Journal of Information Systems Engineering and Management
2025, 10(15s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Lung Cancer Segmentation Using Improved Golden Eagle Optimization with Clustering Model

[1]Sandhya L , [2]Dr.Md.Sameeruddin Khan , [3] Dr. R.Chinnaiyan

[1] Assistant Professor School of CSE and IS,Presidency University, Bengaluru,-560064 Karnataka, India.
[1]saisandhyalax@gmail.com
[2] Professor & Dean School of CSE and IS,Presidency University, Bengaluru-560064, Karnataka, India.
[2] deancse@presidencyuniversity.in
[3] Professor School of CSE Alliance University, Bengaluru-562106 Karnataka, India.
[3]vijayachinns@gmail.com

ARTICLE INFO

Received: 04 Dec 2024

Revised: 22 Jan 2025

Accepted: 04 Feb 2025

ABSTRACT

Introduction: In recent times, lung cancer has emerged as a widespread and concerning ailment, significantly impacting global health. Early detection is currently recognized as the most effective strategy to enhance the survival rate among cancer patients. The conventional methods for diagnosing lung cancer are time-consuming.

Objectives: To develop a Computer Aided Diagnosis (CAD) based lung nodule segmentation and classification.

Methods: Initially, the input lung images are pre-processed by the median filtering. Then, the lung nodules are segmented using the Enhanced Local Information Weighted Intuitionistic with Fuzzy C-means clustering) (ELWI-FCM) with Improved Golden Eagle Optimization) (IGEO) algorithm used for lesion segmentation. The IGEO is the combination of GEO and OBL (opposition based learning). Finally, the Deep Learning (DL) model DenseNet201 is utilized for classifying the lung nodules as normal and abnormal classes.

Results: The experimentation is carried out on the LIDC-IDRI dataset and achieved better accuracy and dice values of 98.7% and 98.9% respectively.

Conclusions: This work presents an automatic earlier detection of lung cancer thus reducing the risks of deaths.

Keywords: Lung Cancer, Computer Aided Diagnosis, Adaptive Golden Eagle Optimization, Densenet201

INTRODUCTION

The emergence of abnormal tissues within the human body leads to the condition known as cancer. Throughout history up to the present day, cancer remains a prevalent and formidable disease that poses a threat to human life. Lung cancer is considered as one of the deadliest disorders [1]. The timely prediction of early stages in cancer can significantly impact and potentially save numerous lives, especially when dealing with tumors in their initial phases.[2] [3]. The development of an early cancer detection technique that can be used to lessen the effects of lung cancer is desperately needed [4][5]. Detecting tumor cells becomes challenging because of the misinterpretation of anatomical structures and the intensity variations in CT scan images [6]. Despite the development of different models for lung cancer detection, achieving excellent and accurate results remains a substantial challenge. Therefore, effective detection of lung cancer can be achieved through the application of image processing approaches [7].

Recent studies in lung CT image segmentation have primarily concentrated on developing segmentation methods that are precise and efficient [8][9]. The utilization of DL as a representation learning model for acquiring hierarchical representation of features has proven to be a significant advancement. The major benefit of employing DL lies in its capacity to generate high level feature representations from image features [10]. Some of the limitations in the existing works are: The conventional thresholding encounters challenges when applied to the bronchus and trachea due to their same grey values, resulting in suboptimal performance [11]. The segmentation is impacted by the similarity in visual behavior among lung nodules it influences the overall performance [12].

This work aims to introduce a method for identifying benign and malignant nodules in CT images. Influenced by optimizing hyperparameters and visualization techniques, we have innovatively implemented a hybrid approach which incorporates an optimized clustering to fine-tune the model's hyperparameters.

Objectives

- To present an efficient pre-processing, segmentation and classification approaches.
- To efficiently segment the lesions using the ELWI-FCM with the IGEO algorithm.
- To overcome convergence issues of standard optimizer, the IGEO is presented.
- To classify the segmented regions as normal and abnormal classes, the DL model DenseNet201 is presented.
- To perform comparative analysis for different DL approaches with respect to the LIDC-IDRI dataset.

