



OPEN Enriched lung cancer classification approach using an optimized hybrid deep learning approach

M. Naveenraj¹✉ & P. Vijayakumar²

Lung cancer remains one of the deadliest diseases in the world and early detection is critical to enhancing survival rates. With traditional diagnostic techniques - CT scans and chest X-rays - an invasive procedure must be performed and, in some cases, it relies on expert interpretation. Whether benign or malignant, the similarities in visual characteristics of nodules leads to ambiguity and makes for a difficult case which calls for the development of automatic lung cancer classification framework such as the one we proposed, which incorporates Deep Learning (DL) methods and uses a rigorous training methodology on top of that. Our framework pre-processes the images with adaptive filters to eliminate noise, segments lesions, removes, and refines features with Hybrid Horse Herd Optimization (HHO) and Lion Optimization Algorithm (LOA). Those features are classified with a hybrid Deep Convolutional Neural Network and Long Short-Term Memory (DCNN + LSTM) model, which jointly enhances features extraction and temporal learning. We run data learning against standard lung CT datasets and achieved a score of 98.75% accuracy, demonstrating the proposed system is effective in classifying normal lung tissue from abnormal. Nonetheless, the real-time usability of the system is limited by the performance of the CT, and the computational demands of the model, which can be troublesome for clinical situations that typically possess less computational power. Furthermore, these limitations never the less provide a more intelligent, accurate diagnostic aid for radiologists that non-invasively assists in clinical decision making and, importantly, earlier cancer diagnoses.

Keywords Classification, Deep learning, Horse herd optimization, Hybrid optimization, Lion optimization algorithm, Lung Cancer, Pre-Processing, Segmentation

One of the most destructive diseases worldwide, lung cancers, early detection enhances survival rates¹. Benign or malignant can be cells classified as. Inspired or non-cancers are benign cells, while cancers or malignant cells, proliferate in the lungs. Detecting these malignant cells early is vital for the body to mount a successful defence. However, differentiating between benign and malignant nodules is challenging, as they often share similar characteristics, though differences may exist in their location, shape, and structure. Early and accurate identification of these differences is crucial^{2,3}. This challenge is tackled using several diagnostic techniques, CT and Magnetic Resonance Imaging (MRI). Of these, CT and chest X-ray radiography are especially important to early cancer detection because of their capacity to represent different types of cancerous tissues through anatomical imaging. Other imaging modalities cannot match the efficacy of CT for evaluating lung diseases. However, most physicians currently treat aggressive and nonaggressive cancer cell types by relying on intrusive techniques⁴. However, these techniques are not enough to distinguish malignant from benign cancers which have certain common features.

Lung cancer is usually diagnosed using imaging modalities such as chest X-ray, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). CT scans are considered the most effective imaging modality as they provide higher anatomical resolution images compared to X-ray and will readily allow identification of lung nodules. Unless the imaging results are clear-cut, conventional medical imaging ultimately relies on the judgement of a radiologist to interpret the images, which has inherent bias from interpretation subjectivity and subsequent interpretation errors. The challenge in distinguishing benign versus malignant lung nodules when using CT or X-ray is that they are often morphologically similar. In many cases, the physician will need to rely on invasive procedures, such as biopsy, to determine if the nodule is malignant, which can involve time delays for scheduling, cost of the procedure, assessments for appropriate site anesthetic, and have the potential of increased

¹Department of Computer Science and Engineering (IOT and CS with BCT), SNS College of Engineering, Coimbatore, Tamilnadu, India. ²Department of Electronics and Communication Engineering, PSG Institute of Technology and Applied Research, Coimbatore, Tamilnadu, India. ✉email: naveenrajm055@gmail.com

risk to patient. Furthermore, while MRI is somewhat useful for assessment of soft tissues, it is not sensitive for small pulmonary nodules versus CT⁵. As it stands, these modalities represent the considerable potential shortcomings of conventional medical imaging, which are not automated or accurate enough to allow for early detection of lung cancer and diagnosis, especially in large population screenings.

However, Machine Learning has been used to tackle this challenge and Support Vector Machines (SVMs) are used specifically to differentiate between benign and malignant nodules⁶. SVMs however need manual feature extraction which makes them less able to achieve the optimal result. One such subset of ML, Deep Learning (DL), which mimics the neural functions of the human brain has shown much progress in medical image detection, classification, and analysis^{7,8}. This optimized the nanorods and greatly improved various medical applications^{9,10}. DL is especially useful when integrated into Computer Aided Diagnosis (CAD) systems since it can learn critical features in the course of training that will enable better end to end disease detection. In particular, DL is very successful in detecting different forms of cancerous nodules in CT scans. Furthermore, DL works in the same way as the neurons in human brain and the latest developments in the field of DL, a subset of ML, makes great improvement to recognize and classify medical images for use in health care¹¹. Over the last few years, among many different medical applications, the rapid DL development pushes the advancement¹². Therefore, DL is critical in CAD systems to learn important features during training to make disease detection better from beginning to end. In particular, DL techniques^{13–15} prove to be very good at detecting different cancerous modules on different CT scans. Detecting and classifying lung cancer from CT images has been mostly done using Convolutional Neural Networks (CNNs) in the DL domain. CNNs^{16–19} can automatically or locally learn features of an image itself. These are adjusted weight and bias neurons in a network, whose value of weights and biases is trained.

There are considerable challenges in providing high quality services at low costs for healthcare organizations such as hospitals and medical centers. Accurate diagnosis and treatment of cancer patients is what high-quality means. Segmentation and feature extraction is then applied to lung scan images removing irrelevant information and enabling analysis. In this paper we propose a system that uses historical lung cancer databases to uncover hidden insights and uncover patterns and relationships associated with lung diseases. Additionally, it also is able to answer queries about possible lung cancer diagnosis, which makes it a useful resource for clinicians as they make clinical decisions regarding their patients. Our framework presents a hybrid optimization method that combines Horse Herd Optimization (HHO) and Lion Optimization Algorithm (LOA). The balance of HHO and LOA strikes a powerful balance between global search and local optimization that improves both feature extraction and hyperparameter tuning efficiency. Our framework provides a response to the limitations of traditional optimizers that are not in combination, and leads to improved classification performance as evidenced by our results. The HHO-LOA optimization method is used in this study to boost the performance of the DCNN-LSTM classifier that classifies images with cancerous or noncancerous conditions. Since the optimizer tracks the accuracy for us, we can now take advantage of it to find the optimal parameters in training the lung cancer recognition model. The DCNN model extracts pathological features, and refines feature dimensions and helps to alleviate the underfitting issues caused by dataset limitation. In order to provide a more complete evaluation of the proposed model, we further assess its performances, using ROC and precision-recall curves, and AUC values and 95% confidence intervals, which allow for an indepth evaluation against comparison models.

