Journal IJCSM

Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421, p-ISSN: 2958-0544

An Automated Lion-Butterfly Optimization (LBO) based Stacking Ensemble Learning Classification (SELC) Model for Lung Cancer Detection

S.Swapna Rani [1], V.Suganya [2], S.Selvi[3]* N.Ashokkumar [4], S.Prema[5]

[1]Department of Electronics and Communication Engineering, Maturi Venkata Subba Rao (MVSR) Engineering college, Hyderabad, Telangana-501510.India.
[2]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai-600062, Tamil Nadu, India.
[3]Department of Electrical and electronics Engineering, Panimalar Engineering College, Chennai - 600 123, India,
[4] Department of ECE, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati-517102. India.
[5]Department of Chemistry, PSNA College of Engineering and Technology, Dindigul - 624622, Tamil Nadu 624622, India.

DOI: https://doi.org/ 10.52866/ijcsm.2023.02.03.008
Received; April 2023; Accepted July 2023; Available online August 2023

ABSTRACT: Lung cancer is one of the most serious and prevalent cancers in the globe. Early detection of lung cancer can increase a patient's chances of life. Computed Tomography (CT) scan images are difficult for clinicians to utilize in order to determine the stages of cancer. Computer-aided systems can assist researchers in more precisely predicting the stages of lung cancer in recent times. This study demonstrates the use of technology that is made possible by machine learning and image processing to accurately classify and predict the lung cancer from CT images. The existing tumor detection frameworks have the major difficulties in terms of high complexity, overfitting and error prediction. Therefore, the proposed work aims to formulate a simple and accurate automated system for the prediction and classification of lung cancer from CT images. Before classifying the input lung scan image, an adaptive median filtering approach is used to improve its contrast and quality. From the segmented lung parts, the histogram and texture features are derived. The most relevant characteristics are chosen using the Lion-Butterfly Optimization (LBO) method for training and testing operations. Eventually, the input CT picture is correctly predicted as either healthy or disease-affected using the Stacking Ensemble Learning Classification (SELC) algorithm. In this study, a thorough performance evaluation is conducted utilizing several measures in order to analyze the outcomes

Keywords: Lung Cancer, Medical Image Processing, Preprocessing, Segmentation, Feature Extraction, Lion-Butterfly Optimization (LBO), Stacking Ensemble Learning Classification (SELC), and Computer Aided Diagnosis (CAD)

1. INTRODUCTION

The most deadly and pervasive disease that causes a significant number of fatalities each year is cancer. Lung cancer is the most common and has the greatest fatality rate of all the major types of cancer [1, 2]. The most frequent and deadly form of cancer that affects both men and women is lung cancer. The development of malignant lung tumors, also known as carcinoma, results from unchecked cell proliferation in the lung tissues. Smoking and tobacco addiction are the two main risk factors for developing malignant lung nodules. In total, just 14% of lung cancer patients survive the disease after all stages, during a period of 5–6 years [3]. The primary issue with lung cancer is that it is typically discovered at advanced stages, which makes treatment more challenging and sharply lowers survival rates. Hence, early detection of lung cancer can lower the death rate by reducing the mortality rate and increasing the survival possibilities up to 70-80% by giving the patients with the essential fast treatments [4, 5]. The two primary kinds of lung cancer are classified according to cell characteristics as small cell lung cancer and non-small cell lung cancer. About 85–88% of all incidences of cancer are non–small cell lung cancer, while small cell lung cancer accounts for approximately 15% of all cancer cases [6]. Researchers have created a variety of lung disease diagnosis models to improve disease detection

87*Corresponding author: *ashoknoc@gmail.com*

methods in the early stages of lung cancer, which will benefit clinicians and medical professionals. The images from a computed tomography (CT) [7, 8] scan are best for identifying lung nodules in lung cancer.

Typically, the CT scans are employed to detect lung cancer because they give a clear image of the tumor inside the body and monitor its development. Despite the fact that CT is favored over other imaging modalities, visual interpretation of these CT scan images may be a lengthy process that increases the chance of error and delays the discovery of lung cancer [9]. Specifically, image processing, machine learning, and artificial intelligence techniques can process the medical field data with the aid of engineering solutions for the purpose of detection and diagnosing the lung disease [8, 10]. These technologies are particularly desirable in the direction of preventing the lung cancer disease and for analyzing the early stage of this disease. Because of new methodologies combining transfer learning techniques like fine-tuning, the automatic segmentation of components existing inside an image, such as the region of interest in diagnostic images, has improved. It took a lot of research to find and classify lung conditions. It is challenging for a professional physician to identify infections, vascular nodules, etc. because of the numerous radiographic images of the lungs and their complex and uneven structure. The computer-assisted diagnosis method enables clinicians to help with the identification of a disease. New methodologies incorporating transfer learning and techniques have enhanced the automatic segmentation of elements existing inside of an image, including the region of interest in diagnostic imaging. Although there are a number of image processing-based approaches for detecting lung cancer [11], each one has drawbacks, as can be shown from the literature. Also, the method's efficacy can be raised by suggesting an improved approach for this use. The diagnosis of lung cancer depends heavily on accurate classification. The approaches described in the literature use various models for accurate classification. The accuracy of the system can therefore be enhanced by implementing an optimal classifier [12, 13]. The objective of this paper is to construct an automated CAD model for the detection and class categorization of lung cancer from CT images. The major research contributions of this paper are as follows:

- An adaptive median filtering technique is employed to preprocess the given medical image for contrast enhancement and quality improvement.
- The segmentation of the Region of Interest (ROI) is carried out to enhance the classification.
- To improve the classifier's training and testing accuracy, the segmented portion's histogram and texture features are retrieved.
- In order to streamline the tumor detection procedure, the most pertinent features are selected using the clever Lion-Butterfly Optimization (LBO) method.
- In addition, the Stacking Ensemble Learning Classification (SELC) algorithm is used to classify photos of healthy and disease-affected individuals in a simple manner.
- The effectiveness and outcomes of the suggested LBO-SELC mechanism's lung cancer detection are investigated and validated utilizing a variety of metrics.

The remaining portions of this paper are structured into the following sections: Section 2 presents the detailed literature review of the existing works relevant to lung cancer segmentation, detection and classification. Also, it examines the pros and cons of several ML algorithms used in the medical imaging field. Section 3 provides the clear description for the proposed LBO-SELC mechanism for an automated lung cancer detection and classification. The performance and comparative results along with the dataset descriptions are presented in Section 4. Finally, the paper is summarized with the findings, outcomes and future scope in Section 5.

2. RELATED WORKS

A study of numerous methods for the classification and detection of cancer using image processing models is presented in this section.

