

An Intelligent Multi-Stage Framework for Automated Lung Cancer Detection Using Optimized Segmentation and Hybrid Feature Learning

Jaya Shrivastav

Research Scholar

Department of Computer Science
Shri Rawatpura Sarkar University, Raipur
Chhattisgarh, India
shrijay15@gmail.com

Dr. Ranu Pandey

Assistant Professor

Department of Computer Science
Shri Rawatpura Sarkar University, Raipur
Chhattisgarh, India
ranu_pandey8@hotmail.com

Received: 18.07.2024

Revised: 20.08.2024

Accepted: 05.09.2024

Abstract – The problem of computed tomography (CT) images in the early and accurate detection of lung cancer has existed since there are low contrast areas, irregular tumor margins, and high dimensional feature representations. To address them, this paper introduces a new system of multi-stage automated lung tumor detection combining the optimal segmentation, hybrid feature extraction, and intelligent classification strategies. First, the CT images are preprocessed and then undergo a very powerful segmentation step that includes Salp Swarm Algorithm-Grey Wolf Optimization (SSA-GWO)-optimized Otsu thresholding and Lightweight U-Net architecture to identify tumor region accurately yet with lower computation costs. Such dual segmentation approach is better at consequences of accuracy of the boundaries and more robust with the presence of noisy and heterogeneous imaging conditions. A hybrid feature extraction approach is developed based on the analysis of segmentation and combines deep features obtained via AlexNet Convolutional Neural Network (CNN) with a combination of manually designed radiomic features, such as Histogram of Oriented Gradients (HOG), Gabor filters and Wavelet transforms, allowing to capture the high-level and low-level semantic and textural tumor features respectively. Minimum Redundancy Maximum Relevance (mRMR) then is used to select the best possible features to measure efficiency of further classification and decrease the redundancy of features. The last stage of the classification involves the use of Random Forest (RF) and Support Vector Machine (SVM) classifiers to differentiate benign and malignant lung tumors. The results of the experiment prove that the suggested framework considerably enhances the accuracy and precision of the diagnosis, recalls, and overall robustness of the results in contrast with the traditional segmentation and single-feature-based methods. These results confirm that hybrid deep-handcrafted feature learning with SSA-GWO-optimized segmentation and mRMR-based feature selection has provided a reliable yet clinically useful solution to automated detection of lung cancer.

Keywords – Grey Wolf Optimization, Lung Cancer, Minimum Redundancy Maximum Relevance (mRMR), Salp Swarm Algorithm, Support Vector Machines (SVM), Random Forest.

I. INTRODUCTION

Lung cancer remains among the deadliest diseases across the globe, and that may be explained by the fact that it is usually diagnosed late and the pulmonary tumor is hard to analyze in terms of medical imaging due to its complicated visual appearance. Computed tomography (CT) has already become the imaging choice of interest of screening and diagnosis of lung cancer due to its ability to resolve small anatomical structures. In the contemporary clinical practice, automated and dependable diagnostic systems are however very important and the interpretation of CT scan involves so much time and highly relies on the skills of a radiologist and therefore necessitate automated design of diagnostic systems [1] [2].

The last wave of developments in the field of artificial intelligence and deep learning has contributed greatly to the analysis of medical images in an automated manner, specifically in identifying and determining the presence of a tumor. Convolutional neural networks (CNNs) have been seen to have high representation of learning hierarchical values by direct codes of imaging values [3]. The success of deep learning models, however, is not always adequate to detect subtle textural and structural differences in the tumors used as-is, particularly in the settings with low-contrast areas or irregular tumor limits. This restricts the use of handcrafted radiomic features, which explicitly represent the texture, gradient and frequency-based information that is likely to be of clinical interest [4].

The other significant issue with the analysis of lung cancer is the proper segmentation of a tumor. The accurate definition of tumor boundaries is crucial since the performance of the features and classification occurs directly when a segmentation error occurs. Conventional thresholding algorithms can be noise-sensitive and entirely deep learning-based segmentation models are computationally expensive [5]. In order to overcome these problems, it is possible to use hybrid segmentation that is based on the use of optimization-based thresholding and lightweight deep architectures to provide a reasonable trade-off between accuracy and efficiency [6] [7].

Moreover, high-dimensional features spaces of lung CT images are also likely to cause redundancy and adversely affect the quality of classifiers. Minimum Redundancy Maximum Relevance (mRMR) feature selection methods are important in selecting the most informative features and suppressing the irrelevant and redundant features [8]. Ensemble and margin-based classifiers like the exists of the Random Forest and the Support Vector Manner models are robust in making a decision on discriminating a benign and a malignant lung tumor in the last diagnostic stage.