RELATED WORKS

Literature works with respect to the lung tumor prediction using different models and the performance achieved are discussed:

Jain et al. [13] presented Generative Adversarial Network (GAN) with Salp Shuffled Shepherd optimizer (SSSO) for segmenting the lung nodules. The integration of GAN with SSSO achieved a better accuracy of 0.93 and a jaccard of 0.80. Navaneetha Krishnan et al. [14] presented CNN with a bat deer hunting model for segmenting the lung nodules. Active contour was used for segmenting the lung lobe and the hybrid segmentation approach fuzzy segNet was used for segmenting the nodules. The accuracy, specificity and sensitivity values achieved were 0.92, 0.89 and 0.94 respectively.

Ren et al. [15] developed the Lung Cancer Data Augmented Ensemble (LCDAE) model to address the issues of overfitting and poor performance in the classification tasks related to lung cancer. Three data augmentation steps like initially, the GAN was utilized for synthesizing images. Then, the data augmentation based ensemble method and hybrid data augmentation were utilized. This approach proved that by utilizing different augmentation the model has less over-fitting issues. Atiya et al. [16] presented non small cell lung cancers classification model using the TL (transfer learning) based CNN model. Initially, the images were resized and normalized and the augmentation process was carried out. Then, the two stages TL based pre-trained were used for classifying different stages of cancer. Finally, the accuracy value achieved by the ResNet50 was 94%.

PROPOSED METHODOLOGY

The advancement of the CAD has played a crucial role in medical analysis, aiding in decision-making regarding diseases of humans. Conventional models for predicting lung cancer faced challenges in accuracy due to poor quality images affecting the segmentation procedure. This paper introduces a novel, optimized approach segmentation model for lung cancer prediction. The collected images undergo pre-processing stage for noise elimination and to enhance the overall lung image quality. Subsequently, the cancerous part is segmented from the noise cleared image using an ELWI-FCM- IGEO. Finally, the features are extracted and classified in the DL model DenseNet201.

Pre-processing

Initially, the Median Filtering (MF) is utilized for removing noise in the LIDC-IDRI dataset. MF serves as a widely adopted technique to filter noise, operating as the low pass filter that retains information about images while effectively eliminating noise. This method involves filtering a neighborhood L with dimensions $m \times n$ by organizing all neighboring elements in ascending order and selecting the center component from the ordered sequence. Subsequently, this chosen middle element replaces the pixel in the center.

Optimal segmentation

In the segmentation process, the lung images are divided into different categories and this process is performed by the algorithm ELWI-FCM-IGEO. In ELWI-FCM, local detail weight k_{lm} is introduced and it influences set of local detail on clustered outcomes. This overcomes local detail relies in less noise regions and avoids the impacts of membership function u_{lm}.

The objective term $\hat{I}(U,V)$ and k_{lm} is given as:

$$\hat{I}(U,V) = \sum_{l}^{c}\sum_{m}^{p} \hat{u}_{lm}\, d^2(x_m, v_l) + \sum_{l=1}^{c} \pi_l^* \exp(1-\pi_l^*) + k_{lm} G_{lm} \quad (2)$$

where x_m is the object, v_l is the center of cluster, c is the cluster, $d(x_m, v_l)$ is the distance of x_m and v_l. Then, the values of k_{lm} is computed by:

$$k_{lm} = \frac{\sigma_m^2 + \alpha}{\sigma^2 + \alpha} \quad (3)$$

where σ_m^2 is the data's variance in the m^{th} window and α is the small parameter. The values of u_{lm} is computed by:

$$u_{lm} = \frac{1}{\sum_{s=1}^{c}\left(\frac{\| x_m - v_l \|^2 + G_{lm}}{\| x_m - v_s \|^2 + G_{lm}}\right)} \quad (4)$$

Here, the optimized term G_{lm} is given as:

$$G_{lm} = \sum \frac{(1-u_{lk})^n \| x_m - v_s \|^2}{d_{mk} + 1} \quad (5)$$

where n is the fuzzified term and k is the is the neighbour pixels in the m^{th} window.