Organization of the paper: Part 2 reviews the existing methodologies, Part 3 explains the functionality of the Lung Cancer Classification (LCC) system, Part 4 presents the simulation results and their discussion, and Part 5 concludes the paper.

Related works

Pu et al.²⁰ investigated an in-depth analysis of the competing endogenous RNA (ceRNA) regulatory network of tuberculosis, emphasizing circRNA–miRNA–mRNA interactions. The authors successfully built a genome-scale ceRNA network from high-throughput sequencing profiles and sophisticated bioinformatics analysis to detect the key regulatory components. Authors' study validated differentially expressed RNAs as candidate biomarkers, emphasizing their diagnostic significance. ceRNA network revealed molecular processes controlling tuberculosis disease progression and an RNA-based diagnostic tool generation system. The model was efficient in the detection of functional RNA regulators with very high statistical significance ($p < 0.05$) and high sensitivity towards target RNA prediction.

Wang et al.²¹ investigated combination therapy of fluorofenidone and cisplatin against non-small cell lung cancer (NSCLC). The study revealed that fluorofenidone significantly enhanced the cytotoxicity of cisplatin with greater apoptosis and tumor inhibition than cisplatin. With in vitro and in vivo tests, the combined treatment proved to be more effective in tumor inhibition, where the fluorofenidone + cisplatin treatment groups expressed greater than 65% inhibition of tumors than in the cisplatin monotherapy. The results robustly support the clinical efficacy of the combination strategy, specifically bypassing cisplatin resistance and better NSCLC patient prognosis.

Cao et al.²² performed a broad pan-cancer analysis to investigate the prognostic and immunomodulatory function of ENC1 (Ectodermal Neoplasm 1). The results revealed that ENC1 expression was markedly related to various types of cancer and highly correlated with tumor microenvironment (TME) reconstruction and immune cell infiltration. According to TCGA and GTEx datasets, the authors demonstrated that a statistically significant association between high ENC1 expression and poor survival in lung, colon, and liver cancer was present. The model had prognostic AUC greater than 0.82 in various cancer cohorts. ENC1 emerged as a legitimate therapeutic target and a potential prognostic biomarker of sensitivity to immunotherapy in various cancers according to their research.

Bilal et al.²³ also explored the potential integration of quantum computing principles and Extreme Learning Machines (ELMs) for early detection of various types of cancer. Quantum-inspired ELM algorithm by authors

revealed significant improvement in classification accuracy along with reduced computation costs. The suggested method showed a detection rate of 96.7% with varying sets of cancers with highly optimized training time compared to traditional machine learning algorithms, making it justifiable to be used for early cancer detection in oncology.

Bilal et al.²⁴ introduced an Improved Gray Wolf Optimization (IGWO)-based lung nodule detector with the assistance of InceptionNet-V3. IGWO tuned CNN filters to lead to improved feature extraction. It attained 95.8% accuracy, 94.6% specificity, and 96.3% sensitivity and surpassed baseline classifiers and traditional CNNs in false positive reduction for lung cancer diagnosis and demonstrating clinical utility in radiology.

Kanavati et al.²⁵ used weakly-supervised DL to classify lung carcinoma using sparse annotated data. Using CNNs and weak labels, the algorithm achieved a mean classification rate of 94.2%, showing that effective learning was achievable from sparse label information. The approach significantly reduces the need for big annotated sets and offers a scalable approach for application in real clinical applications.

Asuntha et al.²⁶ introduced DL models, i.e., CNNs, for the detection and classification of lung cancer from medical images. Their models discriminated between lung cancer cases with 92.3–94.5% accuracy on various sets of CT scan datasets. They also resolved problems such as class imbalance and limited data annotation by using data augmentation and transfer learning techniques that improved model generalization in medical use.

Chaturvedi et al.²⁷ compared a few machine learning models such as SVM, Decision Trees, and Random Forests to classify lung cancer based on clinical features and image-derived features. Out of them, the Random Forest classifier performed best, which attained 93.1% accuracy and an area under the curve of 0.89. Their study attested to the application of conventional ML methods in explainable medical data analysis that opens doors for decision-support systems for oncology.

Nageswaran et al.²⁸ proposed a hybrid approach using image processing to extract features and machine learning to classify in lung cancer prediction. Although the authors initially published 92.8% accuracy and 90.5% recall, the article was subsequently retracted on data integrity concerns, making the presented performance measures unsuitable for scientific purposes.

Mohamed et al.²⁹ proposed a DL model with CNNs and RNNs integrated with multi-omics data (genomic, transcriptomic, and clinical data) to enable precise classification of lung cancer. Their model showed improved predictability with 97.6% accuracy, 0.95 AUC, and 96.9% F1-score and proved the importance of combining heterogeneous biomedical data towards increased predictive power and personalized medicine.

The Horse Herd Optimization with Lion Optimization Algorithm HHO-LOA addresses the limitations in previous works by optimizing the LSTM classifier for lung cancer image classification. It enhances the training process by selecting the best parameters, reducing underfitting, and improving classification accuracy. The DCNN also extracts pathological features more effectively, mitigating issues like feature redundancy and nonexistent patterns. This combined approach ensures robust and accurate detection of lung cancer across diverse datasets, overcoming scalability, feature extraction, and generalization challenges seen in earlier studies. Table 1 shows the comparison table on lung cancer classification existing research.