Nadkarni, et al [14] developed an automated image processing framework for the segmentation and classification of lung tumor from CT images. Here, the morphological operations have been applied on the quality improved lung CT image for accurately segmenting the lung region and tumor affected portions. Moreover, the standard SVM classification algorithm is applied to categorize the class of image as healthy or tumor-affected. However, the suggested technique is not more suitable for handling the large datasets, due to overfitting. Pawar, et al [15] presented a comprehensive study to examine the effectiveness of various machine learning algorithms used for lung cancer prediction. The goal of this work is to examine and evaluate various models that have been put forth in earlier research. This research also clarifies the problems in diagnosing lung disease (LDD) from CT and X-ray images. Moreover, the lung nodule segmentation is also performed with the use of thresholding technique. Nageswaran, et al [16] developed a new automated lung cancer diagnosis model with the use of recent image processing mechanisms. Here, the geometric mean filtering approach is utilized to improve the image quality before classification. Then, the Linear Discriminant Analysis (LDA) mechanism is used to extract the features from the segmented image, which effectively reduces the amount of space required for the construction of data matrix. Moreover, the group of classifiers such as ANN, KNN and RF have been implemented to predict lung cancer according to the extracted feature set. The ANNs are frequently employed in the medical sector to categorize medical images with the intent to detect diseases. The ANN functions similarly to the nervous system in terms of how it does its tasks. By perusing a collection of already categorized images, it is possible to gain the knowledge necessary to make an intelligent prediction about the category to which an image belongs. It is simple to learn about the algorithms used in ML using the KNN technique, which is the strategy that is used the most frequently. It is a supervised learning strategy that doesn't require the usage of any variables. The widely popular random forest approach can be used to create prediction models. Regression and classification are just two of the numerous applications that may be carried out using RF. As long as datasets are correctly modified, it is feasible to create machine learning algorithms that are capable of producing predictions with a high level of accuracy. Mutiullah, et al [17] presented a comprehensive study on various image processing techniques used for lung tumor identification and classification. Nasser, et al [18] utilized an ANN technique for identifying lung tumor from CT images. The purpose of this paper is to analyze the most appropriate factors used for the detection of lung cancer. Here, an ANN is specifically modeled to identify the presence of tumor with high accuracy. Bhatia, et al [19] investigated the performance of various classification models such as deep residual network, XGBoost, RF, and ensemble learning for developing an automated lung cancer detection system. Khumancha, et al [20] used a 3D-CNN model for the detection of lung cancer from CT images. With the aid of image processing and 3D CNN, the authors hope to identify the presence of early or late-stage cancer in CT scan images of the lungs. Whether the patient has cancer or not will be determined by the system using the 3D CT scans as input. Nonetheless, researchers are unable to immediately apply CNN to a CT scan since it would be challenging to identify a small nodule in the extensive 3D CT scan volume. The solution is to crop out small volume 3D cubes from the CT scan that include nodules in order to narrow the search area and train a nodule detector. Once nodules have been found, a second 3D CNN can be taught to distinguish between cancerous and non-cancerous nodules.

Lakshmanaprabu, et al [21] deployed an optimized deep learning model for the detection of lung cancer with increased accuracy. In this study, the Optimal Deep Neural Network (ODNN) and Linear Discriminate Analysis were used to analyze the CT scan lung images (LDA). To categorize lung nodules as malignant or benign, deep features are recovered from CT lung images, and the dimensionality of the features is slightly reduced using LDR. Lung cancer categorization is determined by applying the ODNN on CT scans and then optimizing it using the Modified Gravitational Search Algorithm (MGSA).

From the literature review [22-24], it is obvious that most of the digital image processing techniques utilized in this field have both benefits and limitations, which are illustrated in below:

Advantages:

1. ML techniques are easy to comprehend, compared to the deep learning models.
2. Easy interpretation.
3. Works faster with less training time.
4. Does not require any manual intervention.
5. Highly robust to irrelevant attributes.

Disadvantages:

1. Slow in process while handling the large datasets.
2. Complexity in prototype modeling.
3. Not more compatible.
4. Highly prone to overfitting.
5. Increased number of features degrade the performance rate.
6. Does not have the capability to function with the noisy data.

According to this survey, the proposed work intends to implement a new and advanced image processing algorithms for developing an automated lung tumor detection system.

3. PROPOSED METHODOLOGY

This section gives a brief explanation of the framework for the proposed automated lung cancer diagnosis using machine learning and advanced optimization methods. The major goal of this research is to provide a concise and accurate framework for identifying lung cancer using CT images. It incorporates the following operational components, as depicted in Fig. 1:

- Image acquisition
- Adaptive median filtering-based preprocessing
- ROI segmentation
- Histogram and texture feature extraction
- Lion-Butterfly Optimization (LBO) for feature selection
- Stacking Ensemble Learning Classification (SELC) for disease prediction

