FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

# PSbBO-Net: A Hybrid Particle Swarm and Bayesian Optimization-based DenseNet for Lung Cancer Detection using Histopathological and CT Images

**Saurabh Singh Raghuvanshi[1*], K. V. Arya[2*], and Vinal Patel[3]**

*[1,2]Multimedia and Information Security Research Group, Department of Computer Science and Engineering, ABV-Indian Institute of Information Technology and Management, Gwalior-474015, India; saurabhr@iiitm.ac.in[1]; kvarya@iiitm.ac.in[2]*
*[3]Department of Electrical and Electronics Engineering, ABV-Indian Institute of Information Technology and Management, Gwalior-474015, India; vp@iiitm.ac.in[3]*

***Correspondence:** Saurabh Singh Raghuvanshi and K.V. Arya; *saurabhr@iiitm.ac.in, kvarya@iiitm.ac.in*

**ABSTRACT-** Lung cancer remains a substantial global fatality; early detection is imperative for successful intervention and treatment. Deep learning (DL) models have shown promise in predicting lung cancer from medical images, but optimizing their parameters remains a challenging task. To improve prediction capability, this study introduces an approach by merging Particle Swarm Optimization and Bayesian Optimization (PSbBO) to optimize deep learning parameters. PSO provides an effective way for exploring the hyperparameter space, while Bayesian optimization provides a probabilistic framework for the effective evaluation and refining of a DL network. The simulation study showcases the effectiveness of the proposed model, achieving notable metrics for histopathological images, including an accuracy of 99.5%, precision of 98.3%, recall of 99.2%, F1-score of 99.4%, and an error rate of 1.19%. Furthermore, when applied to lung CT images, the proposed PSbBO demonstrates an accuracy of 98.8%, precision of 97.4%, recall of 98.3%, F1-score of 98.6%, and an error rate of 1.21%.

**Keywords:** Lung cancer detection, Deep learning, Particle Swarm Optimization, Bayesian Optimization.

**ARTICLE INFORMATION**

**Author(s)**: Saurabh Singh Raghuvanshi, K. V. Arya and Vinal Patel;
**Received**: 27/07/2024; **Accepted**: 20/09/2024; **Published**: 25/09/2024;
**e-ISSN**: 2347-470X;
**Paper Id**: IJEER 2707-24;
**Citation:** 10.37391/ijeer.120343
**Webpage-link**:
https://ijeer.forexjournal.co.in/archive/volume-12/ijeer-120343.html

**Publisher's Note:** FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

## 1. INTRODUCTION

Cellular breakdown in the lungs remains one of the most vital reasons for disease-related mortality around the world [1]. The ability to treat lung cancer successfully and completely can be considerably improved with early detection as well as prediction of the disease [2]. Imaging studies, genetic testing, clinical assessment, and predictive modeling are among the array of techniques utilized for the identification and prediction of lung cancer [3]. These techniques can help to identify those who are more likely to develop lung cancer and diagnose the disease earlier when therapy is more effective [4]. As a result, more precise techniques for identifying and forecasting lung cancer have been developed. Examples include machine learning algorithms that can analyze vast volumes of medical data to produce prediction models [5]. People at elevated risk for developing lung cancer may find regular screening for the condition advantageous [6]. Early detection of lung cancer can significantly improve a patient's prognosis and survival chances. This holds as early-stage lung cancer is more manageable than when diagnosed at an advanced stage [7]. Developing a patient-specific treatment plan depends on accurately assessing the likelihood of lung cancer. This can involve figuring out the best course of action, the ideal radiation or chemotherapy dosage, and the length of the course of action, etc. [8]. By reducing the need for costly therapies and lengthy hospital stays, early detection, and accurate lung cancer prediction may be able to lower healthcare expenditures [9]. A major public health concern, lung cancer could be reduced through early detection and prognostication [10]. On a community level, we can improve lung cancer outcomes by identifying high-risk patients and discovering disease at an earlier stage [11].

Even if lung cancer has not manifested any significant symptoms or effects during a person's lifetime, it may still be found and treated. This may result in needless adverse effects and overtreatment [12]. Radiation exposure is a factor in some imaging tests used to find lung cancer, such as CT scans. Repeated radiation exposure can raise the risk of developing cancer, especially in people who are already at a greater risk [12-13]. The stigma and discrimination that come with having lung cancer, especially if the patient is a non-smoker, can be caused by the disease's frequent association with smoking [12-14]. Lung cancer diagnosis and prediction using artificial intelligence (AI) has shown promising results. AI systems can analyze diverse medical data such as imaging scans, clinical notes, and genomic information to discern patterns and make predictions [11-15]. AI has been employed to develop risk prediction models for lung cancer and to analyze radiographic images for the detection of lung nodules and determine how

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

likely they are to be malignant. Large datasets can be used to train machine learning algorithms to find patterns and increase their accuracy over time [10-16].

One of the primary advantages of AI in lung cancer detection and prediction lies in its ability to swiftly and efficiently process large volumes of data. This can assist medical professionals in making better decisions and locating patients who might benefit from additional testing or treatment. Additionally, AI can increase diagnostic precision and lower the possibility of human error [17]. A range of traditional methods were utilized, including Neural Architecture Search (NAS) [18], Recurrent Convolutional Neural Network (RCNN) [19], Radiology Analysis and Malignancy Evaluation Network (R2MNet) [20] to fix these issues, but no workable answers appeared. As a result, this study employs a novel methodology.

The remaining part of the research is summarized as follows: *section 2* outlines the literature and *section 3* constructs the system model and lays out the problem statements. The proposed framework is then described in *section 4*, and the findings are then presented and compared in *section 5*. The paper ultimately concludes in *section 6*.

## 2. RELATED WORKS

Nanglia et al. [21] presented a novel strategy combining a Support Vector Machine and a Feed-Forward Back Propagation (FFBP) Neural Network is introduced to lower the computational cost of categorization. 500 images compose the dataset, of which 75% were used for training and remaining 25% for classification. FFBP neural networks are well suited for handling complicated issues because they can learn non-linear correlations between inputs and outputs. It has been reported that large datasets and intricate structures are computationally expensive to train FFBP neural networks.

It has been investigated that Lung cancer and chronic obstructive pulmonary disease both commonly involve pulmonary emphysema. However, in patients with lung cancer classifications that are ambiguous, quantitative emphysema severity prediction is crucial. Ananthajothi *et al.* [22] introduced an Electric Fish-based Grey Wolf Optimization (EF-GWO) for the early detection of pulmonary emphysema, which holds paramount importance in potentially saving lives. Furthermore, painless positive strain breathing is a fundamental life-saving procedure that focuses on the easing of patients' difficult circumstances. Rational diagnosis is important because it increases the likelihood of survival when non-invasive positive pressure breathing fails.

Lung nodules are important signs that lung cancer is present. Early detection increases a patient's chance of survival by allowing for timely treatment initiation. Scientists have devised algorithms for Computer-Aided Diagnosis (CAD) systems to alleviate the burden of radiologists' time-consuming and challenging task of identifying and classifying malignancy in Computed Tomography (CT) images. To overcome these issues, a CNN-based methodology is deployed to enhance the pulmonary nodules' categorization accuracy in CT scans [23]. This method includes an Energy Layer (EL), which pulls textural characteristics from the convolutional layer. Timely identification of lung cancer is imperative, and radiologists use the results of CT scans to recommend further therapy. The traditional method for looking for nodules in CT images is time-consuming. To address these challenges, a multidimensional Region-based Fully Convolutional Network (mRFCN), developed in [24], has been introduced as a computerized choice guidance system for the identification and categorization of lung nodules. The study explores the application of position-sensitive score mappings (PSSM), employing a distinct multi-layer fusion Region Proposal Network (mLRPN) as the foundation of the image classifier to extract features.

The addition of noise signals alongside creative signals during the image capture stage can potentially compromise the quality of cancer images, leading to a decline in performance and posing a challenge for automated lung disease diagnosis. To tackle these hurdles, *Shakeel et al.* [25] emphasize the reduction of inaccurate categorization to sharpen the cancer prognosis and improve the lung quality of images. They used lung CT pictures from the Cancer Imaging Archive (CIA) dataset, utilizing a weighted mean histogram balance strategy to wipe out commotion effectively. This cycle at last raised picture quality through the Improved Profuse Clustering Technique (IPCT). Alzubi *et al.* [26] presented the Weight Optimized Neural Network with Maximum Likelihood Boosting (WONN-MLB) for exact analysis of Cellular breakdown in lung cancer disease (LCD) in broad datasets. This approach plans to diminish the grouping season of LCD determination by coordinating the Newton-Raphsons MLMR preprocessing model. Furthermore, it uses the Boosted Weighted Optimized Neural Network Ensemble Classification calculation to improve precision and decrease bogus positive rates. It has been found that the method is more sensitive to hyperparameters and has a higher computational cost.