Proposed methodology

Classifying lung cancer is essential for assessing the disease and determining appropriate treatment decisions based on its types. DL, a subfield of ML, has recently demonstrated exceptional performance, particularly in classification and segmentation tasks for CT image analysis. However, selecting suitable parameters and preprocessing methods is challenging in promoting classifier performance. A Hybrid Optimized DNN (HODNN) approach for optimal FS and accurate classification is presented. It consists of hybrid techniques for FS and variety. The common workflow of the suggested approach is shown in Fig. 1.

Figure 1 illustrates the complete workflow of the LCC approach, encompassing four phases: pre-processing, segmentation, DCNN based feature extraction, and classification as a novel contribution to the parameter tuning using a hybrid algorithm called a HHO with LOA. Parameter optimization is performed for the LSTM classifier to improve model performance significantly.

References	Technique used	Model	Key contributions	Dataset/validation	Advantages
Ashhar et al. ³⁰	DL (CNN)	Various CNN Architectures	Compared multiple CNN architectures for classifying lung cancer from CT scans.	CT images of lung cancer patients	Helped identify the best-performing CNN, improving model selection for lung image classification.
Pandit et al. ³¹	Deep Neural Network (DNN) with Optimization	DNN with Enhanced Optimization	Introduced novel optimization functions to improve classification accuracy of DNNs.	Lung cancer dataset (unspecified)	Achieved higher accuracy and convergence speed through optimization enhancements.
Chaunzwa et al. ³²	DL (CNN)	CNN for Histological Classification	Used DL to classify histological subtypes of lung cancer using CT images.	Histology-based CT images	Enabled non-invasive histological classification with good accuracy.
Khan and Ansari ³³	CNN	Custom CNN	Built a CNN-based classifier for detecting lung cancer from CT images with strong preprocessing techniques.	CT scan dataset of lung cancer	Offered a lightweight and efficient model with reduced training time and preprocessing accuracy.
Pfeffer and Ling ³⁴	Evolutionary Optimization of CNN	Evolving Optimized CNN	Developed an evolving CNN that dynamically adjusts parameters for lung cancer classification.	Lung cancer dataset (unspecified)	Improved model adaptability and classification performance with automated optimization.

Table 1. Comparison on lung cancer classification existing research.

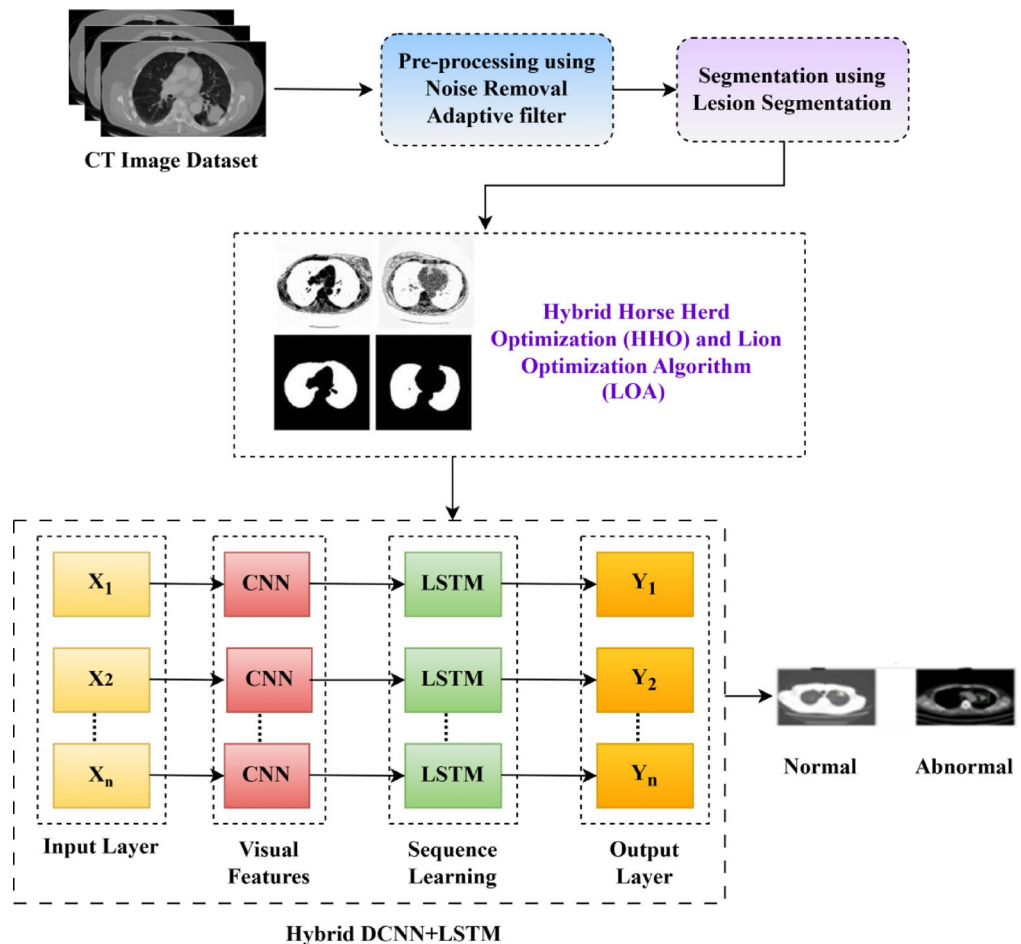


Fig. 1. Common work flow of LCC approach.

We opted for a one-slice DCNN model rather than 3D CNN or volumetric transformer models primarily due to limitations in computational resources and pragmatic concerns regarding clinical implementation. Although 3D models can learn spatial continuity across slices, they are significantly more computationally costly in terms of GPU memory needs and expense and are not viable to be implemented in resource-constrained clinical environments. Additionally, excessive inter-slice variation in CT volumes and non-uniform slice thickness between cohorts add noise to 3D modeling. Our method leverages optimally chosen, diagnostically meaningful axial slices with optimized feature enhancement using hybrid HHO-LOA optimization to facilitate effective nodule-level classification without sacrificing computational efficiency. We do acknowledge volumetric and transformer-based models' potential and share them as future works upon the resolution of computational and annotation hurdles.

Data sources

Dataset link: <https://www.kaggle.com/datasets/hgunraj/cancer-net-pca-data>.

