

Smart Lung Tumor Prediction Using Dual Graph Convolutional Neural Network

Abdalla Alameen*

Department of Computer Science, College of Arts and Sciences, Prince Sattam Bin Abdulaziz University, KSA

*Corresponding Author: Abdalla Alameen. Email: a.alameen@psau.edu.sa

Received: 08 April 2022; Accepted: 21 June 2022

Abstract: A significant advantage of medical image processing is that it allows non-invasive exploration of internal anatomy in great detail. It is possible to create and study 3D models of anatomical structures to improve treatment outcomes, develop more effective medical devices, or arrive at a more accurate diagnosis. This paper aims to present a fused evolutionary algorithm that takes advantage of both whale optimization and bacterial foraging optimization to optimize feature extraction. The classification process was conducted with the aid of a convolutional neural network (CNN) with dual graphs. Evaluation of the performance of the fused model is carried out with various methods. In the initial input Computer Tomography (CT) image, 150 images are pre-processed and segmented to identify cancerous and non-cancerous nodules. The geometrical, statistical, structural, and texture features are extracted from the preprocessed segmented image using various methods such as Gray-level co-occurrence matrix (GLCM), Histogram-oriented gradient features (HOG), and Gray-level dependence matrix (GLDM). To select the optimal features, a novel fusion approach known as Whale-Bacterial Foraging Optimization is proposed. For the classification of lung cancer, dual graph convolutional neural networks have been employed. A comparison of classification algorithms and optimization algorithms has been conducted. According to the evaluated results, the proposed fused algorithm is successful with an accuracy of 98.72% in predicting lung tumors, and it outperforms other conventional approaches.

Keywords: CNN; dual graph convolutional neural network; GLCM; GLDM; HOG; image processing; lung tumor prediction; whale bacterial foraging optimization

1 Introduction

Based on the structure of the cells, lung carcinoma is classified into small-cell carcinomas and non-small-cell carcinomas. More than 85% of lung cancers are caused by long-term tobacco use [1]. There are 15%–20% of lung tumors among people who have never smoked, as a result of factors such as air pollution, asbestos, radon gas, and secondhand smoke. According to the size and location of the lymph nodes on the tumor, lung cancer can be divided into I to IV categories. A timely diagnosis of this lung



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

cancer may reduce the mortality rate, or else the chances of survival are high [2]. Radiology and CT are the traditional methods to detect lung tumors due to their lower distortion levels and the ability to detect malignant growths [3]. The CT image depicts a lung malignancy with irregular and regular lung structures.

An image is preprocessed by normalizing, enhancing quality, and reducing distortion. Various filters in both the frequency and spatial domains are applied to enhance the input image [4]. The adaptive median filter eliminates the noise issues associated with the median filter. In the Wiener filter, low pass filters are used to minimize noise and high pass filters are used for deconvolution on the CT image. In medical imaging, segmentation is considered the most dangerous task for locating the nodule region to extract its features. Different approaches are used for segmentation such as watershed edge-based approaches [5], thresholding [6], and region-based segmentation [7]. Thresholding and morphological operation are used to determine the tumor region from the segmented nodule. The lung nodule region is segmented and statistical, textual, geometric, and shape-based features are extracted [8]. There are several features present in medical neuroimaging, such as gray-level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), local binary pattern (LBP) and histogram, and gray level dependence matrix (GLDM) [9]. These feature sets are used to evaluate the decision-making process.

To reduce the dimension of the feature space, an evolutionary fusion technique referred to as Whale Optimization with Adaptive PSO (WASO-APSO) has been used to remove redundant features [10,11]. To improve the classification stage, the features are grouped using LDA (linear discriminate algorithm) [12,13]. In this study, the whale optimization approach is fused with bacterial foraging optimization to enhance the feature selection for the classification phase. The classification is a major step in determining parameters such as accuracy, sensitivity, and specificity of the classifier that aids in the diagnosis of lung cancer. In this study, support vector machines, convolutional neural networks, and dual graph convolutional neural networks are compared. Based on a linear function from statistical theory, a support vector machine is a multiclass model used to predict medical imaging results [14]. Recently, deep learning has been widely utilized in medical applications and problem-solving techniques. The convolutional neural network and its variants take into account the fully connected neurons with activation functions and use backpropagation to adjust weights [15].

Messay et al. developed a lung tumor detection system using thresholding and a rule-based approach [16]. Support Vector Machine (SVM) and rule-based approach have been implemented here to detect the false positive rate. Ant colony optimization approach has been used to find the edges and given as input to a circular neighbor algorithm to detect the lung nodule in. Bari et al. used a genetic algorithm to detect the relevant lung nodule feature for SVM classification [17]. Kumar et al. analyzed five algorithms for the detection of lung tumors [18]. The algorithms are k-means clustering, k-median clustering, inertia-weighted particle swarm optimization (IWPSO), particle swarm optimization (PSO), and guaranteed convergence particle swarm optimization (GCPSO). With the experimental analysis, they proved that for preprocessing adaptive median filter performs better and for classification GCPSO secured high accuracy of 95.89%.

The remaining section of this smart lung cancer detection is structured as follows: Section 2 discusses the related literature. Section 3 discusses the materials and methods. Experimented results are discussed in Section 4 and Section 5 concludes the proposed lung tumor detection system with its future work.

2 Related Work

The early detection of cancer may increase the chance of a patient's survival. Researchers are rapidly developing computational intelligence techniques to detect disease patterns with medical imaging. In recent research on medical imaging, artificial intelligence [19,20], pattern recognition [21], and computer vision have been applied to data visualization concepts. To extract the abnormalities from CT images,

computer-aided diagnosis is a challenging task for radiologists to take the necessary actions based on their decision-making. This section reviews relevant literature regarding lung tumor detection.

