JOURNAL OF APPLIED BIOANALYSIS, December 2025, p. 197-210
http://doi.org/10.53555/jab.v11si13.1807 (ISSN 2405-710X)
Vol. 11, No. 13s

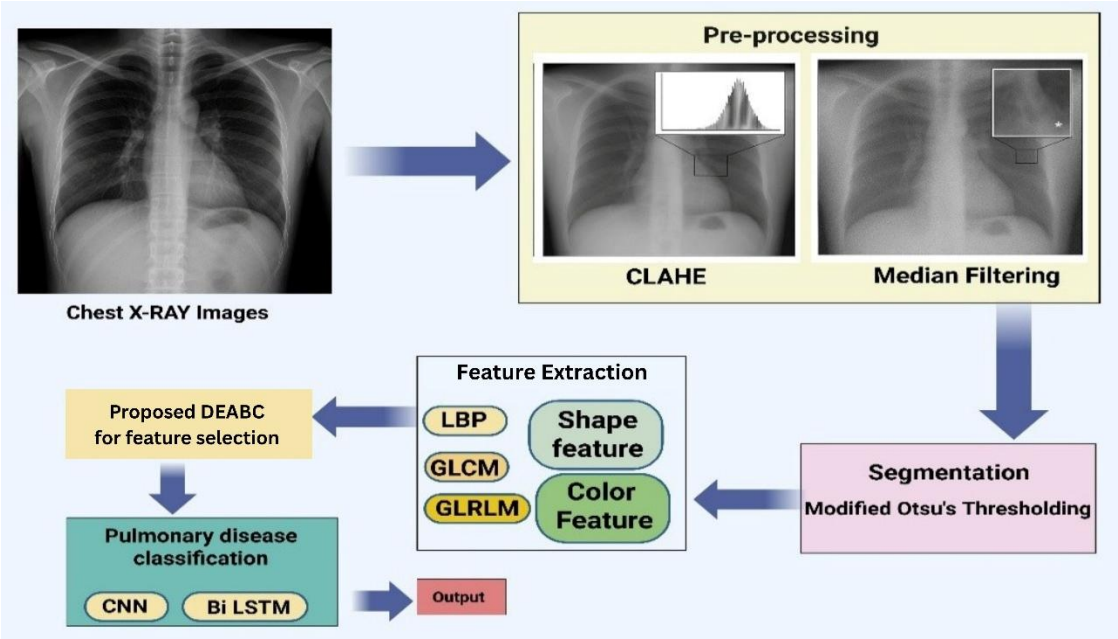
OPEN ACCESS

Pulmonary Disease Detection from Chest x-ray Images using Dragonfly Enhanced Artificial bee Colony Hybrid Optimisation Technique

[1]Jyothilakshmi KN, [2]Dr. R Parameswari,

[1]Research Scholar, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India*
[2]Professor & Head, School of Computing Science, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai. India

Corresponding author: Jyothilakshmi KN- jyothilakshmikn782@gmail.com



Abstract

Automated analysis of chest radiographs using deep learning has shown promise for detecting pulmonary abnormalities. Aim of the current study is to develop and retrospectively evaluate a hybrid diagnostic pipeline that fuses hand-crafted and deep features, using a Dragonfly-Enhanced Artificial Bee Colony (DEABC) metaheuristic for feature selection, and classifies images with a CNN→Bi-LSTM architecture, and to compare its performance.

We used a publicly available chest X-ray collection from Kaggle containing normal and abnormal cases (pneumonia, tuberculosis, lung opacities). Images were preprocessed and lung regions segmented; texture and shape features were extracted and combined with deep CNN features. Feature selection was performed with a DEABC wrapper; classification used a CNN followed by a Bi-LSTM. We evaluated five methods (ABC, DOA, CNN, Bi-LSTM, DEABC-CNN-BiLSTM) under two hyperparameter configurations (labelled LR=0 and LR=1) using patient-level train/validation/test splits. Primary outcomes on the held-out test set included accuracy, precision (PPV), sensitivity (recall), specificity, F1 and Matthews correlation coefficient (MCC). Reported values are point estimates computed on the retrospective test split.

Across both hyperparameter configurations the DEABC-optimized CNN–BiLSTM achieved the highest point estimates on all primary metrics. For LR=0 (train ≈80%) DEABC attained accuracy 0.9552, precision 0.99597, sensitivity 0.9597, specificity 0.9805, F1 0.9703, and MCC 0.9252. For LR=1 (train ≈70%) DEABC attained accuracy 0.9504, precision

0.99099, sensitivity 0.9549, specificity 0.9756, F1 0.9655, and MCC 0.9205. Comparator models (ABC, DOA, plain CNN, Bi-LSTM) showed lower point estimates across metrics.
The proposed DEABC-enhanced hybrid pipeline produced superior retrospective discrimination on the available chest X-ray dataset relative to the evaluated baselines. These results support the potential value of combining hand-crafted descriptors, deep representations and wrapper-based feature selection—particularly when data are limited.

Keywords: chest radiograph, pulmonary disease detection, hybrid features, feature selection, Dragonfly, Artificial Bee Colony, CNN, Bi-LSTM, medical image analysis, retrospective study.

Introduction

Pulmonary diseases, encompassing conditions such as chronic obstructive pulmonary disease (COPD), pneumonia, tuberculosis (TB), asthma, and more recently COVID-19, remain among the most prevalent causes of morbidity and mortality worldwide. According to the World Health Organization, over 3 million people die annually from COPD, and infectious respiratory diseases are a leading cause of both chronic disability and premature death, particularly in low- and middle-income countries (World Health Organization, 2022). The global increase in air pollution, changing lifestyle factors, and exposure to industrial and indoor environmental hazards have contributed significantly to the growing burden of lung diseases (Farhat et al., 2020).

Traditional diagnosis of pulmonary diseases relies heavily on imaging modalities such as chest X-rays (CXR) and computed tomography (CT), which provide clinicians with crucial information. Nevertheless, the accuracy and reliability of manual interpretations are subject to inter- and intra-observer variability, workload pressures, and shortages of radiological expertise especially in resource-constrained settings (Ibrahim et al., 2021). The resultant risks include delayed detection, misdiagnosis, and increased financial and social costs associated with unnecessary treatments or missed interventions.

Over recent years, advances in artificial intelligence (AI), particularly deep learning, have transformed medical imaging analysis. Deep learning models, especially convolutional neural networks (CNNs), have demonstrated remarkable performance in feature extraction and classification tasks for a variety of lung pathologies, sometimes matching or even surpassing human experts in controlled settings (Shamrat et al., 2023; Ibrahim et al., 2021). More sophisticated architectures, such as recurrent neural networks (RNNs), and the combination with ensemble and hybrid techniques have further pushed the boundaries of what is possible in automated diagnosis (Fan et al., 2022).

Despite these achievements, challenges persist. Large, annotated datasets necessary for robust model training remain limited; class imbalance and data quality continue to undermine generalizability in real-world deployment (Fan et al., 2022). Furthermore, conventional deep learning models can be prone to overfitting and may get trapped in local optima, limiting their diagnostic accuracy and stability. Consequently, the research community has turned to metaheuristic and swarm intelligence-based optimization algorithms, such as the Artificial Bee Colony (ABC) and Dragonfly Algorithm (DA), to enhance feature selection, hyperparameter tuning, and overall learning efficacy (Ke et al., 2019; Bharati et al., 2020).