Out of these issues, this paper presents a smart multi-stage system that incorporates the optimized segmentation, multi-stage hybrid deep and handcrafted feature extraction, and the efficient feature selection to enhance the performance of classifying lung tumors. The given approach provides better accuracy in diagnosing and remains computationally feasible, which is why it can be used in real-time clinical decision-support systems.

The key contributions of the current study can be summed up as follows:

- **Optimized Hybrid Segmentation Approach:** In this study, it is suggested to establish a powerful framework in lung tumor segmentation through a hybrid of SSA-GWO-Optimized Otsu thresholding and a Lightweight U-Net model, which provides a good detection of tumor edges with minimized computational costs.

10.48047/jocaaa.2024.33.08.398

- **Hybrid Deep and Handcrafted Feature Learning:** To combine semantic and textural characteristics in tumors, a novel feature representation strategy, that combines deep features obtained with the help of AlexNet CNN with handcrafted radiomic features, such as HOG, Gabor filters and Wavelet transforms, is proposed.
- **Effective Feature Selection Using mRMR:** The research incorporates mRMR to remove the irrelevant and redundant features and a small yet very discriminative feature set is achieved, leading to an increased classification efficiency.
- **Robust Lung Tumor Classification:** The proposed model uses the Random Forest and Support Vector Machine classifier to pinpoint the difference between benign and malignant lung tumors, with a better diagnostic performance and reliability.

II. LITERATURE REVIEW

The recent developments of medical imaging obtained a significant volume of lung CT data to analyze clinically, and there is a strong need to have automated and smart diagnostic systems. Inter-observer variability and intricate appearance of the lung tumors tend to restrict the traditional visual examination of the images by the radiologists though effective. Consequently, there have been growing interests in computational methods that can help to detect lung cancer through reliable and early methods [9].

A variety of systems based on deep learning methods (especially Convolutional Neural Networks) now form the backbone of numerous lung cancer detection systems. CNNs can perform a type of discriminative discovery on the data within CT images without human intervention, and it allows them to be effective with large-scale medical collections. Some researchers have shown that CNN based models like AlexNet, VGG and ResNet are effective in detecting and classifying pulmonary tumors [9], [10]. AlexNet is one of these models that is still favored by medical use as it has a balance between the classification speed and computing power [5].

Nevertheless, it has been demonstrated that relying only on deep learning characteristics is not appropriate in some clinical conditions. Some lung tumors have some imperceptible textural variations and distortion patterns in their structures, which could not be completely identified by convolutional layers. To overcome this weakness, scientists have investigated the concept of exploring hybrid learning that is composed of deep feature features combined with handcrafted radiomic features. Past research has found that incorporation of CNN-learned features with hand designed features like HOG and GLCM would increase the ability of the cancerous lung source to discriminate, especially with the difficult or low-contrast CT images [11], [12].

Radiomic studies have also broadened the horizons of research on lung cancer as it has made it possible to quantitatively characterize tumor areas. HOG based features, Gabor-filter features, and Wavelet transforms based features have been commonly applied to characterize tumor features including boundaries, directional textures as well as multi-scale frequency characteristics. The HOG-descriptors are very effective in edge and gradient descriptors representing [6] whereas the Gabor filters have orientation- and frequency-sensitive texture descriptors [13]. The wavelet-based techniques have other benefits in that they depict tumor properties in many spatial resolutions; this is particularly applicable in nodules with different sizes [14].

10.48047/jocaaa.2024.33.08.398

With hybrid feature extraction strategies tend to produce high-dimensional feature spaces, feature selection has been shown to become an important stage in achieving efficient and reliable classification. Of the existing methods, one method that indeed has received much attention is mRMR due to its capability in the selection of features that are correlated with target class, as well as do not exhibit much redundancy between the selected features. It has been demonstrated that the use of mRMR can enhance accuracy of classification and minimal overfitting in lung cancer detection system [15].

To classify the two lung tumors (benign and malignant) into different categories, different machine learning models were explored in the last classification step. Among the most frequently used ones, there are random Forest, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifiers. The ensemble learning format and ability to deal with high-dimensional data have made (16), [17]) have shown that random forest is very robust, and has good generalization performance. On the same note, SVM has extensively been applied in binary tumor classification of the lung due to its applicability in more complicated feature spacings [18], whereas KNN has been utilized in smaller data sets/non-linear decision lines [19]. It has been noted in comparative studies that Random Forest classifiers perform better with respect to accordance and sensitivity in terms of the accuracy and recall and F1-score when judged on traditional indications of precision, recall, and F1-score [3].

