Extraction Time (s)
SCM Mean HU	MLP	7.119540±0.103273	0.000001±0.000001	
	SVM	25.281826±0.067709	0.006391±0.006366	0.048125±0.006684
	k-NN	**0.000030±0.000001**	0.011761±0.011714	
SCM Gaussian HU	MLP	7.145532±0.081090	0.000001±0.000001	
	SVM	25.425169±0.122533	0.006283±0.006258	**0.044286±0.005243**
	k-NN	0.000032±0.000008	0.011733±0.011686	
SCM Laplacian HU	MLP	7.977795±1.109689	0.000001±0.000001	
	SVM	26.266043±0.148892	0.006566±0.006540	0.045146±0.005200
	k-NN	0.000039±0.000024	0.012832±0.012780	
SCM Sobel HU	MLP	7.158446±0.096704	0.000001±0.000001	
	SVM	25.464731±0.117834	0.006146±0.006122	0.046031±0.005642
	k-NN	**0.000029±0.000001**	0.011495±0.011449	
SCM Mean Grayscale	MLP	7.180762±0.084817	0.000001±0.000001	
	SVM	**24.295231±0.085797**	**0.005814±0.005791**	0.047872±0.007061
	k-NN	0.000031±0.000004	**0.011476±0.011431**	
SCM Gaussian Grayscale	MLP	6.728332±1.267050	0.000001±0.000001	
	SVM	**24.522326±0.083344**	0.006039±0.006014	**0.045036±0.007093**
	k-NN	**0.000029±0.000001**	0.011559±0.011513	
SCM Laplacian Grayscale	MLP	7.080100±0.065734	0.000001±0.000001	
	SVM	25.998016±0.084007	0.006545±0.006519	0.045814±0.007002
	k-NN	**0.000030±0.000002**	0.011703±0.011656	
SCM Sobel Grayscale	MLP	6.616489±1.557682	0.000001±0.000001	
	SVM	25.129902±0.048359	0.006191±0.006166	0.046792±0.007900
	k-NN	**0.000030±0.000001**	0.011334±0.011289	

The training and testing runtimes as well as the extraction runtimes are all very similar. In order to validate the SCM application in the 5-class problem, the results of this later extractor were compared with those of GLCM, LBP, CM and SM in Section III-B.2.

2) COMPARISON WITH FEATURES EXTRACTORS IN THE LITERATURE

In order to make a comparison with the GLCM, LBP, CM and SM extractors, the best results obtained from the SCM (SCM Mean Grayscale and SCM Gaussian Grayscale) were selected.

Table 9 presents the ACC, TPR, SPC, PPV and F-Score average values of 10 iterations for each extractor and classifier. The best results are in bold. The results show that the

TABLE 9. Average rate (10 iterations) of Accuracy (ACC), sensitivity (TPR), specificity (SPC), positive predictive value (PPV) and F-Score of SCM, GLCM, LBP, CM and SM. The firsts and second higher values are in bold.

Extractor	Classifier	SPC(%)	TPR(%)	PPV (%)	F-Score(%)	ACC (%)
SCM Mean Grayscale	MLP	57.50	33.56	24.80	20.05	47.30
	SVM	**82.35**	**54.27**	52.68	52.66	**74.45**
	k-NN	**82.42**	53.76	**52.80**	52.82	74.43
SCM Gaussian Grayscale	MLP	73.30	43.55	39.95	40.58	63.79
	SVM	82.03	54.21	51.76	50.58	73.60
	k-NN	82.23	**54.78**	**52.96**	**53.20**	74.36
GLCM	MLP	58.11	26.13	25.64	24.54	46.69
	SVM	72.12	38.53	38.72	37.52	61.54
	k-NN	73.29	39.48	40.24	39.54	62.88
LBP	MLP	57.12	32.86	24.39	20.58	46.92
	SVM	79.56	48.66	48.60	48.59	70.84
	k-NN	74.84	44.91	42.36	42.53	65.20
CM	MLP	60.03	28.14	27.05	25.49	48.92
	SVM	74.37	40.05	38.80	34.69	61.81
	k-NN	72.81	39.46	39.68	38.94	62.31
SM	MLP	61.57	30.29	28.40	27.29	50.42
	SVM	74.89	41.59	42.00	40.77	64.78
	k-NN	77.28	45.28	45.40	44.88	67.83

SCM, even with values lower than those achieved in binary classification, still achieved the highest values in all metrics. The highest F-Score obtained by the GLCM, LBP, CM and SM methods was 48.59% which was from.