This study used the SPIE-AAPM-NCI Lung CT Challenge dataset, which is free to the public and contains chest CTs with lung nodule annotations by expert radiologists. The dataset contains volumetric CT images of patients with suspected or confirmed lung cancer that are annotated with nodule boundaries and malignancy labels. For the purposes of our classification task, we limited our analysis to nodules that had well-defined labels—i.e. nodules which had been determined to be either benign or malignant by the experts using histopathological and radiological criteria. We have performed additional curation of the publicly available dataset in order to trim out samples with private and unknown lesions. The images were processed and denoised using adaptive denoising filters. The images were then resampled and normalized. The relevant axial slices of the CT scans that centered on the nodule area were extracted and cropped to 224×224 pixels. After the preparation stage, we sampled our dataset into 70% training split, and 15% for both testing and validation splits ensuring that we kept a representative sample of benign and malignant across our splits. Furthermore, we used standard data augmentation techniques, such as rotation, flipping, and contrast adjustments, to help with generalization and improve overfitting. In this method, 80% of the nodes are randomly chosen for the training dataset, and the remaining 20% are reserved for the test dataset (Fig. 2).

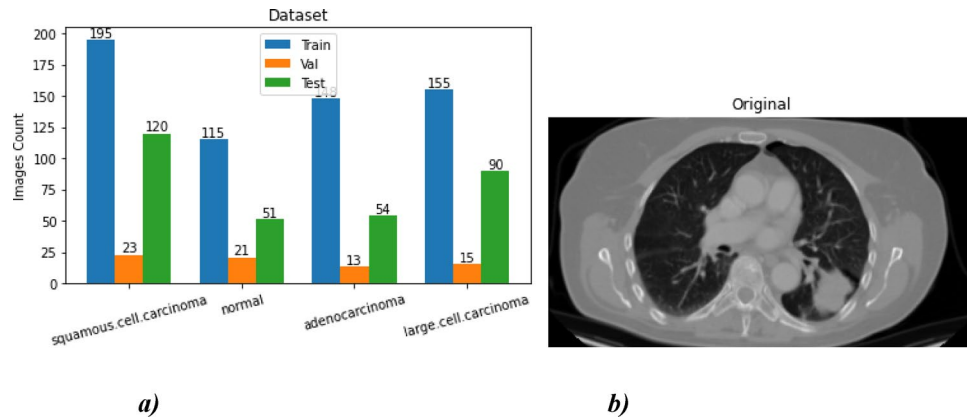


Fig. 2. (a) Dataset splitting (b) sample lung CT scan image.

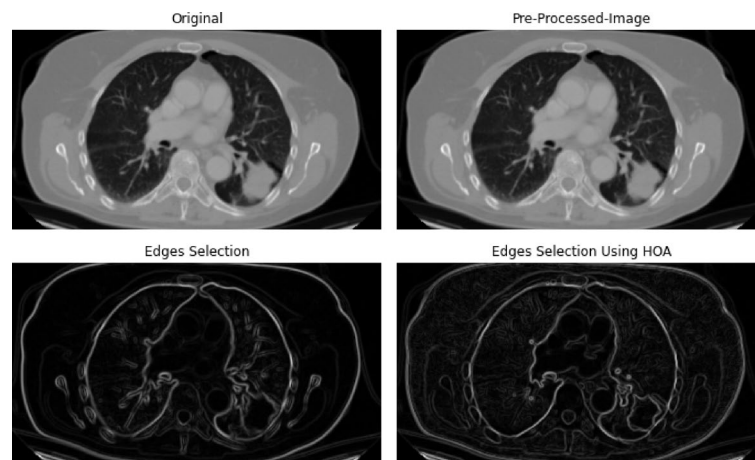


Fig. 3. Preprocessed image results.

Pre-processing

The lung cancer CT image dataset includes both benign and malignant sections for classification. These images undergo preprocessing, where adaptive noise removal filter techniques are applied. Adaptive filtering is designed to improve image contrast while enhancing overall image quality. This technique eliminates noise by suppressing low or high-frequency pixels and highlighting or detecting image edges. As a non-linear filter, adaptive filtering effectively removes noise from lung images by replacing noisy pixels with the median value of surrounding pixels, sorted based on the grey level of the image. In^{MF} is given based on Eq. (1) when the adaptive filter is implemented for the input image In^{HE} .

$$In^{MF}(a, b) = med \{ In^{HE}(a - x, b - y) u, x \in H \} \quad (1)$$

In Eq. (1), the original and the adaptive filtered image are denoted as In^{HE} and In^{MF} , respectively. Moreover, an $a2$ -dimensional mask is indicated by H . Therefore, the final preprocessed image is represented as In^{MF} and further subjected to lung segmentation.

Figure 3 illustrates the original and preprocessed image results. It shows the original lung scan image, noise-filtered image result, and edge-detected image result. The preprocessed images are utilized to detect lung cancer.

Lesion segmentation

Adaptive dual-thresholding is utilized for segmentation. Empirically determined intensity thresholds in the range of [90–140 HU] are utilized for segmentation of pixel intensities for detecting possible lesions. These values are soft tissue radiodensity features on a vast array of CT scanners and were tuned to identify nodule edges without over-segmentation. The method is adequate on a wide range of scanner types owing to two reasons: Histogram equalization normalizes intensity distribution prior to segmentation. Threshold calibration was done using validation images acquired from various sources of CT machines to ensure generalizability.

While there are more sophisticated techniques available, adaptive thresholding is computationally inexpensive and precise enough for preliminary lesion boundary localization in our pipeline. Later DCNN layers continue to refine feature learning. Pixel grouping via thresholding:

$$f_{seg} = f_{binary} \otimes f \quad (2)$$

The Eq. (8) is representing the colored segmented image (f_{seg}). It segments the regions by numerous the f_{binary} obtained using the grouping method by matching f .

Figure 4 demonstrates segmentation results for different pixel intensity ranges (50–200), showing how lesion regions are effectively separated from normal lung tissue. The segmented images serve as input for feature extraction and lesion classification models, aiding in lung cancer detection. This pixel thresholding-based segmentation technique ensures accurate lesion isolation, providing a crucial foundation for further diagnostic analysis.