Finally, the membership function of ELWI-FCM is given as:

$$\hat{u}_{lm} = u_{lm}{}^n + u_{lm}^*{}^n \quad (6)$$

This work presents a metaheuristic algorithm IGEO is presented for optimizing the clustering performance. The fitness function of the IGEO is given as:

$$Fitness = Max(Dice\ value) \quad (7)$$

The mathematical modelling of the IGEO is presented in this section. The proposed IGEO is a novel approach to addressing global optimization problems, and draws inspiration and mathematical formulation from the intelligent behavior of GE (golden eagles), particularly their adept control over the speed of their spiral flight patterns. In emulating the distinctive characteristics of GE, known for their unique swarm behavior during the initial stages of hunting, IGEO efficiently navigates and explores the solution space. The algorithm's effectiveness lies in its adept manipulation of two key components: cruise propensity and attack propensity. GE employs a strategic hunting approach by systematically exploring various areas in search of superior prey. Central to their hunting methodology is a notable characteristic: their possession of a better memory. This unique attribute enables GE to remember and recall information related to both cruise and attack propensities while in flight.

Algorithm 1: Pseudocode of the ELWI-FCM- IGEO
Input: x_m (object), v_l (center of cluster), n (fuzzified term) Population size, Iterations, q_a and q_c
Output: Optimal v_l
Initializing the OBL on the population of GE
Estimate the fitness by expression (7)
Estimate q_a and q_c
for every iteration
Update m_a (initial bias) and m_c (final bias) by expressions (15) and (16)
for all GE j
Randomly define the prey from population memory
Define $\vec{A}_j$ by expression (9)
If length of $\vec{A}_j$ is $\neq 0$
Define C_j by expression (12)
Define Δy_j by expression (13)
Calculate the fitness value for the newly generated positions
If the fitness value exceeds the memory
Utilize the newly calculated position value instead of the position stored in the GE memory.
end
end
end
end

Classification

Finally, for classifying normal and abnormal classes, the DL model DenseNet201 is utilized. DenseNet201 has a similar structure to ResNet, but its notable distinction lies in its approach to information flow across layers. In this network, each layer is directly connected to every other layer in a feed-forward manner. This design maximizes the information exchange among network layers, fostering dense connections throughout the model architecture. That is, the DenseNet's layers are interconnected in a way where every layer receives input from all preceding layers and fed its output to all subsequent layers. This design strengthens the robust information flow throughout the network. DenseNet201 offers numerous benefits, including addressing the gradient vanish issue, enhancing the propagation of features, facilitating feature reusability, and significantly reducing parameter count. Consequently, this network is more lightweight and efficient in terms of computation and memory usage. The architecture of DenseNet201 typically comprises 4-layers of convolutional layers, 3-TL (transition layers), and 1-FC (fully connected) layer and softmax layer.

RESULTS ANALYSIS

The segmentation results of the proposed ELWI-FCM-IGEO are performed in Python. The experimental parameters like batch size (100), epochs (50), maximum iteration (200) and initial population (50) are considered. The experimental outcomes are validated on 10-cross validation.

The LIDC-IDRI [21] is a comprehensive repository of diagnostic and lung tumor monitoring thoracic CT modalities, meticulously annotated with tumor. This dataset comprises subjects of 1018.. The lesions are divided into three distinct classes like nodule ≥3 mm, non-nodule ≥3 mm and nodule <3 mm. Ground truth involves a comprehensive annotation process carried out by expert radiologists.

Experimental measures like Accuracy, Kappa, Jaccard and Dice are computed for evaluating the segmentation performance. These measures are computed using the Ts (true positive), To (true negative), Fs (false positive) and Fo (false negative).

Image analysis

The following section defines the image analysis of the proposed ELWI-FCM- IGEO on the LIDC-IDRI dataset. In Figure 1, the classification of normal and abnormal images is given. The segmented images obtained by the ELWI-FCM-IGEO are compared with the GT image and it is noted that the segmented image is matched with the GT.