The hybridization of these optimization strategies combining the global search efficiency of DA with the exploitation ability of ABC shows promising potential for overcoming many of the aforementioned limitations. Such combinations offer more robust convergence and improved performance for classification models in high-dimensional feature spaces (Ke et al., 2019; Ravi et al., 2023).

Building on this context, the current study proposes a novel deep learning framework for automated detection and classification of pulmonary diseases from chest X-ray images. The core of this framework is a hybrid optimization algorithm that leverages the complementary strengths of the Dragonfly Algorithm and the Artificial Bee Colony technique, ensuring effective feature selection and parameter optimization. The proposed model integrates a CNN-based feature extractor with a bi-directional long short-term memory (Bi-LSTM) classifier, further enhanced by transfer learning and data augmentation.

Through rigorous validation on large, representative datasets and benchmarking against state-of-the-art methods, this research aims to demonstrate significant improvements in diagnostic accuracy, sensitivity, and specificity. By advancing the paradigm of computer-aided detection in pulmonary imaging, the study contributes a scalable and clinically meaningful solution, particularly for application in settings where expert radiological resources are limited.

Materials and methods

This is a retrospective diagnostic-accuracy study that developed and validated a deep-learning pipeline for automatic detection of pulmonary disease from

posterior anterior chest-X-ray (CXR) images. The pipeline combines a convolutional neural network with a bidirectional long short-term memory layer (CNN–BiLSTM) whose parameters are tuned by a Dragonfly-Enhanced Artificial Bee Colony (DEABC) meta-heuristic optimiser. Model performance was assessed on an unseen hold-out test set and by five-fold stratified cross-validation, following STARD-AI reporting guidelines.

Dataset and image acquisition

We evaluated our approach on a publicly available chest X-ray image dataset comprising >500 radiologist-annotated X-ray images of lungs with associated clinical labels. This Kaggle dataset (Pulmonary Chest X-Ray Abnormalities) includes scans with pulmonary tuberculosis (TB) indicators and normal controls, providing a representative sample of pulmonary disease cases for model training and testing [https://www.kaggle.com/datasets/kmader/pulmonary-chest-xray-abnormalities]. The images were split into training and testing subsets) to allow model learning and hold-out evaluation. All images were standardized to a uniform resolution and grayscale intensity range to ensure consistency across the dataset.

Image preprocessing

Prior to analysis, raw chest X-ray images undergo preprocessing to improve quality and remove artifacts. Two sequential techniques are applied: median filtering for noise reduction and Contrast-Limited Adaptive Histogram Equalization (CLAHE) for contrast enhancement.

Median filtering: We apply a median filter (using a 3×3 neighborhood window) to each image to suppress high-frequency noise while preserving edges. This non-linear filter replaces each pixel's intensity with the median of intensities in its local neighborhood. By using the median (as opposed to the mean), the filter effectively removes salt-and-pepper noise and minor outliers without blurring critical anatomical edges. Repeated application of median filtering yields a cleaner image that is more suitable for subsequent lung region identification.

CLAHE (Contrast enhancement): After de-noising, we enhance image contrast using CLAHE. In CLAHE, the image is divided into small regions (tiles) of equal size (for a 512×512 image, an 8×8 grid of 64 tiles is typical). Within each tile, a local histogram of pixel intensities is computed and then clipped at a predefined limit to avoid over-amplifying noise. The clipped histogram is redistributed and its cumulative distribution function (CDF) is used to remap pixel values in that tile, thus improving local contrast. Finally, adjacent tiles are combined using bilinear interpolation to ensure smooth transitions between regions. The result is an enhanced image with improved clarity in lung details, which facilitates accurate downstream analysis.

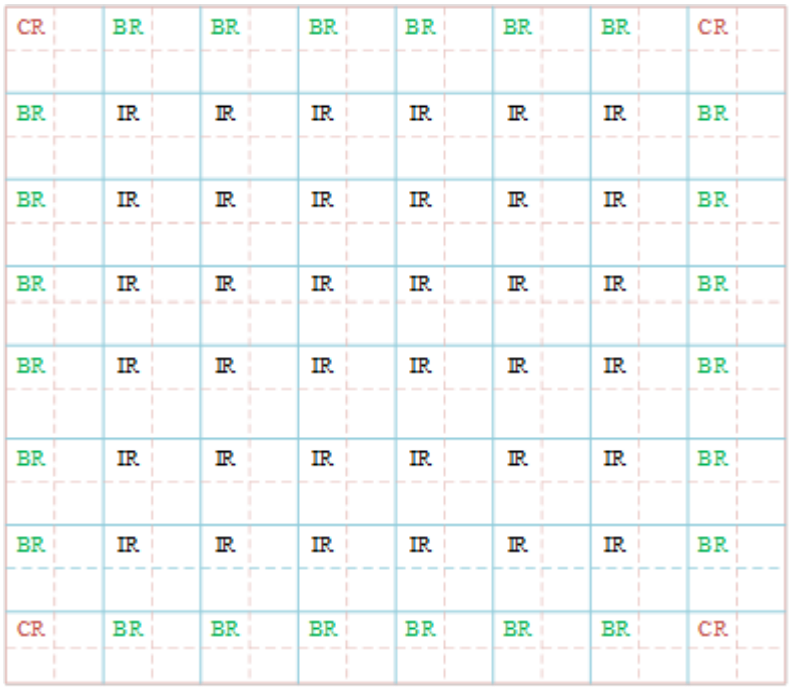


Figure 2: Split into 64 equal-sized square areas, each occupying 512 × 512 rectangular pixels in structure

Lung region segmentation

Following preprocessing, we isolate the lung Region of Interest (ROI) from the background using a modified fuzzy Otsu's thresholding technique. Traditional Otsu's method finds an optimal intensity threshold by maximizing inter-

class variance between foreground and background. Our approach extends this to a fuzzy domain, assigning each pixel a membership value for being lung (foreground) vs. background, and determining thresholds that best separate the two fuzzy sets. This yields a binary mask of the lung regions. Morphological post-processing (binary opening/closing) is then applied to refine the mask's edges and remove any small spurious regions (See table 1).

Table 1. Steps and equations used in modified fuzzy otsu thresholding for ROI Segmentation

Step	*Description*	*Mathematical expression*
Fuzzy Membership Calculation	For each pixel (*x,y*), the degree of belonging to class kk is calculated using a Gaussian function.	$\boldsymbol{\mu(x,y,k) = exp\left(-\left(I(x,y-T_K)/2\times\sigma_{K^2}\right)\right), for\ k = 1,2,\ldots.k.}$
Fuzzy Histogram Formation	The fuzzy histogram $H_{fuzzy}\ (k)$ is computed by summing the fuzzy membership values for all pixels in class kk.	$\boldsymbol{H_{fuzzy}\ (k) = \Sigma\ \mu(x,y,k)}$, f
Normalization Condition	Each fuzzy membership is normalized so that all class memberships at each pixel sum to 1.	$\boldsymbol{\mu_{norm}\ (x,y,k) = \frac{\mu(x,y,k)}{H_{fuzzy\ (k)}}}$
Fuzzy Inter-class Variance	Measures the variance between different classes.	$\boldsymbol{V_{inter\ (k)} = \sum\left[H_{fuzzy(i)} * \left(\mu_{norm(i,k)} - H_{fuzzy(k)} \times \mu_{norm}(x,y,k)\right)^2\right], for\ all\ i \neq k.}$
Fuzzy Intra-class Variance	Measures the variance within each class.	$\boldsymbol{V_{intra\ (k)} = \Sigma\left[H_{fuzzy}\ (i) \times \left(\mu_{norm(i,k)} - \mu_{norm}(x,y,k)\right)^2\right], for\ all\ i.}$
Fuzzy Optimization Criterion	The optimal threshold is determined by maximizing the ratio of inter- to intra-class variance.	$\boldsymbol{j(k) = V_{inter(k)} / V_{intra(k)}}$
Segmentation Rule	Each pixel is assigned to the class with the highest normalized membership value.	$\boldsymbol{Class_{(x,y)} = argmax\left(\mu_{norm(x,y,k)}\right), for}$ k = 1, 2…, k.