Maleki et al. [27] developed a k-Nearest-Neighbors method, utilizing a hereditary calculation for capable component determination, in this way decreasing dataset aspects and speeding up the classifier. Notably, the integration of the kNN approach with a feature selection algorithm significantly enhances classification accuracy.

The main contributions of this paper are as follows:

- A hybrid PSO and Bayesian Optimization approach is integrated into a DenseNet model to fine-tune the model parameters.
- The hybrid optimization approach to the lung cancer detection domain is a new experiment to improve lung cancer diagnosis accuracy.
- The design of Particle swarm-based Bayesian optimization (PSbBO) method which uses above the two optimization techniques that can efficiently explore the parameter space and find optimal hyperparameters, such as the learning rate and epochs of this lung cancer detection model.
- The identification of lung nodules is done using a DenseNet based classifier that classify the malignancy of the nodule

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

on the basis of the features derived from the feature extraction phase.

# 3. PROPOSED METHODOLOGY

A novel particle swarm-based Bayesian optimization (PSbBO) has been designed to predict and classify malignant behaviors in lung modules. PSO is used to adjust the deep learning model's variables to improve accuracy and performance. In the proposed model PSO is used to detect early diagnosis of the cancer and forecasts the malignant behaviors of lung modules. Bayesian optimization is employed to enhance the enactment and precision of the model for predicting cancer of the lung. The GAN neural network can be used to create images, analyze medical images, and find the type of lung cancer. *Figure 1* shows the block diagram of the proposed PSbBO.

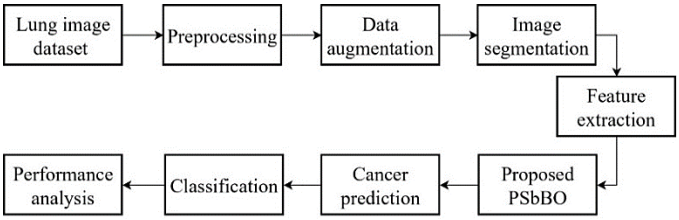



**Figure 1.** Block Diagram of proposed PSbBO

At first, the datasets are sourced from a reputable online platform. Following this, the histopathological lung image dataset is integrated into the system, with particular attention given to excluding any noise and error values during the preprocessing stage. Random cropping and resizing are used for Data Augmentation. This method can be used to create new training images by randomly cropping and resizing the original medical images. This can help to introduce variations in the size and position of the tumor regions, which can enhance models' success in detecting and classifying tumors. Image segmentation is used to recognize and distinguish the nodule region from the surrounding tissues of the images in the context of nodule border identification and ROI extraction. Medical imaging applications are the use of image segmentation techniques for nodule border identification and ROI (Region of Interest) extraction. The separated region then moves on to the feature extraction stage, when different spectral properties are determined, including mean, skewness, standard deviation, and kurtosis. This approach proves effective in identifying features associated with lung cancer. In the proposed work a new Particle swarm-based Bayesian Optimization (PSbBO) is used to obtain the best combination of hyperparameters (batch size, drop-out, epochs, etc.) and model parameters. Moreover, the likelihood of cancer presence in the input images is predicted using the DenseNet-based GAN model. In prediction, the presence of cancer or non-cancer images is predicted. Later, classification is done using the DenseNet model to categorize the input images into distinct cancer types. Finally, the accuracy, recall, and Confusion matrix-based performance measures were computed.

## 3.1. Lung Image Dataset

During initialization, the lung imaging information is gathered from the standard website. Lung image datasets are frequently used for applications such as lung disease diagnosis, lung nodule identification, lung region segmentation, and abnormality categorization. Here, the Histopathological and *CTlungimage* dataset is used. These datasets often include a large number of images that the histopathological images were captured through a medical imaging technique called histopathology, which involves the examination of tissue samples, including lung tissue, under a microscope to study the cellular and structural changes that may indicate disease, such as lung cancer. This makes it possible to create extremely accurate and reliable algorithms that can help medical professionals diagnose lung disorders, find nodules, segment lung regions, and classify abnormalities more effectively. It is commonly employed in tasks involving the recognition, classification, and lung nodule malignant forecasts, crucial for early-stage lung cancer identification. The system utilizes the data initialization function to set up the dataset. *Equation (1)* [37] uses to express it.

$$I(t^*) = (s_1, s_2, s_3 \ldots\ldots s_n) \quad (1)$$

Where $t^*$ represents the collected image dataset, $s$ represents the data present in the lung image dataset and $n$ signifies the total quantity of data in the dataset sample.

## 3.2. Preprocessing

The collected Histopathological lung images were processed in preprocessing. Since the gathered raw images may encompass irregularities and substandard pixels that diminish the accuracy of lung cancer predictions, it is imperative to systematically eliminate noise and uneven pixels through constant pixel adjustment. Preprocessing is performed using the below *equation (2)* [37].

$$P^* = I(t^*) - e(t) \quad (2)$$

Where $P^*$ represents the preprocessing function, $I(t^*)$ represents the dataset, and $e(t)$ represents the error value. Data normalization was done on the images using a min-max technique to scale the input data. It accelerates the procedure and ensures consistent analysis. The normalized image is produced by performing a grayscale image $N = [X \subseteq I^n] \rightarrow [Min, \ldots\ldots, Max]$ and assuming that the normalized picture is $N^* = [X \subseteq I^n] \rightarrow [Min_{new}, \ldots\ldots, Max_{new}]$, with intensity values between $Min_{new}$and$Max_{new}$. The normalized equation is represented in *equation (3)*[28].

$$N^* = Min_{new} + (N - Min) \times \frac{Max_{new} - Min_{new}}{Max - Min} \quad (3)$$

where $Min$ and $Max$ describe the values for the grayscale image intensity whereas $Min_{new}$ and $Max_{new}$ define the values for the normalized image intensity.

## 3.3 Data Augmentation

The performance and generalization of deep learning models are significantly enhanced by data augmentation. This study

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

uses a generation tool and a tool for discrimination to enhance the data using a GAN. The generator creates synthetic visuals that resemble the original data distribution using random noise as input. The discriminator is being taught to distinguish between actual and fake images in the meantime. Extra training samples are generated by applying adjustments like scaling, resizing, and translation to original images. Scaling simulates zoom changes by resizing images and accommodating various object sizes. Translation shifts images horizontally or vertically to train the model for different object positions. Cropping and padding methods remove or add pixels to aid the model's focus on specific regions of interest or handle varied aspect ratios, enhancing detection and classification performance.

### 3.4 Image Segmentation

Here, the augmented data of the lung images are moved to the segmentation stage. It is to identify the boundaries of the images. Segmentation simplifies the image analysis process by bringing down the complexity of the image. Segmenting lung nodules is used to increase the image's resolution. The objective is to precisely locate and separate areas in lung images that might point to the existence of malignant cells or tumors. In this segmentation process to identify the cancer tissues CNN based U-Net model is applied. The histopathological and CT lung images can be segmented into multiple parts using the image segmentation process, allowing the malignant spots to be recognized, and distinguished from healthy lung tissue.

### 3.5 Feature Extraction

The separated part proceeds to the stage of extraction of features. The derived features serve as inputs to the designed model, which assigns them to the class that most accurately represents their characteristics. Through estimating positive attributes, feature extraction aims to minimize the data volume. The extraction of features is performed using the below *equation (4)*[37].

$$F(t) = P(t) - \lambda(t) \tag{4}$$

where $F(t)$ signifies the feature extraction function, $P(t)$ represents the wanted data, and $\lambda(t)$ represents the unwanted data. Later, different otherworldly qualities like mean, standard deviation, fourth, kurtosis, and skewness are calculated and used as a feature to predict and classify cancer in lung cancer image modalities. Thus, reducing the dataset's dimensionality becomes crucial for an overall improvement in lung cancer detection.

## 4. PROPOSED METHOD

The Particle Swarm based Bayesian Optimization (PSbBO) method is a hybrid optimization method used for optimizing a DenseNet model for lung cancer detection. In this work, PSO and BO methods are fused around the GP model which is a core part of the surrogate model in the BO. The GP model is used to approximate an objective function by modelling it as a distribution. In this work, the GP model is used by the PSO for the initialization of the swarm particles. The best candidate of the PSO output is used to refine the GP model. Moreover, particles in the PSO move according to the results from the BO's acquisition function (AQ). The GP model provides predictions and uncertainties for the acquisition function which helps to find the region to be searched next. The logical fusion of the two methods, PSO and BO, is explained in *figure 2*.