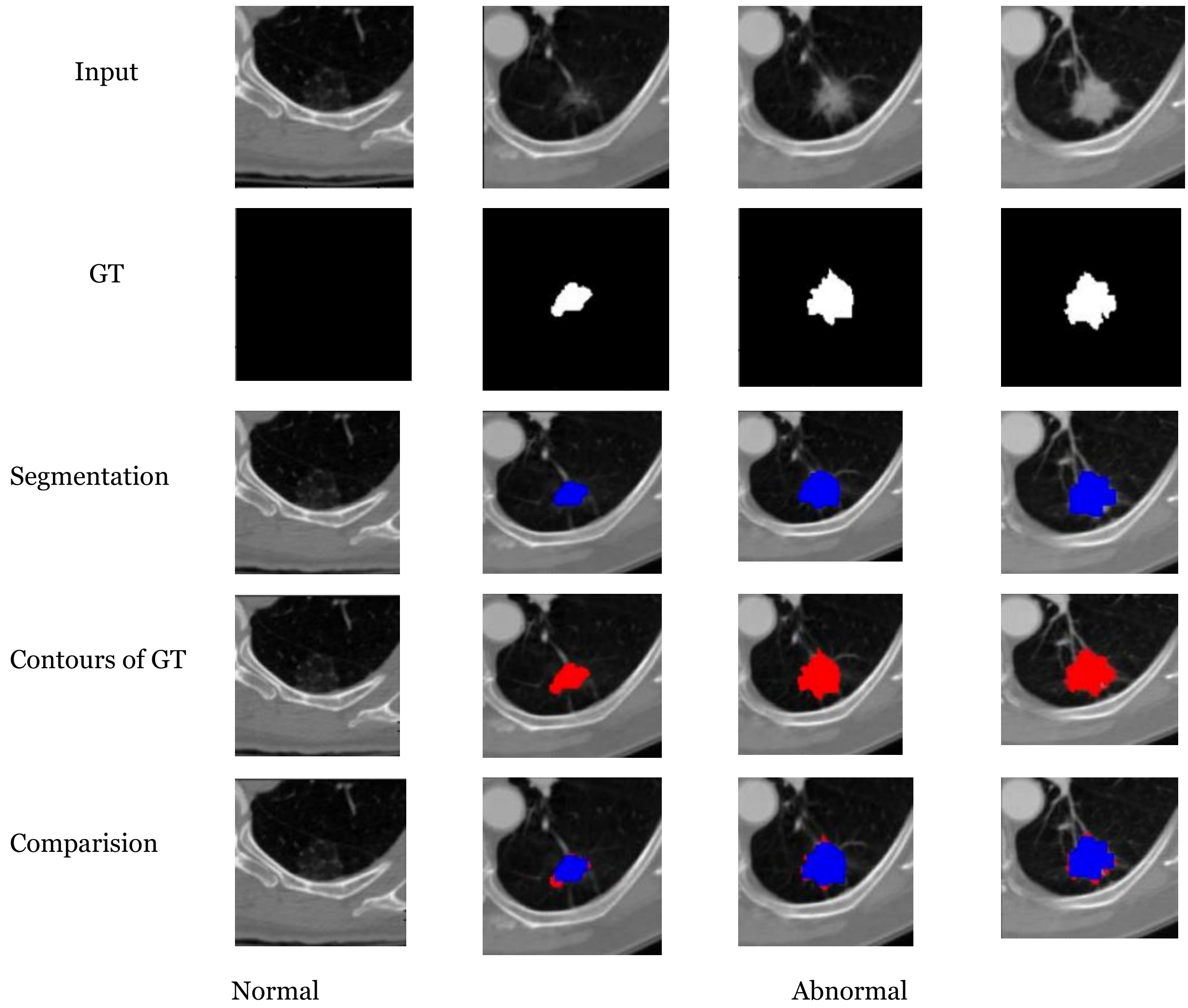
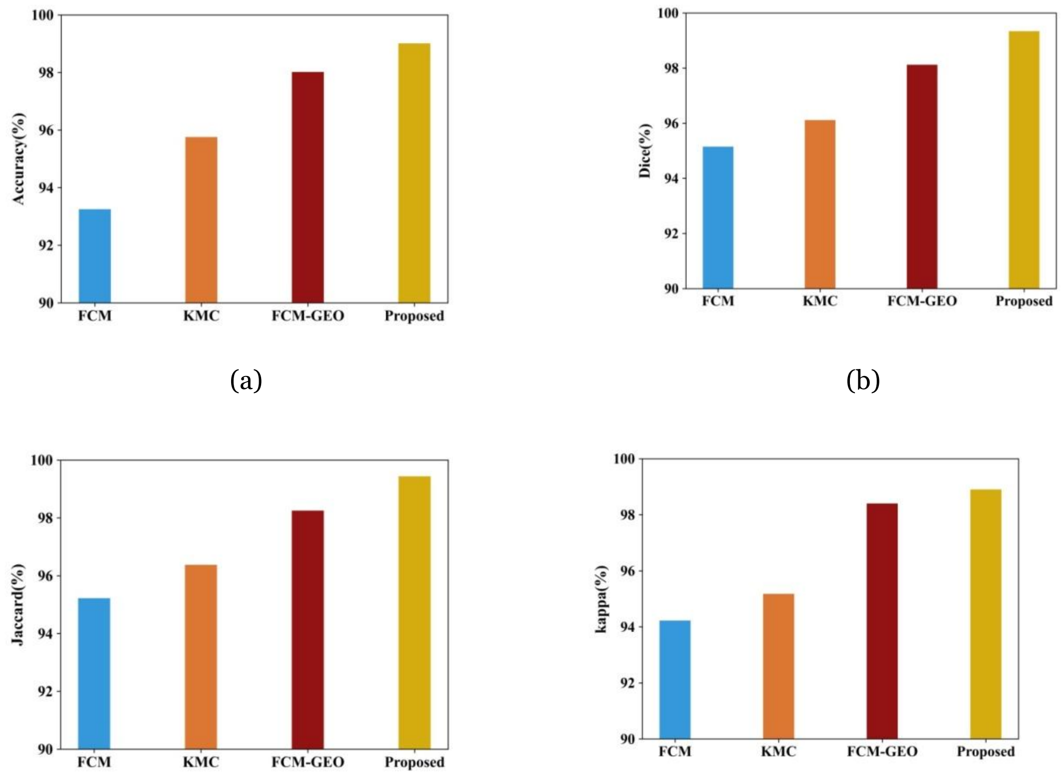


