Mathematical Biosciences and Engineering

MBE, 18(4): 4146–4161.
DOI: 10.3934/mbe.2021208
Received: 15 March 2021
Accepted: 06 May 2021
Published: 12 May 2021

http://www.aimspress.com/journal/MBE

Research article

Intelligent immune clonal optimization algorithm for pulmonary nodule classification

Qi Mao[1,2,*], Shuguang Zhao[2], Lijia Ren[1], Zhiwei Li[1], Dongbing Tong[1], Xing Yuan[3] and Haibo Li[1]

[1] School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
[2] College of Information Science and Technology, Donghua University, Shanghai 201620, China
[3] ABB Electrical Machines Ltd

* **Correspondence:** Email: qimao@sues.edu.cn.

Abstract: Computer-aided diagnosis (CAD) of pulmonary nodules is an effective approach for early detection of lung cancers, and pulmonary nodule classification is one of the key issues in the CAD system. However, CAD has the problems of low accuracy and high false-positive rate (FPR) on pulmonary nodule classification. To solve these problems, a novel method using intelligent immune clonal selection and classification algorithm is proposed and developed in this work. First, according to the mechanism and characteristics of chaotic motion with a logistic mapping, the proposed method utilizes the characteristics of chaotic motion and selects the control factor of the optimal chaotic state, to generate an initial population with randomness and ergodicity. The singleness problem of the initial population of the immune algorithm was solved by the proposed method. Second, considering on the characteristics of Gaussian mutation operator (GMO) with a small scale, and Cauchy mutation operator (CMO) with a big scale, an intelligent mutation strategy is developed, and a novel control factor of the mutation is designed. Therefore, a Gauss-Cauchy hybrid mutation operator is designed. Ultimately, in this study, the intelligent immune clonal optimization algorithm is proposed and developed for pulmonary nodule classification. To verify its accuracy, the proposed method was used to analyze 90 CT scans with 652 nodules. The experimental results revealed that the proposed method had an accuracy of 97.87% and produced 1.52 false positives per scan (FPs/scan), indicating that the proposed method has high accuracy and low FPR.

Keywords: computer-aided diagnosis (CAD); pulmonary nodule; immune clonal selective; feature selection; classification

1. Introduction

Lung cancer is one of the highest rates of morbidity and mortality in patients among the worldwide, and an effective approach for lung nodule detection and classification is very important for the early computer-aided diagnosis of lung cancer [1,2]. In this regard, the methods based on artificial intelligence have been proposed and developed, such as artificial neural network [3], evolutionary computation [4], and artificial immune system (AIS) [5]. Notably, AIS is inspired by human immune system, which has the characteristics of adaptability, robustness, homeostasis, memory, and immunity. Inspired by the human immune system, some new modes, algorithms, technologies and theoretical understandings were presented in the literation [6–8]. Feature selection is one of the most important tasks in many engineering problems. Importantly, AIS has a better performance in feature selection, which has the advantages of global optimization, parallel computing, noise tolerance, etc. It is widely applied in scientific computation and solving engineering problems of the existing literatures [8–10]. The feature selection algorithms based on immune-inspired have the advantage of parallel computing in multi-dimensional space.

Computer-aided lung cancer diagnosis mainly includes lung parenchyma extraction, the region of interest segmentation, the feature extraction and classification. In the image processing, feature selection and classification are important issues for the early detection of lung cancers [11–14]. In the past few years, many different models and approaches were designed for feature selection and classification, each with its own strengths and weaknesses. For instance, the method of kernel optimization was used for feature selection in the high-dimensional data, experimental results demonstrated the efficiency and accuracy [15]. The fuzzy clustering and hybrid intelligent methods were used for the diagnosis of pulmonary nodules [16]. In addition, Froz et al. [4] used an expert system for lung nodule classification, obtained an accuracy of 94.3% and a specificity of 94.78%. Lu et al. [5] used the genetic algorithm for feature selection, and support vector machine (SVM) for pulmonary nodule classification, obtained an accuracy of 99%. Similarly, Deep learning methods also have good performance in feature selection and classification. For example, the deep reinforcement learning method was adopted for lung nodule detection and obtained the accuracy of 64.4% [17]. In order to improve the accuracy of classification, the deep learning and swarm intelligence of computational intelligence techniques were used to develop types of nodule detection and classification, which achieved up to 93.71% accuracy [18]. In other respects, a novel convolutional neural networks with carefully designed loss function regularization was proposed for thyroid nodule detection in the ultrasound images, the proposed model can detect various types of thyroid nodules [19].

From the analysis of literature, it was observed that most of the studies thought that feature selection and classification were the independent processes, whereupon the feature selection and classification were done separately. Although the process is the local optimization, the final results are not global optimum. Actually, the processes are closely related and mutually influential, which could be performed simultaneously. In this regard, many methods have been proposed. For example, the genetic algorithm was used for feature selection and the SVM parameter optimization [20]. But genetic algorithm has an issue of the degradation phenomenon in the solving process of global optimal, which is easily premature convergence. Here, an immune algorithm can solve the problem by produces excellent antibodies through the memory function. However, immune algorithm also has some shortcomings that the initial population is insufficient diversity, and the variation scale is big or small. To solving these issues, tremendous research efforts have done in the past few years. For instance,

Jordehi [21] introduced the chaotic strategies into the immune algorithm and designed to solve the global continuous optimization problems. Wu et al. [22] used the chaotic producers to replace the random generators and used the parallel mutation strategy for the hypermutation of antibodies. Liang et al. [23] used the logistic mapping to generate the initial antibody populations for the diversity of antibodies, etc. While most of the approaches thought that the value of a control factor of the chaotic motion is randomly. Recently, we found that it takes an appropriate value of the control factor will generate the exciting results.