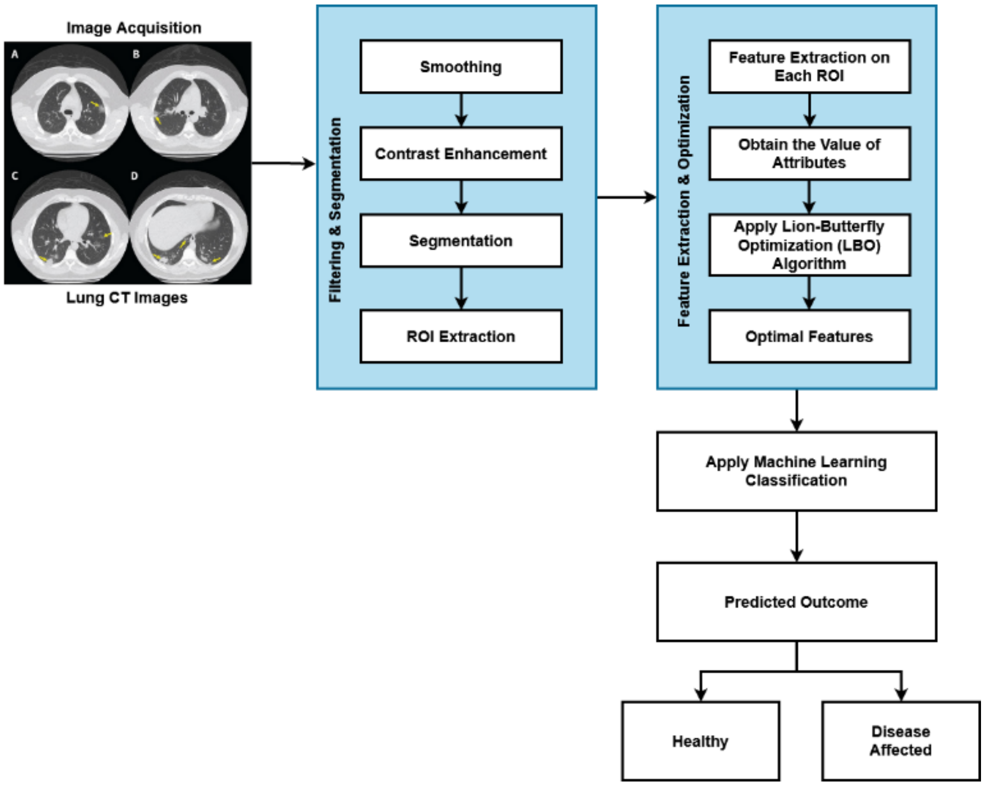

FIGURE 1. Proposed automated lung cancer detection system

A. PREPROCESSING AND SEGMENTATION

In order to optimize the image data for further analysis, preprocessing is highly essential that eliminates undesired distortions or enhance specific visual qualities. It is sometimes used to enhance image features like segments, boundaries, and textures so that we can easily distinguish the desired and undesired components of the images. Moreover, it is necessary to minimize the effects of distortion found in imaging devices such as light fluctuation and coloration, and at the same time preprocessing is required to remove unsolicited regions from the images. Hence, several researchers employ various filtering approaches according to the type of noise for medical applications. The goal of the preprocessing stage of an image is to reduce any undesired distortions already present in the image and to highlight any features that will be helpful during the evaluation phase. Image enhancement and image smoothing are two of the primary phases. To get rid of any unwanted noise that is present in the image, smoothing is used. As salt and pepper noise frequently appears in CT scan images, an adaptive median filtering has been proven to be a very successful technique for removing this impulsive noise while maintaining the edges. The best results for image smoothing come from an adaptive median filtering since it reduces noise without distorting the image. The quality of digital images is improved using image enhancement techniques to produce better results for subsequent operations. As artefacts brought on by contrast fluctuations in the image have an impact on image quality, contrast correction is done to improve the image. By changing input pixel values to new values such that by default 1% of the data gets saturated at low and high intensity of the input image data, contrast adjustment enhances the image's brightness.

The act of separating an image into two halves, one of which is desired and the other undesired, where the ROI is another name for desired component. Our goal in the case of lung cancer is to locate the tumor seen in the lung imaging. Because the lung tissue in the abnormal region has different textural features, it is essential that a machine automatically detect the tumor. The region of interest (ROI) in this case is a particular area of the lung image where lung nodules or tumors were discovered and procedures were performed to extract feature values.

B. FEATURE EXTRACTION

Features are merely certain measured values or characteristics that can be helpful in classifying an element for predictive modeling [25]. In order to limit the amount of data, features must be extracted. Here, the histogram and texture features are used for extracting the features from the normalized lung image, and they are all retrieved from various bands of CT scan images. The image is distorted while describing histogram features. The histogram shows the quantity of pixels in an image at each power value. The histogram approach analyses an entire range of grey levels from the given image. There are 256 levels of grey present, ranging from 0 to 255. It has several traits in common such as variance, mean, skewness, and kurtosis.

- Variance: Gray level variations from the mean value are indicated by the term variance, which also specifies the number of such fluctuations. It is possible to identify reduced texture contrasts using statistical distributions, such as the variance in the length of lines of a particular threshold.
- Mean: The variance indicates how many grey levels deviate from the mean value. It is possible to discern minimal texture contrasts using statistical distributions, such as the variance in the length of lines of a specific limit.
- Skewness: Depending on the histogram's tail, the image skewness is determined. Both the positive and negative tails of the histogram values are separated into two distinct sets.
- Standard Deviation: The square root of the image contrast variance serves as the measure of standard deviation. By comparing high and low variance values, the level of visual contrast is assessed.
- Kurtosis: It represents the degree of abnormality in the image and measures the potential distribution of a real-valued random variable. In statistical analysis, kurtosis and skewness are used to understand the distribution's shape.

Only adjacent to the histogram characteristics are texture features from the input image obtained. When the irregularity is widely dispersed throughout the image, each class's textural orientation is unique, which contributes to higher classification accuracy. The grey level co-occurrence matrix represents a statistical technique for analyzing the surface that considers the spatial relationship of pixels. By calculating the frequency of combinations of pixels with the same values, the GLCM functions describe the texture of an image.

C. LION-BUTTERFLY OPTIMIZATION (LBO) BASED FEATURE SELECTION

In this stage, the most pertinent features are selected from the extracted feature set with the use of LBO algorithm. The traditional works utilized several meta-heuristic optimization models for reducing the dimensionality of features. Due to the increased complexity and low convergence, the proposed work intends to utilize a novel and intelligent optimization algorithm, named as, LBO for feature selection. This technique is developed based on the standard Lion optimization [26] and butterfly optimization [27] algorithms[33]. Typically, butterfly optimization algorithm's foraging and mating behaviors are similar to those of butterflies. The ability of these butterflies to perceive their environment has allowed them to survive for a long time. They use their senses of sight, feel, smell, sound, and taste to find a mate and food. These senses are extremely helpful when moving from one location to another. Moreover, initialization, iteration, and final are the three different phases of butterfly optimization algorithm. The fitness values of each and every butterfly available in the solution space are computed for each iteration as they move to new positions. The technique initially calculates the fitness values for each butterfly at each discrete point in the solution space. These butterflies will release the scent in their spots later. Then, the best solution is computed based on the following equation:

$$S_i^{k+1} = S_i^k + (r^2 \times y^* - S_i^k) \times F_i \tag{1}$$

Where, S_i^{k+1} is the solution vector of i^{th} butterfly, r indicates the random number [0 to 1], y^* defines the current best solution, and F_i is the fragrance of i^{th} butterfly. Then, the local search space is carried out as shown in below:

$$S_i^{k+1} = S_i^k + (r^2 S_j^k - S_h^k) \times F_i \tag{2}$$

Where, j and h are the butterflies. Butterfly mate selection and food search occur on both a local and a global scale. In BOA, switching between the common global search and intense local search is done using the switch probability. The primary driving force behind lion optimization is seen in nature, where lions typically engage in territorial invasion and defense behaviors. It includes the following phases: pride generation, fertility evaluation, mating, territory defense, and termination. During pride generation, the length of lion is represented by using the following model:

$$X = \begin{cases} c & if\ c > 1 \\ d & Otherwise \end{cases} \tag{3}$$

Where, c and d represent the integers used for estimating the length of lions. Consequently, the velocity is estimated with the vector elements by using the following models:

$$\alpha(s_e) \in (s_e^{mn}, s_e^{mx}) \tag{4}$$

Where, α is the velocity, e is the element i.e., $e = 1,2, \dots \alpha$. After pride generation, the fertility evaluation is carried out, where the lioness fertility is updated based on the following equations:

$$s_e^{f+} = \begin{cases} s_g^{f+} & if\ e = g \\ s_e^{f} & Otherwise \end{cases} \tag{5}$$

$$s_g^{f+} = min[s_g^{mx}, max(s_g^{mn}, \Delta_g)] \tag{6}$$

$$\Delta_g = [s_g^f + (0.1r_2 - 0.05)(s_g^M - r_2 s_g^{FM})] \tag{7}$$

Where, g is the random number between the range of 1 to α, then the notation Δ represents the female update function, and r_2 is the random number [0 to 1]. Moreover, s_g^M and s_g^{FM} are the lion and lioness respectively, s_g^{mx} and

s_g^{mn} denotes the maximum and minimum values, and s_e^f denotes the fertile lioness. Furthermore, the mating process is carried out, where the gender clustering is performed as the supplementary step. During territory defense, the nomadic lion is elected, and the process is terminated when the following condition is met:

$$k > k_{max} \tag{8}$$

Based on this process, the best solution is returned as the output function, which can be used to choose the pertinent features from the extracted feature set. The flow of LBO algorithm is shown in Fig 2.