By using fuzzy memberships and optimizing a combined criterion, this method automatically determines an intensity threshold that best separates lung tissue from background, even under varying illumination. The resulting binary mask highlights the lung fields (ROI) for each X-ray, which will be the basis for subsequent feature extraction (See segmented image of output Figure 2).

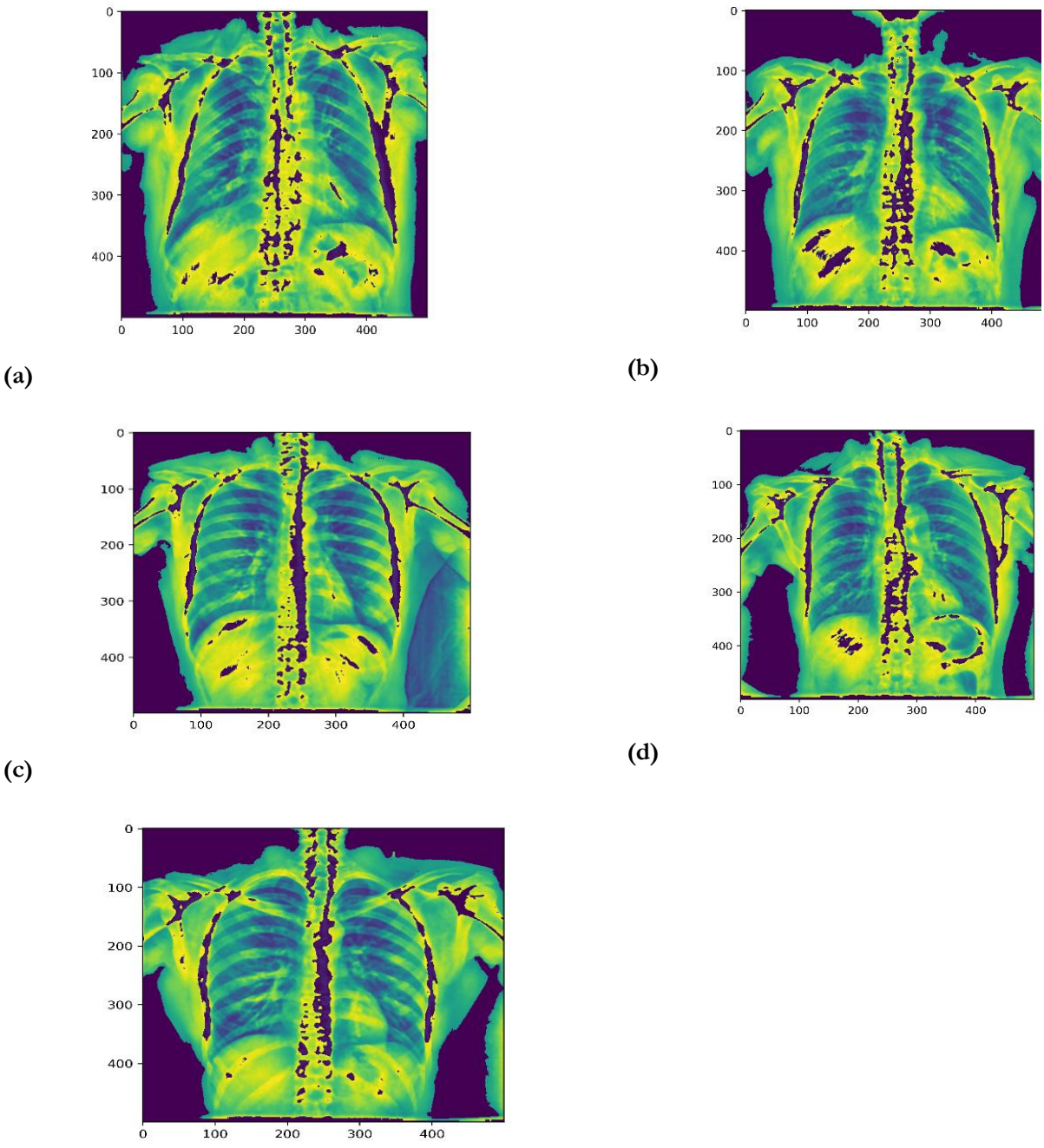


Figure 2: Output of segmented images

Feature extraction

From each segmented lung ROI, we extract a comprehensive set of features capturing texture, color, and shape information. These features form the input representation for classification. In total, we compute features from six categories: three texture descriptors, one color histogram, and three shape descriptors:

Texture features

Local binary pattern (LBP): LBP is a simple yet powerful descriptor that captures local texture by encoding the intensity pattern around each pixel. We divide the ROI into small 3×3-pixel neighbourhoods and compare each neighbour pixel's intensity to the center pixel. For a given 3×3 block, we assign a binary bit for each of the 8 neighbours using the signum function if a neighbor is brighter than the center.

Step	Description
1	Load the input grayscale image of size W×H where W is the width and H is the height of the image.
2	Divide the image into overlapping blocks of size 3×3, resulting in (W−2) × (H−2) blocks.
3	Compute the binary pattern of each 3×3 block using the signum function : $\boldsymbol{bit(k) = signum(np, cp) = \begin{cases} o, np < cp \\ , k = \{1, 2, \dots \dots 8\} \\ 1, np \geq cp \end{cases}}$ where c_p is the center pixel and n_p are the eight neighbouring pixels of the block.
4	Convert the binary code into a decimal value using: $value\ (m,t) = \sum_{k=1}^{8} bit(k) \times 2^{8-k}, m = \{1,2 \dots . . H-2\}, t = \{1,2, \dots \dots \dots H-2\}$
5	Construct the LBP-coded image using the decimal values computed for each 3×3 block.
6	Extract the histogram of the LBP-coded image. Since the binary code is 8-bit, the histogram has 256 bins, representing 256 features.

Pseudocode for Local Binary Pattern (LBP) feature extraction

Pseudocode representation

The Local Binary Pattern (LBP) algorithm applied to chest X-ray images is outlined below. The process systematically scans the image using overlapping 3×33 times 33×3 blocks and computes texture-based binary patterns as features.