The results of the average training and testing runtimes as well as the extraction runtime (average for one sample) are presented in Table 10. The CM and SM extractors obtained the fastest extraction runtimes, followed by the SCM. The SCM reached the fastest training and testing runtimes only when combined with the SVM classifier. The *k*-NN classifier runtimes were close to the SCM, CM and SM results. For MLP, the fastest training runtimes were achieved by the SCM and CM extractors.

TABLE 10. Time of training, test and extraction of SCM, GLCM, LBP, CM, SM and classifiers. The firsts and second higher values are in bold.

Extractor	Classifier	Training Time (s)	Test Time (s)	Extraction Time (s)
SCM Mean Grayscale	MLP	7.180762±0.084817	0.000001±0.000001	
	SVM	**24.295231±0.085797**	**0.005814±0.005791**	0.047872±0.007061
	k-NN	0.000031±0.000004	0.011476±0.011431	
SCM Gaussian Grayscale	MLP	**6.728332±1.267050**	0.000001±0.000001	
	SVM	**24.522326±0.083344**	**0.006039±0.006014**	0.045036±0.007093
	k-NN	0.000029±0.000001	0.011559±0.011513	
GLCM	MLP	7.248782±0.076979	0.000001±0.000001	
	SVM	28.660223±0.061428	0.007197±0.007168	6.0137±0.8831
	k-NN	0.000034±0.000002	0.011990±0.011942	
LBP	MLP	117.122818±0.297861	0.000001±0.000001	
	SVM	164.162704±0.520061	0.043420±0.043246	0.1691±0.0384
	k-NN	0.000251±0.000116	0.047354±0.047165	
CM	MLP	**4.946586±2.802100**	0.000001±0.000001	
	SVM	27.855473±0.067491	0.007331±0.007301	**0.0053±0.0026**
	k-NN	0.000033±0.000001	0.012448±0.012398	
SM	MLP	7.355840±0.667804	0.000001±0.000001	
	SVM	31.248348±0.079760	0.007580±0.007550	**0.0061±0.0021**
	k-NN	0.000034±0.000001	0.013036±0.012984	

Table 9 and 10 show that the SCM extractor even though it does not achieved results as high as in the binary classification, is still the best solution among the methods compared here to handle the classification problem of nodule malignancy levels.

IV. CONCLUSION

This work presented a new approach to classify pulmonary nodules into malignant and benign nodules. This approach applied the SCM extractor combined with a classifier.

To find the best SCM design, eight combinations of SCM and classifier were tested, each one using a filter and a specific input image type. The filters used were: Mean, Gaussian, Laplacian and Sobel, while the input images used were

grayscale and Hounsfield Units. The best results obtained with the SCM extractor were compared with the GLCM, LBP, CM and SM extractors. All extractor tests were performed with the MLP, SVM and *k*-NN classifiers.

After all the comparisons had been analyzed, the best results for nodule classification into malignant and benign were obtained by the SCM extractor configured with the Mean filter and applied to HU images combined with the SVM classifier. All experiments were conducted with the well-known LIDC/IDRI dataset.

In order to prove the efficiency of our proposed approach, the SCM extractor was also used in the classification of nodule malignancy levels for the 5-class classification problem. As in the binary classification, eight variations of the extractor were tested using the same filters and classifiers. The SCM with the Mean and Gaussian filter both applied in grayscale images were the ones that achieved the best results in the comparison among SCM variations as well as in comparison with the GLCM, LBP, CM and SM extractors.

These results validated our SCM extractor approach in handling the malignant classification of pulmonary nodules, and therefore can be of assistance for making medical diagnoses more precise and faster in CAD systems.

Future works should be directed at improving the SCM extractor for classifying malignancy levels of nodules through experiments with various alternatives.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA, Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, Jan. 2016.

[2] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. Tavares, "Automatic 3D pulmonary nodule detection in CT images: A survey," *Comput. Methods Programs Biomed.*, vol. 124, pp. 91–107, Feb. 2016.