For lung CT image segmentation, Kumar et al. developed a hybrid 2D Otsu method with a modified artificial bee colony approach [18]. Uzelaltinbulat et al. applied thresholding and morphological operations to segment lung tumors [22]. Joon et al. utilized k-means clustering with fuzzy c-means to identify cancerous and non-cancerous lung nodules using texture and structural features [23]. Wang et al. used a multilevel image thresholding segmentation method and the best exponential k gravitational search method [24]. Prabukumar et al. developed a hybrid method for segmentation with texture, statistics, and geometrical features called Fuzzy C Mean with Region Growing Method [25]. The features have been enhanced by using a cuckoo search and an SVM classifier has been used to obtain an accuracy of 98.5%. Recently, Naqi et al. proposed a hybrid method called Active Contour Model (ACM), combining 3D neighborhood connectivity with 3D geometric properties for nodule detection [26]. Comparative analysis with naïve Bayes, SVM, and k-nearest neighbors (KNN) is performed., Zhou et al. proposed a novel computer-aided diagnosis method using the Convolutional Neural Network (CNN) for the detection of lung tumors on PET imaging [27]. The system has been evaluated for recall, precision, and F-score.

For lung CT image segmentation, Kumar et al. [18] developed a hybrid 2D Otsu method with a modified artificial bee colony approach. Uzelaltinbulat et al. applied thresholding and morphological operations to segment lung tumors [22]. Joon et al. utilized k-means clustering with fuzzy c-means to identify cancerous and non-cancerous lung nodules using texture and structural features [23]. Wang et al. used a multilevel image thresholding segmentation method and the best exponential k gravitational search method [24]. Prabukumar et al. developed a hybrid method for segmentation with texture, statistics, and geometrical features called Fuzzy C Mean with Region Growing Method [25]. The features have been enhanced by using a cuckoo search and an SVM classifier has been used to obtain an accuracy of 98.5%. Recently, Naqi et al. proposed a hybrid method called Active Contour Model (ACM), combining 3D neighborhood connectivity with 3D geometric properties for nodule detection [26]. Comparative analysis with naïve Bayes, SVM, and k-nearest neighbors (KNN) is performed. Zhou et al. proposed a novel computer-aided diagnosis method using the Convolutional Neural Network (CNN) for the detection of lung tumors on PET imaging [27].

3 Proposed Materials and Method Descriptions

Fig. 1 illustrates the materials and proposed methodologies of the computer-aided diagnosis (CAD) system for the smart detection of lung tumors. This system has five steps, including CT image preprocessing, Image segmentation using thresholding methods and morphological operations, Feature extraction by using GLM, GLRLM, and HOG, optimization with W-BFOA and, Classification by using Dual Graph CNN (DGCNN).

3.1 Image Pre-Processing

Based on 150 CT images of tumorous lung and non-tumor patients obtained by the NCI Lung Cancer Database Consortium, the proposed CAD system called smart lung tumor detection has been implemented [28]. Preprocessing of the input image is considered a significant phase in research related to clinical applications of neuroimaging for improving image quality. In this paper, we perform preprocessing steps such as normalization, noise removal, image enhancement, and augmentation of the input image to produce an improved image,

It has been found that an adaptive median filter (AMF) coupled with a Wiener filter (WF) can be used to remove noise from a normalized input image with a minimum mean square error. An adaptive median filter can address some of the issues associated with the median filter (MF), including how it is only effective on

salt and pepper noise, inefficient when the spatial density of the noise is high, and does not provide proper smoothing for large kernel sizes. Unlike MF, Adaptive MF employs a variable kernel size, and all pixels are not replaced with the median value. It consists of two steps, namely (I) identifying the median value for the kernel, and (II) determining whether the current pixel value is salt and pepper noise (impulse).

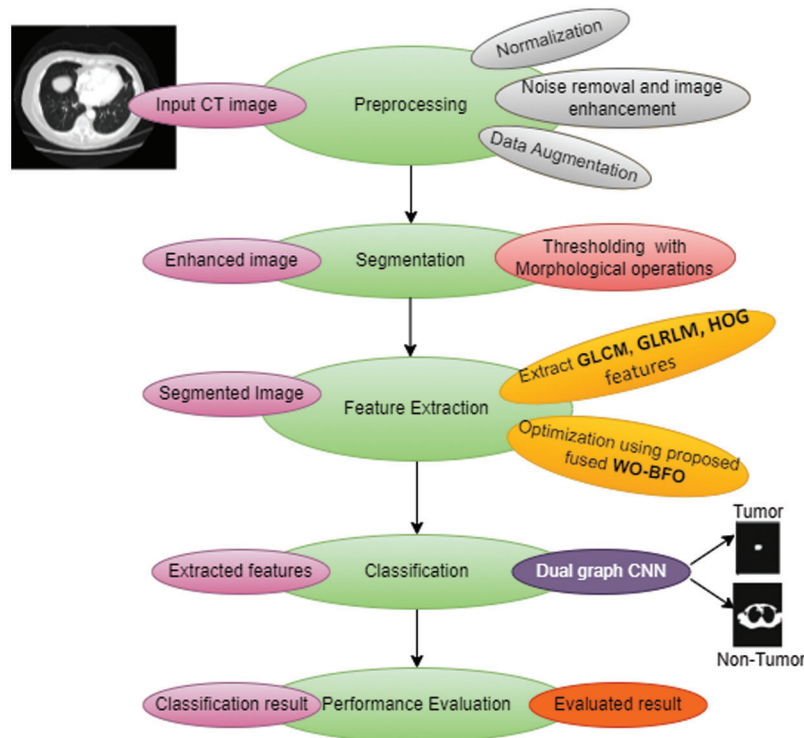


Figure 1: Overview of proposed lung-tumor prediction system

Using AMF, the denoise image is further improved using WF with minimum mean square error. In Eq. (1), WF is a statistical method to reduce image blurring and smoothen the image effect. Fig. 2 illustrates the enhanced CT image using AMF and WF.