In this work, we focused on improving the performance of immune clonal optimization algorithm for pulmonary nodule classification in chest CT images. Therefore, the method of intelligent immune clonal selection is proposed and developed by the population diversity and intelligent mutation. First, according to the mechanism and characteristics of chaotic motion, the proposed method utilizes the characteristics of chaotic motion and selects the control factor of optimal chaotic state, to generate the initial population with randomness and ergodicity. Second, the proposed method solved the singleness problem of the initial population of immune algorithm. Then based on the characteristics of the Gaussian mutation operator with a small scale, and Cauchy mutation operator with a big scale, the intelligent mutation strategy is developed, and the control factor of intelligent mutation is designed. Especially, the intelligent Gauss-Cauchy hybrid mutation operator is designed and implemented. Additionally, the Hessian information and multi-scale reverse LoG (Hessian-MRLoG) method were used to detect and locate the lesion area in the image preprocessing, and the candidate nodule areas are extracted by an algorithm of the lung nodule segmentation. Therefore, the computational amount of feature extraction and classification is greatly reduced, the data quality is improved and the computational complexity is reduced.

The remainder of this paper is organized as follows. Section 2 describes the method and approach of the proposed intelligent immune clonal for feature selection and parameter optimization. Section 3, the experimental results obtained by the proposed method. Lastly, Section 4 presents the final remarks of this work.

2. Materials and methods

2.1. The method of intelligent mutation

2.1.1. Population initialization operator with chaotic maps

The chaotic mapping was adopted for antibody population initialization in the immune clonal selection algorithm. According to the theoretical study of population models, the logistic mapping model is described as follows:

$$x_{n+1} = \mu x_n(1-x_n), \quad n = 1, 2, \ldots \tag{1}$$

where μ is a regulation factor of the chaotic behavioral and $\mu > 0$, $x(n)$ is the chaotic variable and n is the number of iterations. Let the initial condition $x_0 \in [0,1]$, $\mu \in [0,4]$, then the diagram of chaotic motion bifurcation is shown in Figure 1.

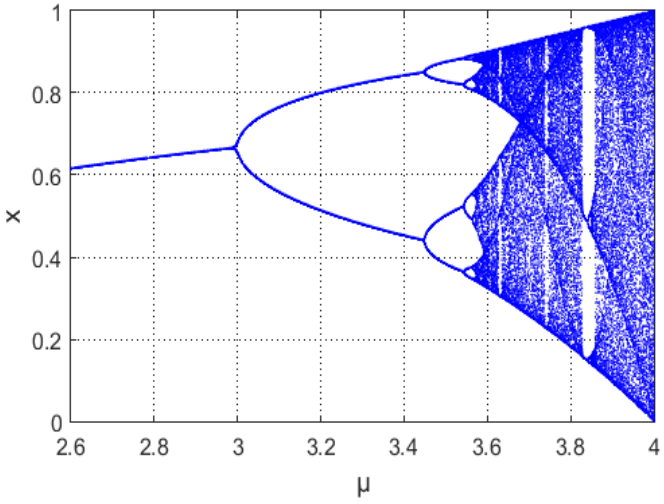

Figure 1. The bifurcation diagram of the logistic model.

From Figure 1, if $0 < \mu < 3$, then the chaotic motion shows single-cycle sequence converges. If $3 < \mu < 3.449$, then the chaotic motion converges with a period of 2 times. If $3.449 < \mu < 3.57$, then the chaotic motion converges with a 4-fold period. If $3.57 < \mu < 4$, the chaotic motion comes into a chaotic state. If $\mu = 4$, then the system enters a completely chaotic state. Based on the analysis, it is shown that a system evolves into chaotic state when the regulation factor $3.57 < \mu \leq 4$. Therefore, let the initial conditions of $x_0 \in [0,1]$, $\mu \in [0,4]$, and $\mu = 4$, $n = 200$, then the random sequence is generated by the logistic model, which is shown in Figure 2.

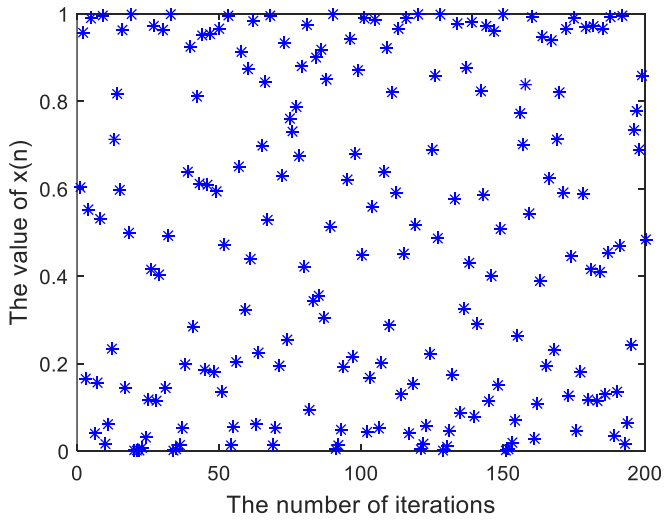

Figure 2. The ergodicity of the logistic mapping.

Figure 2 shows the ergodicity of the logistic mapping with a suitable regulation factor, which was generated an initial antibody population with diversity. Thus, we utilize the characteristics of chaotic motion and select the regulation factor of optimal chaotic state, to generate an initial antibody population with randomness and ergodicity. It solved the singleness problem of the initial population generation of the immune clonal selection algorithm.

2.1.2. The method of intelligent mutation

According to the literature [24], the combination of Gauss-Cauchy is written as:

$$f_{gc} = k \cdot (f_{gaussian}(\mu, \sigma) + f_{cauchy}(x_0, \gamma)) \tag{2}$$

where k is an adjustment factor. $f_{\text{gaussian}}(\mu, \sigma)$ is a Gaussian distribution function, where σ is the standard deviation and $\sigma > 0$, μ is the mathematical expectation. $f_{\text{cauchy}}(x_0, \gamma)$ is a Cauchy distribution function, where x_0 is the length setting of x-axis, γ is a window scale, and $x \in (-\infty, +\infty)$. Let $k = 1/2$, $\mu = 0$, $\sigma = 1$; $x_0 = 0$, $\gamma = 1$, respectively, then the average distribution curve of Gauss-Cauchy combination is shown in Figure 3.