[3] N. Camarlinghi, "Automatic detection of lung nodules in computed tomography images: Training and validation of algorithms using public research databases," *Eur. Phys. J. Plus*, vol. 128, no. 9, p. 110, Sep. 2013.

[4] F. Han *et al.*, "A texture feature analysis for diagnosis of pulmonary nodules using LIDC-IDRI database," in *Proc. IEEE Int. Conf. Med. Imag. Phys. Eng. (ICMIPE)*, Oct. 2013, pp. 14–18.

[5] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *Proc. 12th Conf. Comput. Robot Vis. (CRV)*, Jun. 2015, pp. 133–138.

[6] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *Proc. 24th Int. Conf. Inf. Process. Med. Imag.*, Jun./Jul. 2015, pp. 588–599.

[7] P. P. Rebouças Filho, R. M. Sarmento, G. B. Holanda, and D. de Alencar Lima, "New approach to detect and classify stroke in skull CT images via analysis of brain tissue densities," *Comput. Methods Programs Biomed.*, vol. 148, pp. 27–43, Sep. 2017.

[8] P. P. Rebouças Filho, E. D. S. Rebouças, L. B. Marinho, R. M. Sarmento, J. M. R. Tavares, and V. H. C. de Albuquerque, "Analysis of human tissue densities: A new approach to extract features from medical images," *Pattern Recognit. Lett.*, vol. 94, pp. 211–218, Jul. 2017.

[9] H. Krewer *et al.*, "Effect of texture features in computer aided diagnosis of pulmonary nodules in low-dose computed tomography," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2013, pp. 3887–3891.

[10] W. Shen *et al.*, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognit.*, vol. 61, pp. 663–673, Jan. 2017.

[11] M. Sergeeva, I. Ryabchikov, M. Glaznev, and N. Gusarova, "Classification of pulmonary nodules on computed tomography scans. Evaluation of the effectiveness of application of textural features extracted using wavelet transform of image," in *Proc. 18th Conf. Open Innov. Assoc. Seminar Inf. Secur. Protection Inf. Technol. (FRUCT-ISPIT)*, Apr. 2016, pp. 291–299.

[12] G. L. B. Ramalho, D. S. Ferreira, P. P. R. Filho, and F. N. S. de Medeiros, "Rotation-invariant feature extraction using a structural co-occurrence matrix," *Measurement*, vol. 94, pp. 406–415, Dec. 2016.

[13] P. P. Rebouças Filho, P. C. Cortez, A. C. da Silva Barros, V. H. C. Albuquerque, and J. M. R. Tavares, "Novel and powerful 3D adaptive crisp active contour method applied in the segmentation of CT lung images," *Med. Image Anal.*, vol. 35, pp. 503–516, Jan. 2017.

[14] P. P. Rebouças Filho *et al.*, "Automated recognition of lung diseases in CT images based on the optimum-path forest classifier," in *Neural Computing and Applications*. London, U.K.: Springer, 2017, pp. 1–14. [Online]. Available: https://link.springer.com/article/10.1007/s00521-017-3048-y#citeas

[15] P. P. Rebouc, R. M. Sarmento, P. C. Cortez, and V. H. C. De, "Adaptive crisp active contour method for segmentation and reconstruction of 3D lung structures," *Int. J. Comput. Appl.*, vol. 111, no. 4, pp. 1–8, 2015.

[16] S. G. Armato, III, *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, 2011.

[17] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[18] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proc. IAPR Int. Conf. Pattern Recognit. (ICPR)*, vol. 1. 1994, pp. 582–585.

[19] L. Wang and D.-C. He, "Texture classification using texture spectrum," *Pattern Recognit.*, vol. 23, no. 8, pp. 905–910, 1990.

[20] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[21] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.

[22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NY, USA: Wiley, 2012.

[23] S. S. Haykin, *Neural Networks and Learning Machines*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2008, vol. 3.

[24] V. N. Vapnik, *Statistical Learn Theory*, vol. 1. New York, NY, USA: Wiley, 1998.

[25] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2013.