$$F(m) = \sum_{i=0}^N A_i w[m - i] \quad (1)$$

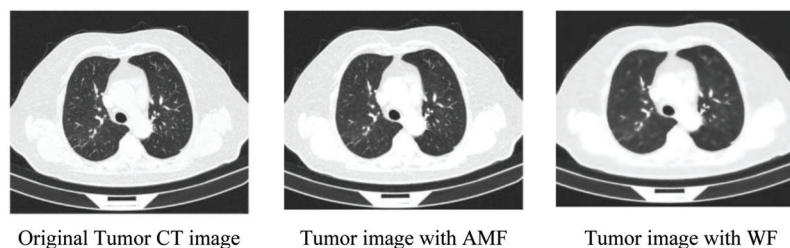


Figure 2: Filtered CT images

To avoid the overfitting issue and satisfy the training process, network augmentation is being utilized. Image segmentation involves the process of separating images into parts based on objects or other relevant

information. Using this method, the image becomes useful data for further analysis. Based on Shakeel et al., the purpose of image segmentation is to identify the boundaries and objects within the image [29]. The proposed work uses thresholding to segment CT images, while morphological operations are used to smooth the region of interest.

One of the simplest forms of segmentation is the threshold which segments grayscale images into binary images. If a pixel value is greater than the threshold value, that pixel is segmented as an object, otherwise as a boundary. Morphological mathematical operations are applied to the possible locating of a particular structuring element to smoothen the region of interest. Assume D is the binary image and B is the structuring element then the morphological operations such as Erosion, Dilation, Opening and Closing are depicted in Eqs. (2)–(5) respectively. Fig. 3 illustrates the thresholding-morphological operation-based segmentation.

$$\text{Erosion: } D \ominus B = \{A|(B)_A \subseteq D\} \quad (2)$$

$$\text{Dilation: } D \oplus B = \{A|(B)_A \cap D \neq \emptyset\} \quad (3)$$

$$\text{Opening: } D \ominus B = D \ominus B \oplus DB \quad (4)$$

$$\text{Closing: } D \oplus B = D \oplus B \ominus B \quad (5)$$

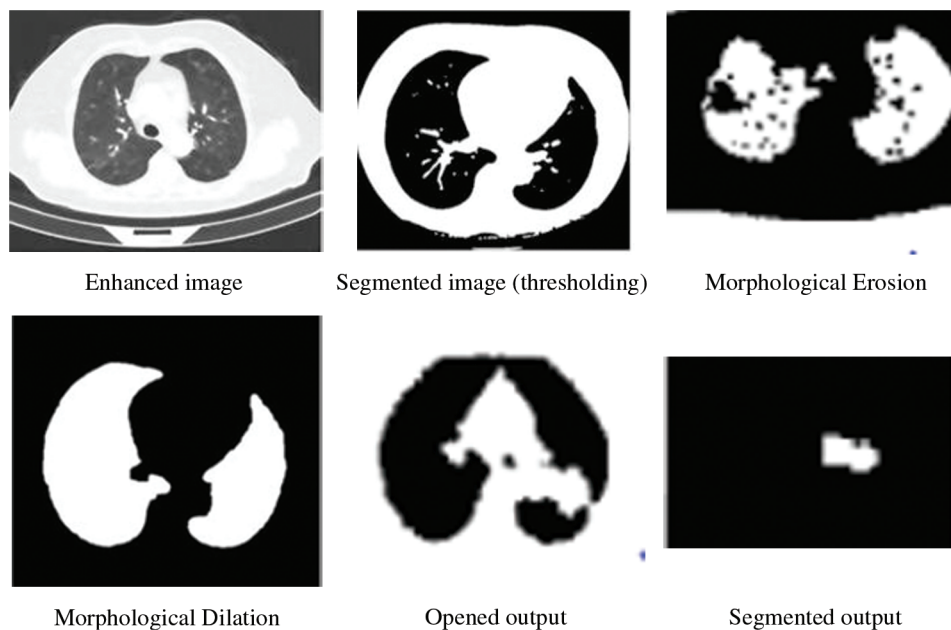


Figure 3: Thresholding-morphological operation based segmentation

3.2 Feature Extraction

After segmentation, lung nodule pattern information is obtained in the feature extraction phase. From each segmented nodule, this proposed feature extraction method extracts various geometrical, texture, statistical and structural features. The techniques such as GLCM, GLRLM, and HOG are used for feature extraction. GLCM is a second-order statistics approach that considers the relationship between couples of pixels. GLRLM obtains a higher-order statistical method that extracts the features with gray levels pixels that are occurring continuously. Metaheuristic approaches mimic the biological phenomenon of nature to

solve the real work applications. In this study, whale optimization and bacterial foraging optimization are combined to select the optimal features to improve the prediction of lung tumor outcomes. Standard Whale Optimization Algorithm (WOA) simulates the behavior of predation by whales through swarm intelligence.

In the WOA approach, first, the prey's location and its surrounding are recognized by the whale. In some situations, the whale didn't find the prey's location. In that case, the current possible prey position is the target prey and the remaining individuals in the prey are moving to that optimal position. Once the best possible solution is found, other individuals also update their positions. This behavior is mathematically expressed as follows in Eqs. (6) and (7).

$$X(t+1) = X^*(t) - P.D \quad (6)$$

$$D = |LX^*(t) - X(t)| \quad (7)$$

where, t : current no. of iterations, $X^*(t)$: vector of prey current position, $X(t)$: vector of prey position, $| |$: an absolute value, and P and L -vectors representing the size of surroundings (or) coefficients. The vectors P and L are calculated as follows in the Eqs. (8) and (9).

$$P = 2a \cdot r - a \quad (8)$$

$$L = 2r \quad (9)$$

where, r : a random number in the range $[0, 1]$, a : control parameter, and for the number of iterations it decreases the value in the range 2 to 0 as stated in Eq. (10), and T_m : maximum number of iterations.

$$a = 2 - \frac{2t}{T_m} \quad (10)$$

The humpback whale's bubble net foraging is within the encirclement as shown in Fig. 4, the whale is moving toward the prey along the spiral path. This behavior is mathematically modeled with two steps such as the shrinking inspiring method and spiral update position respectively.