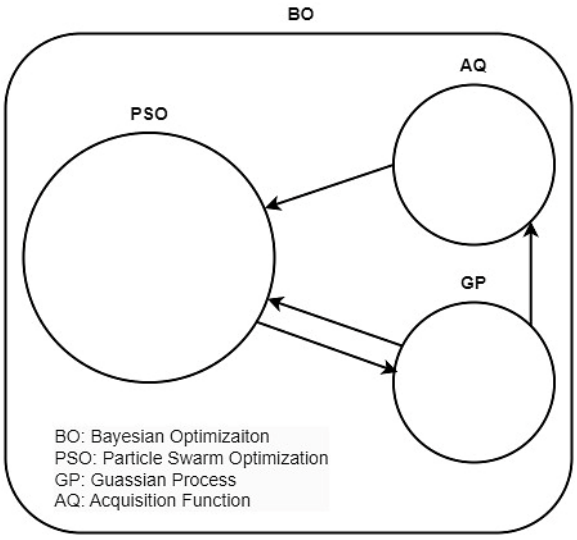



**Figure 2.** The proposed framework showing the logical interaction of the PSO and BO

### 4.1 Particle Swarm Optimization

Kennedy and Eberhart [29] introduced the Particle Swarm Optimization (PSO) algorithm, inspired by the searching patterns of birds. In this method, each "bird" is represented as a massless and volumeless particle, collectively forming the population. These particles symbolize potential solutions to the optimization difficult at hand. Throughout the PSO method, every particle undergoes a comprehensive evaluation based on its individual history as well as the shared experiences of neighbouring particles. After the required features are extracted, the information is gathered by PSO. It takes the data as each individual is massless and volumeless, and the group is formed with these individuals called the population. These are adjusted dynamically to get speed and each position. This process is repeated repeatedly to get the best value. Subsequently, guided by both the local best solution ($p_{best}$)and global best solution ($g_{best}$) continuously modify their flying route and velocities. The following *eq. (5) (6) and (7)* [29] shows the updated speed $S_i(t+1)$ and positions $P_i(t+1)$ of these particles:

$$S_i(t+1) = D.S_i(t) + L_1R_1[O_i(t) - P_i(t)] + L_2R_2[G_i(t) - P_i(t)] \tag{5}$$

$$P_i(t+1) = P_i(t) + S_i(t) \tag{6}$$

$$D = D_{\max} - \frac{D_{\max} - D_{\min}}{t_{\max}} t \tag{7}$$

where $O_i$ stands for the particle's optimal position, $G_i$ for the particle's global optimal position, $L_1$ and $L_2$ for the factors of learning, $R_1$ and $R_2$ for random numbers, $D_{max}$ and $D_{min}$ for the upper and lower bounds of the inertia weight, then $t_{max}$ and for the maximum and current repetitions, respectively.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

## 4.2 Bayesian Optimization

Bayesian Optimization [30] is a method for optimizing expensive black-box functions by building a probabilistic model, usually a Gaussian Process (GP), to approximate the objective function f(x). The GP provides a posterior distribution, updated as new data is collected, with mean μ(x) and variance $\sigma^2(x)$ An acquisition function (e.g., Expected Improvement or Upper Confidence Bound) is used to select the next point to evaluate, balancing exploration and exploitation. This process continues iteratively, optimizing f(x) with minimal evaluations.

Bayesian Optimization uses a probabilistic approach to diagnose lung cancer by optimizing hyperparameters based on diagnostic accuracy. It integrates prior information and extracted features to estimate posterior probabilities using Bayes' theorem, categorizing disease. The process involves a surrogate function created through Bayesian optimization and searched using an acquisition function. Gaussian modelling represents the objective function, aiding the effective search for optimal parameter values.

## 4.3 Proposed Hybrid Optimization

The trained output from the PSO is combined to form hybrid data and it is validated by minimizing the error in the detection process. This will result in globally optimized parameters. These globally optimized hyperparameters are accordingly used for classification. Variations in the hyperparameters are observed during the testing of the trained data.

Suppose $y_n$, is then chosen particle of PSO for optimizing the acquisition function $G$ over the surrogate function for the Gaussian process, as shown in *eq. (8)* [30].

$$y_n = argmax_y\, G(y|D_{1:n-1}) \tag{8}$$

Where $P$ is the set of data points $P = \{y_1, H(y_1), \ldots\ldots, y_n, H(y_n)\}$, and $n$ is the number of samples. Moreover, mean function and covariance functions were evaluated for "n" terms of gaussian process which is approximately equal to the probabilistic function. Consequently, mean function and covariance functions were specified in *eq. (9), (10)*,

$$y_m = e[f(y)] \tag{9}$$

$$(y, y')_k = C[f(y), f(y')] \tag{10}$$

Where, Gaussian process is defined to set the distribution function across the uncertainty of the objective function. Therefore, it validates the posterior distribution over the function, which is mentioned *eq. (11)*

$$f(y)|\{[y_i, f(y_i)]\}_{i=1}^{n} \approx \{\mu(y), \sigma^2(y)\} \tag{11}$$

Where, μ(y), $\sigma^2(y)$ is represented as mean as well as variance of the posterior distribution respectively. Also, posterior distribution is mean (μ(y)) and variance ($\sigma^2(y)$) are computed using *eq. (12), (13):*

$$\mu(y) = a(y) + Q(y,Y)^t[Q(Y,Y) + \sigma_n^2 P]^{-1}(x - a(Y)) \tag{12}$$

$$\sigma^2(y) = q(y,y) - Q(y,Y)^t[Q(Y,Y) + \sigma_n^2 P]^{-1}(Q(Y,y)) \tag{13}$$

Where, observed point of matrix is denoted as Y, observed function value of vector is defined as X, noise variance is represented as $\sigma_n^2$,, covariance matrix among the all observed pair points is denoted as Q(Y, Y) and convenience vector among and all observed pair points in Y is expressed as Q(y,Y). The expected improvement of the acquisition function is expressed in *eq. (14)* as follows:

$$G(y) = E[max(H(y) - H(y^*),0] \tag{14}$$

Where $E$ represents the expectation operator, $H(y)$ is the objective function, $H(y^*)$ indicates the best instance's objective function value and $y^*$ stands for the best instance's location in the search space. The function of objectives is used to assess the chosen samples, and this process is iterated till the lowest point of the objective functions value or the sample with the lowest objective is identified. If the observed objective exceeds a certain threshold, a stooping condition is applied.

The fusion of PSO and BO methods is a key element in the proposed model which improves the accuracy of lung cancer detection. In this proposed model, PSO is used to predict the affected area of the lung and explore the search space to find the optimum values. Then, the results of the PSO are used to train the Gaussian process (GP) for BO. This Gaussian process provides a probabilistic estimation of the optimal value of the objective function. In the proposed method, the PSO is used to train the direct acquisition function's search rather than random sampling. The acquisition function is optimized by PSO to balance the exploration and exploitation. The PSO is integrated with the acquisition function of the BO. The acquisition function $a(x)$ is given below in *eq. (15)*,

$$a(x) \Rightarrow \quad S_i(t+1) = D.S_i(t) + L_1R_1[O_i(t) - P_i(t)] + L_2R_2[G_i(t) - P_i(t)] \tag{15}$$

The fusion involves using PSO for the initial broad search of the hyperparameter space, identifying regions of interest that likely contain optimal solutions. PSO quickly navigates through the solution space, with particles evaluating and updating their positions based on the combined influence of individual and global knowledge. Once PSO identifies a promising region, BO refines the search in this localized space using its surrogate model to fine-tune the hyperparameters efficiently.

The combined model is updated with the new evaluation methods i.e. the PSO optimization and BO acquisition function. Iterate through the process of updating the combined model and optimizing the acquisition function using PSO until the convergence is reached. When PSO and BO are combined, fewer iterations are needed than using each technique separately. PSO effectively searches the space, and BO, using a

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

combined model, focuses the search. A more accurate cancer prediction model emerges from the fusion. By utilizing BO's effective search strategies and PSO's exploration capabilities, PSO with BO improves the efficacy of the cancer prediction model.

### 4.4. Cancer Prediction

The trained images are moved to the cancer prediction module. Here, the likelihood of cancer presence in the images used for training is predicted using the DenseNet-based GAN model. In the DenseNet architecture, a GAN framework is used to generate synthetic lung CT images which improve the prediction performance of the proposed model.

### 4.5. Classification

A DenseNet architecture is used for classification, to accurately distinguish generated candidate nodules from real nodules. To avoid training a model with an unbalanced dataset, the augmentation method is used to stabilize the training dataset by up-sampling the cancer nodules and down-sampling the non-cancer nodules.