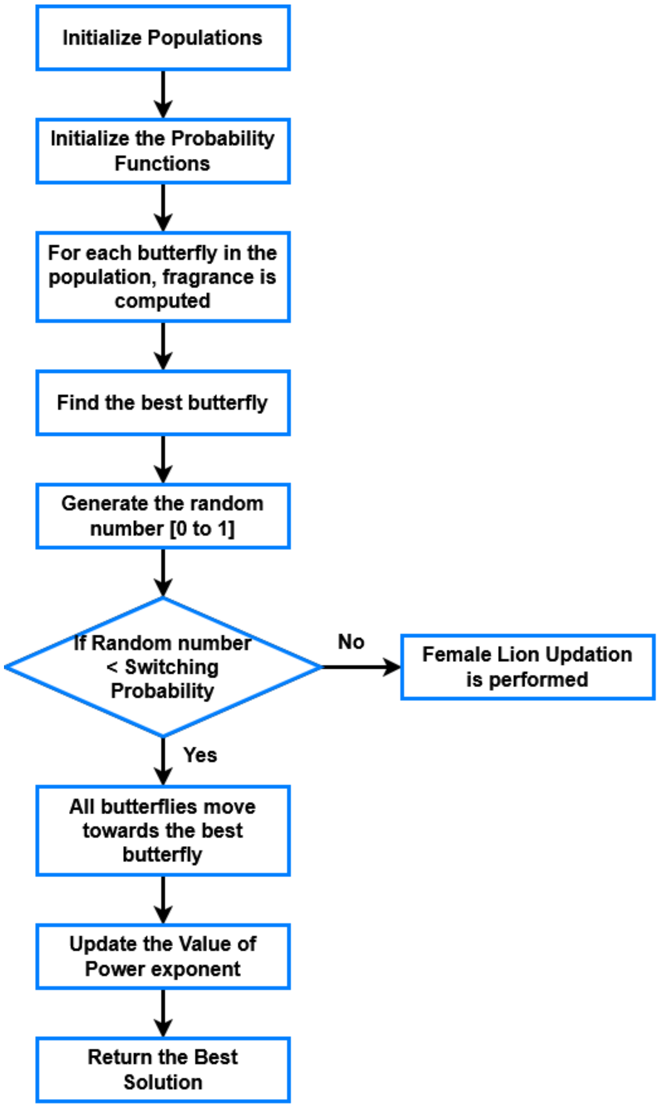

FIGURE 2. **Flow of LBO**

D. STACKING ENSEMBLE LEARNING CLASSIFIER (SELC)

In this stage, the selected features are trained by the classifier for an accurate prediction of lung cancer from the given CT images. A type of heterogeneous ensemble method is known as a stacking ensemble learning model. To improve the strong classifier's capacity for generalization, the heterogeneous ensemble combines various types of basic classifiers into it. After performing training on a number of base classifiers, the meta-classifier is then fed the forecasted results from the basis classifiers. Consequently, the training is repeated for disease prediction. The eventual ensemble approach will improve classification accuracy and performance greatly by making use of the learning capabilities of the base classifier and meta-classifier [28, 29]. The structure of the SELC mode is shown in Fig 3, which includes the base learning models of RF, DT, and gradient boost.

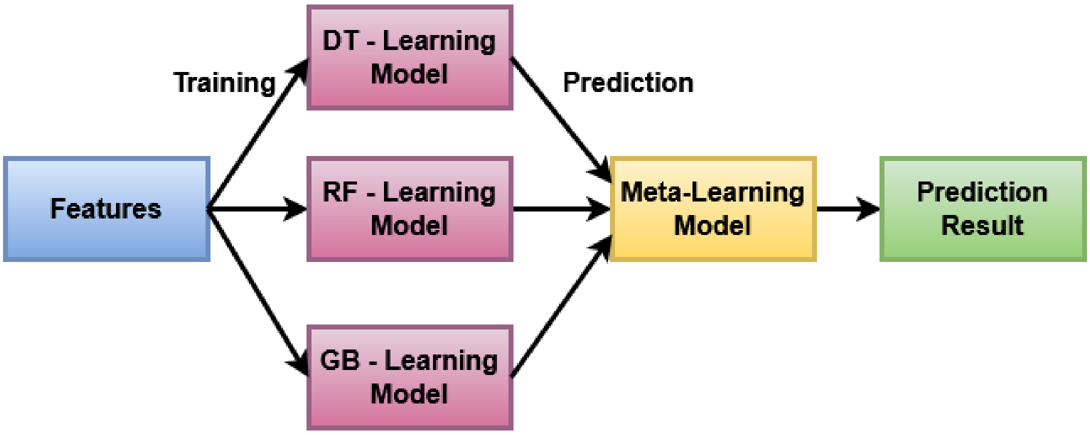

FIGURE 3. **Structure of SELC**

In the context of ensemble learning, RF is a type of algorithmic process termed as bagging concept. The Bagging idea suggests using random sampling to create numerous training subsets from the training set, then training each training subset to create a linear classifier. The prediction result is then generated by aggregating the output of these classification models or by voting. The forecast model is optimistic that it will produce a more precise prediction in this context and be more capable of generalizing. In order to build the Bagging ensemble technique, random forest uses the decision tree as its training sample. In the meantime, it randomly generates the features to improve the prediction solution with low overfitting. A type of algorithm used in ensemble learning called gradient boosting decision tree uses the decision tree as its basis function. The fundamental idea behind this algorithm is that each tree learns by summing its prior inferences from all other trees. By using this classifier, image is accurately categorized as whether healthy or lung-cancer affected.