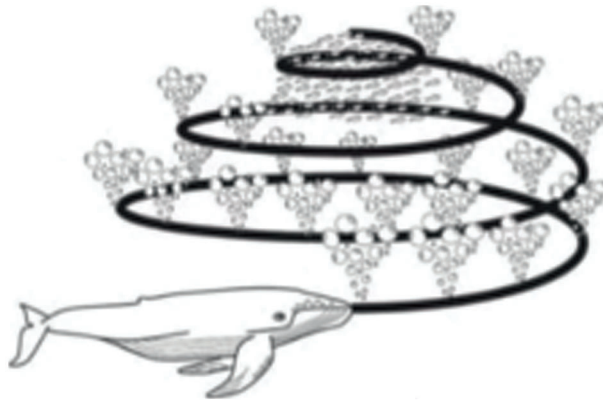


Figure 4: Bubble-net foraging of the whale [30]

In the first step, the convergence factor is reduced using the Eqs. (8)–(10). The second step to updating the position is to calculate the distance from the individual whale to the current position. Simulate the whale to capture food in the spiral mathematically as in Eq. (11).

$$X(t+1) = D^l \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \quad (11)$$

$$D^l = |X^*(t) - X(t)| \quad (12)$$

where D^l : the distance between whale i to the current position, b : constant to define the spiral logarithmic form, l : a random number between $[-1, 1]$. Spiral and contraction envelopment are performed with equal probability to obtain the synchronous model.

Randomly select the whale with the current position if $|A| \geq 1$. It will make the whale far away from the current target and global exploration characteristics are enhanced, it will find the best prey position to replace the current whale as stated in Eq. (13)

$$X(t+1) = X_{rnd} - P.L \quad (13)$$

$$D = |L.X_{rnd} - X(t)| \quad (14)$$

where rnd : random number (vector) of the whale position.

The bacterial Foraging Optimization Algorithm (BFO) was proposed in 2002 to imitate the food searching attitude of a swarm of bacteria. Searching for the nutrients called global solution by bacteria in n -dimensional space. This will use three operations such as chemotaxis, reproduction, and elimination-dispersal events. The bacteria find the best solution with the help of the locomotive flagella linked to it. The swarm is defined that flagella rotation in a counter-clock direction that takes action to bacteria in a forwarding direction. If the flagella rotation is clockwise then each flagellum is behaving independently in a random direction. During the chemotactic progress, bacteria alternate the modes as tumble and run. Tumble is modeled as single chemotactic progress by a run. Among the total population called 'B' bacteria, $\theta(p, c, rp, e)$ denotes the current position of p^{th} bacteria during c^{th} chemotactic, rp^{th} reproduction, and e^{th} elimination. The location of the succeeding bacteria is exhibited as in Eq. (15).

$$\theta(p, c+1, rp, e) = \theta(p, c, rp, e) + C(p, c) \cdot \frac{\phi(p, c)}{\sqrt{\phi^T(p, c) \cdot \phi(p, c)}} \quad (15)$$

where $C(p, c)$: run-length unit of p^{th} bacteria and $\phi(p, c)$: vector of random direction in the range $[-1, 1]$. If p^{th} bacteria improve the fitness function $J(p, c, rp, e)$ as the next step of chemotactic, the bacteria take a similar direction move or else bacteria tumble in a random direction to find the best possible solution. The number of runs is limited to the initialized count Nr . Nutrients searched for Nc chemotactic steps. After this bacteria move to the reproduction phase to work out each bacteria's health with its ability to locate the food during its life span. For each foraging bacteria's health is modeled as in Eq. (16).

$$E_{health}^p = \sum_{c=1}^{Nc+1} E(p, c, rp, e) \quad (16)$$

All 'B' bacteria are ordered based on their health value in ascending order. To make the population of bacteria constant, at least $B/2$ healthy bacteria gets replaced with the remaining bacteria that are healthier. After Nrp reproduction steps, based on the probability of P_{ed} , some of the bacteria are forcefully moved to a random search space which will avoid the local optimal solution to provide the best possible solution in the search area.

Fused WBFO is the novel optimization algorithm that works on both WO and BFO optimization algorithms. Based on WO's best possible solution finding strategy using the bubble net foraging approach, a globally optimal solution is found from the swam population. With this idea to give space to bacteria, the bacteria optimal solution is compared with the global optimal solution in chemotaxis

iteration. The bacteria swim until the best possible solution is found. In BFO, the chemotaxis is the major step and WO is used here to update the arbitrary variables.

Based on the energy, bacteria are sorted in ascending order. Due to a lack of energy, some of the bacteria are dying. The remaining are inherited from the bacterial position of their parents. The bacteria are dispersed as per the Eq. (17)

$$Set(Bacteria_i) = \frac{E_{max}(Bacteria) - E(Bacteria_i)}{E_{max}(Bacteria) - E_{min}(Bacteria)} * P_{ed} \quad (17)$$

The workflow of the proposed optimization algorithm is shown in Fig. 5. Until the maximum number of iterations is reached, the parameters, location, and fitness functions are updated. If the termination condition and elimination dispersion are reached then the best solution is returned. For chemotaxis count, check the parameter value of P, if it is greater than 1 then the current position and bacteria dispersion are calculated, or else the position of the whale is updated. The optimal feature subset using the proposed WBFO has been given as input to the classification phase for the detection of lung tumors.