### 4.6. Segmentation Process

In the segmentation stage, the augmented data of the lung images are fed as an input. This operation is performed to identify the boundaries of the images. Segmentation simplifies the image analysis process by bringing down the complexity of the image. Segmenting lung nodules is used to increase the image's resolution. The objective is to precisely locate and separate areas in lung images that might point to the existence of malignant cells or tumors. Moreover, lung field extraction is the primary segmentation stage in which algorithms separate the lung areas from the rest of the thoracic cavity. To accurately define lung boundaries, the method of thresholding is frequently used. Boundary analysis is carried out after lung field extraction to improve segmentation. This phase is pinpointing the exact shapes of the lung structures and any nodules that might exist. In this segmentation process, to identify the cancer tissues, CNN based U-Net model is applied. The histopathological and CT lung images can be segmented into multiple parts using the image segmentation process, allowing the malignant spots to be recognized and distinguished from healthy lung tissue.

## 5. RESULT AND DISCUSSION

The effectiveness of the proposed methodology is demonstrated through an examination and comparison to alternative approaches in terms of several performance measures e.g., accuracy, f-measure, precision, recall, confusion matrix, and convergence curve with the aid of hybrid optimization (Particle Swarm and Bayesian Optimization). The experiments were performed using Python 3.7 programming language installed in the personal computer (PC) having 16 GB RAM. The tests were kept running on an Intel(R) Core (TM) i7-10750H CPU@2.60GHz processor with the 64-bit operating system, x64-based processor.

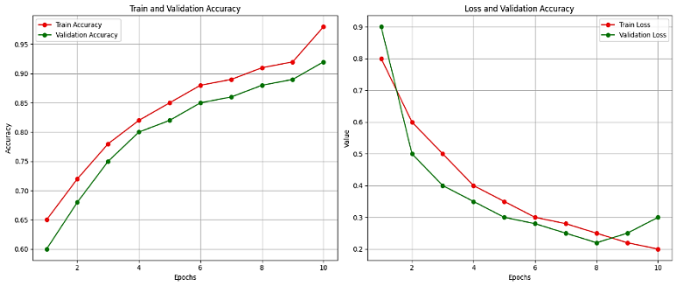
### 5.1. Dataset Description

The dataset "LC25000" is composed of 25,000 histopathological images in JPEG format, all sized at 768x768 pixels. These images were sourced from HIPAA-compliant and validated channels, initially comprising 750 lung tissue images (250 benign, 250 adenocarcinomas, and 250 squamous cell carcinomas), along with 500 colon tissue pictures (250 benign and 250 adenocarcinomas) [42]. Utilizing the Augmentor package for augmentation, the dataset was expanded to a total of 25,000 images. It incorporates three distinct categories, each comprising 5,000 images: Lung adenocarcinoma, Lung benign tissue, and Lung squamous cell carcinoma. The training set constitutes 70% of the data, the testing set comprises 15%, and the validation set accounts for 15%. For deep learning models to be properly trained, we aimed for a large and well-balanced dataset. Here, 25000 histopathology images are a significant number, but we made sure that the data diversity and quality are preserved.

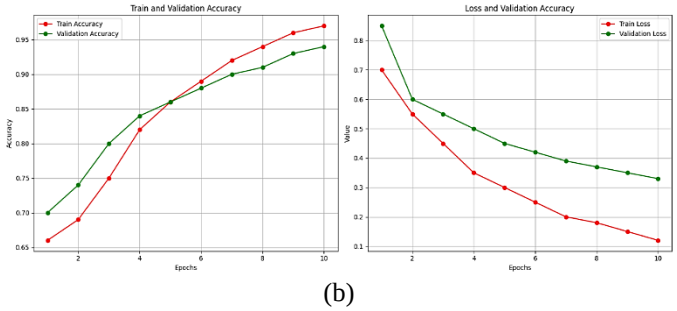
The lung CT scan image dataset (TCGA-LUAD) [40] contains images in JPG or PNG formats rather than DCM format to align with the model requirements [41]. It encompasses three types of lung cancer - Adenocarcinoma, Large Cell Carcinoma, and Squamous Cell Carcinoma, alongside images of normal cells. The training set constitutes 70% of the data, the testing set comprises 15%, and the validation set accounts for 15%.

### 5.2. Train, Validation Accuracy, and Validation Loss of Proposed Model

The accuracy of lung cancer detection is heavily dependent on values such as accuracy and training loss. Figure 3 displays the training validation accuracy and loss values for both the Histopathological and lung CT image datasets.




(a)



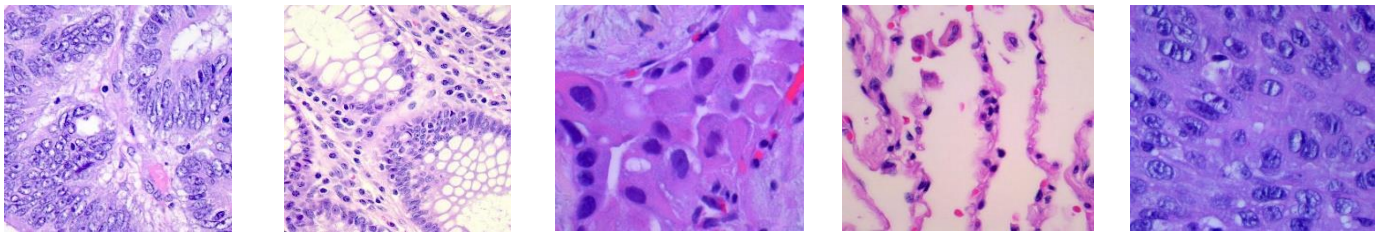

(b)

**Figure 3.** Training performance for accuracy and loss values of (a) Histopathological dataset and (b) lung CT image dataset

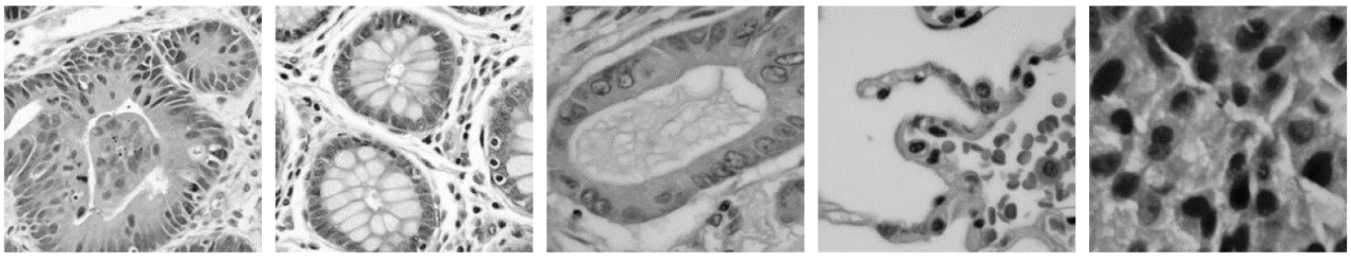
To validate the developed model efficiency, a collage of Histological images at the correspondent outcomes is illustrated in *figure 4*, and a collage of CT images at the correspondent outcomes is shown in *figure 5*. Both of the figures represent the

International Journal of
Electrical and Electronics Research (IJEER)
Open Access | Rapid and quality publishing
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

four stages through which the image samples pass. The first stage is the Input stage, where an image is given as input. The second stage is a grayscale image. The third stage is a morphological image, and the fourth stage shows the segmented output.