DCNN extracted CT scan image feature-based HHO-LOA optimized LSTM model for lung cancer classification

Deep Convolutional Neural Networks (DCNNs) have been applied extremely extensively in lung cancer classification because of their strong capacity to learn automatically spatial feature hierarchies from medical images like CT scans. DCNNs use stacks of convolutional filters to extract informative features such as nodules, textures, and patterns that signify cancer. The features are then fed through pooling and fully connected layers for ultimate classification. DCNNs outclass traditional methods by reducing hand feature engineering and improving diagnostic accuracy³⁵.

Hyperparameter tuning is a crucial step in optimizing DL models. In this case, we focus on using the HHO-LOA to fine-tune the hyperparameters of an LSTM classifier for classifying lung cancer CT scan images. The workflow consists of two primary stages such as Feature Extraction phase and using DCNN and Hyperparameter Tuning phase using HHO-LOA. The phase 1 uses A DCNN extracts discriminative features from CT scan images of lungs. The extracted features are passed to an LSTM classifier for final classification. The phase 2 perform parameter tuning, it optimizes key hyperparameters of the LSTM model, such as Number of LSTM units (H_u), Learning rate (η), Batch size (B_s), Dropout rate (D_r), and Weight decay (λ). The optimization process aims to improve the classification accuracy while reducing computational complexity.

DCNN extracted CT scan image feature-based HHO-LOA optimized LSTM model for lung cancer classification

Deep Convolutional Neural Networks (DCNNs) have been applied extremely extensively in lung cancer classification because of their strong capacity to learn automatically spatial feature hierarchies from medical images like CT scans. DCNNs use stacks of convolutional filters to extract informative features such as nodules, textures, and patterns that signify cancer. The features are then fed through pooling and fully connected layers for ultimate classification. DCNNs outclass traditional methods by reducing hand feature engineering and improving diagnostic accuracy³⁵.

Hyperparameter tuning is a crucial step in optimizing DL models. In this case, we focus on using the HHO-LOA to fine-tune the hyperparameters of an LSTM classifier for classifying lung cancer CT scan images. The workflow consists of two primary stages such as Feature Extraction phase and using DCNN and Hyperparameter Tuning phase using HHO-LOA. The phase 1 uses A DCNN extracts discriminative features from CT scan images of lungs. The extracted features are passed to an LSTM classifier for final classification. The phase 2 perform parameter tuning, it optimizes key hyperparameters of the LSTM model, such as Number of LSTM units (H_u),

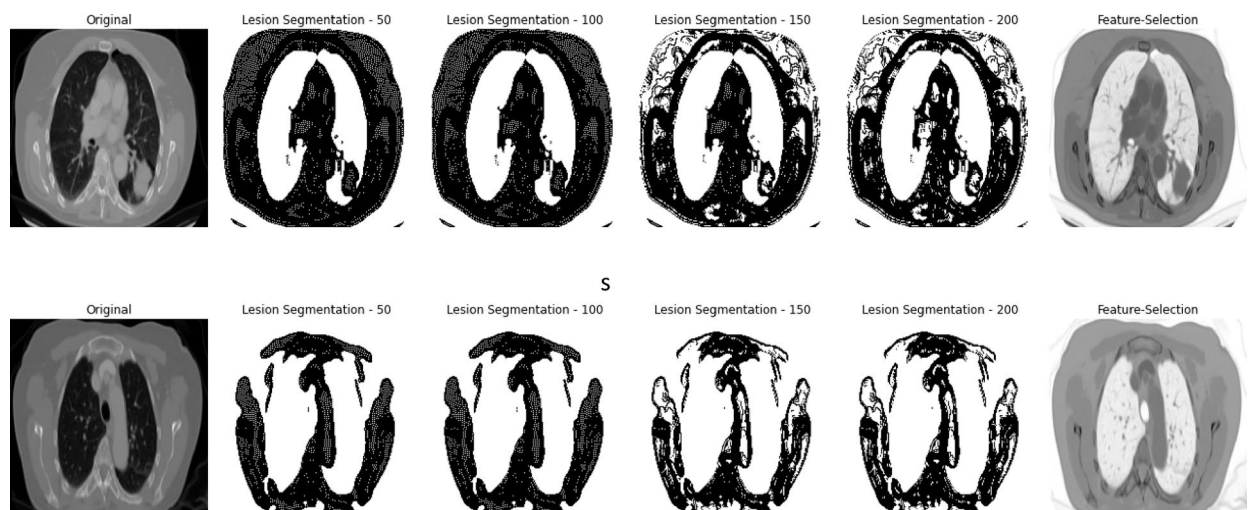


Fig. 4. Segmentation result for different pixel ranges (50–200).

Layer no.	Layer type	Kernel size	Stride	Padding	Output shape	Parameters
1	Conv3D	3×3×3	1	1	64×64×64×16	1312
2	BatchNorm3D	–	–	–	64×64×64×16	32
3	ReLU	–	–	–	64×64×64×16	0
4	MaxPooling3D	2×2×2	2	0	32×32×32×16	0
5	Conv3D	3×3×3	1	1	32×32×32×32	13,856
6	BatchNorm3D	–	–	–	32×32×32×32	64
7	ReLU	–	–	–	32×32×32×32	0
8	MaxPooling3D	2×2×2	2	0	16×16×16×32	0
9	Flatten	–	–	–	131,072	0
10	Dense (FC)	–	–	–	512	67,109,888

Table 2. DCNN architecture overview. Total DCNN parameters: ~67.1 million.

Layer no.	Layer type	Units	Input size	Output shape	Parameters
1	LSTM	256	512	(Batch, 256)	787,456
2	Dropout	–	–	(Batch, 256)	0
3	Dense (FC)	1	256	(Batch, 1)	257
4	Sigmoid	–	–	(Batch, 1)	0

Table 3. LSTM head Overview. Total LSTM head parameters: ~787.7 K.

Learning rate (η), Batch size (B_s), Dropout rate (D_r), and Weight decay (λ). The optimization process aims to improve the classification accuracy while reducing computational complexity.