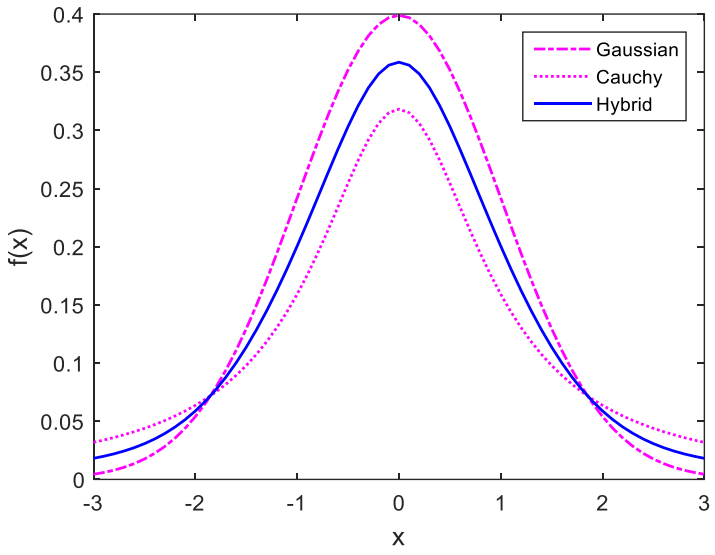

Figure 3. The average distribution curve of the Gauss-Cauchy combination.

Therefore, the hybrid Gauss-Cauchy mutation operator is described as follows:

$$x' = x + k \cdot \left(f_{gaussian}(\mu, \sigma) + f_{cauchy}(x_0, \gamma) \right) \tag{3}$$

More importantly, we designed an intelligent strategy for the mutation, which is described below: In early phase, the mutation rate has the characteristics of large-scale variation; in the intermediate phase, the mutation rate has a middle-scale variation; in the late stage, the mutation rate has a small-scale variation. Thus, the mutation rate reduced gradually with the increase of evolutionary iterations. The mutation rate is designed by:

$$v = 1 - (t / t_{\max}) \tag{4}$$

where t is the current evolutionary iterations, $t_{\max}$ is the maximum iterations. The mutation rate of v is a monotonically decreasing function, and $v \in [0,1]$. Consequently, the intelligent mutation operator (IMO) is given as:

$$x' = x + \nu \cdot k \cdot \left(f_{gaussian}(\mu, \sigma) + f_{cauchy}(x_0, \gamma) \right) \tag{5}$$

where k is an adjustment factor, $f_{\text{gaussian}}(\mu, \sigma)$ is a Gaussian distribution function, $f_{\text{cauchy}}(x_0, \gamma)$ is a Cauchy distribution function. With the guidance of intelligent mutation strategy, the IMO not only has the local convergence characteristics of Gaussian mutation, but also the global convergence characteristics of Cauchy mutation.

2.2. Feature extraction

Table 1. Feature description.

Classification	Feature name	Number	Formulas
I.The intensity features (f_1–f_5)	Mean value	f_1	$\sum_{i=1}^{L}[i \cdot I(i)]$
	Variance	f_2	$\sum_{i=1}^{L}[(i-\mu)^2 \cdot I(i)]$
	Deviation	f_3	$(1/\sigma^2) \cdot \sum_{i=1}^{L}[(i-\mu)^3 \cdot I(i)]$
	Energy	f_4	$\sum_{i=1}^{L} I^2(i)$
	Gray-scale entropy	f_5	$\sum_{i=1}^{L}[I(i) \cdot \log_2 I(i)]$
II. The shape features (f_6–f_{10})	Circularity	f_6	P_0^2/S_0
	Compactness	f_7	$4\pi \cdot \text{Area}/(\text{Circumference})^2$
	Elongation	f_8	$r_{\max} / r_{\min}$
	Convex degree	f_9	Area / Convex_Area
	Rectangularity	f_{10}	W_r / L_r
III.Text features (f_1–f_{14})	Compactness	f_{11}	$r_{\text{inside}} / r_{\text{outside}}$
	Moment	f_{12}	$\sum_{i=1}^{M}\sum_{j=1}^{N}[I^2(i,j)/(1+\lvert i-j\rvert)]$
	Angular second moment	f_{13}	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[p^2(i,j)]$
	Contrast	f_{14}	$\sum_{n=1}^{N_g} n^2\{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\}$
	Correlation	f_{15}	$(\sigma_x\sigma_y)^{-1} \cdot \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\{(ij)p(i,j)-\mu_x\mu_y\}$
	Variance	f_{16}	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\{(i-\mu_x)(i-\mu_y)p(i,j)\}$
	Inverse difference moment	f_{17}	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[1/(1+(i-j)^2) \cdot p(i,j)]$
	Sum average	f_{18}	$\sum_{i=2}^{2N_g}(ip_{x+y}(i))$
	Sum variance	f_{19}	$\sum_{i=2}^{2N_g}[(i-f_8)^2 \cdot p_{x+y}(i)]$
	Sum entropy	f_{20}	$-\sum_{i=2}^{2N_g}[p_{x+y}(i) \cdot \log(p_{x+y}(i))]$
	Entropy	f_{21}	$-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\{p(i,j) \cdot \log(p(i,j))\}$
	Difference variance	f_{22}	$\text{Var}(\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j))$
	Difference entropy	f_{23}	$\sum_{n=1}^{N_g}\{p_{x-y}(i)\log(p_{x-y}(i))\}$
	Information measures of correlation 1	f_{24}	$\{-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log(p(i,j))+ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\ \log(\sum_{j=1}^{N_g} p(i,j) \cdot \sum_{i=1}^{N_g} p(i,j))\} \cdot \{\max(\sum_{j=1}^{N_g} p(i,j), \sum_{i=1}^{N_g}\ p(i,j))\}^{-1}$
	Information measures of correlation 2	f_{25}	$\{1-\exp(2\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(p_x(i)p_y(j) \cdot \log(p_x(i)p_y(j))+ f_9)\}^{1/2}$
	Maximal correlation coefficient	f_{26}	(Second largest eigenvalue of $Q)^{1/2}$, where $Q(i,j) = \sum_k\{ p(i,k)\, p(j,k) / (p_x(i)\, p_y(j))\}$

The compound features were extracted in this work from three aspects, which include the

intensity features, shape features, and texture features [25]. The character representation is listed in Table 1.Though the feature combination, 26 feature vectors were extracted for feature selection. Then the 26-dimensional feature space was constructed.