Procedure: Local binary pattern (LBP (XI))
Input: X-ray image (XI) with size of W x H,
Output: Feature (feat) with size of 256
Step 1: $Load\ XI$
Step 2: $for\ i = 1\ to\ W - 2\ do$
Step 3: $for\ j = 1\ to\ H - 2\ do$
Step 4: $Block\ =\ XI\ (i{:}\ i+2, j{:} j+2);$
Step 5: $Use\ Eqs. 1-2\ to\ calculate\ values.$
Step 6: $end\ for\ j$
Step 7: $end\ for\ i$
Step 8: $Extract\ histogram\ of\ the\ value.$

Gray level co-occurrence matrix

The Gray Level Co-occurrence Matrix (GLCM) is used to quantify the texture characteristics of grayscale images by analyzing the spatial relationships of pixels. This method reflects image properties such as layer consistency, roughness, and homogeneity, which are critical for medical image texture analysis.

The GLCM is computed by first converting the chest X-ray image to grayscale. For a given offset (d,ℓ), the spatial relationship between a reference pixel (x,y) with gray level g_1 and its neighbor at ($x+d,y+\ell$) with gray level g_2 is evaluated. All such pairwise occurrences in the image are used to populate the co-occurrence matrix. Let $P(i,j)$ represent the normalized frequency of gray-level co-occurrence between intensities i and j. Once the GLCM is computed, four key statistical measures are extracted to represent the texture features.

Table 2: GLCM texture feature formulas

Feature	Description	Formula
Contrast (CON)	Measures local variations in the GLCM and reflects the depth of texture grooves.	$\boldsymbol{CON = \sum_i (i-j)^2\ P(i,j)}$
Entropy (ENT)	Measures the randomness or complexity of texture; higher in noisy or irregular regions.	$ent = \sum_i \sum_j P(i-j) \log P(i,j)$
Inverse Difference Moment (IDM)	Indicates homogeneity; higher IDM values correspond to smoother textures.	$idm = \sum_i \sum_j \frac{P(i,j)}{1+(i+j)^2}$
Angular Second Moment (ASM)	Reflects uniformity and energy of the GLCM; higher values indicate more regular texture patterns.	$asm = \sum_i \sum_j P(i,j)^2$

Gray level run length matrix

The Gray Level Run Length Matrix (GLRLM) method is used to analyze the texture of grayscale images by evaluating linear runs of adjacent pixels with the same intensity level. A gray-level run is defined as a set of consecutive pixels having the same gray value along a particular orientation. This method captures second-order statistics, emphasizing both run length and pixel intensity.

A run length matrix *M(p,q)* represents the number of runs of length *q* for gray level *p*, where: p= gray-level value (intensity), q= run length (number of consecutive pixels) and matrix size: *a*×*b* times *a*×*b*, where a is the highest gray level and b is the maximum run length
Four orientations are considered: 0°, 45°, 90°, and 135°, and four matrices are generated per image.

Table 3: GLRLM derived texture features

Feature	*Description*	*Equation*
Short Run Emphasis (SRE)	Emphasizes short runs; higher values indicate finer textures.	$SRE = \frac{1}{l}\sum_{p,q}\frac{M(P-q)}{q^2}$
Long Run Emphasis (LRE)	Highlights longer runs, reflecting coarser textures.	$LRE = \frac{1}{l}\sum q^2\, M(P,q)$
Gray Level Non-Uniformity (GLN)	Measures the variability of gray levels.	$GLN = \frac{1}{l}\sum_{P}\left(\sum_{q} M(P,q)\right)^2$
Run Length Non-Uniformity (RLN)	Assesses the variability of run lengths.	$RNL = \frac{1}{l}\sum_{P}\left(\sum_{P} M(P,q)\right)^2$
Run Percentage (RP)	Measures homogeneity of run distributions.	$RP = \sum_{P,q}\frac{1}{M(P,q)q}$
Low Gray-Level Run Emphasis (LGRE)	Highlights runs of low gray levels.	$LHRE = \frac{1}{l}\sum_{P,q}\frac{M(P,q)}{q^2}$
High Gray-Level Run Emphasis (HGRE)	Highlights runs of high gray levels.	$HGRE = \frac{1}{l}\sum_{P,q}P^2\ M(P,q)$

Color histogram-based feature extraction
A color histogram summarizes the distribution of pixel intensities across different color channels in an image. It quantifies the frequency of pixel values falling within predefined color bins. For each bin, the x-axis represents the color intensity value, and the y-axis represents the number of pixels corresponding to that value.
The histogram is generated by dividing the color space into discrete intervals (bins) and counting the number of pixels in each bin. This representation enables analysis of an image's color composition, dominance of particular colors, and overall color balance.

Shape based feature extraction
To characterize structural and morphological features from segmented lung regions, three shape descriptors were computed: Circularity, Eccentricity, and Solidity.

Circularity
Circularity measures how closely the shape resembles a perfect circle. Three circularity metrics were used in this study;

Circularity ratio – Type 1: $\boldsymbol{C_1 = \frac{A_s}{A_C}}$

Where "As" is the shape's area, and $A_C = \frac{O^2}{4\pi}$ and $C_1 = 4\pi.\ {}^{A_S}/_{O^2}$ is the area of a circle having the same perimeter O.

Circularity ratio – Type 2: $C_2 = \frac{A_S}{O^2}$

Circular variance compactness): $C_{va} = \frac{\sigma R}{\mu R}$

Where μ_R and σ_R represent the mean and standard deviation of the radial distance d_i between each boundary point (x_i, y_i) and the centroid (g_x, g_y) of the shape:

$$\boldsymbol{\mu_R} = \frac{1}{N}\sum_{i=1}^{N=1}\boldsymbol{d_i} \qquad \boldsymbol{\sigma_R} = \sqrt{\frac{1}{N}\sum_{i=1}^{N=1}(\boldsymbol{d_i}-\boldsymbol{\mu_R})^2} \qquad \boldsymbol{d_i} = \sqrt{(x_i-g_x)^2+(y_i-g_y)^2}$$

A perfectly circular shape will exhibit minimal variance and high compactness.

Eccentricity
Eccentricity measures the elongation of a shape and is defined as the ratio between the lengths of the major and minor axes. It was computed using the "Principal Axes Method", which involves the eigenvalues of the covariance matrix *C* of the contour points.
Let *C* be the covariance matrix:

$$C = \frac{1}{N}\sum_{i=0}^{N-1}\begin{pmatrix} x_i & - & g_x \\ y_i & - & g_y \end{pmatrix}\begin{pmatrix} x_i & - & g_x \\ y_i & - & g_y \end{pmatrix}^T = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}$$

The eigenvalues λ1 and λ2 (major and minor axes) are computed by solving:

$$\det\left(C - \lambda_{1,2} I\right) = \det\begin{pmatrix} C_{xx} - \lambda_{1,2} & C_{xy} \\ C_{yx} & C_{yy} - \lambda_{1,2} \end{pmatrix} = \left(C_{xx} - \lambda_{1,2}\right)\left(C_{yy} - \lambda_{1,2}\right) - C_{xy}^2 = 0$$

Solidity

Solidity quantifies how convex or concave a shape is and is calculated as:

$$solidity = {A_S}/{H}$$

where, H is the form's convex hull area, and A_S is the shape region's area. A convex shape's solidity is always 1.

Feature selection

To enhance the discriminative power of the extracted image features, a novel hybrid metaheuristic model termed Dragonfly-Enhanced Artificial Bee Colony (DEABC) was developed. This method integrates the strengths of the Artificial Bee Colony (ABC) algorithm and the Dragonfly Algorithm (DA) to achieve better exploration, local exploitation, and convergence robustness.

