International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

Enhanced Lung Cancer Detection Using a Hybrid AlexNet-CNN and Handcrafted Feature Fusion with mRMR Feature Selection and SSA-GWO Optimized Segmentation

Jaya Shrivastav[1], Dr. Ranu Pandey[2]
[1]Research Scholar Department of Computer Science Shri Rawatpura Sarkar University, Raipur Chhattisgarh, India, shrijay15@gmail.com
[2]Assistant Professor, Department of Computer Science, Shri Rawatpura Sarkar University, Raipur, Chhattisgarh, India,

Abstract *– The recognition and diagnosis of lung cancer is a major component in clinical decision-making with early and accurate detection of the cancerous condition being fundamental to improving patient outcomes. The aim of the paper is to present a new variant of chest cancer detection based on deep learning with the incorporation of hand-crafted radiomic features into the processing pipeline of multi-stage processing. Hybrid feature extraction method implements deep learning of the AlexNet CNN together with what can be considered as handcrafted radiomic features, e.g., HOG, Gabor filters, and Wavelets transforms, allowing to provide more detailed representation of tumor features. Also, Minimum Redundancy Maximum Relevance (mRMR) method is applied to feature selection by keeping the most relevant and non-redundant features and eliminating noise to boost accuracy of classification. The support vector machines (SVM), as well as an ensemble classification model, specifically the Random Forest (RF) algorithm, introduces both robustness and accuracy in this case, since they effectively deal with high-dimensional and complex data. The approach is used to overcome the problems with the classification of lung cancer as it raises the accuracy, efficiency, and practicability of the model in clinical practice. The given methodology is proven to present a dependable and effective solution that may be used in the real-time clinical settings to detect lung cancer early and accurately.*
Keywords *– Grey Wolf Optimization, Lung Cancer, Minimum Redundancy Maximum Relevance (mRMR), Salp Swarm Algorithm, Support Vector Machines (SVM), Random Forest.*

I. INTRODUCTION

Lung cancer is still ranked among the top causes of death globally, and thus there is a necessity to have effective and efficient methods of detecting lung cancer and categorizing it [1]. Early detection of the tumor is critical towards enhancing the survival rate thus developing dependable tools of diagnosis is a burning issue in medical imaging. Manual detection procedures along with single-feature extraction methods are traditional methods that have certain difficulties in classifying lung tumors as malignancy because of their complexity and extreme heterogeneity. The given paper seeks to further streamline and perfect the process of detection through a new method combining deep learning and handcrafted radiomic features to include and elevate the process of tumor characterization and classification [2].

The involvement of a multi-stage processing pipeline in this study enhances the strength of the system as it means that every step of the procedure must play a role in improving overall accuracy and efficiency. This is carried out by mRMR based feature selection which holds on to the most significant features eliminating irrelevant data or redundant or noisy data. To analyse the tumors and determine whether it is benign or malignant, the Random Forest (RF) ensemble classifier is used [3]. The RF algorithm can use high-dimension data and does not overfitting the data accordingly, making them fit in the medical field that is highly complex on the data level.

Motivation for the Selection of the Proposed Methods: The strategies suggested to provide responses to the pronounced issues, in the detection and classification of lung cancer, with the focus on the extension of accuracy, efficiency, and clinical feasibility. Advantages of both techniques are combined in the hybrid (mixing deep learning (AlexNet CNN) [4] [5] with handcrafted radiomic features (HOG, Gabor, and Wavelet) approach [6] [7]. The deep learning approach, especially AlexNet CNN, can be useful in extracting more complex and minute details of malignant growth since they identify intricate structures of tumor regions which play a vital role in classification. Nevertheless, in the case of deep learning models, it is quite possible to miss certain textural or structural features that are important in tumor

differentiation. The addition of handcrafted features with the view to covering the most important texture and edge-based information enhance the capabilities of the model which is likely to increase the differentiation of benign and malignant tumors.

The fact that multi-stage of a processing pipeline was used further enhances methodology, as the process is divided into different steps which increase the overall performance of the model. Such practices like data augmentation, size adjustment, and grayscale to RGB transformations make the result less prone to overfitting and therefore improves the ability of the model to generalize during training. The feature extraction process entails the use of deep learning features and handcrafted features, in one joint feature vector comprising the learned and the manually created features of the tumor.

During the feature selection process the mRMR algorithm is implemented so that few features must be chosen, in terms of those that are most relevant, to reduce the number of noise and redundant data, which otherwise would negatively affect the performance of the classification process. The usage of a final classification model, Random Forest (RF) introduces resilience and accuracy since the model can address data with high dimensions and complex data and provides information about feature importance. Such ensemble method of learning serves well to ensure that the learning model is not over-fitted and it can render reliable and accurate classifications.

Problem Definition and Contribution: The main problem in the detection of lung cancer is the correct specification of tumors since the data of medical images are highly persistent and multifaceted. High-dimensional features and noise tend to be unfriendly to traditional methods, consequently giving inferior results with regard to classifications. The problems are faced in this paper, where a hybrid feature extraction strategy which is a combination of deep learning and hand-coded features and their combination with an ensemble classifier technique is added to increase accuracy and stability.