2.3. Immune clonal optimization

(1) Radial basis kernel function (RBF)

According to the study of lung nodule model in our previous study [26], we know that the intensity distribution of a lung nodule in the CT image, gradually increases from the periphery to the center. Therefore, this paper adopts RBF as the kernel function, which has a better performance on nonlinear mapping. A 2-dimensional (2D) Gaussian kernel function is represented by:

$$K(\mathbf{x},\mathbf{x}_i)=\exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2\sigma^2}\right) \tag{6}$$

Let $\gamma = 1/(2\sigma^2)$, then the Eq (6) is instead by a RBF

$$K(\mathbf{x},\mathbf{x}_i)=\exp\left(-\gamma\|\mathbf{x}-\mathbf{x}_i\|^2\right) \tag{7}$$

(2) Chromosome coding method

The antibody coding is the combining process of feature subsets and SVM parameters. An antibody is a result of the combination of feature subsets and SVM parameters. The synchronous optimization of the feature selection and SVM parameter is to find the optimal antibodies. It is noteworthy that an antibody has multiple gene positions, the number of gene positions is equal to the length of the coding string. In the chromosome coding, 1 is a dominant gene, which indicates the selected feature; 0 is a recessive gene, which indicates the unselected feature. In this work, the chromosome coding includes the penalty factor of C, the kernel parameter of r, and the feature set of f. While, the corresponding number of genes are represented by I, J and K, respectively. L denotes the string length, the encoded string is given by ($x_{c1}, x_{c2},\ldots, x_{ci}, \ldots, x_{cI}; x_{\gamma 1}, x_{\gamma 2}, \ldots, x_{\gamma j}, \ldots, x_{\gamma J}; x_{f1}, x_{f2}, \ldots, x_{fk}, \ldots, x_{fK}$), and $I + J + K = L$. The chromosome encoding is shown in Figure 4.

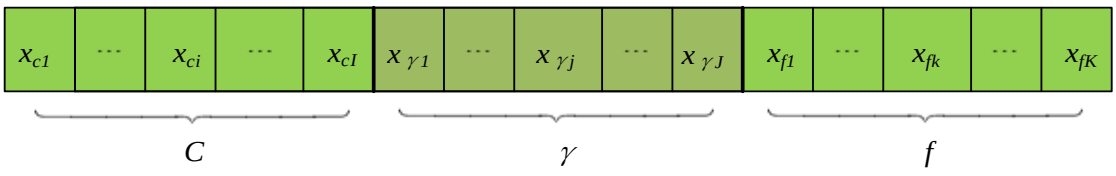

Figure 4. Chromosome coding.

From Figure 4, it should be noted that the first eight bits of ($x_{c1}, x_{c2},\ldots, x_{ci}, \ldots, x_{cI}$) are the coding of the penalty factor C, the following eight bits of ($x_{\gamma 1}, x_{\gamma 2}, \ldots, x_{\gamma j}, \ldots, x_{\gamma J}$) are the coding of the kernel parameter γ, the last 26 bits of ($x_{f1}, x_{f2}, \ldots, x_{fk}, \ldots, x_{fK}$) are the coding of the feature set f.

(3) The affinity function

$$\text{Affinity} = \text{SVM\_Accuracy} \tag{8}$$

(4) Antibody selection and cloning

If the population size is N, the probability of antibody fitness is:

$$p_{fit} = \frac{Fitness}{\sum_{i=1}^{N} Fitness(i)} \tag{9}$$

The number of clone is:

$$N_c(i) = Round\,(p_{fit} \cdot N / i) \tag{10}$$

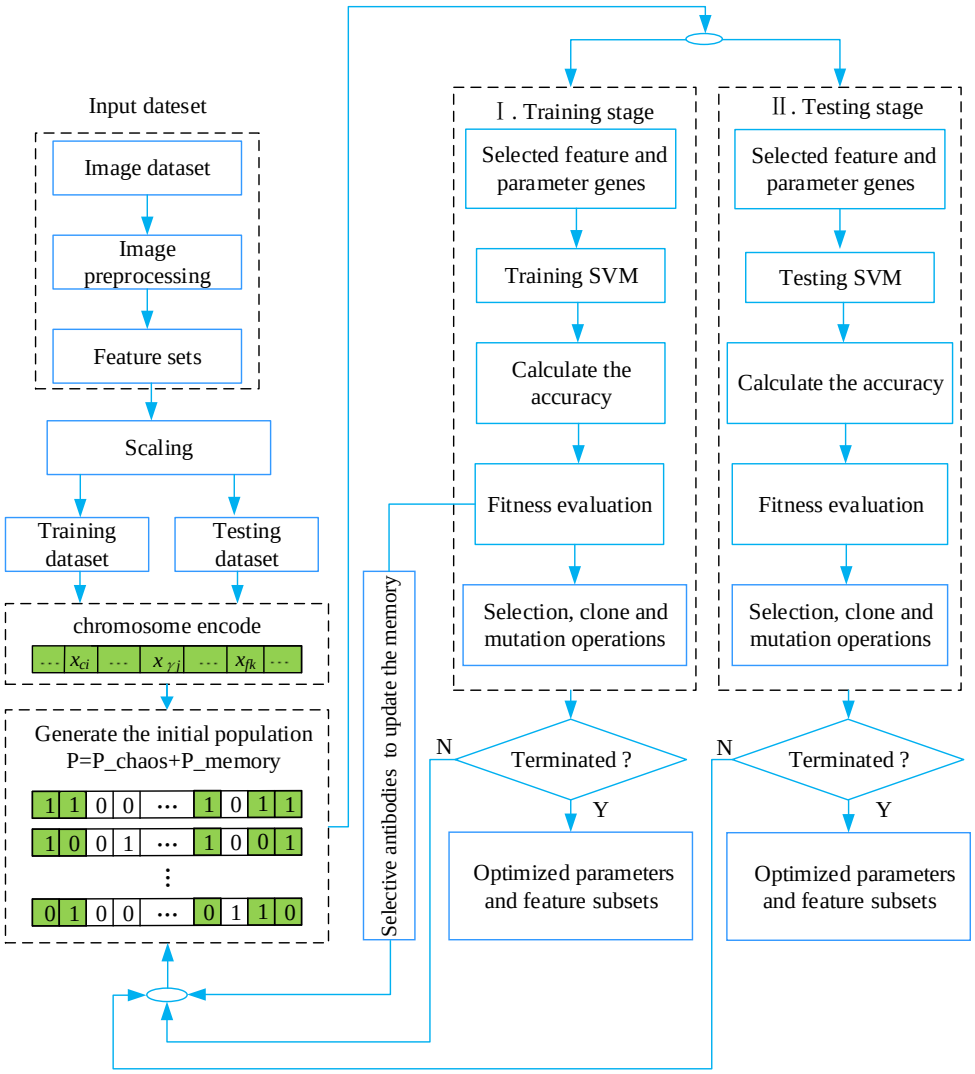

Figure 5. Flowchart of the ICOA-SVM.

Additionally, the intelligent mutation was implemented by the proposed intelligent mutation operator. From the above analysis, the method of immune clonal optimization algorithm (ICOA) for parameter optimization in support vector machine (ICOA-SVM) was proposed and developed. The flowchart of the developed method is shown in Figure 5.

It should be noted that the implementation procedure can be obtained from the Figure 5. The input is the candidate feature set of F: F = $\{(x_i, y_i) \mid x \in F_n^L\}$, $y \in (1,2,\ldots,L)$, $i = 1, \ldots, n\}$. The output are the optimized parameters of (C, γ), and the selected feature subsets of $f = \{f_1, f_2, \ldots, f_k\}$, $1 \le k \le K$. The implementation of the proposed algorithm includes the training and the testing stages. In the process of immune clonal optimization, the antibodies with high affinity were selected and cloned, and generated the offspring of $Ab'(l) = \{\mathrm{Ab}(l), Ab_1'(l), \ldots, Ab_i'(l), \ldots, Ab_n'(l)\}$, $Ab_i'(l) = (x_{i1}, x_{i2}, x_{i(Nc-1)})$, $Nc(i) = \mathrm{Round}(p_{fit} \cdot N/i)$. Additionally, the others new individuals were generated by the low affinity antibodies though the mutation of $Ab' = Ab + 0.5 \cdot (1 - t/t_{\max})\ (f_{\mathrm{gaussian}}(0,1) + f_{\mathrm{cauchy}}(0,1))$.

3. Results and discussion

The proposed method was implemented in Matlab (R2015a; MathWorks, Inc., Natick, MA, USA) on a machine with Core i3 M 350 @ 2.27GHz, 6G RAM, 64-bit operating system, the CT images that we adopted were obtained from the international public LIDC database [27]. The 90 scans of CT images with the number of 652 lung nodules were randomly selected from the LIDC database. One scan has the number of 100~400 CT images. Thus, more than 20,000 images were used in the experiments of this study.

3.1. Classification of lung nodules

There are several steps for the image processing. First, medical images are usually noisy, especially the low-dose spiral CT images, which mainly have the quantum noise, electronic noise, Gaussian noise, and speckle noise [28]. Therefore, a hybrid filtering strategy was developed, which used the median filter for the isolated noise reduction, and the multi-scale windowed Fourier filtering for the noise reduction [29]. Second, the multi-scale elliptic rolling-ball method, which was proposed and developed in our previous work [30], was used for pulmonary parenchyma extraction. Third, the Hessian-MRLoG method [2] was used for the detection and location of pulmonary nodule candidates, then the candidate nodule areas were extracted by an algorithm of the lung nodule segmentation [29]. Four, the features of the regions of interest (ROI) were extracted to construct the 42-dimensional coding space. Especially, there are 26-dimensional feature vectors in feature subsets, which include 5 gray features, 7 shape features, and 14 texture features. The experimental dataset was used to verify the effectiveness of the proposed method, and the experimental adjustment parameters of the proposed method were shown in Table 2.

Especially, the penalty factor of $C \in [0.001, 2.2 \times 10^4]$, the kernel parameter of $\gamma \in [1 \times 10^{-5}, 5]$. The initial population was generated by the chaotic mapping. The feature subsets and SVM parameters were selected and optimized by the developed method of ICO. The optimal feature subsets and the best matching parameters were obtained by five times of 5-fold cross-validation. According to the number of selected feature dimensions, the optimal feature subsets and best match parameters are shown in Table 3.

Table 2. Parameter settings of the ICOA-SVM.

Parameters	Name of the parameters	Value
N	Population size	60
t_{max}	Maximum iterations	100
L	The length of chromosome genes	34
m	The number of elites	5
Nc_max	Maximum number of clones	120
w	Mutation probability	0.5
ξ_i	The relaxation variable	0.5
k	The number of k-fold cross	5

Table 3. Optimal feature subsets and best match parameters.