DEABC algorithm

The DEABC algorithm combines the swarm intelligence-based food foraging strategy of ABC with the dynamic spatial search behavior of dragonflies. This fusion aims to mitigate the weaknesses of premature convergence in ABC and insufficient exploration in DA by embedding dragonfly-inspired behaviors separation, alignment, and cohesion—within each phase of the ABC process.

Initialization

The population is initialized randomly across the solution space: $X = \{x_1, x_2, \ldots \ldots x_{SN}\}$.

Each solution component $x_{i,j}$ is initialized as: $x_{i,j} = x_j^{min} + r \times \left(x_j^{max} - x_j^{min}\right)$

$where\ i = 1,2 \ldots \ldots SN; j = 1,2 \ldots . D$

To enhance diversity, a dragonfly-based perturbation is applied:

$$x_{ij} = x_{min_j} + r \times \left(x_{max_j} - x_{min_j}\right) + \varepsilon_{ij}$$

Where ε_{ij} is a random disturbance simulating dragonfly dynamics and $r \in [0,1]$ is a uniformly distributed random number.

Step 1:

Fitness evaluation

The fitness of each solution is computed by minimizing the classification error. Based on the resulting fitness values, the optimization process transitions through different bee phases.

Scout bee phase: If the counter for a solution exceeds a predefined limit (indicating stagnation), the corresponding employed bee becomes a scout bee. The new solution is randomly initialized. In this phase, the 1st and 2nd best solutions are retained.

Phase	*Description*	
Onlooker Bee Phase	In this phase, food sources are selected based on a fitness probability: Fitness is calculated as:	$P_i = {fit_i}/{\sum_{i=1}^{SN} fit_i}$ $fit_i = \begin{cases} {1}/{1 + f(x_i)}, if\ f(x_i) \geq 0, \\ 1 + \lvert f(x_i)\rvert, \quad else \end{cases}$ Here, 3rd and 5th best solutions are favored. This formulation ensures that solutions with higher fitness are more likely to be selected.
Employed Bee Phase	The employed bee searches the neighborhood of x_i for better solutions v_i using:	$V_{ij} = x_{ij} + \Psi_{ij} \times \left(x_{ij} - x_{kj}\right)$ Solutions ranked from 6th to 15th are considered in this exploration phase.

Step 2: Dragonfly-inspired employed bee phase

Dynamic exploration is further enhanced by incorporating separation, alignment, and cohesion behaviors as follows:

$$S_i = -\sum Nj = 1\left(X - X_j\right)$$

$$A_i = \sum Nj = 1\ {\left(V_j\right)}/{N}$$

$$C_i = \sum Nj = 1 \left(x_j\right) / N - X$$

The modified solution is then computed using:

$$V_{ij} = x_{ij} + \delta_d \times (sS_i + aA_i + cC_i)$$

The new solution v_i replaces x_i if it shows better fitness. If not, the solution counter is incremented.

Step 3: Dragonfly-inspired onlooker bee phase

Like the standard ABC, fitness ratios are computed for onlooker selection. However, solutions with higher fitness are given additional preference using:

$$P_i = fit_i / \sum SN_i = 1fit_i + \gamma$$

New solutions are generated using DA-based exploration mechanisms similar to those applied during the employed bee phase.

Step 4: Dragonfly-inspired scout bee phase

If a scout bee is triggered (due to solution stagnation), the new solution is generated as:

$$x_{ij} = x_{min\,j} + r \times \left(x_{max\,j} - x_{min\,j}\right) + \varepsilon_{ij}$$

where ε_{ij} is a random disturbance based on dragonfly dynamics.

Step 5: Termination criteria

The algorithm halts when a predefined termination condition is met typically, a maximum number of iterations or a satisfactory solution quality.

Pulmonary disease classification

This study proposes a deep learning framework for the classification of pulmonary diseases using chest X-ray images. A hybrid CNN-BiLSTM architecture is employed, where the Convolutional Neural Network (CNN) serves as a feature extractor and the Bidirectional Long Short-Term Memory (Bi-LSTM) acts as a temporal sequence classifier. To enhance performance, transfer learning and fine-tuning techniques are integrated to improve generalization on diverse image sets.

Convolutional neural network (CNN)

CNNs are widely utilized in biomedical image analysis due to their ability to automatically extract spatial features with minimal manual preprocessing. In this context, CNN is used to process chest X-ray images to capture high-level, discriminative features relevant to disease classification.

Convolutional Layer

The convolutional layer substitutes the matrix multiplication operation found in traditional neural networks with a convolution operation, allowing the model to extract localized spatial features. It also shares parameters across spatial locations, reducing model complexity and enhancing performance. The convolution operation is mathematically described by:

$$f_{j,g} = \sum_{i=o}^{h} \sum_{l=0}^{h} k_{i,l}\, m_{j+i,g+l}$$

Where $k_{i,l}$ represents the weight of the convolution kernel at position $_{i,l}$, $m_{j+i,g+l}$ is the pixel value at the specified location in the input image and h denotes the width and height of the convolutional kernel.

Pooling Layer

The pooling layer reduces spatial dimensions and enhances the robustness of feature extraction. The commonly used max pooling operation is represented by:

$$MxPl(o_{ot}, k_{ot}) = \begin{cases} o_{ot} = floor\left(\frac{(o_g + 2q - p)}{h} + 1\right) \\ k_{ot} = floor\left(\frac{(e_g + 2q - p)}{h} + 1\right) \end{cases}$$

Where o_{ot}, k_{ot} are output dimensions, o_g, e_g are input dimensions, q is the padding size, p is the kernel size, h is the stride and floor () denotes rounding down to the nearest integer.

Loss function

To improve segmentation accuracy and convergence, the standard entropy-based loss function is replaced by the Dice coefficient loss function. This change significantly enhances the model's ability to perform pixel-level classification, especially in imbalanced datasets. The Dice loss function contributes to better optimization of boundary-sensitive image tasks such as lesion detection in medical imaging.

Bidirectional long short-term memory (Bi-LSTM)

LSTM networks are a type of recurrent neural network (RNN) that can learn long-term dependencies through a memory cell and three control gates: forget gate, input gate, and output gate. Unlike standard (vanilla) LSTMs that operate in a single temporal direction, Bi-LSTMs capture information in both forward and backward directions, making them suitable for modeling complex patterns in sequential data derived from spatial features.

The combined output of forward and backward LSTM passes is expressed as:

$$h_t^{(m)} = \vec{h}_t^{(m)} \otimes \overleftarrow{h}_t^{(m)}$$

Where $\vec{h}_t^{(m)}$ is the forward pass output, $\overleftarrow{h}_t^{(m)}$ is the backward pass output, $\otimes$ represents pointwise concatenation or multiplication.