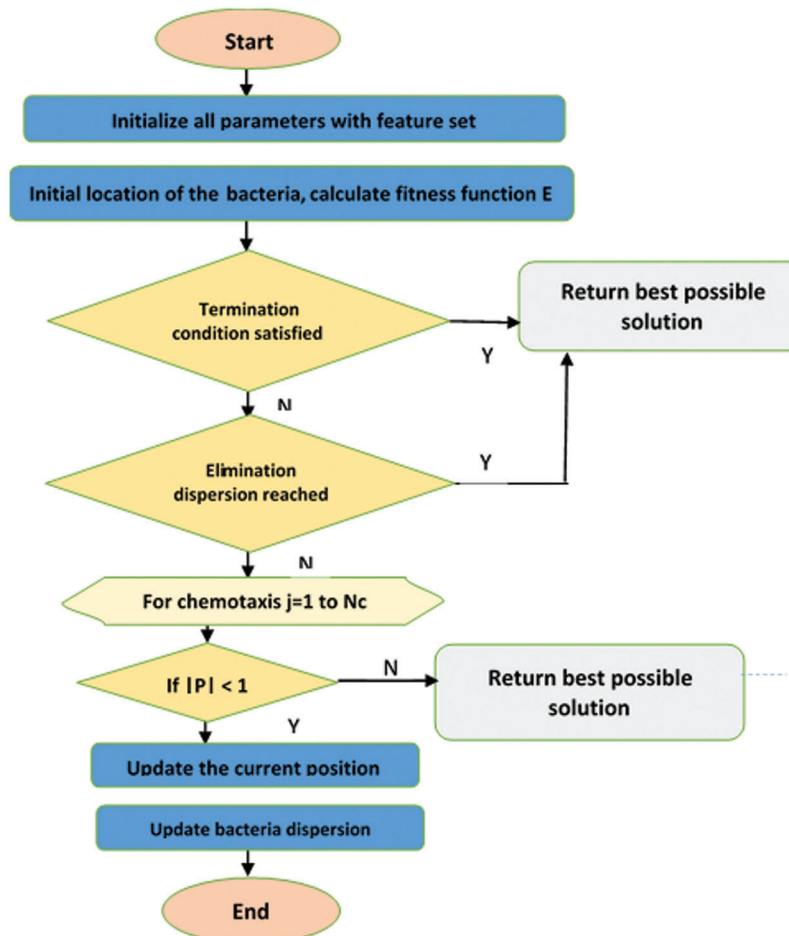


Figure 5: Proposed WBFO workflow

Each node of the NN graph is represented with the features extracted. The graph local and global consistency is considered by dual graph CNN with the implementation of two CNN. In Dual

graph CNN, the dataset is represented as a matrix $D \in R^{m \times n}$, and data points are represented as $D = d_1, d_2, \dots, d_i, d_{i+1}, \dots, d_n$. For the initial points, the label y is represented as $y = y_1, y_2, \dots, y_l$. The structure of the graph is denoted as an adjacency matrix called $A \in R^{n \times n}$. Dual graph CNN is constructed with local and global consistency (GL) and ensemble. Local consistency (LC) called $conv_{lc}$ is a type of feed-forward network. Feature matrix F and adjacency matrix A are the input of this model. The output of LC is z for hidden layer 'i' is denoted in Eq. (18).

$$conv_{lc}^i(X) = z^i = \sigma\left(X^{-\frac{1}{2}}\bar{A}X^{-\frac{1}{2}}z^{i-1}w^i\right) \quad (18)$$

where, $\bar{A} = A \times I_n$, I-identity matrix, \bar{A} -adjacency matrix with self-loops, X -normalized adjacency matrix in normalized form, z -output, w -trainable parameter, and σ -activation function. The node feature vector is represented with the addition of the neighboring node feature vector.

The frequency matrix f_m is calculated as follows: The random path is chosen from the sequence of nodes visited by the random user. If the random user on the node x_i at time t , the state is described as $s(t) = x_i$. The transition from one node to the neighbor x_j is declared as, $r(s(t+1) = x_j | s(t) = x_i)$. These transition states are assigned to the given adjacency matrix A as Eq. (19). The random walk used here is to calculate the similarity between the nodes.

$$r(s(t+1) = x_j | s(t) = x_i) = A_{i,j} / \sum_j A_{i,j} \quad (19)$$

The frequency matrix f_m is created using the Eq. (3) for all the pairs of nodes. The path between the nodes is calculated using a random walk. Once the f_m is created, the i^{th} node of FM is the i^{th} row vector of f_{mi} and the j^{th} column of f_m is the j^{th} column vector of f_{mj} . it is also called context c_j . the entry of the frequency the matrix $f_{mi,j}$ is the number of times x_i occurs in c_j . then the f_m has been used to calculate the PPMI matrix $P \in R^{n \times n}$ as in the Eqs. (20)–(23).

$$P_{i,j} = \frac{f_{mi,j}}{\sum_{i,j} f_{mi,j}} \quad (20)$$

$$P_{i,*} = \frac{\sum_j f_{mi,j}}{\sum_{i,j} FM_{i,j}} \quad (21)$$

$$P_{*,j} = \frac{\sum_i f_{mi,j}}{\sum_{i,j} f_{mi,j}} \quad (22)$$

$$P_{i,j} = \max\{pmi_{i,j} = \log\left(\frac{P_{i,j}}{P_{i,*} P_{*,j}}\right), 0\} \quad (23)$$

where, $P_{i,j}$: the probability of node x_i occurring in context c_j , $P_{i,*}$: the probability of node x_i and, $P_{*,j}$: the probability of context c_j .