Serial numbers	Parameters of SVM	Selected subsets of features	Accuracy
1	$C = 3985$, $\gamma = 0.0432$	$f_6, f_7, f_3, f_{25}, f_9$	0.920
2	$C = 6591$, $\gamma = 0.0332$	$f_6, f_7, f_4, f_{11}, f_9, f_{26}, f_{14}$	0.965
3	$C = 6730$, $\gamma = 0.0436$	$f_6, f_7, f_4, f_5, f_8, f_9, f_{26}, f_{14}$	0.978
4	$C = 8390$, $\gamma = 0.0529$	$f_6, f_7, f_3, f_5, f_9, f_{26}, f_{25}, f_2, f_{17}, f_{14}$	0.940
5	$C = 8534$, $\gamma = 0.0623$	$f_6, f_7, f_3, f_4, f_5, f_8, f_9, f_{26}, f_{25}, f_{14}, f_{11}, f_1, f_{18}$	0.935

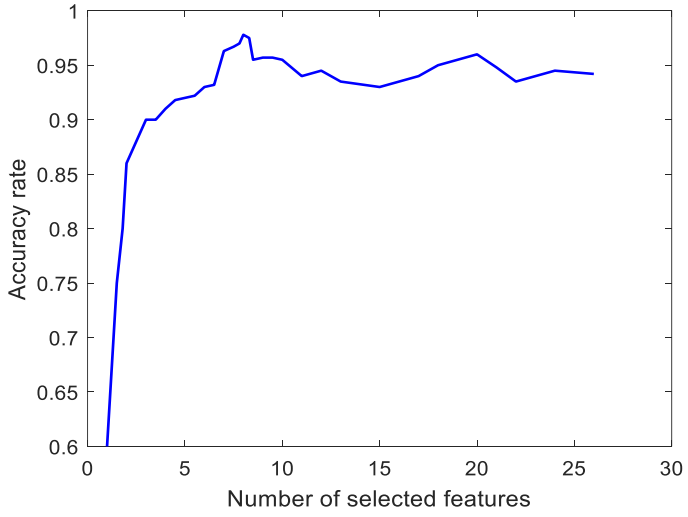

Figure 6. The relation curve.

In this experiment, the eight features of $f_6, f_7, f_4, f_5, f_8, f_9, f_{26}$, and f_{14} constructing an optimal feature subset, and the corresponding parameters of SVM are $C = 6730$, $\gamma = 0.0436$, which obtained the highest accuracy of 97.8%. The experimental results revealed that the accuracy rate is not proportional to the dimensions of selected features. After ten times of 5-fold cross-validation, the relationship between the number of selected features and the classification accuracy is shown in Figure 6.

From Figure 6, the results indicate that it is not the more selected feature dimensions, the higher of the classification accuracy. Actually, it has a better match between the classification accuracy and the number of selected features. In this work, when the dimensions of selected features are 8, the accuracy of classification is the highest.

3.2. Performance evaluation

The combination of the two-classification problems can be represented by a two-class confusion matrix [26]. If the practical situation is a nodule, and the prediction result is also a nodule, then it belongs to the nodule, this situation is the true positive (TP). If the practical situation is a non-nodule, and the prediction result is a nodule, then it belongs to the non-nodule, this situation is the false positive (FP). If the practical situation is a nodule, and the prediction result is a non-nodule, then it belongs to the situation of false negative (FN). If the practical situation is a non-nodule, and the prediction result is a non-nodule, then it belongs to the situation of true negative (TN). Generally speaking, the performance of classification algorithms are commonly evaluated by the receiver operating characteristic (ROC) curve and its area under the curve (AUC) value. The evaluation indicators are shown in Table 4.

Table 4. Evaluation indicators and the corresponding meanings.

Indexes	Implication	Formulas
Sensitivity	True positive rate	Sens = TP / (TP + FN)
Specificity	True negative rate	Spec = TN / (TN + FP)
Accuracy	Accuracy	Acc = (TP + TN) / (TP + TN + FP + FN)
FPR	False positive rate	FPR = TP / (TP + FP)
AUC	The area under the characteristic curve	AUC = $\sum_i\{(1\text{-}3y(i) / 2\text{-}y(i\text{-}1) / 2) * (x(i)\text{-}x(i\text{-}1))\}$

Table 5. The comparison among different approaches.

Methods	Accuracy (%)	Sensitivity (%)	AUC (%)
GA-SVM	91.37	92.25	94.92
ICSA-SVM	93.25	94.75	96.54
CNN	93.13	92.50	96.64
The proposed	**97.87**	**98.75**	**99.09**

The bigger of the value of sensitivity, specificity, accuracy, and AUC, the better the performance of a classifier. Additionally, the lower the value of false positive rate (FPR), the better the performance of a classifier. In this work, at the same experimental conditions, we analyzed the properties of genetic algorithm (GA), immune clone selective algorithm (ICSA), and the proposed ICOA-SVM, respectively. Additionally, the approach of convolutional neural network (CNN) was used for the pulmonary nodule classification at the same experimental dataset. Especially, the initial learning rate is 0.001. After that, each generation of learning rate decays to 0.9 of the previous generation learning rate. After experiments, the experimental results were shown in Table 5.

From Table 5, the proposed method has the best performance among the four approaches. Therefore, the proposed method is superior to the mentioned approaches. Besides, to more clearly represent the classification performance, the ROC was plotted in Figure 7.

Table 6. Comparison of the results obtained in this study with those of other works.

Literatures	Scans	Nodules	Accuracy (%)	Sensitivity (%)	FPs/scan	Dataset
Lü et al. [3]	50	435	84.6	88.9	--	Private
Liu et al. [31]	80	978	93.2	92.4	4.5	LIDC
Liu et al. [32]	107	252	--	89.4	2.0	LIDC
Kaya [33]	--	439	84.7	67.4	--	LIDC
Khan et al. [34]	84	103	97.5	98.9	--	LIDC
Zhang et al. [35]	--	353	88.6	86.3	--	LIDC
Jiang et al. [36]	--	--	90.24	92.04	11.06	LIDC
Santos et al. [14]	28	252	--	90.6	1.17	LIDC
The proposed	90	652	**97.87**	**98.75**	**1.52**	LIDC

As can be seen from Figure 7, the area under the cure value of GA-SVM, ICSA-SVM, CNN, and the proposed are 94.92, 96.54, 96.64, and 99.09%, respectively. The proposed method has the biggest AUC value among the four approaches. According to its analysis, the proposed method is superior to the mentioned methods in the performance of lung nodule classification. Additionally, the performance of the proposed method was also compared with other related approaches, which reported in the literature in the last five years, as summarized in Table 6.