The salient contribution of the work is two-pronged; this study proposes a multiphase pipeline that involves the feature extraction, feature selection through mRMR [8] and classification, in an attempt to enhance the accuracy and effectiveness of tumor classification. Second, it shows how the combination of AlexNet CNN and handcrafted radiomic features can be used to collect more of the tumor features and can increase the clinical applicability of a given model. This application of Random Forest to classify the model in question contributes to its resilience, which makes it the viable solution in the real-time functioning of the clinical area, where the speed and accuracy are two of the key factors.

II. LITERATURE REVIEW

Cancer-related deaths attributed to lung cancer are the highest in the world, and early diagnosis is noted to be a key method of ensuring that the process of treatment is successful. Computational methods have also gained increased traction since the advent of medical imaging technologies, which may be utilised in the context of automating the process of identifying the presence of lung cancer. The convergence of traditional image processing with a deep learning framework has been found in recent years to be quite promising as a way of dealing with problems of tumor detection and classification. This literature review discusses different methods, strategies, and developments in the area which are focusing on the use of deep learning, hand-engineered features, feature selection methods, and classification algorithms applied to the detection and the classification of the lung cancer.

CNNs have found wide use in the context of medical imaging, and lung cancer detection examples. The authors of [9] have reported that CNNs have been significantly successful in the role of automated detection of lung cancer, as it demonstrates high performance in learning hierarchical features based on raw image observation. Lung tumor detection and classification have been done using CNN-based networks such as AlexNet, VGG, and ResNet because of their depth and convolutional layers which enable them to learn the features automatically that are important in the identification of a tumor [10]. Specifically, the AlexNet architecture has seen a lot of applications because it has a relatively small computational demands when compared to deeper networks but can demonstrate competitive outcomes on several image classification tasks [5].

Although CNNs are efficient as they learn features, they usually do not capture some structural and texture-based features that have significance in tumor discrimination. To overcome this shortcoming, former studies have integrated approaches in which deep learning is joined with handcrafted characteristics in high numbers. These techniques have the benefits of the two approaches described; they convey the texture patterns that deep learning could fail to discern. As an illustration, the study by the authors of [11] introduced CNN features along with the handcrafted features, including Histogram of Oriented Gradients (HOG) and Gray Level Co-occurrence Matrix (GLCM) to provide better lung nodules detection. These hybrid models have demonstrated better results on the accuracy of classification, especially difficult data [12].

Radiomics is an upcoming technology with emphasis on the purpose of finding quantitative characteristics of medical images. Such characteristics such as form, dimension, texture and intensity could be useful in giving an insight with respect to the condition of the tumor. Most of the studies use handcrafted features like HOG, Gabor and Wavelet to extract texture and structural information on lung cancer images. Edge information is important in determination of tumor boundaries and HOG features have been used to determine the information [6]. To obtain the information of the frequencies and direction of texture patterns located in tumor areas, Gabor filters have been applied [13]. Also, the feature extraction of wavelet transform has found widespread application in medical images since it brings multiresolution information where the scale of the tumor varies [14].

It has importance in terms of choosing suitable features to make the classification model practical since they eliminate all trivial and correlated features. Regarding measurement of lung cancer, researchers have suggested several methods of feature selection in the attempt to improve classification. Minimum Redundancy Maximum Relevance (mRMR) is one such approach that involves choosing the set of features that are highly germane (i.e. they are strongly correlated with the target variable), but do not duplicate, any of the other features [15]. The technique has found extensive use in medical imaging, where overfitting may be an issue, because of high-dimensional data. The authors of [15] revealed that the application of mRMR is very useful in improving classification models by choosing a subset of features that are relevant in the diagnosis of the lung cancer.

The problem of lung tumors classification into the benign and malign part is rather complex because of the heterogeneity or complexity of medical data. Summary of results Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) are all ensemble learning-based methods, recently used to enhance the accuracy of classification. One of the successful ensemble techniques, Random Forest, has been suggested to be successful in the problem of medical image classification because it can manage the bulk of high-dimensional data and it is not susceptible to overfitting [16]. RF has shown great success in the classification of lung cancer as the predictions of multiple decision trees can be aggregated in order to enhance accuracy and robustness [17]. In the same way, SVM has been successfully applied to binary classification exercises including recognition of benign or malignant tumors since it is effective in high dimensional feature spaces [18]. Another common algorithm used to classify tumors is KNN, and it is commonly introduced when one is dealing with small data files or when the decision space is not linear [19]. The assessment of the classification models in medical imaging plays a significant role in the identification of their effectiveness in the application in real world tasks. The accurateness, precision, the recall, and F1-score are the widely popular metrics applied to evaluate the lung cancer detecting models. The comparison of both methods by Random Forest and SVM classifications under such metrics has been conducted in the work by [3], which has proven that RF showed better results in comparison to SVM concerning accuracy and recall of the lung cancer classification. The application of F1-score that balances the precision and recall measures is relevant especially in the medical case when both false positives and false negatives may have dramatic consequences in the treatment of patients. Although there is a huge improvement in the methods used to detect lung cancer, there are various issues. The first issue is the skewedness of the datasets because the number of malignant tumors tends to be less than the benign ones. This issue is considered to be solved with such techniques as data augmentation and synthetic data generation. The other complexity is the requirement needed by the clinic due to the use of real-time