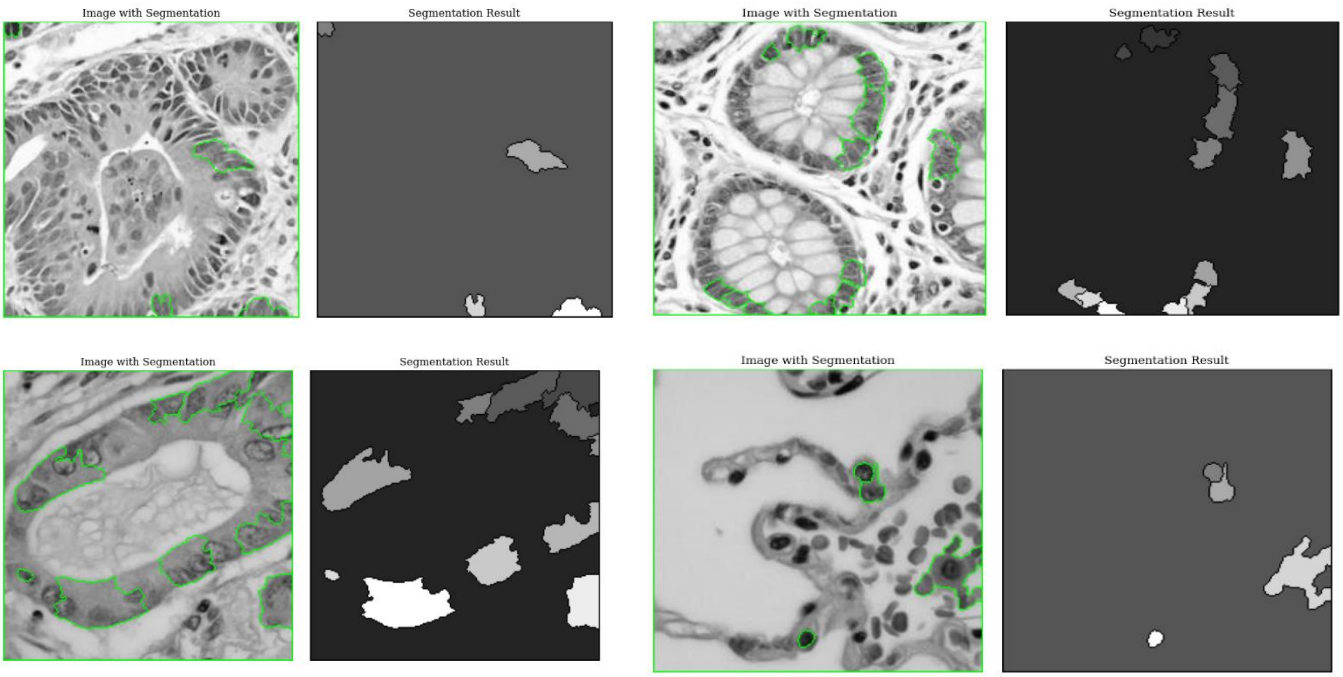


a. Histological images (Input images)

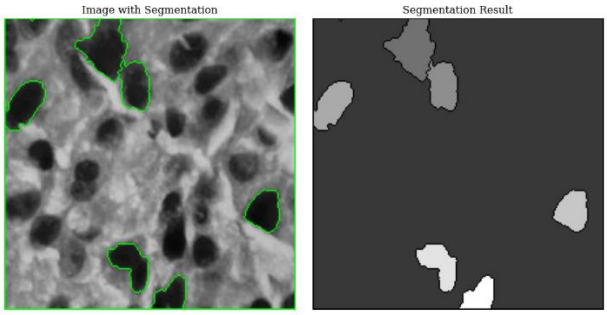


b. Grayscale images


c. Morphology images

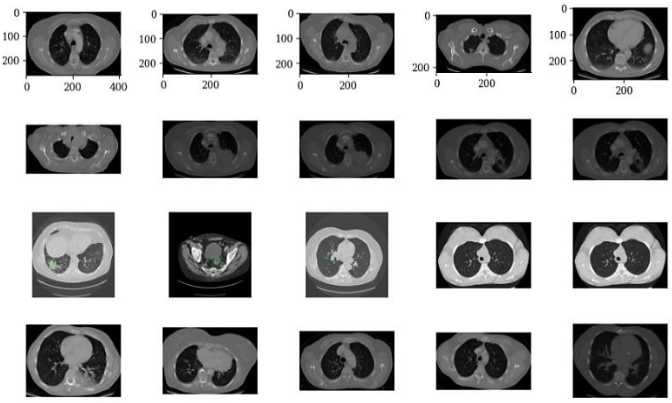


FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

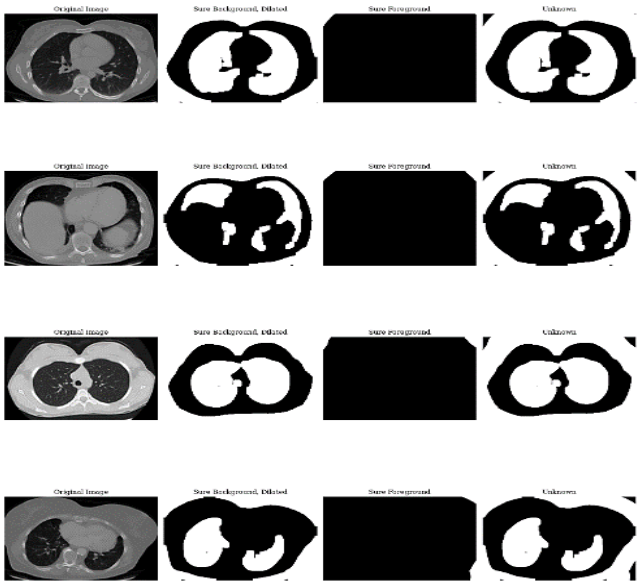



d. Segmented images

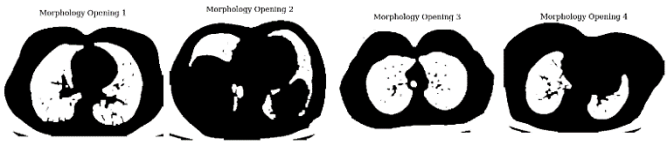
**Figure 4.** Simulation results for the histological dataset images a. histological images, b. Grayscale images, c. morphology image, d. segmented images



a. CT images (input image)



b. Grayscale image




c. Morphology image


d. Segmented images

**Figure 5**. Simulation results for CT scan dataset images a. CT images, b. Grayscale images, c. morphology image, d. segmented images

## 5.3. Performance Matrixes

For the validation of our proposed particle Swarm-based Bayesian Optimization (PSbBO) model, we assess its effectiveness by measuring parameters including accuracy, recall, F-measure, and precision, confusion matrix, convergence curve, computation time, and complexity.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

#### 5.3.1. Accuracy ($A_C''$)

Accuracy represents the overall percentage of accurate predictions, encompassing both positive and negative outcomes. It is the proportion of correct predictions with all the predictions which in mentioned in *eq. (11)*

$$A_C'' = \frac{H_{tp}'' + H_{tn}''}{H_{tp}'' + H_{tn}'' + H_{fp}'' + H_{fn}''} \quad (11)$$

Where $A_C''$ indicates the accuracy, $H_{tp}''$ is denoted as real positive, $H_{tn}''$ is real negative, $H_{fp}''$ indicates the untrue positive, *and* $H_{fn}''$is untrue negative.

#### 5.3.2. Precision ($P_C''$)

The average of the relevant samples that were retrieved is called precision, which is also known as positive forecasting accuracy. It can be employed to evaluate the precision of the model's optimistic predictions by dividing the entire count of true positives by all positive predictions. This is computed using the *eq. (12)* given below.

$$P_C'' = \frac{H_{tp}''}{H_{tp}'' + H_{fp}''} \quad (12)$$

#### 5.3.3. Recall ($R_C''$)

Recall measures the fraction of the model correctly recognized positives out of all potential positives, calculated by dividing actual positives by the entire count of actual positives. The *eq. (13)* below provides a formula for the recall.

$$R_C'' = \frac{H_{tp}''}{H_{tp}'' + H_{fn}''} \quad (13)$$

#### 5.3.4. F-Measure ($F_1''$)

The F-Measure represents the harmonic average of recall and precision, effectively incorporating untrue negatives as well as untrue positives. Its calculation is based on the following *eq.(14).*

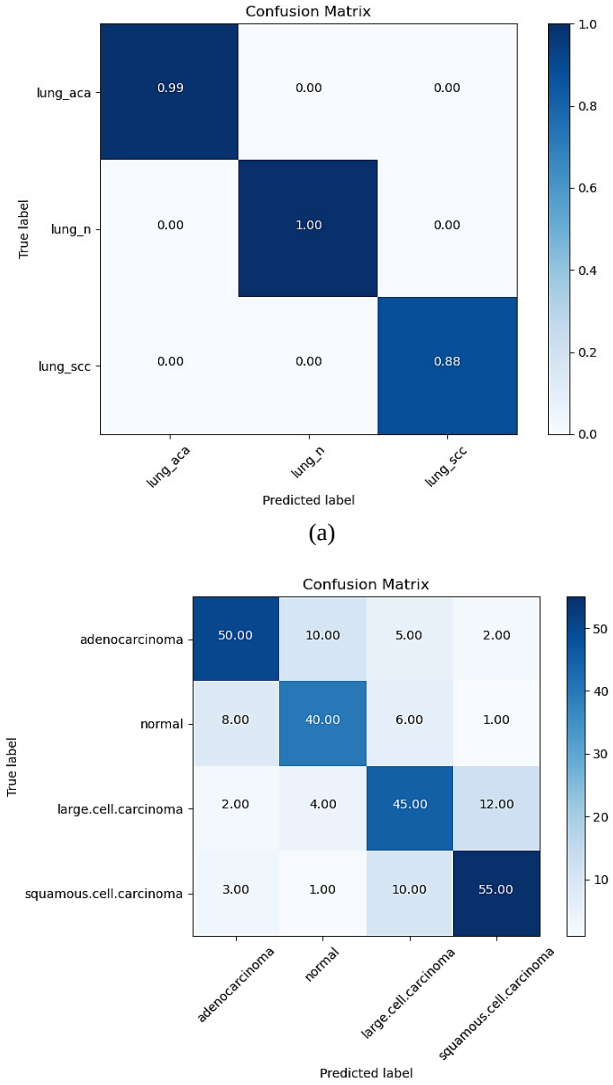
$$F_1'' = \frac{2(P_C'' * R_C'')}{P_C'' + R_C''} \quad (14)$$

#### 5.3.5. Error Rate

Error rate is the proportion of incorrect outcomes compared to total attempts or observations. It is used to assess accuracy and performance in various fields. Lower error rates indicate higher precision and effectiveness in the respective domain. Our model obtained very low error rates of 1.19 and 1.21 for both the Histopathological and lung CT image datasets respectively.