The comparison reveals that the proposed method in this work has the highest sensitivity, while the approach reported in another study has the lowest FPR [14]. But, regarding the overall performance, the proposed method in this study has the best performance. Thus, the proposed method is superior to the traditional approaches. Additionally, the Hessian-MRLoG method was used to detect and locate the lesion area, then the candidate nodule areas were extracted by an algorithm of the lung nodule segmentation. As a result, the computational amount of feature extraction and classification is greatly reduced, the data quality is improved and the computational complexity is reduced. Therefore, this study gives a reference for a novel light-weight method of computer-aided diagnosis of pulmonary nodules.

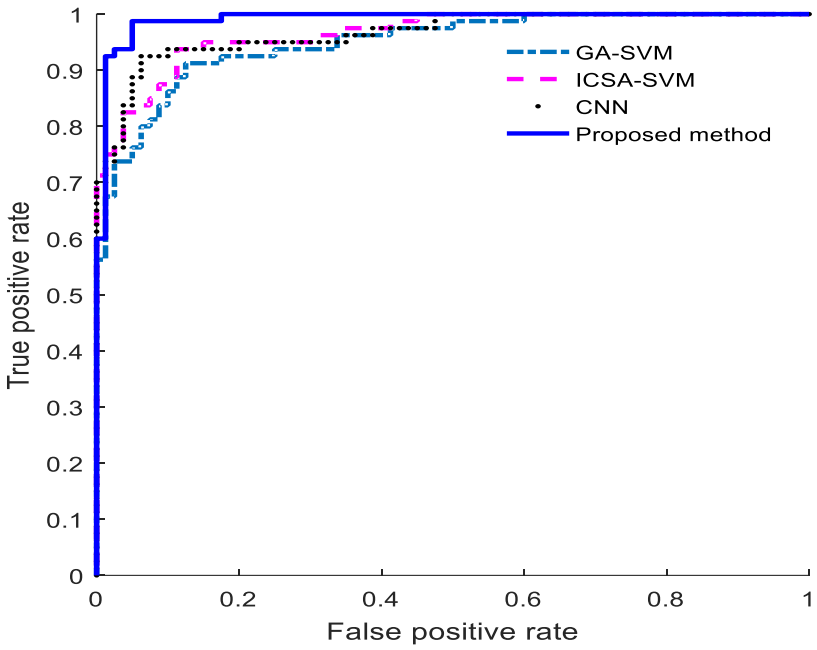

Figure 7. Comparison of the area under the curve of ROC.

4. Conclusions

Computer-aided pulmonary nodule diagnosis is an effective approach for the early detection of lung cancers. In this paper, an intelligent immune clonal optimization algorithm is proposed and developed to find the optimal feature subset and best match parameters of the classification algorithm, which makes the feature selection and classification be performed simultaneously. The initial population was generated by the randomness and ergodicity of the chaotic mapping. The singleness problem of the initial population of the immune clonal selection algorithm was solved in this work. The proposed method has the global and local convergence characteristics by the intelligent mutation operator. Therefore, a novel method based on an intelligent immune clonal optimization algorithm for pulmonary nodule classification was proposed and developed in this work. The proposed approaches of this work were used for the classification of 90 scans with 652 nodules, and the results revealed an accuracy of 97.87% with the 1.52 FPs/scan. Consequently, the experimental results demonstrated that the proposed approach can improve the accuracy and reduce the false positive rate in pulmonary nodule classification. This work has the potential for deployment in the early computer-aided diagnosis of lung cancer.

Acknowledgments

Supported by the grants from the National Natural Science Foundation of China (Grant No. 61673257, 61705127), the construction project of applied undergraduate of Shanghai for the major of electrical engineering and automation of SUES, and the doctoral training foundation of SUES (Grant No. 0232-A3-0100-21-0931).

Conflict of Interest

No potential conflicts of interest were disclosed.

References

1. W. L. Bi, A. Hosny, M. B. Schabath, M. L. Giger, N. J. Birkbak, A. Mehrtash, Artificial intelligence in cancer imaging: clinical challenges and applications, *CA: Cancer J. Clin.*, **69** (2019), 127–150.
2. Q. Mao, S. Zhao, D. Tong, S. Su, Z. Li, X. Cheng, Hessian-MRLoG: Hessian information and multi-scale reverse LoG filter for pulmonary nodule detection, *Comput. Biol. Med.*, **131** (2021), 104272.
3. X. Q. Lü, L. Wu, Y. Gu, W. L. Zhang, J. Li, Detection of low dose CT pulmonary nodules based on 3D convolution neural network, *Opt. Precis. Eng.*, **26** (2018), 1211–1218.
4. B. R. Froz, A. O. de C. Filho, A.C. Silva, A. C. de Paiva, R. A. Nunes, M. Gattass, Lung nodule classification using artificial crawlers, directional texture and support vector machine, *Expert Syst. Appl.*, **69** (2017), 176–188.
5. C. Lu, Z. Zhu, X. Gu, An intelligent system for lung cancer diagnosis using a new genetic algorithm based feature selection method, *J. Med. Syst.*, **38** (2014), 97.
6. A. Abid, M. T. Khan, M. S. Khan, Multidomain features-based GA optimized artificial immune system for bearing fault detection, *IEEE Trans. Syst., Man Cybern. Syst.*, **50** (2020), 348–359.
7. J. Gao, Q. Jiang, B. Zhou, D. Chen, Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview, *Math. Biosci. Eng.*, **16** (2019): 6536–6561.
8. Z. Wang, X. Tang, H. Liu, L. Peng, Artificial immune intelligence-inspired dynamic real-time computer forensics model, *Math. Biosci. Eng.*, **17** (2020), 7221–7233.
9. T. Gong, T. Fan, L. Pei, Z. Cai, Magnetic resonance imaging-clonal selection algorithm: An intelligent adaptive enhancement of brain image with an improved immune algorithm, *Eng. Appl. Artif. Intell.*, **62** (2017), 405–411.
10. D. González-Patiño, Y. Villuendas-Rey, A. J. Argüelles-Cruz, O. Camacho-Nieto, C. Yáñez-Márquez, AISAC: an artificial immune system for associative classification applied to breast cancer detection, *Appl. Sci.*, **10** (2020), 1–22.
11. L. Liu, Q. Dou, H. Chen, J. Qin, P. A. Heng, Multi-task deep model with margin ranking loss for lung nodule analysis, *IEEE Trans. Med. Imaging*, **39** (2020), 718–728.
12. Z. Zhou, S. Li, G. Qin, M. Folkert, S. Jiang, J. Wang, Multi-objective-based radiomic feature selection for lesion malignancy classification, *IEEE J. Biomed. Health Inf.*, **24** (2020), 194–204.
13. J. Z. Farkas, G. T. Smith, G. F. Webb, A dynamic model of CT scans for quantifying doubling time of ground glass opacities using histogram analysis, *Math. Biosci. Eng.*, **15** (2018), 1203–1224.
14. A. M. Santos, A. Filho, A. C. D. Paiva, R. A. Nunesb, M. Gattassc, Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, tsallis entropy and SVM, *Eng. Appl. Artif. Intell.*, **36** (2014), 27–39.
15. N. Bi, J. Tan, J. Lai, J. H. Lai, C. Y. Suen, High-dimensional supervised feature selection via optimized kernel mutual information, *Expert Syst. Appl.*, **108** (2018), 81–95.