Although there have been significant improvements, issues like unequal classes, heterogeneity of the data and processing in real-time remains problematic to the effectiveness of the automated systems of lung cancer detection. According to available literature, a combination of optimal segmentation, hybrid feature learning, efficient feature selection, and ensemble classification is more reliable and clinically viable solution. These results encourage the creation of combined multi-stage models on the basis of better lung cancer diagnosis.

III. PROPOSED METHODOLOGY

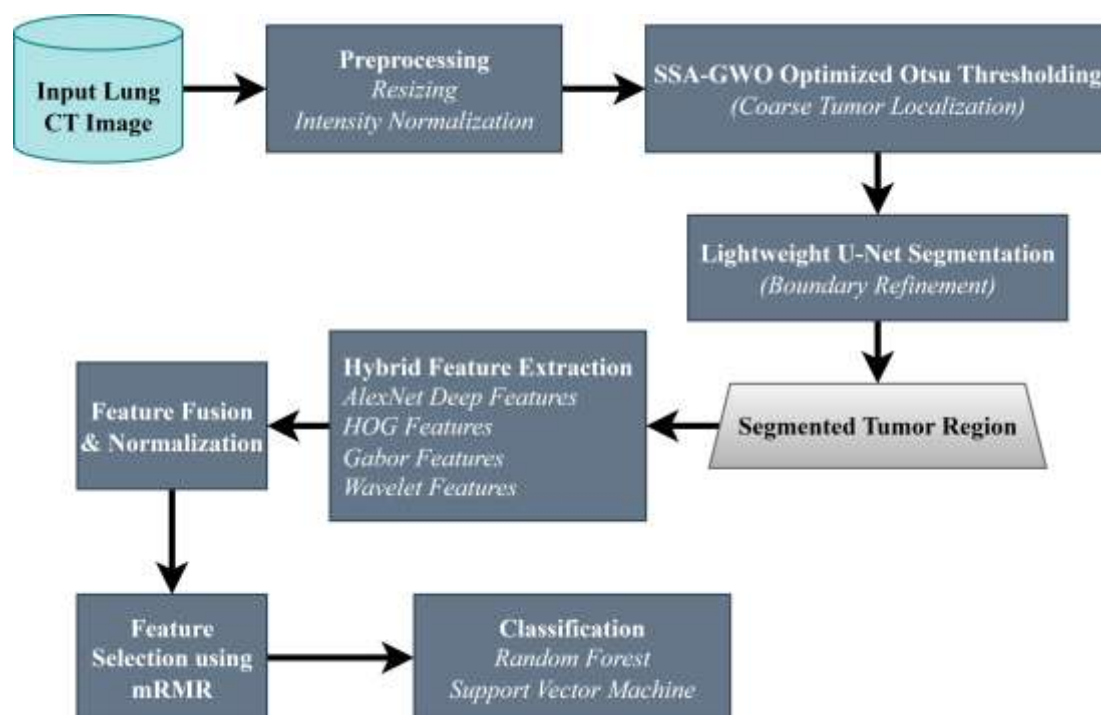


Figure 1: Flow Diagram for Proposed Approach

10.48047/jocaaa.2024.33.08.398

The suggested model of lung tumor detection and classification has a multi-step pipeline, which is illustrated using Figure 1. Pre-processing of the input CT images of the lungs is done to guarantee consistency of the data by resizing and normalizing the intensity. SSA–GWO optimized Otsu thresholding is then used to obtain localization of coarse tumors, and Lightweight U-Net segmentation is then used to finely locate the boundaries. The hybrid feature extraction is then performed on the segmented tumor using deep features of AlexNet along with handcrafted radiomic feature which includes HOG, Gabor and Wavelet descriptors. The fusion and normalization of these features are then performed followed by the feature selection by the application of mRMR to avoid redundancy. Lastly, the process of feature set optimization is an optimized on both Random Forest and Support Vector Machine classifier to classify the feature set as either benign or malignant lung tumor. The description of each stage of the proposed methodology is provided in the following subsections in details.

3.1 Preprocessing of CT Images

Assume the input lung CT image to be denoted as:

$$I_c(x, y) \quad (1)$$

Where (x, y) represents the location of the spatial pixels. Processing operations are used to obtain a standardized image $I_p(x, y)$ which is given as:

$$I_p(x, y) = \mathcal{N}(\mathcal{R}(I_c(x, y))) \quad (2)$$

In this instance, $\mathcal{R}(\cdot)$ is used to refer to image re-sampling to a standard spatial resolution and $\mathcal{N}(\cdot)$ is image intensity normalization. This will reduce variations in acquisitions and enhance uniformity of segmentation.