4. RESULTS AND DISCUSSION

By using a variety of performance metrics, this section confirms the effectiveness and outcomes of the suggested LBO-SELC procedures. Also, a few contemporary machine learning models that are popular in the medical field have been taken into comparison. The Computed Tomography (CT) images with annotated lesions are included in the Lung Image Database Consortium image collection (LIDC-IDRI) [30-32], which is used for lung cancer diagnosis and screening. It comprises of more than a thousand DICOM picture scans from high-risk patients. Each scan includes a collection of pictures that show the chest cavity in various axial slices. The number of 2D slices in each scan varies depending on the subject and the imaging device used. The DICOM files comprise a header that lists the patient id information as well as other scan-related information like slice thickness. The performance measures used to assess the results are calculated using the following models:

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \tag{9}$$

$$Sensitivity = \frac{Tp}{Tp+Fn} \tag{10}$$

$$Specificity = \frac{Tn}{Tn+Fp} \tag{11}$$

$$F1 - score = \frac{2\times Precision\times \mathrm{Recall}}{\mathrm{Precision+Recall}} \tag{12}$$

$$\mathrm{Precision} = \frac{\mathrm{Tp}}{\mathrm{Tp+Fp}} \tag{13}$$

$$\mathrm{Recall} = \frac{\mathrm{Tp}}{\mathrm{Tp+Fn}} \tag{14}$$

Where, Tp – True positive, Tn – True negative, Fp – False positive, and Fn – False negative. Fig 4 and Table 1 presents the overall comparative analysis of the existing and proposed classification methodologies used for lung cancer detection. The major goal of this project is to create a system for lung tumor automated diagnosis. There is discussion of the model estimations and results. The suggested model's measure indicators have been assessed based on their precision, accuracy, recall, and F1-score. A measurement used to assess system accuracy is called precision. The accuracy is a significant metric that is mostly evaluated for system performance analysis. This indicator is calculated using the ratio of accurate estimation values to all available estimates. The proportion of all relevant instances that were recovered is known as the recall. This indication demonstrates the classifier's capacity to identify all of the positive pixels. A test's accuracy can be determined by the F1-score. The recall and precision indications serve as the foundation for this metric.

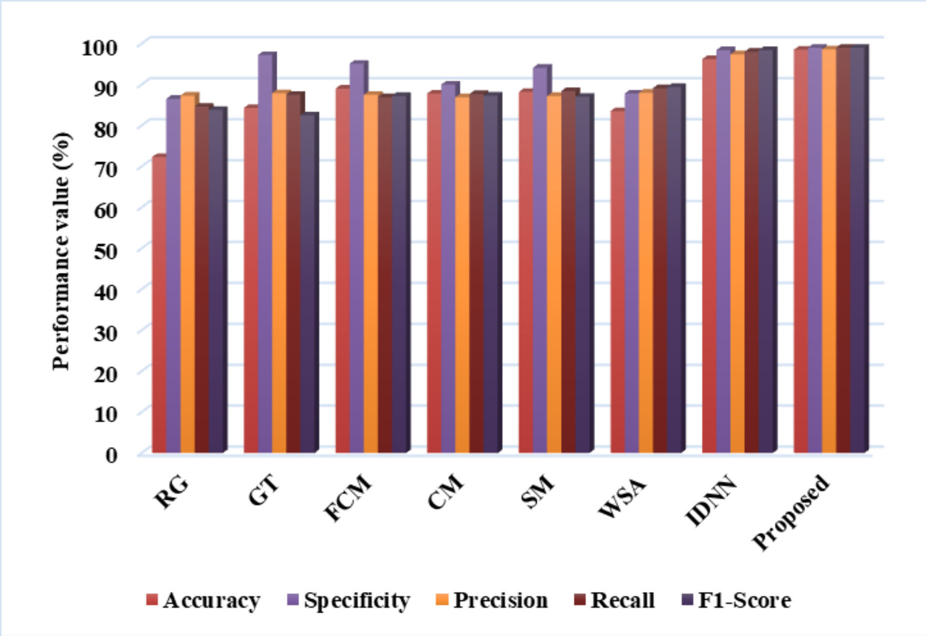

FIGURE 4. Performance comparative analysis using LIDC dataset

Table 1. Performance evaluation

Methods	Accuracy	Specificity	Precision	Recall	F1-Score
RG	72.3	86.5	87.3	84.6	83.8
GT	84.3	97.2	87.9	87.5	82.5
FCM	89	95.1	87.5	86.9	87.2
CM	87.8	90	86.9	87.7	87.3
SM	88.2	94.1	87.2	88.4	87
WSA	83.5	87.8	88	89.1	89.4
IDNN	96.2	98.4	97.4	98	98.4
Proposed	98.5	99	98.6	99	99

Fig 5 to Fig 8 presents the comparative analysis between the proposed LBO+SELC and existing optimization + ML models, which includes the classifiers of NB, IBK, J48 and ensemble learning model. In order to prove the superiority of the proposed lung cancer detection system, the results are validated and compared with many ML algorithms incorporated with optimization techniques. Overall, the results proved that the LBO-SELC overwhelms the existing classifiers with highly improved performance outcomes.

Table 2. Comparative analysis between the proposed and existing optimization + NB classification models

Methods	Accuracy	Precision	Recall	F1-Score
NB + GAWA	81.8	89.3	87.9	82.8
NB + ACO	78.6	79.6	75.6	76
NB + PSOMS	73.8	72.5	73	74.8
NB + HSOGR	68.1	66.2	69.3	65.81
Proposed	98.5	98.3	98.5	98.4

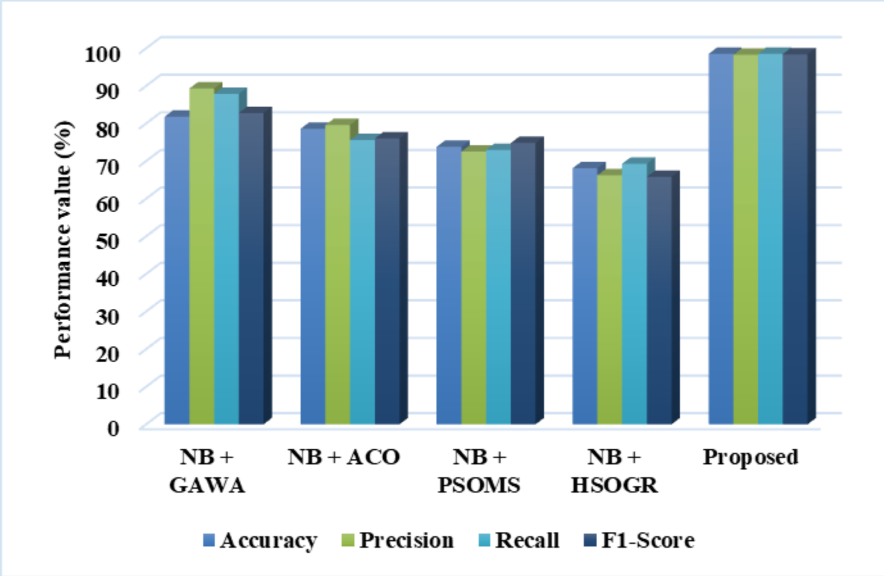

FIGURE 5. **Comparative analysis among the LBO-SELC and optimization + NB classification models**

Table 3. **Comparative analysis between the proposed and existing optimization + IBK classification models**

Methods	Accuracy	Precision	Recall	F1-Score
IBK + GAWA	88.9	88	88.9	91
IBK + ACO	79.6	78.6	76	83.6
IBK + PSOMS	78	76.8	77.3	76.5
IBK + HSOGR	73.1	72.1	71.3	70.2
Proposed	98.5	98.4	98.5	98.3