16. F. V. Farahani, A. Ahmadi, M. H. F. Zarandi, Hybrid intelligent approach for diagnosis of the lung nodule from CT images using spatial kernelized fuzzy c-means and ensemble learning, *Math. Comput. Simul.*, **149** (2018), 48–68.
17. I. Ali, G. R. Hart, G. Gunabushanam, Y. Liang, W. Muhammad, B. Nartowt, et al., Lung nodule detection via deep reinforcement learning, *Front. Oncol.*, **8** (2018), 108.
18. C. A. de P. Pinheiro, N. Nedjah, L. de M. Mourelle, Detection and classification of pulmonary nodules using deep learning and swarm intelligence, *Multimedia Tools Appl.*, **79** (2020), 15437–15465.
19. F. Abdolali, J. Kapur, J. L. Jaremko, M. Noga, A. R. Hareendranathan, K. Punithakumar, Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks, *Comput. Biol. Med.*, **122** (2020), 103871.
20. M. R. G. Raman, N. Somua, K. Kirthivasan, R. Liscano, V. S. S. Sriram, An efficient intrusion detection system based on hypergraph-Genetic algorithm for parameter optimization and feature selection in support vector machine, *Knowl.-Based Syst.*, **134** (2017), 1–12.
21. A. R. Jordehi, A chaotic artificial immune system optimisation algorithm for solving global continuous optimisation problems, *Neural Comput. Appl.*, **26** (2015), 827–833.
22. S. Wu, H. D. Wan, S. K. Shukla, B. Li, Chaos-based improved immune algorithm (CBIIA) for resource-constrained project scheduling problems, *Expert Syst. Appl.*, **38** (2011), 3387–3395.
23. H. Liang, F. Kang, Adaptive chaos parallel clonal selection algorithm for objective optimization in WTA application, *Optik*, **127** (2016), 3459–3465.
24. K. Chellapilla, Combining mutation operators in evolutionary programming, *IEEE Trans. Evol. Comput.*, **2** (1998), 91–96.
25. Z. Wang, J. Xin, P. Sun, Z. Lin, Y. Yao, X. Gao, Improved lung nodule diagnosis accuracy using lung CT images with uncertain class, *Comput. Methods Programs Biomed.*, **162** (2018), 197–209.
26. Q. Mao, S. Zhao, Q. Zheng, S. Su, L. Li, X. Zhang, Modified Gaussian models for pulmonary nodule simulation in chest tomosynthesis, *J. Med. Imaging Health Inf.*, **8** (2018), 1718–1725.
27. S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, et al., Data from LIDC–IDRI, The cancer imaging archive, 2015. Available from: http://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX.
28. M. Hilts, C. Duzenli, Image filtering for improved dose resolution in CT polymer gel dosimetry, *Med. Phys.*, **31** (2004), 39–49.
29. Q. Mao, S. Zhao, T. Gong, Q. Zheng, An effective hybrid windowed Fourier filtering and fuzzy c-mean for pulmonary nodule segmentation, *J. Med. Imaging Health Inf.*, **8** (2018), 72–77.
30. Q. Mao, S. Zhao, Modified rolling-ball method for pulmonary parenchyma segmentation, *J. Med. Imaging Health Inf.*, **10** (2020), 364–369.
31. J. k. Liu, H. Jiang, M. Gao, C. He, Y. Wang, P. Wang, et al., An assisted diagnosis system for detection of early pulmonary nodule in computed tomography images, *J. Med. Syst.*, **41** (2017), 30.
32. X. L. Liu, F. Hou, H. Qin, A. Hao, A CADe system for nodule detection in thoracic CT images based on artificial neural network, *Sci. China Inf. Sci.*, **60** (2017), 072106.
33. A. Kaya, Cascaded classifiers and stacking methods for classification of pulmonary nodule characteristics, *Comput. Methods Prog. Biomed.*, **166** (2018), 77–89.
34. S. A. Khan, S. Hussain, S. Yang, K. Iqbal, Effective and reliable framework for lung nodules detection from CT scan images, *Sci. Rep.*, **9** (2019), 1–14.

35. G. Zhang, X. Liu, D. Zhu, P. He, L. Liang, Y. Luo, et al., 3D spatial pyramid dilated network for pulmonary nodule classification, *Symmetry-Basel*, **10** (2018), 376.
36. H. Jiang, F. Gao, X. Xu, F. Huang, S. Zhu, Attentive and ensemble 3D dual path networks for pulmonary nodules classification, *Neurocomputing*, **398** (2020), 422–430.

AIMS Press