As compared to the adjacency matrix, the $PPMI$ matrix has reduced the hub nodes and increased the relationship among the data points. The global consistency convolution network $conv_G$ based on the $PPMI$ matrix is denoted in Eq. (24).

$$Conv_{gc}^i(D) = Z^i = \sigma\left(X^{-\frac{1}{2}}PX^{-\frac{1}{2}}z^{i-1}w^i\right) \quad (24)$$

where, the P - $PPMI$ matrix and $X_{i,i} = \sum_j P_{i,j}$

To combine the local and global consistency convolution for a dual graph convolutional network, a regularize is used. Loss function with the regularize is represented in Eqs. (25) and (26)

$$Loss = L_0(Conv_{lc}) + \lambda(t)L_{reg}(Conv_{lc}, Conv_{gc}) \quad (25)$$

$$L_0(Conv_{lc}) = \frac{1}{|y_L|} \sum_{l \in y_L} \sum_{i=1}^c y_{l,i} z_{l,i}^A \quad (26)$$

where y_L -labeled data for training and $Y \in R^{n \times n}$

4 Experimental Results and Analysis

The 150 lung CT images for tumor and non-tumor categories are collected from the lung cancer consortium for the proposed system evaluation. The input feature sets are split into a 7:3 ratio of training and testing datasets. The proposed system is implemented using python Sci-kit-learn and the results are compared with existing preprocessing filters, feature extraction methods, and classification systems in terms of accuracy, sensitivity, and specificity.

4.1 Evaluation of Preprocessing

In the preprocessing stage, the filters such as Adaptive median and wiener filter combination are used in the proposed system. This is compared with other filters to prove the performance of the proposed preprocessing system in terms of the Speckle Suppression Index (*SSI*) and Speckle Suppression and Mean Preservation index (*SMPI*). *SSI* value is a linear minimum mean-square error filter value, and it can be reassessed similarity between pixels as shown in Eq. (27). *SMPI* value determines the lower values that give better performance of the filter in terms of mean preservation and noise reduction as shown in Eqs. (28) and (29).

$$SSI = \frac{\sqrt{Var(I_f)}}{Mean(I_f)} * \frac{Mean(I_o)}{\sqrt{Var(I_o)}} \quad (27)$$

$$SMPI = Q * \frac{\sqrt{Var(I_f)}}{\sqrt{Var(I_o)}} \quad (28)$$

$$Q = 1 + |Mean(I_o) - Mean(I_f)| \quad (29)$$

where I_f is the filtered image and I_o is the noisy image. This *SSI* tends to be less than 1 if the filter performance is efficient in reducing the speckle noise. Tab. 1 shows the *SSI* and *SMPI* index values for sample 5 images in terms of filter results using proposed and existing filters. For image 3, the proposed combination of filters obtained *SSI* and *SMPI* values of 0.4141 and 0.5671 respectively. Other filters such as median, adaptive median, and Wiener filters obtained the *SSI* value as 0.5111, 0.5072, and 0.4901 respectively. *SMPI* values of these filters are 0.6401, 0.6721, and 0.6504 sequentially. The statistical measurements of the results prove that the proposed adaptive median with a Wiener filter obtained less value, which indicates the best filter to remove the noise compared to other filters.

4.2 Evaluation of Feature Extraction Method

The evaluation of the proposed feature extraction with optimization using WBFO is shown in Tab. 2. The optimized threshold value obtained by WBFO is compared with the standard Whale optimization algorithm (WOA) and Bacterial foraging optimization (BFO). Based on the proposed WBFO optimization, we selected first-order texture features kurtosis, skewness, and uniformity, and for gray level

matrix, we also selected Run-length Nonuniformity Low gray-level to run emphasis High gray-level run emphasis for Gray level run length matrix features.

4.3 Evaluation of Classification System

The proposed smart lung tumor detection using Dual graph CNN with an optimized feature set using WBFO has been evaluated with the parameters such as accuracy, sensitivity, and specificity. The proposed classification system with WBFO is compared with standard classification with traditional classifiers such as support vector machine (SVM), convolutional neural network (CNN), and Dual graph CNN without WBFO. The evaluated results are shown in [Tabs. 3–6](#). The overall comparative analysis in terms of the accuracy of the classification system is shown in [Fig. 6](#).

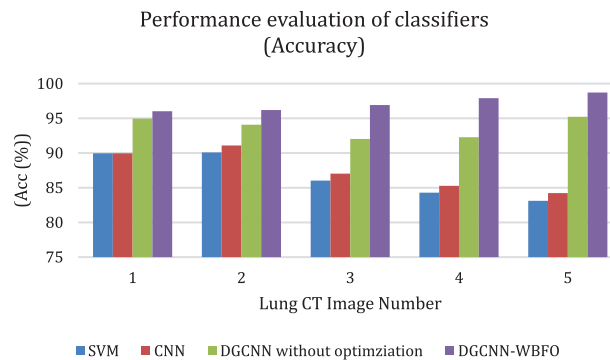


Figure 6: Accuracy comparison of classifiers

As a result of various evaluations, it has been demonstrated that the proposed lung cancer detection system is effective when the proposed WBFO-based optimization of feature selection is combined with a Dual Graph CNN classifier. Contrary to other classifiers such as SVM, CNN, and DGCNN without optimization, the proposed system ensured high accuracy. The accuracy of sample image 5 was determined by the machine learning approaches to be 83.11%, 84.22%, and 95.22%, respectively. The proposed optimization approaches enhance the result of the Dual graph CNN classifier and increase the accuracy percentage to 98.72%. Consequently, the proposed WBFO evolutionary approach improves the classification accuracy of the detection of lung tumors and non-lung tumor images.

The proposed DGCNN with WBFO secures the index percentage as shown in [Tab. 5](#). CNN achieved 87.91%, 87.87%, and 89.96% for TPR, TNR, and AUC for the sample CT images number 1. Whereas SVM achieved 85.91% 85.86% and 89.96% and DGCNN generated 92.9%, 90.12% and 96.02% respectively. The sample result analysis for lung tumors is shown in [Tab. 6](#). Thus, the proposed WBFO improves the similarity index values, thereby proving the efficiency of the proposed system. In all types of statistical analyses, the proposed WBFO enhances the lung CT image feature selection and classification system for efficient lung tumor prediction.