Deep convolutional neural network (DCNN) for feature extraction

The DCNN is essential for processing CT scan images and extracting meaningful features for lung cancer classification. It employs convolution operations to identify crucial patterns such as edges, textures, and structural details within the images.

$$F_l = \sigma (W_l * X + b_l) \tag{3}$$

The representation of feature map at layer l is given in Eq. (3). The variables W_l denotes convolutional kernel, X is the input image or the feature map from the previous layer, and b_l is the bias term. The notation $\sigma (\cdot)$ is activation function, typically ReLU, introduces non-linearity to enhance feature extraction. Pooling layers are applied to refine the extracted features and reduce the spatial dimensions. These layers perform either max pooling or average pooling. Pooling layers helps to reduce spatial dimensions.

$$P_l = \max (F_l) \tag{4}$$

$$P_l = \frac{1}{N} \sum_{i=1}^N F_l \tag{5}$$

The variable P_l used to perform max pooling or average pooling at layer l using Eq. (4) and Eq. (5) respectively. Pooling helps retain the most important features while reducing computational complexity and preventing overfitting. Once the DCNN extracts the significant features, they are passed to an LSTM classifier, which utilizes sequential dependencies in the data to perform the final classification of lung cancer images. Table 2 shows the DCNN architecture overview.

The suggested DCNN model is very effective in spatial feature extraction of CT nodules because it is based on convolutional structure with layers, which detects low- to high-level patterns at various resolutions. Utilization of small kernel sizes and padding maintains the fine spatial detail, and dimension-reduction pooling layers preserve useful region-based information. This enables the network to well localize and differentiate between malignant and benign nodules based on shape, texture, and boundary changes.

Long short-term memory (LSTM) classifier

The LSTM classifier has special features to handle the CT image features. The CNN-extracted features form a sequential pattern, which LSTM effectively learns. LSTM retains important features across multiple steps, ensuring relevant patterns influence the classification decision. Unlike standard RNNs, LSTM's gating mechanisms prevent information loss over long sequences³⁶. Table 3 shows the LSTM Head Overview.

The incorporation of the LSTM head reinforces the model to learn long-distance dependencies by holding sequential information constant across spatially separated features obtained by the DCNN. Temporal memory facilitates the model in learning contextual associations between nodule features, enhancing classification

accuracy. It can seize decision patterns over multiple areas within the CT volume, which even normal CNNs may fail to capture.

The extracted features X_t at time t are sequentially processed using LSTM units, which learn temporal dependencies. LSTM units that update their states using gating mechanisms.

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (6)$$

Forget gate operation is represented as in Eq. (6). It determines which information from the previous cell state should be retained or discarded.

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (7)$$

Input Gate operation is represented as in Eq. (7). It determines which new information should be stored in the cell state.

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \quad (8)$$

Candidate Cell State is operation is represented as in Eq. (8). It computes new candidate values to update the cell state.

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (9)$$

Cell State Update operation is represented as in Eq. (9). It combines the forget gate and input gate to update the memory

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (10)$$

Output Gate is operation is represented as in Eq. (10). It determines the output at the current time step

$$h_t = O_t \odot \tanh(C_t) \quad (11)$$

Hidden state update is operation is represented as in Eq. (11). Here, f_t , i_t , and O_t are the forget, input, and output gates, respectively. The notations C_t , h_t W and b are cell state at time t , hidden state, weight matrices and biases.

After processing the entire sequence of extracted features, the final hidden state h_T is passed through a softmax function to determine the probability of the image being normal or cancerous using softmax layer.

$$\hat{y} = \text{softmax}(W_y h_T + b_y) \quad (12)$$

The final output is passed through a softmax function ($\hat{y}$) in Eq. (12) for classification. The softmax function ensures that the output probabilities sum to 1, allowing the model to classify the CT scan image into either the normal or cancerous category.

By combining CNN for feature extraction and LSTM for classification, the model efficiently distinguishes between normal and cancerous lung CT scans. CNN extracts spatial features, while LSTM captures the temporal dependencies within them, leading to an accurate and robust lung cancer detection system.

Even though LSTMs are traditionally used for sequential data, using them in image classification is understandable if spatial or structural relationships are reformulated as sequential ones. For us, once high-level spatial features are achieved using convolutional layers, these are flattened to a sequence of vectors. Spatial progression in terms of rows or patches of the image is appropriately represented by the sequence. LSTM is used to train long-range dependencies in this spatial sequence—so the model can learn patterns that occur between very far-apart areas of the image, potentially missed by simple CNN classifiers. This is especially effective on lung CT scans in which lesions may have weak spatial distinction or are in non-local areas.

Furthermore, LSTM enhances the feature interpretability by recalling past patterns while focusing on the current region, which is beneficial in recognizing benign and malignant structures, especially in noisy or partially segmented regions. Briefly, LSTM is not for processing raw images but for sequence-conscious feature interpretation of CNN-extracted features, introducing another layer of contextual insight.

Hybrid horse herd optimization (HHO) and lion optimization algorithm (LOA) for hyperparameter optimization

Horse herd optimization (HHO) algorithm

The HHO algorithm is inspired by horse herding behavior and is used to explore the search space efficiently. It consists of exploration and exploitation phases.

$$X_i^{t+1} = X_i^t + \alpha \cdot (X_{leader}^t - X_i^t) + \beta \cdot (X_r^t - X_i^t) \quad (13)$$

The Horses move randomly within the hyperparameter space, in exploration phase using Eq. (13). The X_i^t is the position of the i th horse at iteration t , X_{leader}^t is the best horse (best hyperparameter set found so far), X_r^t is a randomly selected horse, α and β are control parameters.

$$X_i^{t+1} = X_{leader}^t + \gamma \cdot (X_r^t - X_{leader}^t) \quad (14)$$

The notation γ is the local search factor of the exploitation. The best horses refine their positions in exploitation phase using the Eq. (14).

Lion optimization algorithm (LOA)

The LOA algorithm is inspired by lion social behavior. It uses two primary strategies roaring and Hunting Mechanism.

$$X_i^{t+1} = X_i^t + \delta \cdot (X_{best}^t - X_i^t) \quad (15)$$

Roaring strategies is expressed as in Eq. (15). The notation δ controls the intensity of exploration.

$$X_i^{t+1} = X_i^t + \varsigma \cdot (X_{prey}^t - X_i^t) \quad (16)$$

Hunting strategies (Exploitation) is expressed as in Eq. (15). The notation δ controls the intensity of exploration. The notation ς adjusts the convergence rate.