Figure 1 Image analysis of the proposed segmentation

Quantitative analysis

The following section defines the Quantitative analysis of the various segmentation models on the LIDC-IDRI dataset. Comparison is made for different segmentation models like FCM, KMC (k means clustering), FCM-GEO and the proposed segmentation model.



(a) (b)

(c) (d)

Figure 2 Comparison of (a) Accuracy, (b) Dice, (c) Jaccard and (d) Kappa

Figure 2 depicts the comparison of accuracy, dice, Jaccard and Kappa measures. It is observed from the graphical analysis that the proposed ELWI-FCM-IGEO achieved better accuracy of 98.7%, dice of 98.9%, Jaccard of 99% and Kappa of 98.3% respectively on the LIDC-IDRI dataset.

CONCLUSION

In this study, a novel hybrid optimizer named GEO and OBL was devised for effectively segmenting lung nodules. Initially, the CT images were subjected to preprocessing using a MF to eliminate noise. Subsequently, segmentation of the affected region was carried out and this process was performed by the ELWI-FCM. The segmentation performance was enhanced by the optimizer IGEO and finally, the classification of lung disease was carried out by the DenseNet201. The performance of the ELWI-FCM-IGEO method is rigorously evaluated based on various metrics to assess its effectiveness in accurately segmenting lung cancer.Future work should focus on integrating more adaptable deep learning models and further optimizing computational efficiency to address these limitations and improve overall performance.

REFERENCES

[1] Sang, Jun, Mohammad S. Alam, and Hong Xiang. "Automated detection and classification for early stage lung cancer on CT images using deep learning." In Pattern recognition and tracking XXX, vol. 10995, pp. 200-207. SPIE, 2019.