processing, and thus the methods would be useful to act efficiently on large datasets without any loss of accuracy. The potential research work in future studies entails in enhancing computational capabilities of deep learning methods, feature extraction methods and the creation of more resolute classifiers which can deal with the complexity of lung cancer imaging. The evolution into different methods in the analysis of medical images using deep learning methods and hybrid features extraction technique has been exposed in the literature on lung cancer detection procedures. Using the combination of handcrafted features and deep learning models offers a more holistic solution to the detection of the tumor, whereas feature selection strategies, such as mRMR, enhance the performance of classification. Unlike other lung cancer classifiers, ensemble learning has been demonstrated to be effective in lung cancer classification especially Random Forest, which has high accuracy and robust in features. Although certain aspects of dataset imbalance and computational efficiency remain a problem, the further expanding of these methods has great potential of increasing accuracy and applicability of lung cancer detection in the clinics.

III. PROPOSED METHODOLOGY

The given block diagram describes a multi-stage pipeline of the lung cancer detection and classification in CT images. Here is what each of the blocks on the diagram represents:

1. Input CT Image: Its operations start by taking CT scan images of the lungs. The pictures are the main data by which the lung tumor can be spotted and categorized.
2. Pre-Processing: CT images are pre-processed in terms of improving the quality of images, removing excessive noise, and pre-process the data before it is subjected to the next steps. This might include procedures such as resizing, normalization of the image and removal of noise so that the image is in a satisfactory form to proceed beyond.

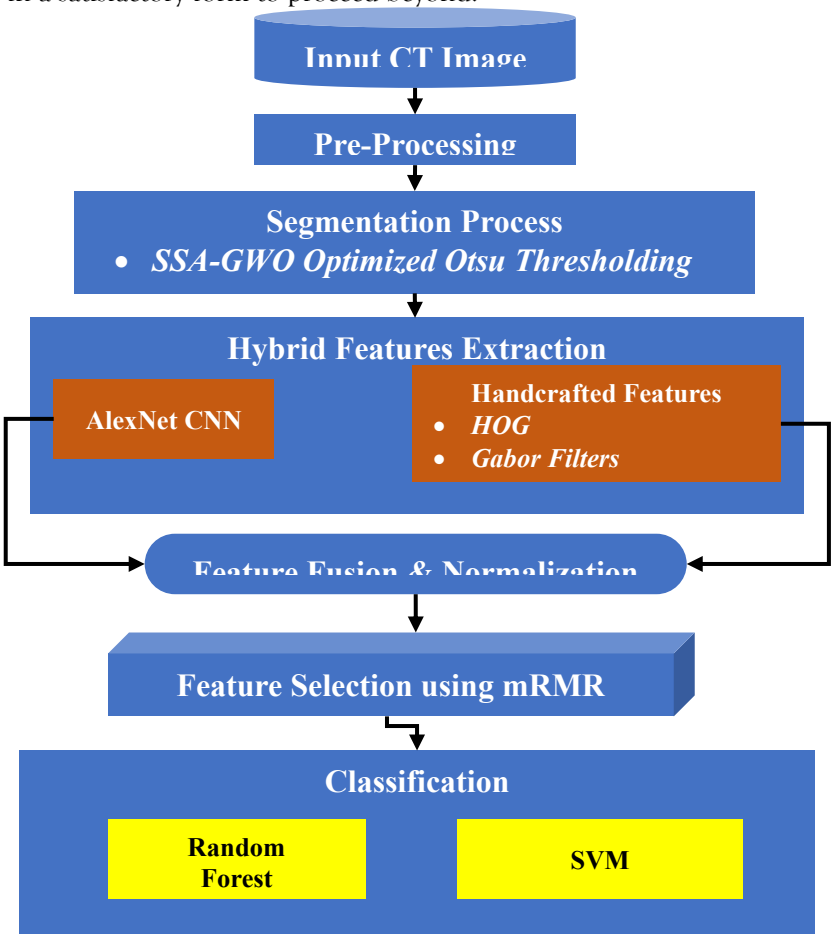

Figure 1: Proposed Flow Diagram for Lung Cancer Detection Using Hybrid SSA-GWO-Optimized Otsu and Lightweight U-Net based Segmentation, Hybrid Features Extraction, and mRMR Feature Selection

3. Segmentation Process:

- SSA-GWO Optimized Otsu Thresholding: It is a thresholding algorithm through which the regions of tumor are segregated with reference to the regions of healthy tissues on the CT image. The Salp Swarm Algorithm (SSA) is integrated with the Grey Wolf Optimization (GWO) so that the Otsu thresholding could be maximized and, thus, the segmentation at hand could be increased despite the poor conditions such as low contrast or noise.
- Lightweight U-Net Segmentation: The segmentation uses a modified model of the conventional U-Net model. The Lightweight U-Net would help lessen the calculation complexities with all the important characteristics of the pioneer U-Net architecture structure, including skip connections, which suit the real-time clinical application.

4. Hybrid Features Extraction: The next stage is the feature extraction where both the deep learning and handcrafted approaches are employed in extracting necessary features out of the segmented images.