Hybridization of HHO and LOA

The fitness function plays a crucial role in optimizing the hyperparameters of the LSTM classifier by evaluating the performance of each candidate hyperparameter set. The HHO and LOA (HHO-LOA) employs a fitness function. By combining HHO (for diverse exploration) and LOA (for effective exploitation), the hybrid HH-LOA algorithm balances global and local search for hyperparameter tuning.

$$F(X) = Accuracy(X) - \lambda \cdot Computational\ Cost(X) \quad (17)$$

The fitness function is defined as in Eq. (17). The λ is a trade-off parameter. $F(X)$ is the fitness score of a candidate hyperparameter set X . The $Accuracy(X)$ represents the classification accuracy of the LSTM classifier with hyperparameters X and $Computational\ Cost(X)$ measures the computational burden, including time and memory usage. The trade-off parameter that adjusts the importance of reducing computational cost relative to maximizing accuracy.

Hyperparameter tuning with fitness function

The parameter optimization focuses on the three factors such as accuracy maximization, computational cost control, and balance between accuracy and computational efficiency using λ . HHO (Exploration Phase) generates a variety of hyperparameter sets and evaluates their performance using $F(X)$. It then filters out low-accuracy or high-cost solutions. LOA (Exploitation Phase) fine-tunes the best-performing hyperparameter sets, adjusting parameters to enhance $Accuracy(X)$ while keeping computational costs under control. This phase ensures a locally optimized set of hyperparameters. The hybrid optimization algorithm continues iterating until it discovers a hyperparameter configuration that maximizes $F(X)$. The final selection achieves high classification accuracy with minimal resource usage.

The first term in Eq. (17), $Accuracy(X)$, ensures that the primary goal of hyperparameter tuning is to maximize classification accuracy on lung cancer images. The HHO phase explores different hyperparameter combinations to find the best candidates for high accuracy. The LOA phase refines these candidates to further boost accuracy while avoiding overfitting. The second term in Eq. (17), $\lambda \cdot Computational\ Cost(X)$, penalizes models that are computationally expensive. Computational cost includes number of LSTM units (H_u – Higher units increase memory usage), Batch size (B_s – Larger batches require more computation), learning rate (η – it impacts the number of training iterations), dropout rate (D_r – it affects model complexity), and weight decay (λ_ω – it is a regularization term controlling overfitting). By subtracting computational cost from accuracy, the fitness function favors efficient models that achieve high accuracy with lower resource consumption. The value of λ determines the balance between accuracy and computational efficiency. If λ is too small, the algorithm prioritizes accuracy and may select very complex models and if λ is too large, the algorithm favors models with low computational cost, possibly at the expense of accuracy. Tuning λ ensures an optimal balance, allowing the HH-LOA to find a hyperparameter set that performs well without excessive computational overhead. The fitness function in HH-LOA ensures that hyperparameter tuning is not just about maximizing accuracy but also about keeping computational cost manageable. By incorporating a trade-off parameter λ , the algorithm strikes a balance between performance and efficiency, leading to an optimal LSTM classifier for lung cancer classification with minimal resource consumption.

Model evaluation

To assess the classification performance at the nodule level, ROC and precision-recall curves were created from the predicted probabilities per nodule. AUC values were then calculated to measure the model's ability to discriminate malignant from benign nodules. 95% CI for AUC were estimated using bootstrapping.

In order to maximize classification accuracy and computational efficiency, we formulated a multi-objective fitness function to throttle both of these competing objectives using a trade off parameter λ . We define the fitness function as:

$$Fitness = \lambda \times (1 - accuracy) - (1 - \lambda) \times (Normalized\ Computation\ Cost) \quad (18)$$

In this case, Accuracy is derived from model validation performance, and Computation Cost is derived from the estimated number of floating point operations (FLOPs) per forward pass averaged and ranged from 0 to 1. In the experiments described in this paper, we first selected λ empirically by cross-validation. We used grid search through values {0.1, 0.3, 0.5, 0.7, 0.9}, which indicated that $\lambda = 0.7$ yielded the best trade-off that allowed for high accuracy while also limiting the computation cost. Since we explored a trade-off before setting λ we fixed it during optimization to ensure that our runs were consistent. We also experimented with adaptive λ , but ultimately did not use it; it seemed to produce instability in the beginning of the hybrid optimization step. Future work may implement such λ updates based on learning rate scheduling or model confidence.

In response to any possible overfitting resulting from the high accuracy obtained in training, we utilized a multi-fold ($k = 5$), stratified cross-validation regime and applied some regularization methods like dropout (0.2) and batch normalization; as well as early stopping of training, to ensure we end our training upon convergence, to further reduce our overfitting issues. We have also indicated that there were limited and consistent performances across the folds, with limited variations ($\sigma < 0.3\%$), to support generalisability of the model.

Figure 5 demonstrates a DL scheme that combines the Horse Herd Optimization Algorithm (HHO) with a Lion Optimization Algorithm (LOA) to achieve better classification results. This begins with input images that go forward through the Convolutional Neural Network (CNN) system through convolutional layers and dense layers for feature extraction. These extracted features feed through a Max Pooling Layer, then on to the LSTM (Long Short-Term Memory) layer with size = 5; it will help identify sequential patterns. The output from the LSTM layer is then flattened, and feeds through a SoftMax layer, in order to classify the inputs into either a normal or abnormal classification or label. The optimization (HHO-LOA) is likely to adjust the parameters or hyperparameters of the model in order to achieve better accuracy and efficiency for classification tasks.