[2] Pedrosa, João, Guilherme Aresta, Carlos Ferreira, Gurraj Atwal, Hady Ahmady Phoulady, Xiaoyu Chen, Rongzhen Chen et al. "LNDb challenge on automatic lung cancer patient management." Medical image analysis 70 (2021): 102027.

[3] Gunjan, Vinit Kumar, Ninni Singh, Fahimudin Shaik, and Sudipta Roy. "Detection of lung cancer in CT scans using grey wolf optimization algorithm and recurrent neural network." Health and Technology 12, no. 6 (2022): 1197-1210.

[4] Ali, Zeeshan, Aun Irtaza, and Muazzam Maqsood. "An efficient U-Net framework for lung nodule detection using densely connected dilated convolutions." The Journal of Supercomputing 78, no. 2 (2022): 1602-1623.

[5] Chaunzwa, Tafadzwa L., Ahmed Hosny, Yiwen Xu, Andrea Shafer, Nancy Diao, Michael Lanuti, David C. Christiani, Raymond H. Mak, and Hugo JWL Aerts. "Deep learning classification of lung cancer histology using CT images." Scientific reports 11, no. 1 (2021): 5471.

[6] Bhattacharjee, Ananya, R. Murugan, and Tripti Goel. "A hybrid approach for lung cancer diagnosis using optimized random forest classification and K-means visualization algorithm." Health and Technology 12, no. 4 (2022): 787-800.

[7] Meraj, Talha, Hafiz Tayyab Rauf, Saliha Zahoor, Arslan Hassan, M. IkramUllah Lali, Liaqat Ali, Syed Ahmad Chan Bukhari, and Umar Shoaib. "Lung nodules detection using semantic segmentation and classification with optimal features." Neural Computing and Applications 33 (2021): 10737-10750.

[8] Agnes, Sundaresan A., and Jeevanayagam Anitha. "Efficient multiscale fully convolutional UNet model for segmentation of 3D lung nodule from CT image." Journal of Medical Imaging 9, no. 5 (2022): 052402-052402.

[9] Kido, Shoji, Shunske Kidera, Yasushi Hirano, Shingo Mabu, Tohru Kamiya, Nobuyuki Tanaka, Yuki Suzuki, Masahiro Yanagawa, and Noriyuki Tomiyama. "Segmentation of lung nodules on ct images using a nested three-dimensional fully connected convolutional network." Frontiers in artificial intelligence 5 (2022): 782225.

[10] Mahesh, Shanthi. "Hybrid optimized MRF based lung lobe segmentation and lung cancer classification using Shufflenet." Multimedia Tools and Applications (2023): 1-30.

[11] de Margerie-Mellon, Constance, and Guillaume Chassagnon. "Artificial intelligence: A critical review of applications for lung nodule and lung cancer." Diagnostic and Interventional Imaging 104, no. 1 (2023): 11-17.

[12] Tsivgoulis, Michail, Thomas Papastergiou, and Vasilis Megalooikonomou. "An improved SqueezeNet model for the diagnosis of lung cancer in CT scans." Machine Learning with Applications 10 (2022): 100399.

[13] Jain, Supiksha, Sanjeev Indora, and Dinesh Kumar Atal. "Lung nodule segmentation using salp shuffled shepherd optimization algorithm-based generative adversarial network." Computers in Biology and Medicine 137 (2021): 104811.

[14] Navaneethakrishnan, M., M. Vijay Anand, G. Vasavi, and V. Vasudha Rani. "Deep Fuzzy SegNet-based lung nodule segmentation and optimized deep learning for lung cancer detection." Pattern Analysis and Applications (2023): 1-17.

[15] Ren, Zeyu, Yudong Zhang, and Shuihua Wang. "LCDAE: data augmented ensemble framework for lung cancer classification." Technology in Cancer Research & Treatment 21 (2022): 15330338221124372.

[16] Atiya, Shaik Ummay, N. V. K. Ramesh, and B. Naresh Kumar Reddy. "Classification of non-small cell lung cancers using deep convolutional neural networks." Multimedia Tools and Applications (2023): 1-30.