[26] C.-W. Hsu, C.-C. Chang, and C. Lin, "A practical guide to support vector classification," Dept. Comput. Sci. Inf. Eng., Nat. Taiwan Univ., Taipei, Taiwan, Tech. Rep., 2003. [Online]. Available: https://www.bibsonomy.org/bibtex/2c04ef97dc3c3de168e684c3e4abe061b/zeno and https://www.csie.ntu.edu.tw/~cjlin/papers.html

RAUL VICTOR M. DA NÓBREGA received the bachelor's and master's and degrees in computer science from the Federal Institute of Education, Science and Technology of Cear, in 2016 and 2018, respectively. He is currently a member of the Laboratório de Processamento de Imagens e Simulação Computacional. He has research in the area of classification of pulmonary nodules through techniques of digital image processing and artificial intelligence.

SHARA SHAMI A. ALVES received the bachelor's and master's degrees in computer science from the Federal Institute of Education, Science and Technology of Cear, in 2015 and 2017, respectively. She is currently a member of the Laboratório de Processamento de Imagens e Simulação Computacional. He has research in the area of classification of pulmonary nodules through techniques of digital image processing and artificial intelligence.

PEDRO PEDROSA REBOUÇAS FILHO received the Ph.D. degree in teleinformatics engineering from the Federal University of Ceará, Fortaleza, Brazil, in 2013. He is currently a Professor with the Instituto Federal de Educação, Maracanau, CE, Brazil. His current research interest is applications in computational vision, mainly using medical images.

MURILLO B. RODRIGUES received the bachelor's and master's degrees in computer science from the Federal Institute of Education, Science and Technology of Cear. He is currently a member of the Laboratório de Processamento de Imagens e Simulação Computacional. He has research in the area of classification of pulmonary nodules through techniques of digital image processing and artificial intelligence.

JOÃO BATISTA F. DUARTE received the B.Sc. degree in physics from the Universidade Federal de Santa Maria, RS, Brazil, in 1993, and the M.Sc. and Ph.D. degrees in physics the Federal University of Ceará, CE, Brazil, in 1995 and 1998, respectively. He has been a Professor with the Graduate Program in Applied Informatics, Universidade de Fortaleza, since 2001. His research interests include heat transference, thermodynamic and thermal processes, acting on the following subjects: combustion, porous media, gaseous fuels, combustion, and biomass.

ARUN K. SANGAIAH received the M.E. degree from Anna University in 2007 and the Ph.D. degree from the Vellore Institute of Technology in 2014. He was a Visiting Professor with the School of Computer Engineering, Nanhai Dongruan Information Technology Institute in China from 2016 to 2017. He is currently an Associate Professor with the School of Computing Science and Engineering, Vellore Institute of Technology, Vellore, India. He has authored over 130 scientific papers in high standard SCI journals like the IEEE-TII, the IEEE-*Communication Magazine*, the IEEE Systems Journal, the IEEE-IoT, the IEEE TSC, the IEEE ETC, and so on. In addition he has authored/edited over eight books (Elsevier, Springer, Wiley, Taylor and Francis) and 50 journal special issues such as the IEEE-*Communication Magazine*, the IEEE-IoT, the IEEE *Consumer Electronic Magazine*, and so on. His area of interest includes software engineering, computational intelligence, wireless networks, bio-informatics, and embedded systems. Also, he was registered a one Indian patent in the area of computational intelligence. Besides, he is responsible for Editorial Board Member/Associate Editor of various international SCI journals.

VICTOR HUGO C. DE ALBUQUERQUE received the Ph.D. degree in mechanical engineering with emphasis on materials from the Federal University of Paraíba in 2010, the M.Sc. degree in teleinformatics engineering from the Federal University of Ceará in 2007, and the degree in mechatronics technology from the Federal Center of Technological Education of Cear in 2006. He is currently an Assistant VI Professor with the Graduate Program in Applied Informatics, Universidade de Fortaleza. He has experience in computer systems, mainly in the research fields of: applied computing, intelligent systems, visualization and interaction, with specific interest in pattern recognition, artificial intelligence, image processing and analysis, and automation with respect to biological signal/image processing, image segmentation, biomedical circuits and human/brain-machine interaction, including augmented and virtual reality simulation modeling for animals and humans. Additionally, he has research at the microstructural characterization field through the combination of non-destructive techniques with signal/image processing and analysis, and pattern recognition.

● ● ●