#### 5.3.6. Confusion Matrix

A crucial tool for evaluating the model's ability to classify carcinoma of the lungs is the confusion matrix. This matrix offers a structured breakdown of the model's predictions, classifying them into four essential outcomes: true positives (accurate predictions of cancer cases), false positives (erroneous predictions of cancer cases), true negatives (accurate predictions of non-cancer cases), and false negatives (erroneous predictions of non-cancer cases).




(a)

(b)

**Figure 6.** Confusion Matrix of proposed PSbBO (a) Histopathological dataset(b) lung CT image dataset

The confusion matrices corresponding to the Histopathological and CT image datasets are given in *figure 6*. In the histopathological dataset, adenocarcinoma labels are classified with an accuracy of 99% while 88% of the squamous cell carcinoma labels are correctly classified. Similarly, for the CT image dataset, 98% of squamous cell carcinoma labels are correctly classified using the proposed model.

### 5.4. Comparative Analysis

Combining the benefits of both optimization techniques BO and PSO can increase the accuracy of cancer predictions by optimizing predictive models. Additionally, by fusing BO's capacity to manage uncertainty with the PSO's ability to handle noisy data, the fusion of PSO and BO improves resilience and provides more accurate cancer forecasts. The overall result of this combination is an improvement in the accuracy of cancer prediction through improved feature selection, hyperparameter tuning, model training, and validation processes. A more dependable and efficient optimization strategy tailored to the complexities of cancer prediction can be developed by

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

integrating these techniques. Comparing the efficacy of the suggested strategy to alternative approaches, like as Genetic Algorithm with Wrapper Approach (GAWA), Particle Swarm Optimization-based Multiobjective Selection (PSOMS), Ant Colony Optimization (ACO), and Hybrid intelligent Spiral Optimization-based Generalized Rough set approach (HSOGR) [31], Whale Optimization (WO), Artificial Bee Colony (ABC), BAT optimization (BAT) algorithm, Particle Swarm Optimization (PSO) [32], Moth Flame optimization (MFO) strategy [33], grey wolf optimization (GWO) algorithm [34], Ant Lion Optimization (ALO) [35] , and Bayesan Optimization (BO) [36]. This comparison is conducted to determine system accuracy, recall, F1-score, and precision.

**5.4.1 Comparing the Proposed Method with Existing Approaches, Focusing on the Performance Metrics**

The outcomes of the proposed model are compared with those of alternative methods to assess their respective performance levels. Our proposed method PSbBO has a score of 0.995 and 0.988 accuracy higher than all other methods for both the Histopathological and lung CT image datasets respectively as given in *table 1 and 2*.

**Table 1. Comparison between proposed models with existing optimization methods-based models using F1-Score, Recall, Precision, and Accuracy for Histopathological images**

| Methods | F-Measure | Recall | Precision | Accuracy |
|---|---|---|---|---|
| GAWA | 0.828 | 0.879 | 0.893 | 0.818 |
| WO | 0.972 | 0.971 | 0.969 | 0.971 |
| ABC | 0.969 | 0.969 | 0.967 | 0.968 |
| PSO | 0.959 | 0.956 | 0.954 | 0.959 |
| ACO | 0.76 | 0.756 | 0.793 | 0.786 |
| BAT | 0.968 | 0.965 | 0.963 | 0.966 |
| HSOGR | 0.658 | 0.693 | 0.662 | 0.681 |
| PSOMS | 0.748 | 0.73 | 0.725 | 0.738 |
| MFO | 0.89 | 0.91 | 0.911 | 0.90 |
| GWO | 0.93 | 0.934 | 0.94 | 0.93 |
| ALO | 0.92 | 0.92 | 0.925 | 0.92 |
| BO | 0.916 | 0.913 | 0.92 | 0.916 |
| Proposed PSbBO | **0.994** | **0.992** | **0.983** | **0.995** |

*Table 1* presents the performance comparison of various optimization methods used for lung cancer detection, evaluated through key metrics such as F-measure, Recall, Precision, and Accuracy. Among the tested methods, PSbBO (Particle Swarm-based Bayesian Optimization) is found to get better results across all metrics, achieving an F-measure of 0.994, Recall of 0.992, Precision of 0.983, and Accuracy of 0.995. This suggests that PSbBO is highly effective in identifying lung cancer-related features with minimal errors, with compared to all other optimization techniques.

WO (Whale Optimization) and ABC (Artificial Bee Colony) also performed strongly, with WO slightly surpassing ABC in F-measure (0.972 vs. 0.969), Recall (0.971 vs. 0.969), and Accuracy (0.971 vs. 0.968). This indicates that WO provides excellent balance in feature detection and classification accuracy, making it another competitive approach. BAT, GWO (Grey Wolf Optimizer), and ALO (Ant Lion Optimizer) follow closely, with all three showing consistently high performance, demonstrating their effectiveness in the same task. BAT achieved an F-measure of 0.968, while GWO and ALO scored 0.93 and 0.92 respectively, reinforcing their utility in this context.

On the other hand, GAWA (Genetic Algorithm with Weighted Averaging), despite moderate performance, has relatively lower Accuracy (0.818) compared to the top-performing methods. ACO (Ant Colony Optimization), HSOGR (Harmony Search with Opposition-based Gravitational Rule), and PSOMS (Particle Swarm Optimization with Multi-Strategy) demonstrate weaker results, particularly HSOGR, with the lowest performance across all metrics (F-measure of 0.658 and Accuracy of 0.681). This indicates that these methods may struggle with the complexity of the data, likely due to suboptimal parameter tuning or feature extraction.
Overall, the results clearly demonstrate the effectiveness of hybrid and nature-inspired optimization techniques for lung cancer detection, with PSbBO leading by a significant margin. While methods such as WO and ABC are also competitive, techniques like HSOGR and PSOMS show room for improvement.

**Table 2. Comparison between proposed models with existing optimization methods-based models using F1-Score, Recall, Precision, and Accuracy for Lung CT images**

| Edvrv Methods | F-Measure | Recall | Precision | Accuracy |
|---|---|---|---|---|
| GAWA | 0.828 | 0.879 | 0.893 | 0.818 |
| WO | 0.972 | 0.971 | 0.969 | 0.971 |
| ABC | 0.969 | 0.969 | 0.967 | 0.968 |
| PSO | 0.959 | 0.956 | 0.954 | 0.959 |
| ACO | 0.76 | 0.756 | 0.793 | 0.786 |
| BAT | 0.968 | 0.965 | 0.963 | 0.966 |
| HSOGR | 0.658 | 0.693 | 0.662 | 0.681 |
| PSOMS | 0.748 | 0.73 | 0.725 | 0.738 |
| MFO | 0.89 | 0.91 | 0.911 | 0.90 |
| GWO | 0.93 | 0.934 | 0.94 | 0.93 |
| ALO | 0.92 | 0.92 | 0.925 | 0.92 |
| BO | 0.919 | 0.927 | 0.912 | 0.926 |
| Proposed PSbBO | **0.986** | **0.983** | **0.974** | **0.988** |

*Table 2* shows the comparative performance of various optimization methods in detecting lung cancer, evaluated using metrics like F-measure, Recall, Precision, and Accuracy. A particularly high-performing method is PSbBO (Particle Swarm-based Bayesian Optimization), applied to lung CT images. It achieves high scores across all metrics, with an F-measure of 0.986, Recall of 0.983, Precision of 0.974, and Accuracy of 0.988. This indicates that the PSbBO method is highly efficient at detecting relevant features for lung cancer, boasting a nearly perfect balance between sensitivity (Recall) and specificity (Precision). Its Accuracy indicates that the model's ability to correctly classify a high proportion of lung cancer cases approves its better performance among the methods.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

Similarly, WO (Whale Optimization) and ABC (Artificial Bee Colony) also demonstrate strong performance, with WO showing an F-measure of 0.972 and Accuracy of 0.971, and ABC following closely with an F-measure of 0.969 and Accuracy of 0.968. Both methods show balanced performance across all metrics which makes them highly reliable in detecting lung cancer features with minimal misclassifications.