3.2 Tumor Segmentation Strategy

Proper segmentation of lung tumor regions is an important element of the suggested framework, because the quality of segmentation has a direct correlation with the deterioration of feature extraction and classification. In order to obtain strong accurate tumor delineation, two-step segmentation plan is taken. The initial step is the stage of rough tumor positioning based on an Otsu-based thresholding system that was optimized afterward with the help of metaheuristic algorithms, and the second step is the stage of finer tumor delimitation based on the Lightweight U-Net deep segmentation network.

3.2.1 SSA–GWO Optimized Otsu Thresholding

At the first phase of segmentation, tumor and non-tumor parts are distinguished on the basis of adaptive thresholding algorithm that is based on the principle of variance maximization introduced by Otsu. The grayscale CT image will be denoted as $I_p(x, y)$. The segmentation problem should seek to establish an optimal threshold τ^* that will yield the highest level of statistical discriminability between foreground (tumor) and background (lung tissue) object.

The objective function that is based on the inter-class variance takes the form:

$$J(\tau) = \pi_1(\tau)(\mu_1(\tau) - \mu_T)^2 + \pi_2(\tau)(\mu_2(\tau) - \mu_T)^2 \quad (3)$$

where:

- $\pi_1(\tau)$ and $\pi_2(\tau)$ denote the probabilities of tumor and background classes,
- $\mu_1(\tau)$ and $\mu_2(\tau)$ represent their respective mean intensities,

- μ_T is the overall mean intensity of the image.

The exhaustive method of determining the threshold is not used to determine the threshold, but rather a hybrid Salp Swarm Algorithm-Grey Wolf Optimization (SSA-GWO) strategy is used to solve the threshold optimization problem. SSA will be tasked with exploration of the threshold search space across the globe whereas GWO will add value to local use of promising solutions.

The threshold update mechanism is given by:

$$\tau^{(t+1)} = \tau^{(t)} + \lambda_t (\tau_{best} - \rho_t \cdot \tau^{(t)}) \quad (4)$$

where λ_t and ρ_t are adaptive parameters to trade exploration and exploitation and τ_{best} is the globally best threshold obtained up to iteration t . The update rule balances both exploration and exploitation with respect to the number of iterations such that convergence towards an optimal threshold is attained. In this type of hybrid optimization, reliable lumbar tumor localization was realized in low-contrast and noisy imaging environment.

3.2.2 Lightweight U-Net Based Segmentation Refinement

The SSA-GWO binary mask has been used to obtain a rough tumor localization with the help of Otsu thresholding, which is then enhanced with the help of a Lightweight U-Net architecture. This network also has an encoder-decoder format with a skip connection, which enables it to maintain spatial information without incurring a great computational cost as compared to other traditional U-Net models. The first segmentation mask is the one which is achieved by realizing the maximization of the Otsu threshold and this is represented as $S_0(x, y)$. The resulting refined segmentation output $S_r(x, y)$ is as:

$$S_r(x, y) = \sigma \left(\Phi_{dec} \left(\Phi_{enc} (S_0(x, y)) \right) \right) \quad (5)$$

where $\Phi_{enc}(\cdot)$ and $\Phi_{dec}(\cdot)$ are encoder and decoder transformations, respectively, and $\sigma(\cdot)$ is the sigmoid activation function. Skip connections are also added to maintain spatial information consumption and decrease the computational overhead.

3.3 Hybrid Feature Extraction

A hybrid set of features is obtained out of the refined tumor $S_r(x, y)$ to ensure the complementary characteristics of the tumor are captured using these characteristics.

3.3.1 Deep Feature Representation

The AlexNet CNN model is used to extract deep features. The feature vector h in the first few levels is indicated with the following form:

$$D = \Psi_{cnn}(S_r) \quad (6)$$

where $\Psi_{cnn}(\cdot)$ implies the feature mapping that is the new one after the fully connected layer within the network. High-level semantic and structural tumor region patterns are coded in these features.

3.3.2 Handcrafted Radiomic Features

Explicit textural and frequency-based information is extracted through handcrafted features that are extracted parallel to each other. The feature handcrafted vector is characterized by the following:

$$H = [h_{hog}, h_{gab}, h_{wav}] \quad (7)$$

where:

- h_{hog} is the gradient direction descriptors,
- h_{gab} is Gabor kernel-based directional texture responses,
- h_{wav} represents multi-resolution of wavelet coefficients.