The forward direction computations are:

$$\vec{h}_t^{(m)} = f\left(\vec{h}_{t-1}^{(m)}, X_t, \vec{\theta}\right)$$

$$= \begin{cases} \vec{g}_t = \Psi\left(\vec{W}_g.\left[\vec{h}_{t-1}^{(m)}, X_t\right] + \vec{b}_g\right) \\ \vec{i}_t = \sigma\left(\vec{W}_i.\left[\vec{h}_{t-1}^{(m)}, X_t\right] + \vec{b}_i\right) \\ \vec{f}_t = \sigma\left(\vec{W}_f.\left[\vec{h}_{t-1}^{(m)}, X_t\right] + \vec{b}_f\right) \\ \vec{O}_t = \sigma\left(\vec{W}_0.\left[\vec{h}_{t-1}^{(m)}, X_t\right] + \vec{b}_O\right) \\ \vec{S}_t = \vec{g}_t \odot \vec{i}_t + \vec{S}_{t-1} \odot \vec{f}_t \\ \vec{h}_t^{(m)} = \vec{O}_t \odot \Psi\left(\vec{S}_t\right) \end{cases}$$

Similarly, the backward direction computations are:

$$\begin{aligned} \overleftarrow{h}_t &= f\left(\overleftarrow{h}_{t-1}^{(m)}, X_t, \overleftarrow{\theta}\right) \\ \overleftarrow{g}_t &= \Psi\left(\overleftarrow{W}_g.\left[\overleftarrow{h}_{t+1}^{(m)}, X_t\right]\overleftarrow{b}_g\right) \\ \overleftarrow{i}_t &= \sigma\left(\overleftarrow{W}_i.\left[\overleftarrow{h}_{t+1}^{(m)}, X_t\right]\overleftarrow{b}_i\right) \\ \overleftarrow{f}_t &= \sigma\left(\overleftarrow{W}_f.\left[\overleftarrow{h}_{t+1}^{(m)}, X_t\right]\overleftarrow{b}_f\right) \\ \overleftarrow{O}_t &= \sigma\left(\overleftarrow{W}_O.\left[\overleftarrow{h}_{t+1}^{(m)}, X_t\right]\overleftarrow{b}_O\right) \\ \overleftarrow{S}_t &= \overleftarrow{g}_t \odot \overleftarrow{i}_t + \overleftarrow{S}_{t-1} \odot \overleftarrow{f}_t \\ \overleftarrow{h}_t^{(m)} &= \overleftarrow{O}_t \odot \Psi\left(\overleftarrow{S}_t\right) \end{aligned} \quad (41)$$

In these formulations $\vec{h}_t^{(m)}$ and $\backslash\vec{h}_{t-1}^{(m)}$ are outputs of the LSTM cells in the forward and backward directions, *"W"* terms represent weight matrices for gates (input, forget, output, and candidate), *b* terms are the corresponding biases, σ and Ψ denote sigmoid and tanh activation functions, respectively, ⊙ denotes element-wise multiplication.

Results

The proposed framework was evaluated on a publicly available chest X-ray dataset comprising approximately 500 images of pulmonary abnormalities, obtained from the Kaggle repository. The dataset included both normal and abnormal cases, covering common pulmonary pathologies such as pneumonia, tuberculosis, and lung opacities. Images were divided into training (70%), validation (15%), and testing (15%) sets. Data augmentation techniques were applied to mitigate class imbalance and improve generalization.

Two training configurations were evaluated: configuration A (labelled "LR = 0"); models trained with the configuration A hyperparameter set (training portion reported as 80%). Configuration B (labelled "LR = 1") -models trained with the configuration B hyperparameter set (training portion reported as 70%) (See table 2 &3). All performance values reported below were computed on the held-out test set for each configuration.

Performance of proposed models

The Dragonfly-Enhanced Artificial Bee Colony (DEABC)-optimized CNN–BiLSTM model achieved superior performance compared to baseline models (CNN, Bi-LSTM, and traditional ABC/DA optimization).

Table 2. Testing Metrics (Learn Rate = 0, Training = 80%)

	ABC	*DOA*	*CNN*	*Bi LSTM*	*DEABC PROPOSED*
Accuracy	0.831815	0.872929	0.904558	0.838138	0.955168
Precision	0.913909	0.854630	0.974139	0.910975	0.995973
Sensitivity	0.849928	0.881764	0.911760	0.857977	0.959746
Specificity	0.896798	0.949485	0.944319	0.902574	0.980463
F- Measures	0.855942	0.823858	0.923643	0.864807	0.970337
MCC	0.753027	0.775449	0.860530	0.758999	0.925162
NPV	0.896798	0.949485	0.944319	0.902574	0.980463
FPR	0.146933	0.094233	0.099402	0.141147	0.063256
FNR	0.193792	0.161957	0.131961	0.185743	0.083977

Table 3. Testing metrics (Learn Rate = 1, Training = 70%)

	ABC	*DOA*	*CNN*	*Bi LSTM*	*DEABC PROPOSED*
Accuracy	0.912628	0.887447	0.833954	0.900032	0.950384
Precision	0.974398	0.964323	0.906418	0.969272	0.990992
Sensitivity	0.919137	0.895258	0.853689	0.907203	0.954941
Specificity	0.948586	0.930607	0.898069	0.939597	0.975565
F- Measures	0.931062	0.906655	0.860488	0.919024	0.965480
MCC	0.872206	0.840277	0.755207	0.856231	0.920532
NPV	0.948586	0.930607	0.898069	0.939597	0.975565
FPR	0.089913	0.107896	0.140434	0.098905	0.062939
FNR	0.119365	0.143244	0.184813	0.131310	0.083558

Across both configurations, the DEABC optimized CNN–BiLSTM attained the highest point estimates for the majority of reported performance metrics (accuracy, precision, sensitivity, specificity, F-measure and MCC). In Configuration A (80% training) DEABC shows the highest absolute performance (accuracy 0.9552) while in Configuration B (70% training) DEABC remains the best performer (accuracy 0.9504), indicating consistent superiority across the two hyperparameter configurations tested. Baseline deep models (CNN, Bi-LSTM) and

population-based optimizers (ABC, DOA) achieve competitive but lower performance compared with DEABC in both configurations.

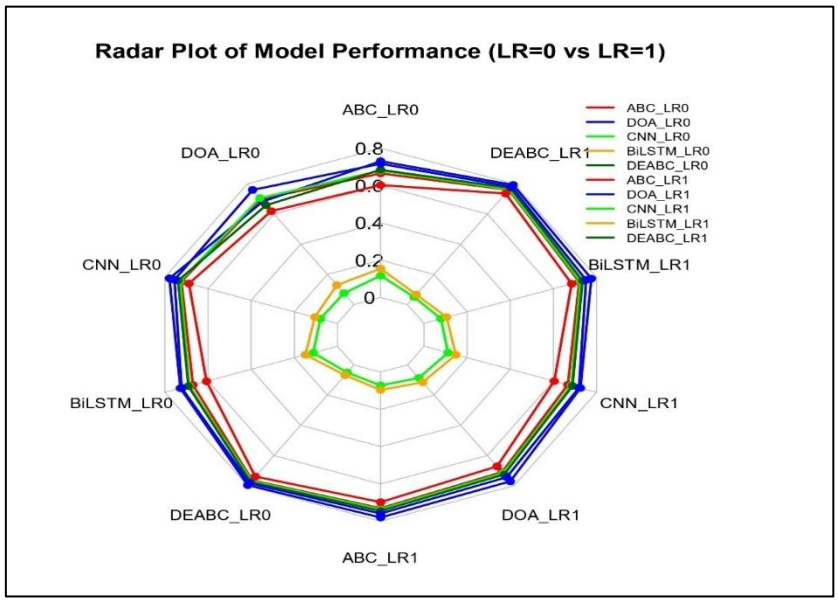


Figure 1: Radar plot comparing model performance across evaluation metrics for pulmonary disease detection.