BAT, GWO (Grey Wolf Optimizer), and ALO (Ant Lion Optimizer) exhibit commendable results as well. For instance, BAT achieves an F-measure of 0.968 and Accuracy of 0.966, while GWO and ALO deliver slightly lower yet competitive values, suggesting they are robust methods for classification tasks, though not quite on par with PSbBO, WO, or ABC.

In contrast, GAWA (Genetic Algorithm with Weighted Averaging), despite offering moderate performance, lags behind with an Accuracy of 0.818, indicating weaker generalization ability. Likewise, ACO (Ant Colony Optimization) and PSOMS (Particle Swarm Optimization with Multi-Strategy) show suboptimal results, particularly in Precision and Accuracy, which could imply that these methods struggle with overfitting or feature extraction challenges. Notably, HSOGR (Harmony Search with Opposition-based Gravitational Rule) performs the worst across all metrics, with an F-measure of 0.658 and Accuracy of 0.681, underscoring its limitations in handling complex datasets like lung CT images. MFO (Moth Flame Optimizer), although not among the top performers, yields reasonably strong results with an F-measure of 0.89 and Accuracy of 0.90, placing it in the mid-tier range of optimization methods.

In summary, the *table 2* indicate that PSbBO, especially when applied to lung CT images, significantly outperforms the other methods in terms of both precision and overall accuracy. Meanwhile, methods like WO, ABC, and BAT are also competitive, though not quite at the same level. On contrary, optimization techniques like HSOGR, PSOMS, and GAWA exhibit weaker performance and may require further enhancements in parameter tuning and feature extraction to improve their efficacy in lung cancer detection tasks.

### 5.4.2. Comparing the Proposed Method with Existing Approaches, Focusing on Error Rate

*Table 3* presents a comparative analysis of error rates for various models across histopathological and CT image datasets, with the DenseNet with PSbBO (Particle Swarm-based Bayesian Optimization) emerging as the top performer. The proposed method achieves the lowest error rates, with 1.19% for the histopathological dataset and 1.21% for the CT image dataset. This highlights the effectiveness of the DenseNet architecture combined with hybrid optimization techniques in accurately identifying features relevant to lung cancer, making it superior in reducing misclassification compared to other models.

When comparing these results to traditional models like LeNet and AlexNet, significant differences can be observed. For instance, LeNet (RMSprop) yields relatively low error rates of 2.224% and 2.145% for histopathological and CT image datasets, respectively, while LeNet (ADAM) shows higher error rates at 2.731% and 2.626%. Similarly, AlexNet (SGD) and AlexNet (SGD-Drop) exhibit improved performance with the dropout technique, where the error rate drops from 3.842% and 3.5% (without dropout) to 1.725% and 1.65% (with dropout), respectively, indicating the importance of regularization techniques like dropout in improving model performance.

Among the deep learning models, VGG-16 also demonstrates competitive results, with an error rate of 1.412% for histopathological data and 1.52% for CT images, performing close to the proposed DenseNet with PSbBO but still slightly higher in both datasets.

In contrast, traditional machine learning models like Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN) show higher error rates. For example, SVM and RF have error rates of 3.2% and 2.119% for the histopathological dataset, respectively, with KNN slightly outperforming both with a 1.8% error rate. However, for CT images, SVM (3.171%) and RF (2.72%) show reduced accuracy, whereas KNN reflects competitive performance with a 1.9% error rate.

Methods such as BoVW (Bag of Visual Words) and CRNN (Convolutional Recurrent Neural Network) also perform moderately, with BoVW having one of the highest error rates (3.791% and 3.69%) across both datasets. Meanwhile, CRNN fares better with error rates of 1.820% and 1.85%, yet still falling behind the top-performing deep learning models.

In summary, the *table 3* clearly indicate that the DenseNet with PSbBO method outperforms all other techniques in terms of minimizing error rates for both histopathological and CT image datasets. The inclusion of hybrid optimization strategies contributes to the fine-tuned performance. While deep learning models like VGG-16 and AlexNet (SGD-Drop) also show competitive results, traditional machine learning approaches, such as SVM, RF, and KNN, lag behind, in handling the complexity of the data.

**Table 3. Comparison values of error rate of the proposed model with the existing classification models**

| Methods | Error Rate for Histopathological Dataset | Error Rate for CT image dataset |
|---|---|---|
| LeNet (RMSprop) [34] | 2.224 | 2.145 |
| LeNet (ADAM) [34] | 2.731 | 2.626 |
| AlexNet (SGD) [34] | 3.842 | 3.5 |
| AlexNet (SGD-Drop) [34] | 1.725 | 1.65 |
| VGG-16 [34] | 1.412 | 1.52 |
| BoVW [35] | 3.791 | 3.69 |
| CRNN [35] | 1.820 | 1.85 |
| CNN [35] | 2.50 | 2.73 |
| SVM [45] | 3.2 | 3.171 |
| RF [46] | 2.119 | 2.72 |
| KNN [46] | 1.8 | 1.9 |
| Proposed DenseNet with PSbBO | **1.19** | **1.21** |

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

## 6. CONCLUSION

This study introduces an innovative approach that merges PSO and BO methodologies to fine-tune DL parameters to predict cancer in the lung. By leveraging both techniques, we observed significant enhancements in model accuracy and reliability. The integration allowed efficient exploration and exploitation of hyperparameters, leading to optimal configurations. Performance improvement stemmed from fine-tuned parameters, highlighting their importance in lung cancer detection. The study's outcomes hold crucial implications for early lung cancer identification, potentially aiding radiologists in making swift, accurate decisions to improve clinical outcomes and reduce mortality. The model demonstrated high performance with an accuracy of 99.5%, precision of 98.3%, recall of 99.2%, F1-score of 99.4%, and an error rate of 1.19% for histopathological images. Similarly, for lung CT images, it achieved an accuracy of 98.8%, precision of 97.4%, recall of 98.3%, F1-score of 98.6%, and an error rate of 1.21%. The limitations of the proposed hybrid framework are its high computational complexity which limits the applicability for very large datasets but it shows a good improvement in the accuracy on the average from 3.3% and 7.6% respectively over the corresponding constituent methods PSO and BO used individually. Moreover, there is a risk of overfitting during hyperparameter tuning, which needs to be addressed carefully to achieve the generalization.

This work could be extended for diverse lung image datasets to further improve the prediction accuracy. Moreover, nature-inspired algorithms would also be tested against different sets of features to optimize the accuracy of the lung cancer prediction. To overcome the high computational complexity, a Knowledge-Distillation based light weight models parametric hypercomplex convolution approach could be developed.

**Availability of data and materials:** All the datasets used during this study are available online in opensource.

**Conflict of interest:** The authors declare no competing interests.

## REFERENCES

[1] Sharma, P., Mehta, M., Dhanjal, D. S., Kaur, S., Gupta, G., Singh, H, & Satija, S. (2019). Emerging trends in the novel drug delivery approaches for the treatment of lung cancer. Chemico-biological interactions, 309, p.108720.

[2] Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. Cancer Letters, 471, pp.61-71.

[3] Dhiman, P., Ma, J., Navarro, C. A., Speich, B., Bullock, G., Damen, J. A.,& Collins, G. S. (2021). Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. Journal of Clinical Epidemiology, 138, pp.60-72.

[4] Ibrahim, D. M., Elshennawy, N. M., & Sarhan, A. M. (2021). Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. Computers in biology and medicine, 132, p.104348.

[5] Toh, T. S., Dondelinger, F., & Wang, D. (2019). Looking beyond the hype: applied AI and machine learning in translational medicine. EBioMedicine, 47, pp.607-615.

[6] Vaidya, P., Bera, K., Gupta, A., Wang, X., Corredor, G., Fu, P.,& Madabhushi, A. (2020). CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, II resectable non-small cell lung cancer: a retrospective multicohort study for outcome prediction. The Lancet Digital Health, 2(3), pp.116-128.

[7] Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems, 92, pp.374-382.

[8] Zade, A. E., Haghighi, S. S., & Soltani, M. (2022). Deep neural networks for neuro-oncology: Towards patient individualized design of chemo-radiation therapy for Glioblastoma patients. Journal of Biomedical Informatics, 127, p.104006.

[9] Ilhan, B., Guneri, P., & Wilder-Smith, P. (2021). The contribution of artificial intelligence to reducing the diagnostic delay in oral cancer. Oral oncology, 116, p.105254.