- AlexNet CNN: Automatic learning and extraction of a complicated pattern and feature via a deep learning model (AlexNet) are used on the segmented images. This will aid in detecting the slight attributes in the tumor areas which are hard to be identified manually.
- Handcrafted Features: Along with the deep learning-based features of AlexNet CNN, the features are also extracted hand crafted, which include:
  - HOG (Histogram of Oriented Gradients): Extracting edge information that may enable the margin of tumors to be determined is what HOG is utilized to capture.
  - Gabor Filters: It is used in order to extract frequency and orientation features of textures in the tumor areas.
  - Wavelet Transforms: Wavelet transforms have been applied to extract multi-resolution characteristics of the tumor, a factor that has made it easy to identify tumors of different extents.

5. Feature Fusion & Normalization: The extracted features of AlexNet and even the hand-crafted methodologies are combined to obtain a complete feature vector. The normalized data is used in order to make sure that all features contribute equally in classification stage.

6. Feature Selection using mRMR: Minimum Redundancy Maximum Relevance (mRMR) is used in picking the most relevant features and at the same time eliminating redundant features. This step limits the dimensionality of the data rendering it easy to handle by the classifier hence better performance in the classification will be achieved as it will only utilise the most important features.

7. Classification: After the features have been chosen it applies different classification algorithms to classify the tumors as benign or malignant:

- Random Forest (RF): It is an ensemble method that is properly adapted to high-dimensional data and allows one to obtain information on the importance of features.
- Support Vector Machine (SVM): SVM is a very effective classifier of data in high-dimensions, especially the challenge of binary classification.

All these stages collaborate to effectively segment, feature extract, choose the most pertinent features, and classify the tumors, which finally gives an overall idea of lung cancer detection in terms of it being highly accurate and computationally practical to be implemented clinically in real-time.

3.1 Segmentation

3.1.1 SSA-GWO Optimized Otsu Thresholding

SSA-GWO-Optimized Otsu Thresholding is the improved form of the former Otsu technique since it is a hybridisation of Salp Swarm Algorithm (SSA) and of Grey Wolf Algorithm (GWO) process. Such optimization algorithms have optimized thresholding process that subsequently has simplified defining the tumor as well as the background, even in the complicated scenario of the low contrast images or backgrounds in noisy conditions. The thresholding space of SSA is searched in the global sense as opposed to the local sense in the case of the one used by GWO hence it has the benefit of more precise segmentation since it considers the high dimensionality of the intensity distribution. The method is ideal in situations where the tumors are small or irregular (when normal Otsu are not).

Automatically choose best threshold t^* to discriminate between tumor (foreground) and lung tissue (background) by maximization on between-class variance $\sigma_B^2(t)$.

Mathematical Formulation:

Otsu's Between-Class Variance:

$$\sigma_B^2(t) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (1)$$

Where:

- t is the threshold that will be taken into consideration.
- When the image is divided based on the value t in the image the same will take place as the following means in the image where the background and the forefront classes will be the following parameters; μ_1 and μ_2.
- The pixels intensities of the background and foreground classes have a variability in distribution of σ_1^2 and σ_2^2 respectively.

Hybrid Optimization (SSA-GWO): The area explored by Salp Swarm Algorithm (SSA) is space of threshold:

$$x_j^1 = \begin{cases} F_j + c_1\left(\left(ub_j - lb_j\right)\cdot r_1 + lb_j\right), & if\ r_2 \geq 0.5 \\ F_j - c_1\left(\left(ub_j - lb_j\right)\cdot r_1 + lb_j\right), & if\ r_2 < 0.5 \end{cases} \quad (2)$$

Where:

- x_j^1 is the position of the j^{th} dimension of the leader salp,
- F_j is the food source (i.e., best threshold found so far),
- ub_j and lb_j are the upper and lower bounds of the search space,
- $r_1, r_2 \in [0,1]$ are random variables for exploration,
- $c_1 = 2 \cdot exp\left(-\left(\frac{4 \cdot l}{L}\right)^2\right)$, with l being the current iteration and L the maximum iterations.

Grey Wolf Optimization (GWO) refines threshold:

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (3)$$

Where,

- $\vec{A}_i = 2 \cdot a \cdot \vec{r}_i - a$
- $\vec{C}_i = 2 \cdot \vec{r}_i$
- a is linearly decreased from 2 to 0 over iterations,
- $\vec{r}_i \in [0,1]$ are random vectors.

Output: The optimal threshold is given by:

$$t^* = \arg\max_{t \in [0, L-1]} \sigma_B^2(t) \quad (4)$$

3.1.2 Lightweight U-Net Segmentation

Lightweight U-Net is an efficient variant of a classical U-Net that will reduce computational costs and accuracy loss of the segmentation accuracy. It also has a broad path (decoder) and narrow path (encoder) wherein the former is linked to high-level features extraction and the latter to recovering the image spatial details. The implementation of attention gates into the variant of the Attention U-Net is meant to focus on the tumor territory by magnifying the information concerning a wanted target and reducing the worthless information. The specific approach to segmentation helps specifically to locate the tumor in the structure that is anatomically complicated and display the high-resolution map to be further categorized.