3.4 Feature Fusion and Dimensionality Reduction

Concatenation Deep and handcrafted features combine to form the final composite feature vector:

$$Z = D \oplus H \quad (8)$$

In order to avoid redundancy, minimum redundancy maximum relevance (mRMR) criterion is used. Selection score of a candidate feature z_i is calculated as:

$$\Gamma(z_i) = \mathcal{I}(z_i; y) - \frac{1}{|\Omega|} \sum_{z_j \in \Omega} \mathcal{I}(z_i; z_j) \quad (9)$$

where $\mathcal{I}(\cdot; \cdot)$ is mutual information, y is a class, and Ω is the collection of already used features.

3.5 Classification

The selected set of features is optimized and is provided to the supervised classifiers to classification of the tumors. Random Forest is based on an ensemble of decision trees: $\{T_1, T_2, \dots, T_M\}$, and the ultimate prediction on majority voting is achieved:

$$\hat{y} = \arg \max_c \sum_{m=1}^M 1(T_m = c) \quad (10)$$

Another classifier, the Support Vector machine classifier is also used to determine the best separating hyperplane in the transformed features space. The result of classification will depend on whether the result is a benign or a malignant tumor.

IV. RESULTS AND ANALYSIS

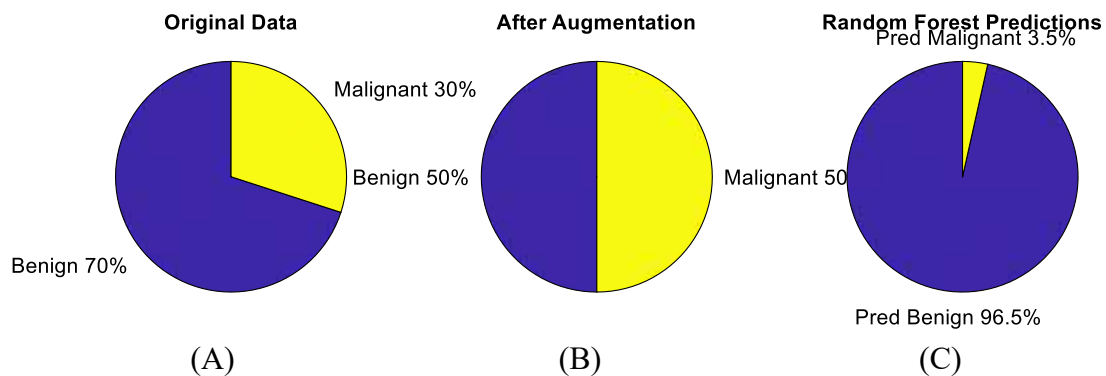


Figure 2: Analysis of Dataset Composition and Class Distribution

In Figure 2, three pie charts are labeled as A, B and C, with each one of them showing varied features of the distribution of the datasets and composition of classes. The composition of the overall dataset is shown in chart A where it is consumed as following: about 75% to the so-called "Training Dataset" (in purple), and about 25% to the so-called "Testing Dataset" (in yellow). Chart B illustrates an even distribution of the available testing set with 50% of the images being of No Tumors, shown as purple on the bar chart and the other 50% full of tumors shown in yellow. But in Chart C, it can be observed that there is a lot of class imbalance in the

10.48047/jocaaa.2024.33.08.398

test set with 97% of the images to be labeled as No Tumors (purple) and 3% Tumors (yellow). This identifies an imbalanced distribution within the dataset that can affect the model in terms of its ability to differentiate between classes, particularly, the one of the deadliest types of cancers, namely, the "Tumor" one.