5 Conclusion

Early diagnosis of lung tumors is very important in determining the type of treatment that should be administered to the patient as well as taking preventative measures to lengthen the life of the patient. Using evolutionary and deep learning methods, we present a novel approach to the diagnosis of lung tumors. Preprocessed images are normalized and, adaptive median filters with wiener filters are applied to the input images. After the image has been normalized and enhanced, the tumor region is segmented using thresholding and morphological operations. To improve the accuracy of decision-making, the

various features of lung CT images are extracted using GLCM, GLRLM, and HOG to obtain statistical information on the features. To enhance the feature selection process, the fused WFO method is used. With the proposed system, lung tumor detection accuracy was evaluated at 98.72% using graph dual CNN. It is compared with support vector machines, standard CNNs, and DGCNNs without optimization. In comparison to existing algorithms, the proposed approach is shown to be effective through performance analysis using various metrics. To enhance the performance of the proposed system, various medical imaging and other meta-heuristic approaches will be added to the proposed work in the future.

Acknowledgement: The author thanks the Deanship of Research, Prince Sattam Bin Abdul-Aziz University, KSA, for providing an opportunity to conduct research.

Funding Statement: The author received no specific funding for this study.

Conflicts of Interest: The author declares that he has no conflicts of interest to report regarding the present study.

References

- [1] M. Kurkure and A. Thakare, "Introducing automated system for lung cancer detection using evolutionary approach," *International Journal of Engineering and Computer Science*, vol. 5, no. 5, pp. 16736–16739, 2016.
- [2] S. Makaju, P. Prasad, A. Alsadoon, A. K. Singh and A. Elchouemi, "Lung cancer detection using CT scan images," *Procedia Computer Science*, vol. 125, no. 1, pp. 107–104, 2018.
- [3] C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed *et al.*, "Going deeper with convolutions," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 1–9, 2015.
- [4] N. Mesanovic, M. Grgic, H. Huseinagic, M. Males, E. Skejic *et al.*, "Automatic CT image segmentation of the lungs with region growing algorithm," in *18th Int. Conf. on Systems, Signals and Image Processing-IWSSIP*, Maribor, Slovenia, pp. 395–400, 2011.
- [5] J. John and M. G. Mini, "Multilevel thresholding based segmentation and feature extraction for pulmonary nodule detection," *Procedia Technology*, vol. 24, pp. 957–963, 2016.
- [6] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *IEEE Access*, vol. 7, pp. 8975–9000, 2019.
- [7] C. Zu, Y. Wang, L. Zhou, L. Wang and D. Zhang, "Multi-modality feature selection with adaptive similarity learning for classification of Alzheimer's disease," in *2018 IEEE 15th Int. Symp. on Biomedical Imaging ISBI 2018*, Washinton, USA, pp. 1542–1545, 2018.
- [8] J. Nalepa and M. Kawulok, "Selecting training sets for support vector machines: A review," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 857–900, 2019.
- [9] D. C. R. Novitasari, A. Lubab, A. Sawiji and A. H. Asyhar, "Application of feature extraction for breast cancer using one order statistic, GLCM, GLRLM, and GLDM," *Advances in Science, Technology and Engineering Systems Journal (ASTESJ)*, vol. 4, no. 4, pp. 115–120, 2019.
- [10] D. Zhang, J. Hu, F. Li, X. Ding, A. K. Sangaiah *et al.*, "Small object detection via precise region-based fully convolutional networks," *Computers, Materials and Continua*, vol. 69, no. 2, pp. 1503–1517, 2021.
- [11] J. Zhang, J. Sun, J. Wang and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8427–8440, 2021.
- [12] A. Teramoto, H. Fujita, O. Yamamuro and T. Tamaki, "Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique," *Medical Physics*, vol. 43, no. 6 Part 1, pp. 2821–2827, 2016.
- [13] N. Zheng, G. Loizou, X. Jiang, X. Lan and X. Li, "Computer vision and pattern recognition," *International Journal of Computer Mathematics*, vol. 84, no. 9, pp. 1265–1266, 2007.

- [14] S. Vijh, D. Gaur and S. Kumar, "An intelligent lung tumor diagnosis system using whale optimization algorithm and support vector machine," *International Journal of System Assurance Engineering and Management*, vol. 11, no. 2, pp. 374–384, 2020.
- [15] P. Kavitha and G. Ayyappan, "Lung cancer detection at early stage by using SVM classifier techniques," *Int J Pure Appl Math*, vol. 119, no. 12, pp. 3171–3180, 2018.
- [16] T. Messay, R. C. Hardie and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Medical Image Analysis*, vol. 14, no. 3, pp. 390–406, 2010.
- [17] M. Bari, A. Ahmed, M. Sabir and S. Naveed, "Lung cancer detection using digital image processing techniques: A review," *Mehran University Research Journal of Engineering & Technology*, vol. 38, no. 2, pp. 351–360, 2019.
- [18] S. Kumar, T. K. Sharma, M. Pant and A. K. Ray, "Adaptive artificial bee colony for segmentation of CT lung images," *Int J Comp App iRAFIT*, vol. 5, pp. 1–5, 2012.
- [19] G. T. Herman, "Image reconstruction from projections," *Real-Time Imaging*, vol. 1, no. 1, pp. 3–18, 1985.
- [20] U. Reddy, B. Reddy and B. Reddy, "Recognition of lung cancer using machine learning mechanisms with fuzzy neural networks," *Traitement du Signal*, vol. 36, no. 1, pp. 87–91, 2019.
- [21] S. Vijh, S. Sharma and P. Gaurav, "Brain tumor segmentation using OTSU embedded adaptive particle swarm optimization method and convolutional neural network," in *Data Visualization and Knowledge Engineering*, Springer, Cham, pp. 171–194, 2020.
- [22] S. Uzelaltinbulat and B. Ugur, "Lung tumor segmentation algorithm," *Procedia Computer Science*, vol. 120, pp. 140–147, 2017.
- [23] P. Joon, S. B. Bajaj and A. Jatain, "Segmentation and detection of lung cancer using image processing and clustering techniques," in *Progress in Advanced Computing and Intelligent Engineering*, Springer, Singapore, vol. 713, pp. 13–23, 2019.
- [24] J. Wang, Y. Wu, S. He, P. K. Sharma, X. Yu *et al.*, "Lightweight single image super-resolution convolution neural network in portable device," *KSI Transactions on Internet and Information Systems (TIIS)*, vol. 15, no. 11, pp. 4065–4083, 2021.
- [25] M. Prabukumar, L. Agilandeswari and K. Ganesan, "An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, pp. 267–293, 2019.
- [26] S. M. Naqi, M. Sharif and I. U. Lali, "A 3D nodule candidate detection method supported by hybrid features to reduce false positives in lung nodule detection," *Multimedia Tools and Applications*, vol. 78, no. 18, pp. 26287–26311, 2019.
- [27] S. R. Zhou and B. Tan, "Electrocardiogram soft computing using hybrid deep learning CNN-ELM," *Applied Soft Computing*, vol. 86, pp. 1–11, 2020.
- [28] C. Bhuvaneshwari, P. Aruna and D. Loganathan, "A new fusion model for classification of the lung diseases using genetic algorithm," *Egyptian Informatics Journal*, vol. 15, no. 2, pp. 69–77, 2014.
- [29] P. M. Shakeel, M. A. Burhanuddin and M. I. Desa, "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks," *Measurement*, vol. 145, pp. 702–712, 2019.
- [30] Y. G. Kim, S. M. Lee, K. H. Lee, R. Jang, J. Seo *et al.*, "Optimal matrix size of chest radiographs for computer-aided detection on lung nodule or mass with deep learning," *European Radiology*, vol. 30, no. 9, pp. 4943–4951, 2020.