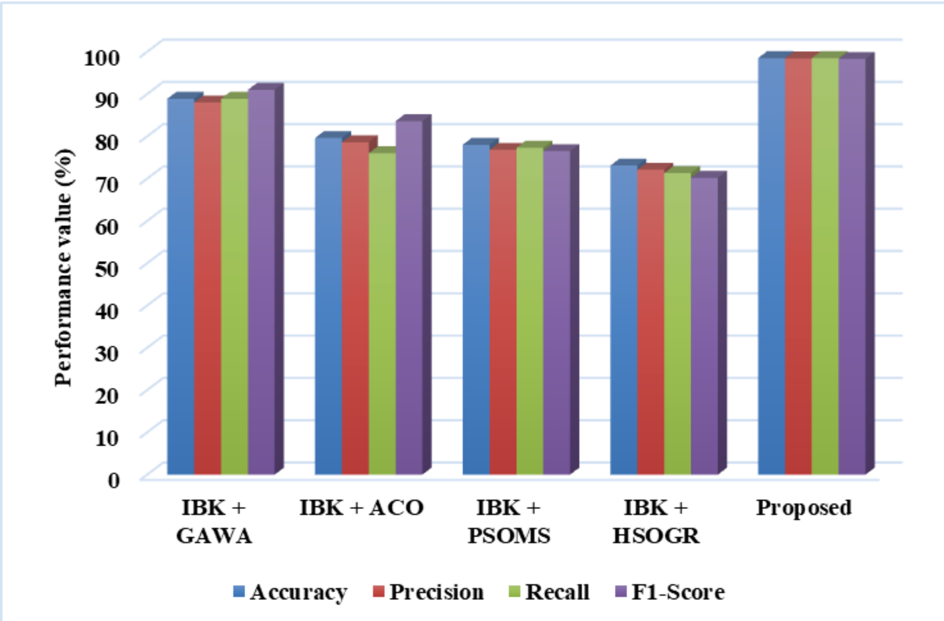

FIGURE 6. **Comparative analysis among the LBO-SELC and optimization + IBK classification models**

Table 4. **Comparative analysis between the proposed and existing optimization + J48 classification models**

Methods	Accuracy	Precision	Recall	F1-Score
J48 + GAWA	88	90.3	88.9	89
J48 + ACO	80	83.6	78	79

J48 + PSOMS	79.3	77.2	78	78.2
J48 + HSOGR	76.2	71.4	72	75.4
Proposed	98.5	98.4	98.5	98.3

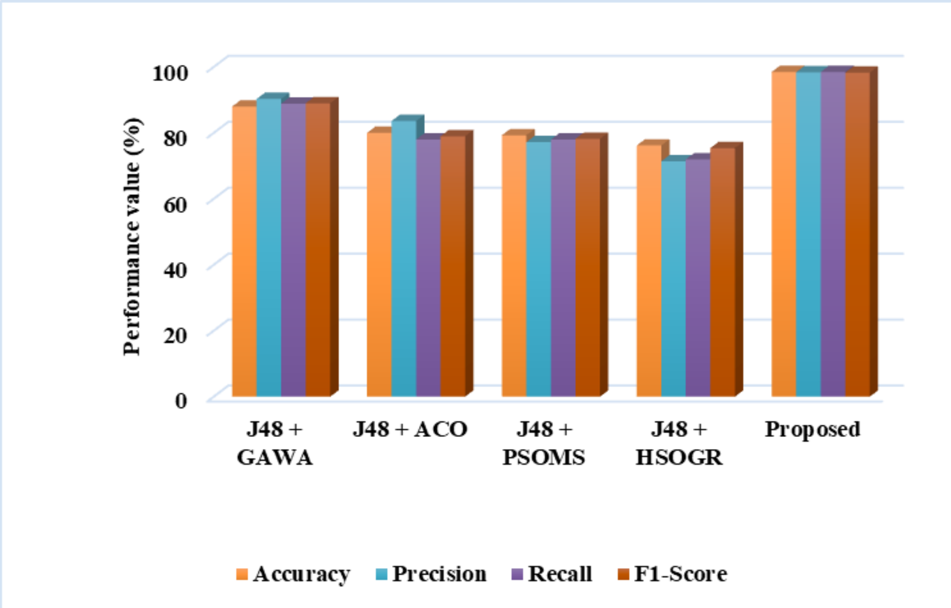

FIGURE 7. Comparative analysis among the LBO-SELC and optimization + J48 classification models

Table 5. Comparative analysis between the proposed and existing optimization + Ensemble classification models

Methods	Accuracy	Precision	Recall	F1-Score
Ensemble+ GAWA	97.61	97.2	97.53	97.54
Ensemble + ACO	94.8	93.5	94.4	94.8
Ensemble + PSOMS	93.2	92.52	93.6	93.4
Ensemble + HSOGR	92.11	91.3	92.3	92.1
Proposed	98.5	98.4	98.5	98.3

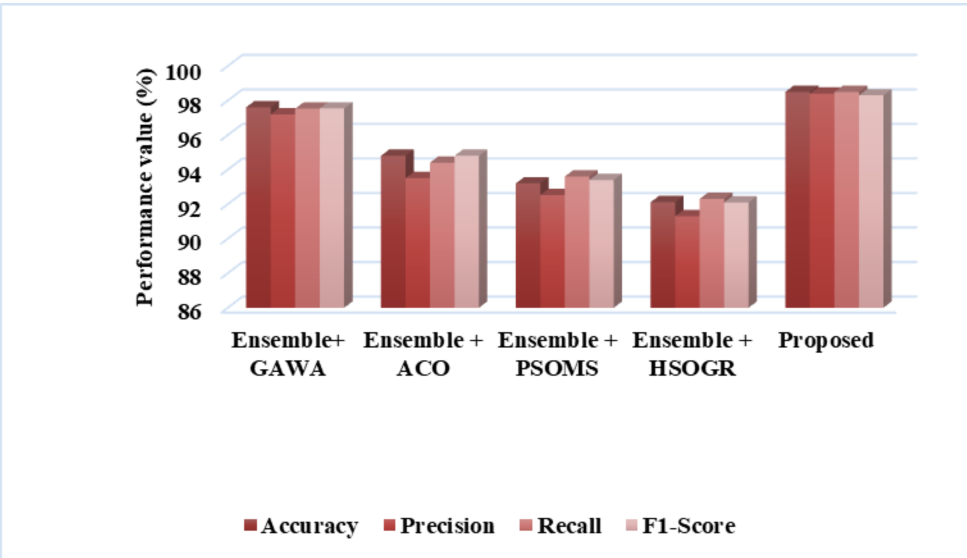

FIGURE 8. Comparative analysis among the LBO-SELC and optimization + Ensemble classification models

Moreover, the general execution results such as accuracy, sensitivity, and specificity of the proposed LBO-SELC technique and other state of the art models are shown in Fig 9 to Fig 11. As observed from the results, it is clear that the combination of LBO-SELC outperforms the other classification approaches with improved accuracy, sensitivity, and specificity values. Due to the proper image preprocessing and feature extraction operations, the overall lung cancer detection performance is greatly improved in the proposed system.