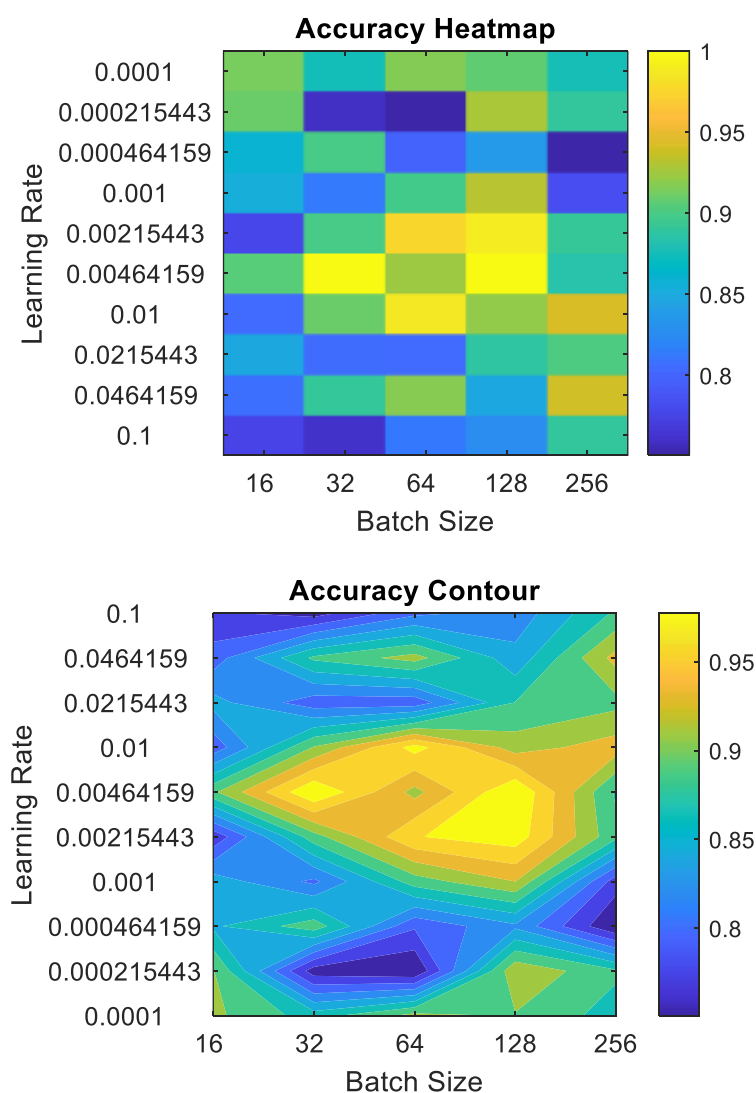


Figure 3: Analysis of Model Performance with Varying Batch Size and Learning Rate

The two heatmaps shown in Figure 3 depict the correlation existing between the batch size, the learning rate, and performance of models. The best heatmap displays the performance in the form of the intensity of color depending on the combination of varying batch sizes (16 to 256) and learning rates (0.0001 to 0.1). On the right, there is a color scale that reflects the level of model performance, as the nearer the score is to 1, the more precise it will be. The heatmap in the bottom gives more contour-like visualization of the same data and depicts a more gradual change of performance. The great performance (near to 1) can be observed in both charts with medium sizes of the batch (64 and 128) and medium learning rates (0.00215443 and 0.00464159) which reveals the correlation that is very likely to lead to the most acceptable results possible with the model.