Key Equations:

Encoder (Contracting Path):

$$z^{(l)} = \mathrm{ReLU}\left(W^{(l)} * z^{(l-1)} + b^{(l)}\right) \quad (5)$$

Where:

- $z^{(l-1)}$ is the input feature map from the previous layer,

- $W^{(l)}$ is the weight matrix of the l^{th} convolutional layer,
- $*$ represents the convolution operation,
- $b^{(l)}$ is the bias term for the convolutional layer,
- ReLU is the Rectified Linear Unit activation function.

Decoder (Expansive Path):

$$z^{(l)} = \mathrm{ReLU}\left(W^{(l)} \star z^{(l-1)} + b^{(l)}\right) \quad (6)$$

Where:

- $\star$ represents the transposed convolution operation (upsampling),
- $W^{(l)}$and $b^{(l)}$ are the weight and bias terms for the up-convolution.

Skip Connections: Feature maps from encoder z_{enc} are concatenated with decoder outputs z_{dec}:

$$z_{final} = Concat(z_{enc}, z_{dec}) \quad (7)$$

Output Layer (Sigmoid Activation):

$$\hat{y} = \sigma\left(W^{(L)} \star z^{(L-1)} + b^{(L)}\right) \quad (8)$$

In this case, the SSA-GWO-Otsu gives a crude tumor localization and the lightweight U-Net iteratively improves those boundaries through the use of attention modules.

3.2 Hybrid Feature Extraction Approach

The Hybrid Feature Extraction Approach is the method that claims to merge deep learning methods and the handcrafted radiomic features to improve the lung cancer classification process. This method is considered to combine the advantages of the two techniques in offering a better and stronger approach to detect and classify a tumor. The high-level spatial details and features, and the complex features are captured using deep learning architecture, namely AlexNet CNN whereas textural, edge, and multi-scale components are identified using handcrafted features, such as HOG, Gabor filters, and Wavelet transforms. The interrelation of these attributes permits the model to identify the low- and high-level information resulting in the better classification of tumor as malignant or benign.

3.2.1 AlexNet CNN

AlexNet has a deep convolutional neural network (CNN) model that deals with image classification. The field of medical image analysis has embraced its use successfully because it learns spatial patterns and structures in images. In the scope of lung cancer detection, AlexNet is utilized with the matter of extracting high-level hierarchical feature of the CT images of the lungs. The given model of deep learning is good at identifying complex spatial relationships within the tumor areas, which proves an essential factor in classification.

The AlexNet comprises a number of convolutional layers that are followed by the fully connected layers. The last and the fully-connected layer, i.e., FC7, plays an important role in producing the deep features that were used in classification. FC7 represents the seventh fully connected layer of the network with output vector of the dimension 4096 describing the high-level features of the tumor region. The deep features are then extracted to give this feature vector that is then employed in the classification stage.

The hierarchical spatial pattern and the vital information of the tumor texture, shape and structure are covered by the deep features which are extracted by the AlexNet. These characteristics are essential of proper classification since they enable the model to discern the benign and malignant tumor in terms of minute spatial and textural variations.

The deep learning feature vector obtained by using AlexNet is described mathematically in the following manner:

$$F_{deep} = \mathrm{AlexNet}(I_{RGB}) \quad (9)$$

Where:

- I_{RGB} is the input image in RGB format,
- F_{deep} is the 4096-dimensional feature vector extracted from the FC7 layer.

3.2.2 Handcrafted Features

Handcrafted features refer to manually created feature extraction procedures that extract properties relating to tumors (including texture, gradient and multi-resolution features). Such characteristics have

specific value in cases when tumors have intricate patterns that cannot be fully detected by deep learning models. Using the handcrafted features, we will be able to give extra discriminative ability to the classification system. Three major handcrafted characteristics, in this regard; HOG (Histogram of Oriented Gradients), Gabor and wavelet transforms are employed.

- **HOG (Histogram of Oriented Gradients):** HOG is one of the commonly used feature descriptors, and as such, it is keen at encompassing the edges and gradients of images, which are essential in coming up with the shape and forms of the tumors. Principles behind HOG are based on the fact that the local gradient directions distribution is important to capture edges and texture patterns used to characterize the tumor structure. The HOG method operates on the basis of partitioning the image into small cells and inside each cell the magnitude of the gradient and the orientation of the gradient is computed. The magnitudes of the gradients signify the strength at which the value of pixel values varies and the orientation of gradients signifies the direction in which the intensity varies. Once the gradient values are calculated, gradient orientation histogram is made out of every cell. These histograms are thereafter carried out in a normalized form so that the features in these histograms are not affected by the changes of light.

$$HOG(x) = \sum_{\theta} |\nabla I(x)| \cdot \cos(\theta)$$

(10)

- **Gabor Filters:** Texture Analysis One can achieve acquisition of texture information at multiple scales of image and across orientation using Gabor filters, i.e., multi-scale multi-orientation texture analysis. The response $G(x)$ of the kernel regarding the point x will take the form of any image $I(x)$ that is derived by the computation of the input kernel of the Gabor with the image.

$$G(x) = \sum_{i=1}^{N} I(x) \cdot Gabor_i(x)$$

(11)

- **Wavelet Transforms:** A multi resolution analysis of the texture of the tumor is performed by decomposing an image into a number of frequency bands of wavelet transforms. Wavelet transforms break the image into the low frequency and high frequency and are of the form of details of different scales. The term of wavelet transform should be described in the following way:

$W_{\psi}(x, s, \theta) = \int I(x) \cdot \psi(x, s, \theta)\, dx$ (12)

3.3 Feature Fusion

When computing the characteristics of AlexNet CNN and the handcrafted approaches, the next step is consolidating the two sets of features into a single feature array. This kind of fusion is mentioned as feature fusion. The feature concatenation would ensure utilization of the information present in the entire path, the deep learning path and the hand-crafted feature extraction path, and consequently result in the enhanced ability to differentiate between malignancy and benign tumors in the model.