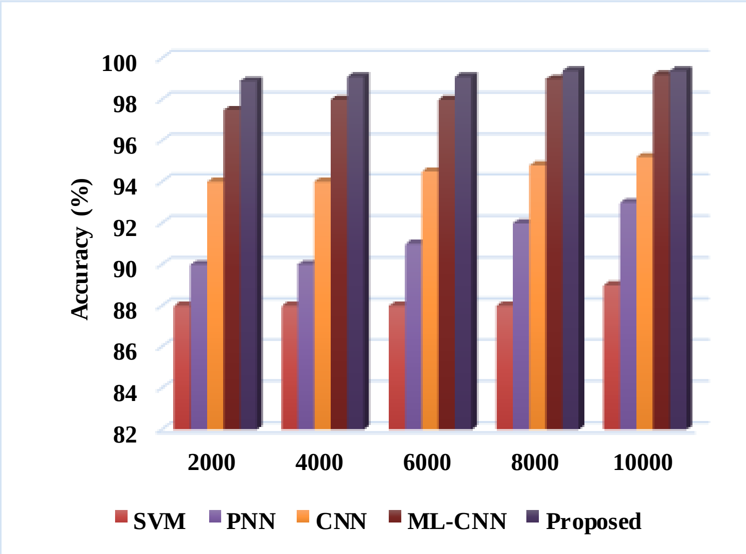

FIGURE 9. Accuracy analysis

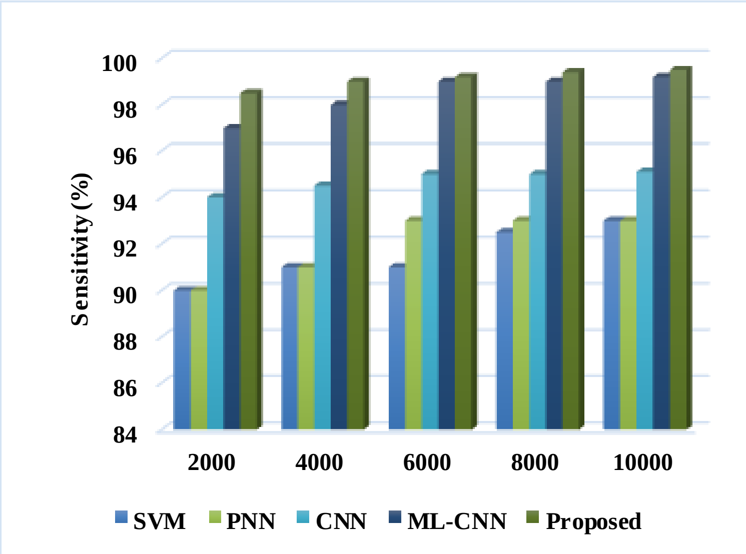

FIGURE 10. Sensitivity analysis

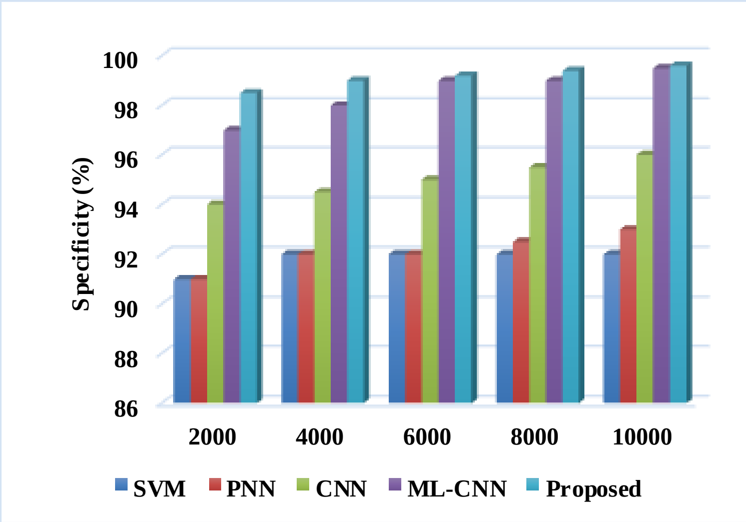

FIGURE 11. **Specificity analysis**

5. CONCLUSION

A million people die from lung cancer each year, making it one of the worst forms of the disease. The detection of lung nodules on chest CT images is crucial in the field of human medicine. The adoption of CAD systems is essential for the early diagnosis of lung cancer as a consequence. An essential task that is used across many different economic sectors is image processing. To find areas of the lung that have been affected by cancer, image processing techniques such noise reduction, feature extraction, and classification are utilized. This work uses an advanced optimization and machine learning based image processing methodologies to demonstrate precise lung cancer identification and prognosis. To improve contrast and quality, an adaptive median filtering approach is used to preprocess the provided medical image. Then, the ROI is segmented in order to improve classification. The histogram and texture properties of the segmented area are obtained to enhance the classifier's training and testing accuracy. By using the LBO method, the most important features are chosen in order to speed the tumor diagnosis process. Moreover, the SELC method is employed to quickly categorize images of persons with healthy and diseased states. By using a range of measures, the efficiency and results of the proposed LBO-SELC mechanism's lung cancer detection are examined and validated. According to the results, it is obvious that the proposed LBO-SELC provides an improved disease detection performance with increased accuracy up to 99%, when contrasted to the other classifiers. In future, the deep learning-based segmentation and classification models can be implemented for developing a lung cancer diagnosis framework.

Funding

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

None

REFERENCES

[1] S. Huang, J. Yang, N. Shen, Q. Xu, and Q. Zhao, "Artificial intelligence in lung cancer diagnosis and prognosis: current application and future perspective," in Seminars in Cancer Biology, 2023.

[2] A. heba mehdi, & F. Rabee. (2022). Breast Cancer Diagnosis By IoB Implanted Tag Design And Machine Learning . Al-Salam Journal for Engineering and Technology, 2(1), 1–12. https://doi.org/10.55145/ajest.2023.01.01.001

[3] N. S. Reddy and V. Khanaa, "Intelligent deep learning algorithm for lung cancer detection and classification," Bulletin of Electrical Engineering and Informatics, vol. 12, pp. 1747-1754, 2023.

[4] K. S. Pradhan, P. Chawla, and R. Tiwari, "HRDEL: High ranking deep ensemble learning-based lung cancer diagnosis model," Expert Systems with Applications, vol. 213, p. 118956, 2023.