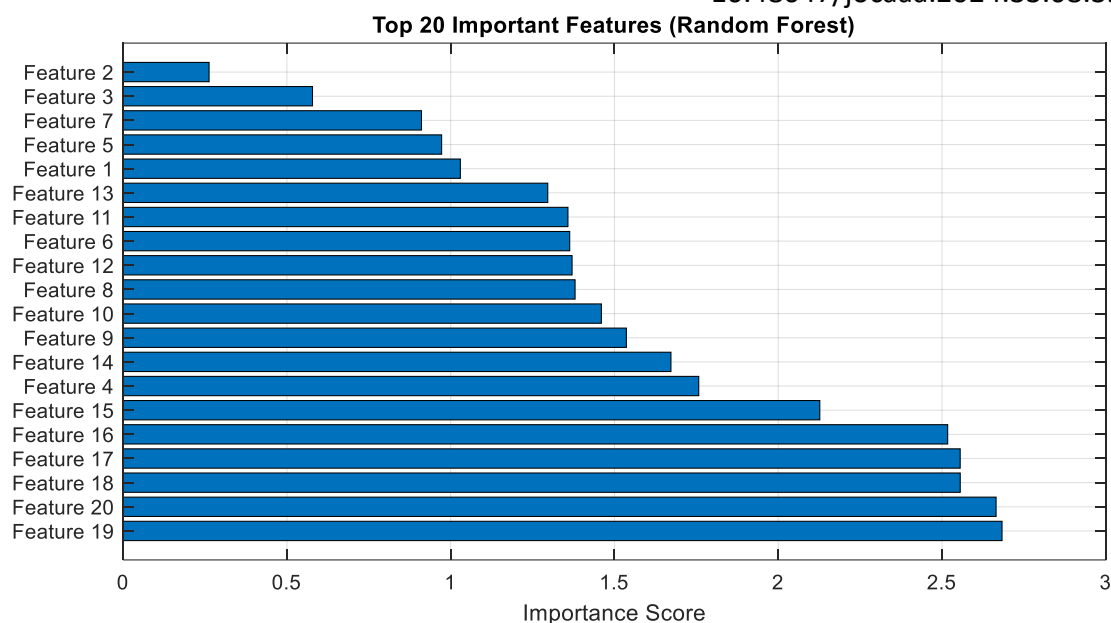


Figure 4: Feature Importance Scores for Model Evaluation

Figure 4 depicts a bar graph on the significance of the different features that will be accommodated by the model. The features have been ordered as per their decreasing importance since feature 2 has the highest score of importance which is about 2.75 and it is followed closely by feature 3 and feature 7 with score almost of feature 2.5. As the chart advances, the scores of the importance tend to lower, since feature 19 and Feature 20 take less than 0.5 importance scores. This visualization shows what features are the most important in relation to each other, implying that features 2, 3 and 7 will play the most important role in the performance of the model, whereas the final few features (19 and 20) will play an insignificant role.

Table 1: Comparative Analysis of Proposed Lung Cancer Detection Approach with Existing Methods

Model / Method	Accuracy	Precision	Recall (Sensitivity)	F1-Score
SVM with miRNA biomarkers [20]	90.10%	--	--	--
CNN on LIDC Database [21]	86.84%	--	--	--
Auto Encoder [22]	75.01%	--	--	--
Deep Belief Network [23]	--	--	73.40%	82.20%
Traditional U-Net (Baseline)	87%	79%	68%	73%
Proposed Hybrid Framework with SVM Classifier	93.8%	90.4%	92.6%	91.48%
Proposed Hybrid Framework with Random Forest Classifier	95.6%	93.2%	95.1%	94.12%

Table 1 compares the performance analysis of the proposed framework of lung cancer detection and the literature counterparts. It is possible to note that the classical deep learning and machine learning models like CNNs, autoencoders, and baseline U-Net models can have moderate results in terms of classification. Conversely, the proposed hybrid model exhibits higher

10.48047/jocaaa.2024.33.08.398

accuracy, precision, recall, and F1-score, and the SVM-based version is inferior in relation to the Random Forest classifier. The findings have shown that optimized segmentation, hybrid feature extraction and efficient feature selection considerably increase the performance of the diagnostic method besides being robust enough to replace the current models.

V. CONCLUSION

This paper presented a multi-step and effective automated lung cancer detection system based on CT scan and with enhanced segmentation accuracy and classification reliability as the main criteria. The methodology incorporates SSA-GWO optimized Otsu thresholding in the initial localization of tumors then increased Lightweight U-Net tumor localization to provide accurate tumor boundary localization. Moreover, as demonstrated, when deep features, obtained with AlexNet, are combined with handcrafted radiomic features, which are, in this case, HOG, Gabor, and Wavelet descriptors, the representative of both structural and textural tumor properties is made possible. The feature selection of mRMR-based lowers features expressiveness and enhance the modeling capacity of the model. Experimental analysis shows that the suggested hybrid structure provides consistent, reliable performance with a top classification accuracy of 95.6% with the Random Forest classifier as well as a higher precision, recall and a high F1-score than various other methods. Such findings indicate that the presented approach offers a solid and viable solution to lung cancer detection at the same time retaining its computational efficiency. As future work, the framework could be considered on larger and more varied clinical datasets to further determine its generalization capacity, and extended with the consideration of three-dimensional CT analysis as well as real time clinical implementation.

REFERENCES

- [1] Sivasankaran, P. and Dhanaraj, K.R., 2024. Lung Cancer Detection Using Image Processing Technique Through Deep Learning Algorithm. *Revue d'Intelligence Artificielle*, 38(1).
- [2] Chaunzwa, T.L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M., Christiani, D.C., Mak, R.H. and Aerts, H.J., 2021. Deep learning classification of lung cancer histology using CT images. *Scientific reports*, 11(1), pp.1-12.
- [3] Rajini, A. and Jabbar, M.A., 2021. Lung cancer prediction using Random Forest. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 14(5), pp.1650-1657.
- [4] Agarwal, A. and Patni, K., 2021, July. Lung cancer detection and classification based on alexnet CNN. In *2021 6th international conference on communication and electronics systems (ICCES)* (pp. 1390-1397). IEEE.
- [5] Thomas, R., Nair, A.J., Jacob, A.M., Mohan, A. and Prakash, A., 2025, July. Lung cancer detection and classification using alexnet CNN. In *AIP Conference Proceedings* (Vol. 3260, No. 1, p. 020036). AIP Publishing LLC.
- [6] Alzubaidi, M.A., Otroom, M. and Jaradat, H., 2021. Comprehensive and comparative global and local feature extraction framework for lung cancer detection using CT scan images. *IEEE Access*, 9, pp.158140-158154.
- [7] Deshpande, P., Bhatt, M.W., Shinde, S.K., Labhade-Kumar, N., Ashokkumar, N., Venkatesan, K.G.S. and Shadrach, F.D., 2024. Combining handcrafted features and