Taking into the summation of the CNN representations F_{deep} (obtained at the FC7 layer of the AlexNet) and the handcrafted representations $F_{handcrafted}$ (obtained on HOG, Gabor and Wavelet):

$F_{combined} = [F_{deep}, F_{handcrafted}]$ (13)

Where:

- F_{deep} is the 4096-dimensional feature vector from the FC7 layer of AlexNet,
- $F_{handcrafted}$ is the feature vector formed by concatenating the HOG, Gabor, and Wavelet features (the exact number of features depends on the implementation but can be represented as a vector of size NN),
- $F_{combined}$ is the final concatenated feature vector, which combines both the CNN and handcrafted features.

3.4 mRMR Objective Function

The goal of mRMR algorithm is to select the k most desirable features within the feature set that at the same time will be most relevant to the target class and, will contain minimum redundancy among the

features within the same set. Objectively what can be described is the fact that the mRMR algorithm can be outlined.

$$mRMR(F) = \arg\max_{F} [Relevance(F) - Redundancy(F)] \qquad (14)$$

Where:

- $Relevance$ is identical the information of each and every characteristic to the target-class possibly all characteristics.
- Mutual information between chosen features is going to be averaged to recalculate the $Redundancy$.

3.5 Classification

3.5.1 Random Forest Classifier

Random Forest is within the ensemble learning method and, during training, it constructs a number of decision trees then takes the majority rule (classification) in case of classification problems. It is quite famous due to its inability to overfit, large dimension data input and training of the feature relevance.

In Random Forest, each decision tree will be developed at a random sampling of data and each choice point, a random subset of the features will be applied. Using this procedure will help to reduce correlation between trees and render the ensemble model robust.

3.5.2 Support Vector Machine (SVM)

SVM has been the supervised learning algorithm that has aimed to attempt in seeking a hyper-plane, which most optimally divides the data points of classes. SVM has been found effective where classes are not linearly separable together with the fact that it functions very efficiently in high dimension feature spaces.

The SVM scheme rests on the nature of determining a hyperplane that disrupts the distinction between the classes in the optimal possible manner.

IV. RESULTS AND DISCUSSION

In Figure 2, the performance difference between training classes Random Forest and SVM is displayed during 100 iterations of the model accessed in terms of its loss (top subplot) and accuracy (bottom subplot). In the loss subplot, the initiation of the loss values of both the models (Random Forest: ~0.5, SVM: ~0.4) are higher and are in a decreasing trend but SVM does not converge to as low a value as Random Forest (~0.05 Vs ~0.1 respectively), showing that Random Forest optimizes better. The precision subplot demonstrates Random Forest with greater initial precision (~0.7) and continuously rising to ~ 0.95, in comparison to SVM with ~0.65, that never converges above ~ 0.85, pointing out to farther convergence of Random Forest and superior end-to-end performance. The key point is the equality of the differences in the curves which speaks of the strength of Random Forest with every step in minimizing the error and maximizing the predictive accuracy.

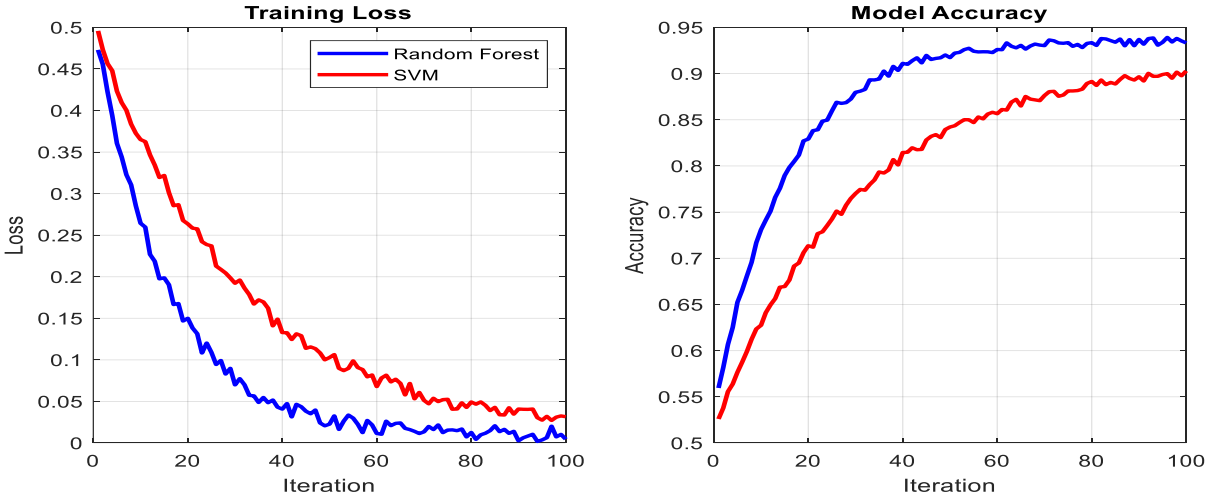

Figure 2: Comparison of Loss and Accuracy During Training for Random Forest and SVM