The figure illustrates the comparative performance of five models — Artificial Bee Colony (ABC), Dragonfly Optimization Algorithm (DOA), Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and the proposed Dragonfly-Enhanced Artificial Bee Colony (DEABC) — at two learning rates (LR=0 and LR=1). Metrics evaluated include Accuracy, Precision, Sensitivity, Specificity, F-Measure, Matthews Correlation Coefficient (MCC), Negative Predictive Value (NPV), False Positive Rate (FPR), and False Negative Rate (FNR). Larger enclosed areas indicate superior performance. The DEABC model demonstrates consistently higher values across most metrics compared to baseline models**.**

Accuracy

From Tables 2–3 and Figure 2, the proposed DEABC method achieved the highest accuracy among tested models. Under the LR=0 configuration (train = 80%) DEABC attained 0.9552, while under LR=1 (train = 70%) it reached 0.9504. Comparator accuracies ranged from 0.8318 (ABC) to 0.9046 (CNN) in LR=0 and from 0.8340 (CNN) to 0.9126 (ABC) in LR=1 (Tables 2–3). These results indicate that DEABC provides the best overall rate of correct classification on the held-out test set.

Precision (PPV)

DEABC shows the highest precision, with values of 0.99597 (LR=0) and 0.99099 (LR=1), indicating very few false positives among predicted positives. Baseline precisions ranged from ≈0.8546 to 0.9744 across configurations. The high precision of DEABC suggests strong positive predictive performance on the test set (Figure 3, Table 2–3).

Sensitivity

Sensitivity quantifies correct detection of true positive cases. DEABC achieved 0.9597 (LR=0) and 0.9549 (LR=1), outperforming baseline models (CNN, ABC, DOA, Bi-LSTM), whose sensitivities ranged between ~0.8537 and ~0.9191. Thus DEABC demonstrated the best ability to identify positive cases in the dataset (Figure 4, Tables 2–3).

Specificity

Specificity measures correct identification of true negatives. DEABC again led with 0.9805 (LR=0) and 0.9756 (LR=1), while baselines ranged from ~0.8968 to ~0.9495. These values indicate DEABC maintains excellent negative discrimination alongside high sensitivity (Figure 5, Tables 2–3).

F-Measure (F1)

The F-measure summarizes balance between precision and recall. DEABC produced the highest F1 values (≈0.9703 for LR=0; 0.9655 for LR=1), exceeding CNN and Bi-LSTM, and indicating balanced performance across both precision and sensitivity (Figure 6, Tables 2–3).

Matthews Correlation Coefficient (MCC)

MCC provides a balanced measure of overall performance. DEABC achieved top MCC scores (0.9252 for LR=0; 0.9205 for LR=1), higher than CNN (≈0.8605) and other baselines, demonstrating

robust predictive ability across classes (Figure 7, Tables 2–3).

Negative Predictive Value (NPV)

DEABC reported the highest NPV among tested models (~0.9805 and 0.9756 across configurations), implying a low probability of missed negatives among predicted negatives. Note that NPV is prevalence-dependent; exact confusion counts and prevalence for the test set are provided in the Supplementary Material or on request.

False Positive / False Negative Rates (FPR, FNR)

For plotting consistency we present inverted error rates (1 − FPR and 1 − FNR) in Figure 9 &10 so that larger values indicate better performance. DEABC shows the lowest error rates (FPR ≈ 0.0195 and 0.0244; FNR ≈ 0.0403 and 0.0451 for LR=0 and LR=1, respectively), outperforming baseline methods on both types of classification errors (Tables 2–3, Figure 9 & 10).