[5] S. Maurya, S. Tiwari, M. C. Mothukuri, C. M. Tangeda, R. N. S. Nandigam, and D. C. Addagiri, "A review on recent developments in cancer detection using Machine Learning and Deep Learning models," Biomedical Signal Processing and Control, vol. 80, p. 104398, 2023.
[6] A. Choudhury, S. Balasubramaniam, A. Pradeep kumar, S. Karthikeyan, and S. N. Prasad Kumar, "PSSO: Political Squirrel Search Optimizer driven Deep learning for severity level detection and classification of Lung cancer," International Journal of Information Technology & Decision Making, 2023.
[7] A. M. Sebastian and D. Peter, "Identifying the Predictors from Lung Cancer Data Using Machine Learning," in Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022, ed: Springer, 2023, pp. 691-701.
[8] M. Varchagall, N. Nethravathi, R. Chandramma, N. Nagashree, and S. M. Athreya, "Using Deep Learning Techniques to Evaluate Lung Cancer Using CT Images," SN Computer Science, vol. 4, p. 173, 2023.
[9] M. Aluka and S. Ganesan, "A Comparative Study on Pre-Training Models of Deep Learning to Detect Lung Cancer," International Journal of Intelligent Systems and Applications in Engineering, vol. 11, pp. 148-155, 2023.
[10] K. Elgohary, S. Ibrahim, S. Selim, and M. Elattar, "A CAD System for Lung Cancer Detection Using Chest X-ray: A Review," in Advances in Model and Data Engineering in the Digitalization Era: MEDI 2022 Short Papers and DETECT 2022 Workshop Papers, Cairo, Egypt, November 21–24, 2022, Proceedings, 2023, pp. 12-25.
[11] R. S.JEBUR, (2023). Image Denoising using Mean Filter . Al-Salam Journal for Engineering and Technology, 2(2), 173–181. https://doi.org/10.55145/ajest.2023.02.02.020
[12] D. Hu, S. Li, N. Wu, and X. Lu, "A Multi-modal Heterogeneous Graph Forest to Predict Lymph Node Metastasis of Non-small Cell Lung Cancer," IEEE Journal of Biomedical and Health Informatics, 2023.
[13] S.-O. Shim, M. H. Alkinani, L. Hussain, and W. Aziz, "Feature ranking importance from multimodal radiomic texture features using machine learning paradigm: A biomarker to predict the lung cancer," Big Data Research, p. 100331, 2022.
[14] N. S. Nadkarni and S. Borkar, "Detection of lung cancer in CT images using image processing," in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 863-866.
[15] V. J. Pawar, K. D. Kharat, S. R. Pardeshi, and P. D. Pathak, "Lung cancer detection system using image processing and machine learning techniques," Cancer, vol. 3, p. 4, 2020.
[16] R. Sharma, , Rai, B., & Sharma, S. (2023). Classification on Unsupervised Deep Hashing With Pseudo Labels Using Support Vector Machine for Scalable Image Retrieval . Wasit Journal of Computer and Mathematics Science, 2(2), 40–55. https://doi.org/10.31185/wjcms.147.
[17] M. Islam and R. Rab, "Analysis of CT scan images to predict lung cancer stages using image processing techniques," in 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019, pp. 0961-0967.
[18] I. M. Nasser and S. S. Abu-Naser, "Lung cancer detection using artificial neural network," International Journal of Engineering and Information Systems (IJEAIS), vol. 3, pp. 17-23, 2019.
[19] S. Bhatia, Y. Sinha, and L. Goel, "Lung cancer detection: a deep learning approach," in Soft Computing for Problem Solving: SocProS 2017, Volume 2, 2019, pp. 699-705.
[20] M. B. Khumancha, A. Barai, and C. R. Rao, "Lung cancer detection from computed tomography (CT) scans using convolutional neural network," in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-7.
[21] A. Sabih Shaker, F. Omar Youssif, Mohammad Aljanabi, ABBOOD, Z., & Mahdi S. Mahdi. (2023). SEEK Mobility Adaptive Protocol Destination Seeker Media Access Control Protocol for Mobile WSNs. Iraqi Journal For Computer Science and Mathematics, 4(1), 130–145. https://doi.org/10.52866/ijcsm.2023.01.01.0011
[22] L. N. Gumma, R. Thiruvengatanadhan, L. Kurakula, and T. Sivaprakasam, "A Survey on Convolutional Neural Network (Deep-Learning Technique)-Based Lung Cancer Detection," SN Computer Science, vol. 3, pp. 1-7, 2022.
[23] Y. Li, X. Wu, P. Yang, G. Jiang, and Y. Luo, "Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis," Genomics, Proteomics & Bioinformatics, 2022.
[24] G. Chassagnon, C. De Margerie-Mellon, M. Vakalopoulou, R. Marini, T.-N. Hoang-Thi, M.-P. Revel, et al., "Artificial intelligence in lung cancer: current applications and perspectives," Japanese Journal of Radiology, pp. 1-10, 2022.
[25] S. A. Althubiti, S. Paul, R. Mohanty, S. N. Mohanty, F. Alenezi, and K. Polat, "Ensemble learning framework with GLCM texture extraction for early detection of lung cancer on CT images," Computational and Mathematical Methods in Medicine, vol. 2022, 2022.
[26] K. Pradhan, P. Chawla, and S. Rawat, "A deep learning-based approach for detection of lung cancer using self adaptive sea lion optimization algorithm (SA-SLnO)," Journal of Ambient Intelligence and Humanized Computing, pp. 1-15, 2022.
[27] P. Sinthia, M. Malathi, and M. Suresh Anand, "Improving lung cancer detection using faster region-based convolutional neural network aided with fuzzy butterfly optimization algorithm," Concurrency and Computation: Practice and Experience, vol. 34, p. e7251, 2022.

[28] H. Sabah Salih, M. Ghazi and M. Aljanabi, . (2023) “Implementing an Automated Inventory Management System for Small and Medium-sized Enterprises”, Iraqi Journal For Computer Science and Mathematics, 4(2), pp. 238–244. doi: 10.52866/ijcsm.2023.02.02.021.
[29] M. Mamun, A. Farjana, M. Al Mamun, and M. S. Ahammed, "Lung cancer prediction model using ensemble learning techniques and a systematic review analysis," in 2022 IEEE World AI IoT Congress (AIIoT), 2022, pp. 187-193.
[30] P. M. Shakeel, M. Burhanuddin, and M. I. Desa, "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier," Neural Computing and Applications, pp. 1-14, 2022.
[31] Y. Hussain Ali, V. Sabu Chooralil, K. Balasubramanian, R. R. Manyam, S. Kidambi Raju, A. T. Sadiq, et al., "Optimization System Based on Convolutional Neural Network and Internet of Medical Things for Early Diagnosis of Lung Cancer," Bioengineering, vol. 10, p. 320, 2023.
[32] Q. Tian, Y. Wu, X. Ren, and N. Razmjooy, "A new optimized sequential method for lung tumor diagnosis based on deep learning and converged search and rescue algorithm," Biomedical Signal Processing and Control, vol. 68, p. 102761, 2021.
[33] Chitra, T., Sundar, C., & GOPALAKRISHNAN, S. (2022). Investigation and Classification of Chronic Wound Tissue images Using Random Forest Algorithm (RF). *International Journal of Nonlinear Analysis and Applications*, *13*(1), 643-651. doi: 10.22075/ijnaa.2021.24438.2744