[10] Xie, Y., Meng, W. Y., Li, R. Z., Wang, Y. W., Qian, X., Chan, C., & Leung, E. L. H. (2021). Early lung cancer diagnostic biomarker discovery by machine learning methods. Translational oncology, 14(1), p.100907.

[11] Seijo, L. M., Peled, N., Ajona, D., Boeri, M., Field, J. K., Sozzi, G., & Montuenga, L. M. (2019). Biomarkers in lung cancer screening: achievements, promises, and challenges. Journal of Thoracic Oncology, 14(3), pp.343-357.

[12] Dunn, B. K., Woloshin, S., Kramer, B. S., & Xie, H. (2022). Cancer overdiagnosis: a challenge in the era of screening. Journal of the National Cancer Center.

[13] Curigliano, G., Lenihan, D., Fradley, M., Ganatra, S., Barac, A., Blaes, A., & ESMO Guidelines Committee. (2020). Management of cardiac disease in cancer patients throughout oncological treatment: ESMO consensus recommendations. Annals of Oncology, 31(2), pp.171-190.

[14] Dalamaga, M., Christodoulatos, G. S., Karampela, I., Vallianou, N., & Apovian, C. M. (2021). Understanding the co-epidemic of obesity and COVID-19: current evidence, comparison with previous epidemics, mechanisms, and preventive and therapeutic perspectives. Current obesity reports, 10(3), pp.214-243.

[15] Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. Artificial Intelligence in Healthcare, pp. 25-60.

[16] Bode, G., Thul, S., Baranski, M., & Müller, D. (2020). Real-world application of machine-learning-based fault detection trained with experimental data. Energy, 198, p.117323.

[17] Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. Z., Paul, R., Smriti, K., & Somani, B. K. (2022). The role of data science in healthcare advancements: applications, benefits, and prospects. Irish Journal of Medical Science (1971- ), 191(4), pp.1473-1483.

[18] Jiang, H., Shen, F., Gao, F., & Han, W. (2021). Learning efficient, explainable, and discriminative representations for pulmonary nodules classification. Pattern Recognition, 113, p.107825.

[19] Kumar, R., Wang, W., Kumar, J., Yang, T., Khan, A., Ali, W., & Ali, I. (2021). An integration of blockchain and AI for secure data sharing and detection of CT images for hospitals. Computerized Medical Imaging and Graphics, 87, p.101812.

[20] Zheng, S., Shen, Z., Pei, C., Ding, W., Lin, H., Zheng, J.,& Huang, L. (2021). Interpretative computer-aided lung cancer diagnosis: From radiology analysis to malignancy evaluation. Computer Methods and Programs in Biomedicine, 210, p.106363.

[21] Nanglia, P., Kumar, S., Mahajan, A. N., Singh, P., & Rathee, D. (2021). A hybrid algorithm for lung cancer classification using SVM and Neural Networks. ICT Express, 7(3), pp.335-341.

International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 3 | Pages 1074-1086 | e-ISSN: 2347-470X

[22] Ananthajothi, K., Rajasekar, P., & Amanullah, M. (2023). Enhanced U-Net-based segmentation and heuristically improved deep neural network for pulmonary emphysema diagnosis. Sādhanā, 48(1), p.33.

[23] Ali, I., Muzammil, M., Haq, I. U., Khaliq, A. A., & Abdullah, S. (2020). Efficient lung nodule classification using transferable texture convolutional neural network. IEEE Access, 8, pp.175859-175870.

[24] Masood, A., Sheng, B., Yang, P., Li, P., Li, H., Kim, J., & Feng, D. D. (2020). Automated decision support system for lung cancer detection and classification via enhanced RFCN with multilayer fusion RPN. IEEE Transactions on Industrial Informatics, 16(12), pp.7791-7801.

[25] Shakeel, P.M., Burhanuddin, M.A. and Desa, M.I., (2019). Lung cancer detection from CT images using improved profuse clustering and deep learning instantaneously trained neural networks. Measurement, 145, pp.702-712.

[26] ALzubi, J.A., Bharathikannan, B., Tanwar, S., Manikandan, R., Khanna, A. and Thaventhiran, C., (2019). Boosted neural network ensemble classification for lung cancer disease diagnosis. Applied Soft Computing, 80, pp.579-591.

[27] Maleki, N., Zeinali, Y. and Niaki, S.T.A., (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Systems with Applications, 164, p.113981.

[28] Tian, Q., Wu, Y., Ren, X. and Razmjooy, N., (2021). A new optimized sequential method for lung tumor diagnosis based on deep learning and converged search and rescue algorithm. Biomedical Signal Processing and Control, 68, p.102761.

[29] Ding, L., Zhang, X.Y., Wu, D.Y. and Liu, M.L., (2021). Application of an extreme learning machine network with particle swarm optimization in syndrome classification of primary liver cancer. Journal of Integrative Medicine, 19(5), pp.395-407.

[30] Atteia, G., Alhussan, A.A. and Samee, N.A., (2022). Bo-Tallinn: Bayesian-based optimized CNN for acute lymphoblastic leukemia detection in microscopic blood smear images. Sensors, 22(15), p.5520.

[31] Shakeel, P.M., Burhanuddin, M.A. and Desa, M.I., (2022). Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. Neural Computing and Applications, 34, pp.9579-9592.

[32] Ghoniem, R.M., Algarni, A.D., Refky, B. and Ewees, A.A., (2021). Multi-modal evolutionary deep learning model for ovarian cancer diagnosis. Symmetry, 13(4), p.643.

[33] Sebastian, A.E., Dua, D., (2023). Lung Nodule Detection via Optimized Convolutional Neural Network: Impact of Improved Moth Flame Algorithm. Sens Imaging, 24, p.11.

[34] Bilal, Anas, et al. (2022). Lung nodules detection using grey wolf optimization by weighted filters and classification using CNN. Journal of the Chinese Institute of Engineers, 45(2), pp.175-186.

[35] Vidhya, R., and T. T. Mirnalinee (2022). Hybrid Optimized Learning for Lung Cancer Classification. Intelligent Automation & Soft Computing 34(2).

[36] Mkindu, Hassan, Longwen Wu, and Yaqin Zhao (2023). Lung nodule detection in chest CT images based on vision transformer network with Bayesian optimization. Biomedical Signal Processing and Control 85, p.104866.

[37] Toğaçar, M., Ergen, B. and Cömert, Z., (2020). Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks. Biocybernetics and Biomedical Engineering, 40(1), pp.23-39.

[38] SU, A., PP, F.R., Abraham, A. and Stephen, D., (2022). Deep Learning-Based BoVW–CRNN Model for Lung Tumor Detection in Nano-Segmented CT Images. Electronics, 12(1), p.14.

[39] Guan, X., Du, Y., Ma, R., Teng, N., Ou, S., Zhao, H. and Li, X., (2023). Construction of the XGBoost model for early lung cancer prediction based on metabolic indices. BMC Medical Informatics and Decision Making, 23(1), pp.1-16.

[40] Albertina, B., Watson, M., Holback, C., Jarosz, R., Kirk, S., Lee, Y., Lemmerman, J. (2016). Radiology Data from The Cancer Genome Atlas Lung Adenocarcinoma [TCGA-LUAD] collection. The Cancer Imaging Archive available at URL: http://doi.org/10.7937/K9/TCIA.2016.JGNIHEP5

[41] Dansana, J., Kabat, M.R. and Pattnaik, P.K., (2023). A Novel Optimized Perturbation-Based Machine Learning for Preserving Privacy in Medical Data. Wireless Personal Communications, 130(3), pp.1905-1927.

[42] Borkowski AA, Bui MM, Thomas LB, Wilson CP, DeLand LA, Mastorides SM. Lung and colon cancer histopathological image dataset (LC25000). The LC25000 dataset is available at URL: https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images

[43] Hany M. Chest CT-scan images dataset. Kaggle. Retrieved November. 2020; 13:2022.

[44] Akhand ZE, Rahman AI, Sarda A, Fahim MZ, Tushi LT, Azad K, Tahiat HT (2023). Lung Cancer Detection Using Ensemble Technique of CNN. In Proceedings of International Conference on Information and Communication Technology for Development: ICICTD, pp. 497-507.

[45] Ashraf, Syed Faaz, Yin K, X. Meng Cindy, Qang Qi, Wang Qiong, Pu Jiantao, Dhupar Rajeev, (2022). Predicting benign, preinvasive, and invasive lung nodules on computed tomography scans using machine learning. The Journal of Thoracic and Cardiovascular Surgery 163.4, pp.1496-1505.

[46] El-Askary, Nada S., Mohammed A-M. Salem, and Mohamed I. Roushdy (2022). "Features processing for random forest optimization in lung nodule localization." Expert Systems with Applications 193, p.116489.