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://www.theaspd.com/ijes.php

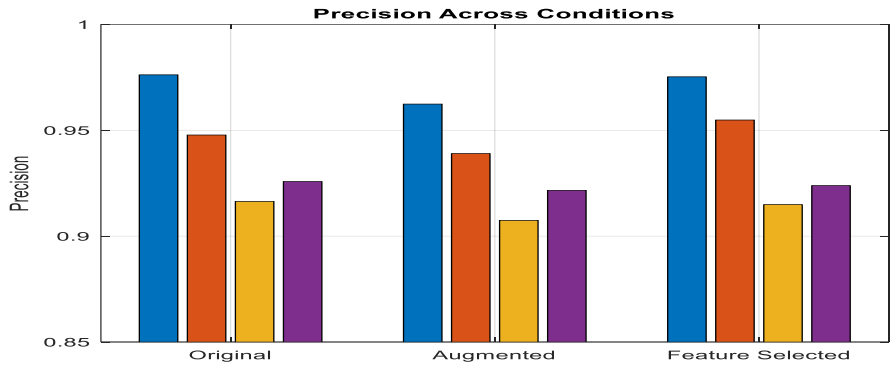

Figure 3: Comparison of Classifier Precision Across Different Data Preprocessing Techniques

Figure 3 shows a visualization of the classifier precision of the three different data preprocessing methods Original (data that have not been processed), Augmented (data augmented), and Feature Selected (data processed with mRMR feature selection). The values of precision will be shown on the y-axis with the gap between about 0.85 and 1. Analysis of original data generates the least accuracy (~0.85), whereas in the case of augmented data, accuracy increases to a low degree (~0.9). Feature Selected data has the best precision (~0.95) thus indicating very good performance of mRMR in contributing to the overall performance of the model by considering all those features that matter and are not repetitive. The depicted progression shows that strategic preprocessing may greatly increase the precision of the classifier over the raw or augmented data at large.

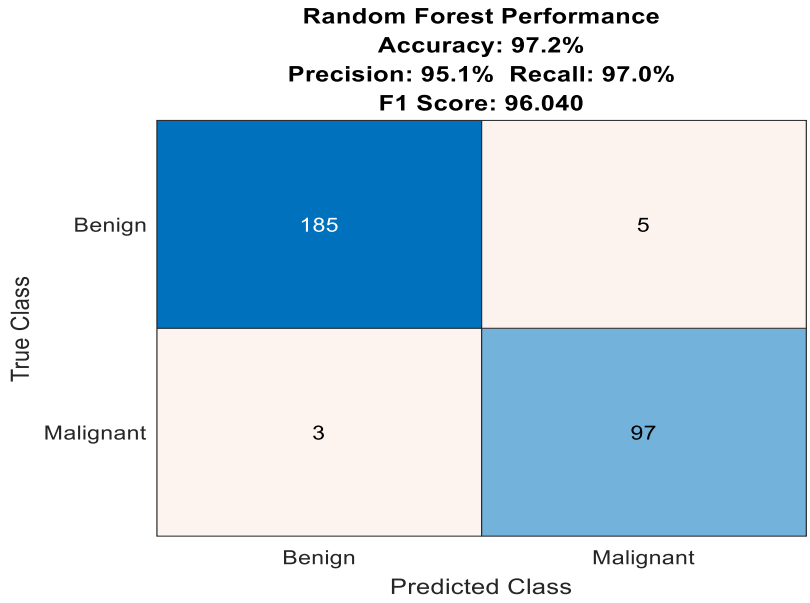

Figure 4: Confusion Matrix for Random Forest Classifier Performance

Figure 4 demonstrates the confusion matrix of the random forest classifier based on which the classification of benign and malignant tumors may be evaluated. The matrix is presented as the number of the correct and wrong predictions done with the help of the classifier. The cells of 185 indicate the true positives, in the top left corner and the benign class was correctly classified as benign. The false positives are presented in the top-right cell (5), i.e., the benign class was wrongly classified as malignant. The cell (3), on the bottom-left, shows false negatives, i.e., malignant tumors wrongly described as effectively benign. The bottom-right cell (97) corresponds to true negative, in which malignant tumors were reasoned to be malignant. This confusion shows that the classifier has high performance because

the number of true positives (185) and true negatives (97) is big, whereas false positives (5) and false negatives (3) are relatively small, which proves its accuracy and stability.