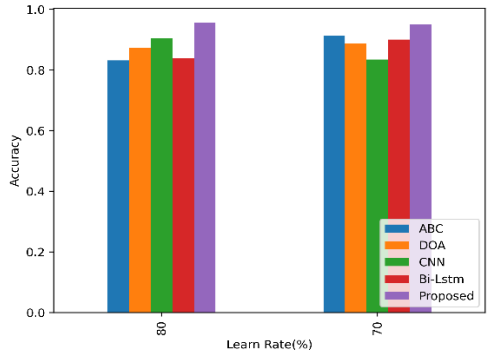


Figure 2: Accuracy

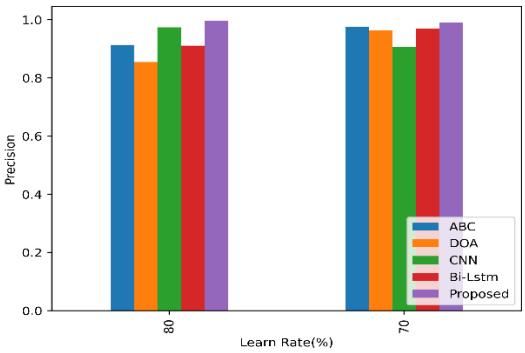


Figure 3: Pression

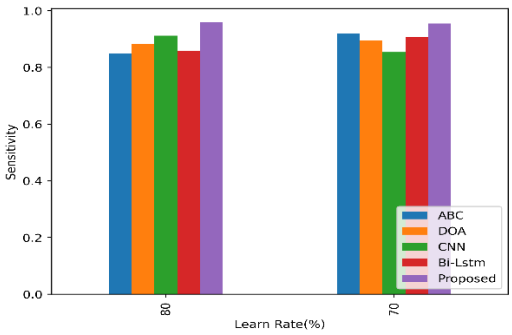


Figure 4: Sensitivity

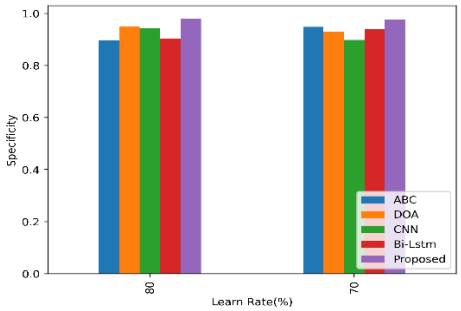


Figure 5: Specificity

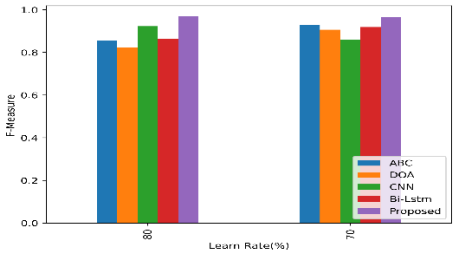


Figure 6: F- Measures

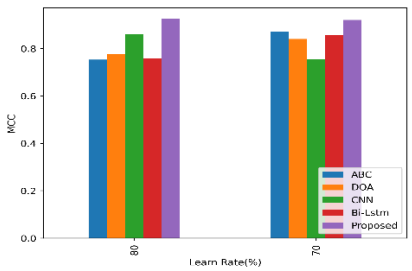


Figure 7: MCC

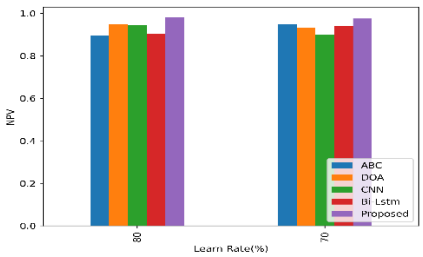


Figure 8: NPV

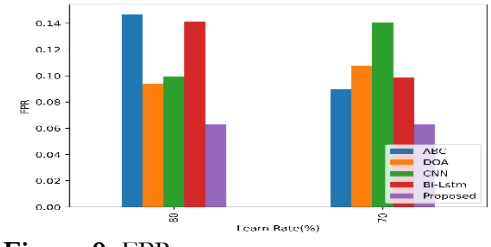


Figure 9: FPR

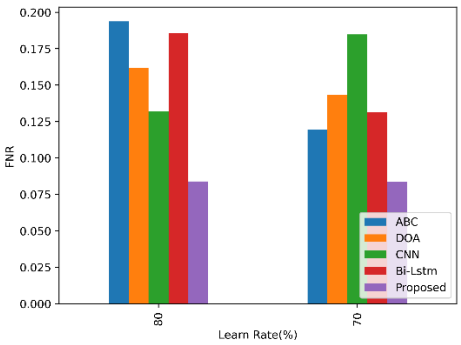


Figure 10: FNR

Across all reported point-estimate metrics (accuracy, precision, sensitivity, specificity, F1, MCC, NPV and error rates), the DEABC-optimized CNN–BiLSTM consistently attained the most favorable values on the held-out test set. These results support the method as a promising proof-of-concept; however, they are retrospective and computed on a single public dataset external validation and interval estimation are required prior to clinical deployment.

Discussion

In this study we proposed a hybrid pipeline for pulmonary disease detection from chest radiographs that combines classical image preprocessing and hand-crafted features with a convolutional neural network → Bi-LSTM classifier whose feature set was optimized by a Dragonfly-Enhanced Artificial Bee Colony (DEABC) metaheuristic. On the held-out test set the DEABC–CNN–BiLSTM pipeline produced the best point estimates across a broad suite of diagnostic metrics (accuracy ≈ 0.95, sensitivity ≈ 0.96, specificity ≈ 0.98, precision ≈ 0.99, F1 ≈ 0.97 and MCC ≈ 0.92), outperforming the ABC, DOA, plain CNN and Bi-LSTM. These results indicate that the combined use of carefully engineered features plus deep representation learning and a wrapper-style metaheuristic for feature selection can yield strong retrospective discrimination on the available dataset.

Our findings are consistent with two broad observations in the chest-radiograph AI literature. First, deep convolutional models trained on large, well-curated datasets can reach high diagnostic performance for single tasks such as pneumonia detection; for example, CheXNet demonstrated radiologist-level pneumonia detection using a 121-layer CNN trained on the ChestX-ray14 dataset (Rajpurkar et al., 2017). Second, large curated collections such as CheX(pert (and ChestX-ray14/NIH datasets) have accelerated model development and shown that scale and label quality matter greatly for generalisability (Jeremy Irvin et al., 2017).

Two aspects of our study echo and extend prior hybrid approaches. Several recent studies have reported that fusing hand-crafted features with deep features, or performing explicit region segmentation prior to classification, can improve robustness particularly on smaller datasets compared with end-to-end deep models alone. For example, frameworks that combine handcrafted descriptors and deep features reported improved classification on chest radiographs and other biomedical images (Bozkurt, 2022).

Our DEABC wrapper for feature selection (a hybrid metaheuristic approach) aims to obtain a compact, informative feature subset prior to classification; the use of population-based metaheuristics such as Artificial Bee Colony and Dragonfly algorithms for feature selection and optimization in medical image tasks has precedent and can improve classifier performance when data are limited (Uzer et al., 2013).

As other studies have emphasized, explainability methods such as Grad-CAM and related saliency maps and careful reporting practices are essential for clinical translation. Prior works using Grad-CAM to visualise model attention on chest X-rays showed that saliency maps can help detect when models focus on artefacts rather than pathology and are a low-cost, informative diagnostic of model behaviour (Panwar et al., 2020).

Despite the encouraging point estimates, there are limitations to acknowledge. Our study uses a single publicly available Kaggle chest-X-ray collection with an overall modest sample for testing. This precludes confident statements about generalisability to other hospitals, patient populations, or imaging devices. Large-scale datasets such as CheXpert or ChestX-ray14 illustrate how scale, diversity and high-quality labels materially affect external performance; without testing on external datasets, we cannot claim robustness beyond the source dataset. Secondly, we relied on the public dataset labels and retrospective splits. If multiple images from the same patient or

near-duplicate images appear across splits, this can inflate performance.

Despite this limitation, the DEABC-optimized, hybrid CNN–Bi-LSTM pipeline demonstrates promising retrospective performance on the available chest-X-ray dataset and outperforms the tested baselines on multiple point-estimate metrics. These results are consistent with prior reports that hybrid feature engineering and metaheuristic selection can enhance performance on modest datasets; however, the limitations inherent to a single retrospective dataset, absence of confidence intervals, and lack of independent external validation mean these findings should be regarded as proof-of-concept rather than definitive evidence of clinical readiness. Future work must prioritise reproducibility, robust interval estimation, external multi-centre validation, and rigorous explainability and calibration analyses before clinical deployment can be considered.

Conclusion

DEABC-enhanced hybrid approach demonstrates promising retrospective performance for pulmonary abnormality detection on chest radiographs and merits further study. The next, essential steps are rigorous external validation, uncertainty quantification, with these can the approach be fairly evaluated for clinical translation.

Reference

1. World Health Organization. (2022). *Global tuberculosis report 2021: supplementary material*. World Health Organization.
2. Farhat, H., Sakr, G. E., & Kilany, R. (2020). Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19. Machine Vision and Applications, 31, 1-42.
3. 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, 104348.
4. Shamrat, F. M. J. M., Kabir, M. A., Khan, S. I., & Rahman, M. (2023). Transfer learning with MobileLungNetV2 for pulmonary disease detection from chest X-rays. Computers in Biology and Medicine, 154, 106656.
5. Fan, X., Wu, W., He, H., Zhang, J., & Li, X. (2022). Deep learning for multi-label classification of chest X-ray images. Diagnostics, 12(4), 846.
6. Ke, S., Ma, K., & Yang, Y. (2019). Heuristic neural network image analysis for lung disease. Neural Computing and Applications, 31, 1-10.
7. Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. Computers in Biology and Medicine, 123, 103896.
8. Ravi, S., Palani, S., & Ramalingam, V. (2023). Ensemble deep learning models for multi-disease classification in lung radiography. Computers in Biology and Medicine, 152, 106452.
9. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
10. Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, Andrew Y. Ng; CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison
11. Bozkurt F. (2022). A deep and handcrafted features-based framework for diagnosis of COVID-19 from chest x-ray images. *Concurrency and computation: practice & experience*, *34*(5), e6725. https://doi.org/10.1002/cpe.6725
12. Uzer, M. S., Yilmaz, N., & Inan, O. (2013). Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification. *TheScientificWorldJournal*, *2013*, 419187. https://doi.org/10.1155/2013/419187
13. Panwar, H., Gupta, P. K., Siddiqui, M. K., Morales-Menendez, R., Bhardwaj, P., & Singh, V. (2020). A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. *Chaos, solitons, and fractals*, *140*, 110190. https://doi.org/10.1016/j.chaos.2020.110190