Appendix A:**Table 1:** Evaluation result for preprocessing noise removal filters

Sample CT image number	SSI				SMPI			
	Median filter	Adaptive median filter	Wiener filter	Proposed adapt	Median filter	Adaptive median filter	Wiener filter	Proposed adaptive median
1	0.6512	0.6362	0.6102	0.5311	0.7501	0.7916	0.7182	0.6711
2	0.6321	0.6282	0.6121	0.5311	0.7621	0.8036	0.7302	0.6891
3	0.5111	0.5072	0.4901	0.4141	0.6401	0.6721	0.6504	0.5671
4	0.6712	0.6672	0.6501	0.5641	0.8011	0.8326	0.7692	0.7261
5	0.6212	0.6163	0.7002	0.5241	0.7512	0.7816	0.7183	0.6762

Table 2: Selected optimal feature subset using proposed WBFO optimization

Sum average	Long run emphasis	Run-length non- uniformity	Low gray-level run emphasis	High gray- level run emphasis	Mean	Skewness	Kurtosis	Class
2.31	18231.12	97.43	101.332	5782.123	1.14	2.10	4.29	0
2.01	27918.29	162.38	98.167	5317.27	1.06	2.57	8.26	0
2.13	25262.22	163.48	99.267	5318.37	1.72	3.67	9.36	0
2.01	30562.18	80.27	78.286	4261.81	1.02	2.27	11.27	0
3.21	30563.38	81.47	79.486	4263.01	2.22	3.47	12.47	0
2.03	30562.20	80.29	78.309	4261.833	1.05	2.30	11.30	0
2.10	30562.27	80.36	78.379	4261.903	1.12	2.37	11.37	0
2.18	110188.2	198.27	48.26	2736.29	1.08	5.28	26.26	1
2.09	110188.0	198.08	48.07	2736.1	0.89	5.09	26.07	1
2.01	1001298.	196.06	46.051	2734.081	1.02	3.07	24.05	1
2.01	1223231.	280.28	27.187	2461.18	1.09	7.28	71.27	1

Table 3: Classification result of SVM classifier

#Sample CT image	SVM classification (%)				
	TPR (SN)	TNR (SP)	FPR	FNR	ACC
1	85.91	85.86	8.54	9.84	89.96
2	88.31	91.11	10.88	12.65	90.08
3	84.28	88.08	11.41	13.61	86.03
4	82.01	87.17	11.98	13.88	84.28
5	84.29	88.21	12.02	13.92	83.11

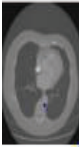
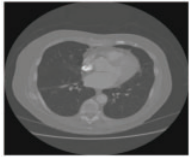
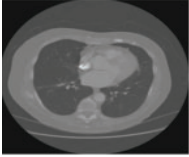




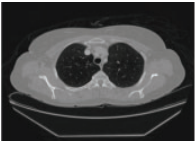
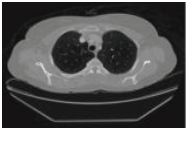



Table 4: Classification result of CNN classifier

CNN classification (%)					
#Sample CT image	TPR (SN)	TNR (SP)	FPR	FNR	ACC
1	87.91	87.87	7.54	8.84	89.96
2	89.31	92.11	9.88	11.65	91.08
3	85.28	89.08	10.41	12.61	87.03
4	83.01	88.17	10.98	12.88	85.28
5	85.39	89.21	11	12.82	84.22

Table 5: Classification result of Dual graph CNN with proposed WBFO optimization

Dual graph CNN with proposed WBFO (%)					
#Sample CT image	TPR (SN)	TNR (SP)	FPR	FNR	ACC
1	92.9	90.12	0.012	3.81	96.02
2	91.32	95.21	0.021	5.54	96.2
3	93.81	92.63	0.018	5.32	96.91
4	95.02	96.72	0.0011	3.72	97.9
5	96.91	95.022	0.0001	4.21	98.72

Table 6: Sample lung nodule detection as tumor or non-tumor

Input image	Normalization	Filters	Thresholding	Morphological opening & closing	Detection of lung nodule
					 Tumor
					 Non-tumor