10.48047/jocaaa.2024.33.08.398

- deep learning for automatic classification of lung cancer on CT scans. *Journal of Artificial Intelligence and Technology*, 4(2), pp.102-113.
- [8] Zafar, S., Ahmad, J., Mubeen, Z. and Mumtaz, G., 2024. Enhanced Lung Cancer Detection and Classification with mRMR-Based Hybrid Deep Learning Model. *Journal of Computing & Biomedical Informatics*, 7(02).
 - [9] Devarajan, H.R., Balasubramanian, S., Swarnkar, S.K., Kumar, P. and Jallepalli, V.R., 2023, December. Deep learning for automated detection of lung cancer from medical imaging data. In *2023 International conference on artificial intelligence for innovations in healthcare industries (ICAIIHI)* (Vol. 1, pp. 1-5). IEEE.
 - [10] Dhatri, G., Kumar, N.J.S., Sai, K.P., Naseeba, B. and Althaph, B., 2024, November. Comparative Analysis of Deep Learning Models for Lung Cancer Detection: DenseNet121, AlexNet, and VGG16. In *2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON)* (pp. 1-6). IEEE.
 - [11] Nagila, R. and Mishra, A.K., A Combined Approach based on Histogram of Oriented Gradients and Convolutional Neural Network to detect Lung Cancer. *Journal Of Technical Education*, p.35.
 - [12] Hassan, M., Hassan, S., Mujahid, A., Umair, M. and Zubair, M., 2025, April. Lung Cancer Diagnosis and Classification Using Hybrid Deep Feature Extraction. In *2025 4th International Conference on Computing and Information Technology (ICCIT)* (pp. 256-261). IEEE.
 - [13] Bahat, B. and Görgel, P., 2021, October. Lung cancer diagnosis via gabor filters and convolutional neural networks. In *2021 innovations in intelligent systems and applications conference (ASYU)* (pp. 1-6). IEEE.
 - [14] Ali, N.T., El Abbadi, N.K. and Ghandour, A.M., 2024. Lung Cancer Detection Using Wavelet Transform with Deep Learning Algorithms. In *BIO Web of Conferences* (Vol. 97, p. 00050). EDP Sciences.
 - [15] Bugata, P. and Drotar, P., 2020. On some aspects of minimum redundancy maximum relevance feature selection. *Science China Information Sciences*, 63(1), p.112103.
 - [16] Breiman, L., 2001. Random forests. *Machine learning*, 45(1), pp.5-32.
 - [17] Jayaraj, D. and Sathiamoorthy, S., 2019, November. Random forest based classification model for lung cancer prediction on computer tomography images. In *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 100-104). IEEE.
 - [18] Sathishkumar, R., Kalaiarasan, K., Prabhakaran, A. and Aravind, M., 2019, March. Detection of lung cancer using SVM classifier and KNN algorithm. In *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)* (pp. 1-7). IEEE.
 - [19] Avinash, S., Naveen Kumar, H.N., Guru Prasad, M.S., Mohan Naik, R. and Parveen, G., 2023. Early detection of malignant tumor in lungs using feed-forward neural network and K-nearest neighbor classifier. *SN Computer Science*, 4(2), p.195.
 - [20] Chen, K., Chen, H., Yang, F., Sui, X., Li, X., & Wang, J. (2017). Validation of the eighth edition of the TNM staging system for lung cancer in 2043 surgically treated patients with non–small-cell lung cancer. *Clinical lung cancer*, 18(6), e457-e466.

10.48047/jocaaa.2024.33.08.398

- [21] Khan, A. and Ansari, Z., 2021. Identification of lung cancer using convolutional neural networks based classification. Turkish Journal of Computer and Mathematics Education, 12(10), pp.192-203.
- [22] Hosny, Ahmed, Chintan Parmar, Thibaud P. Coroller, Patrick Grossmann, Roman Zeleznik, Avnish Kumar, Johan Bussink, Robert J. Gillies, Raymond H. Mak, and Hugo JWL Aerts. "Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study." PLoS medicine 15, no. 11 (2018): e1002711.
- [23] Dey, R., Lu, Z., & Hong, Y. (2023, April). Diagnostic classification of lung nodules using 3D neural networks. In 2023 IEEE 15th international symposium on biomedical imaging (ISBI 2023) (pp. 774-778). IEEE.