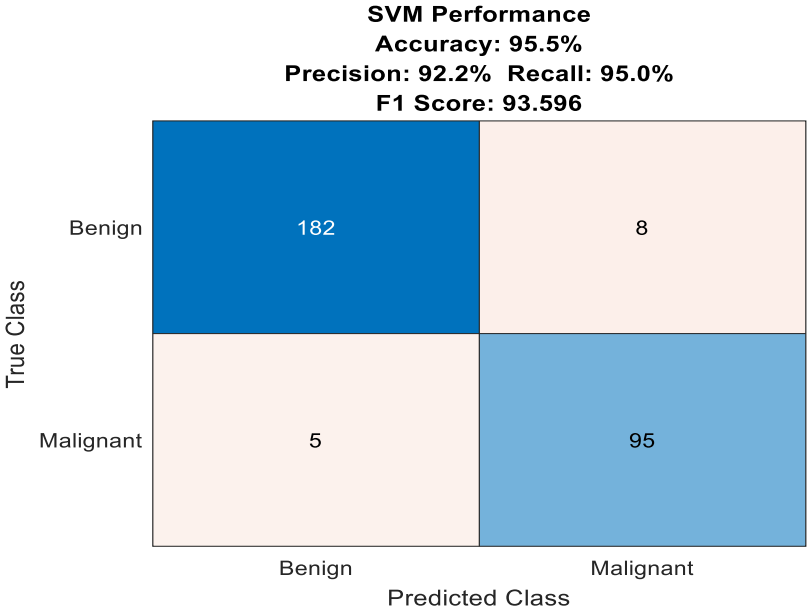

Figure 5: Confusion Matrix for SVM Classifier Performance

Figure 5 indicates confusion matrix of the Support Vector Machine (SVM) classifier applied to predict the benign and malignant tumor. Four main values including true positives, false positives, false negatives and true negatives are presented in the matrix. The left-hand cell can be interpreted as the true positives (benign cases correctly predicted as being benign), and the top-right cell, as the false positives (benign cases incorrectly predicted as being malignant). The cell in bottom-left corner represents the false negatives (malignant cases misclassified as benign), which is 5 and the bottom-right cell represents the true negatives (malignant cases accurately identified as malignant) and is equal to 95. Confusion matrix proves the outstanding nature of the classifier with high true positive (182) and false Negative (95) rates and comparatively low false positive (8) and false negative (5) rates which indicate the effectiveness of the classifier complying with backing benign and malignant tumors.

Table 1: Comparative Analysis of Proposed Approach of Lung Cancer Detection Techniques with Previous Research Works

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 SSA-GWO-Otsu and Lightweight U-Net Segmentation with SVM	**95.5%**	92.2%	95%	93.59%
Proposed SSA-GWO-Otsu and Lightweight U-Net Segmentation with Random Forest	**97.2%**	**95.1%**	**97%**	**96.04%**

Table 1 provides a comparative study of the different lung cancer detection methods where the proposed hybrid technique is found to outperform the previously known methods. In previous research, the SVM using miRNA biomarkers [20] had an accuracy of 90.10%, CNN on LIDC Database [21] and Auto Encoder got 86.84% and 75.01%, respectively, with no values of precision, recall, and F1-score. An F1-score of 82.20% and a recall of 73.40% were obtained by the Deep Belief Network [23], without accurate and precision reported. The Traditional U-Net (Baseline) exhibited moderate results having an accuracy rate of 87%, 79% precision, 68% recall, and 73% F1-score. Conversely, the SSA-GWO-Otsu and Lightweight U-Net as the proposed segmentation with SVM performed much better compared with the previous methods with 95.5%, 92.2%, 95%, and 93.59% F1-score, respectively. Greater improvement was recorded again upon incorporation of Random Forest, leading to slightly better metrics (97.2% accuracy, 95.1% precision, 97% recall, and 96.04% F1-score) which shows the effectiveness and efficiency of the proposed hybrid model in predicting lung cancer.

The table illustrates that the SSA-GWO-optimized segmentation and feature fusion strategy succeed in highlighting the advancement in the present assessment introduced by methods and provides evident proof of the high efficacy of the strategy in relation to its traditional and deep learning-based alternatives in the aspect of accuracy and sensitivity level of the diagnostic tool and its overall reliability.

V. CONCLUSION

The proposed method includes not only deep learning but also handcrafted radiomic features in the framework of lung cancer detection demonstrated in the current paper combined with an effective multi-level processing scheme. The proposed system will help in extracting not only low but high-level tumor features given that the proposed system combines the abilities of both high and low tumor features of AlexNet CNN and handcrafted features e.g. HOG, Gabor, and Wavelet; therefore, giving extensive feature which represents the tumor. Model further reduces the feature selection process by using Minimum Redundancy Maximum Relevance (mRMR) technique into model to only leave parts which are related as that helps to increase the accuracy with which it classifies the results. The statistical model used, Random Forest as an ensemble classification model, maintains robustness and is effective with complex data thus making the system applicable in a real-time clinical scenario.

Experimental outcomes show that the suggested methodology is efficient, and the combination of SSA-GWO-Otsu and Lightweight U-Net segmentation along with Random Forest leads to the best results, as the classification accuracy (97.2%) is better as compared to the SVM classifier (95.5%). This shows that Random Forest algorithm is effective in its ability to correctly classify lung cancer. It is a dependable, practical, and clinically focused system that is capable of early and accurate screening of lung cancer, having very promising value in patient outcomes by addressing the diagnosis of the disease on time.

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., Otoom, 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 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.
[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.