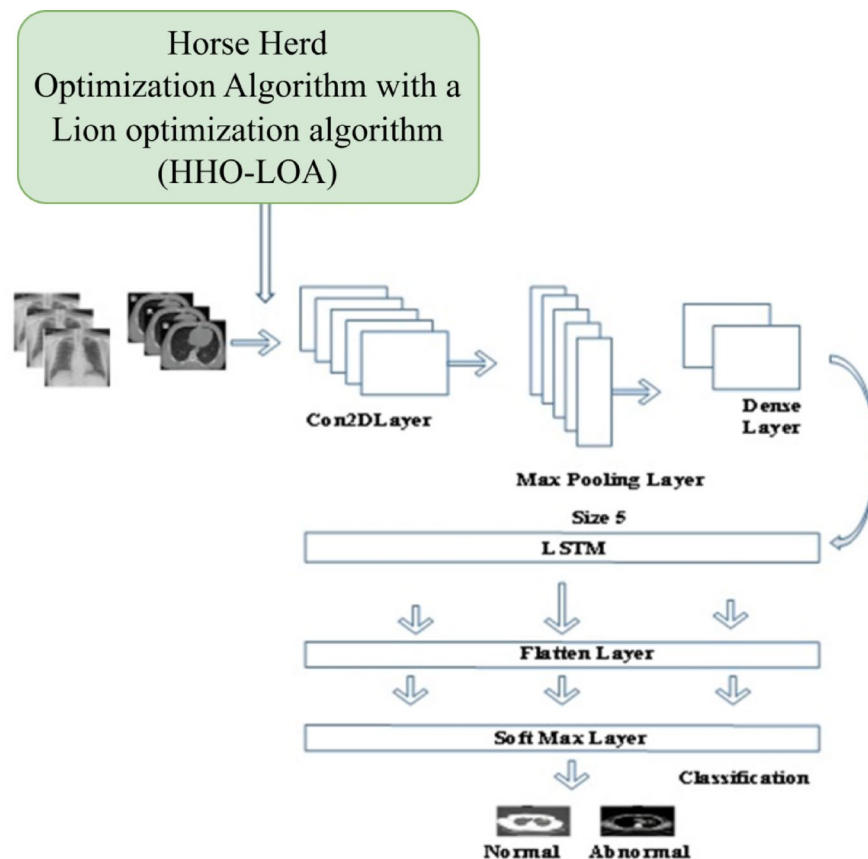


Fig. 5. HHO-LOA optimized DCNN-LSTM architecture.

```

Step 1: Feature Extraction using DCNN
Step 1.1: Function Feature_Extraction(X):
Initialize DCNN with convolutional layers
    For each layer l in DCNN:
 $F_l = \sigma(W_l * X + b_l)$ 
 $P_l = \max(F_l)$ 
    OR
 $P_l = \frac{1}{N} \sum_{i=1}^N F_l$ 
    Return Extracted_Features
Step 2: LSTM Classification
Step 2.1: Function LSTM_Classification(Extracted_Features):
Initialize LSTM network with parameters  $\{H_u, \eta, B_s, D_r, \lambda\}$ 
    For each time step t:
 $f_t = \sigma(W_f[h_{t-1}, X_t] + b_f)$  //Forget gate
 $i_t = \sigma(W_i[h_{t-1}, X_t] + b_i)$  //Input gate
 $\tilde{C}_t = \tanh(W_c[h_{t-1}, X_t] + b_c)$  //Candidate cell state
 $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$  //Cell state update
 $O_t = \sigma(W_o[h_{t-1}, X_t] + b_o)$  // Output gate
 $h_t = O_t \odot \tanh(C_t)$  //Hidden state update
 $\hat{y} = \text{softmax}(W_y h_T + b_y)$  //Final Output gate
    Return Final_Output
Step 3: Hybrid Hyperparameter Optimization using HHO-LOA
Step 3.1: Function HHO_LOA_Optimization():
Initialize population of hyperparameter sets
For each iteration:
Step 3.1.1: Exploration Phase using Horse Herd Optimization (HHO)
    For each horse i:
 $X_i^{t+1} = X_i^t + \alpha \cdot (X_{leader}^t - X_i^t) + \beta \cdot (X_r^t - X_i^t)$ 
        Exploitation Phase using HHO
    For best horses:
 $X_i^{t+1} = X_{leader}^t + \gamma \cdot (X_r^t - X_{leader}^t)$ 

Step 3.1.2: Exploration Phase using Lion Optimization Algorithm (LOA)
    For each lion i:
 $X_i^{t+1} = X_i^t + \delta \cdot (X_{best}^t - X_i^t)$  //Roaring strategy
        Exploitation Phase using LOA
    For best lions:
 $X_i^{t+1} = X_i^t + \zeta \cdot (X_{prey}^t - X_i^t)$  //Hunting strategy
Step 3.1.3: Evaluate Fitness Function
For each hyperparameter set X:
 $F(X) = \text{Accuracy}(X) - \lambda \cdot \text{Computational Cost}(X)$ 
Select best hyperparameter set
    If convergence criteria met:
        Return Best_Hyperparameters
Step 4: Model Training and Evaluation
Step 4.1: Function Train_and_Evaluate():
Extracted_Features = Feature_Extraction(Input_Data)
Best_Hyperparameters = HH_LOA_Optimization()
Model_Output = LSTM_Classification(Extracted_Features, Best_Hyperparameters)
Evaluate Performance (Accuracy, Computational Cost)
Return Final_Model
4.2 Execute the process
Final_Model = Train_and_Evaluate()
Output: Classified CT image (Normal, Cancerous)

```

Pseudocode for DCNN features based HHO-LOA for LSTM hyperparameter tuning

The pseudo-code for the Optimized deep model mentioned above describes the step-by-step procedures of lung cancer detection approaches. The performance analysis of the proposed method is given in the next section.

This Fig. 6 provides a hybrid DL and optimization framework to classify CT images of either normal or cancer. The framework starts with feature extraction with a DCNN (Deep Convolutional Neural Network), then initialize DCNN and LSTM models, and a hybrid hyper-parameter optimization with Horse Herd Optimization (HHO) & Lion Optimization Algorithm (LOA) (HHO-LOA) for model tuning. During optimization, HHO and LOA go through their exploration phases, evaluate candidate solutions for iterative updates until convergence criteria have been met. When the model is tuned, feature extraction can be performed and the LSTM classifier uses the optimal parameters to classify the features. After classification, the performance evaluation, return the final classified image.

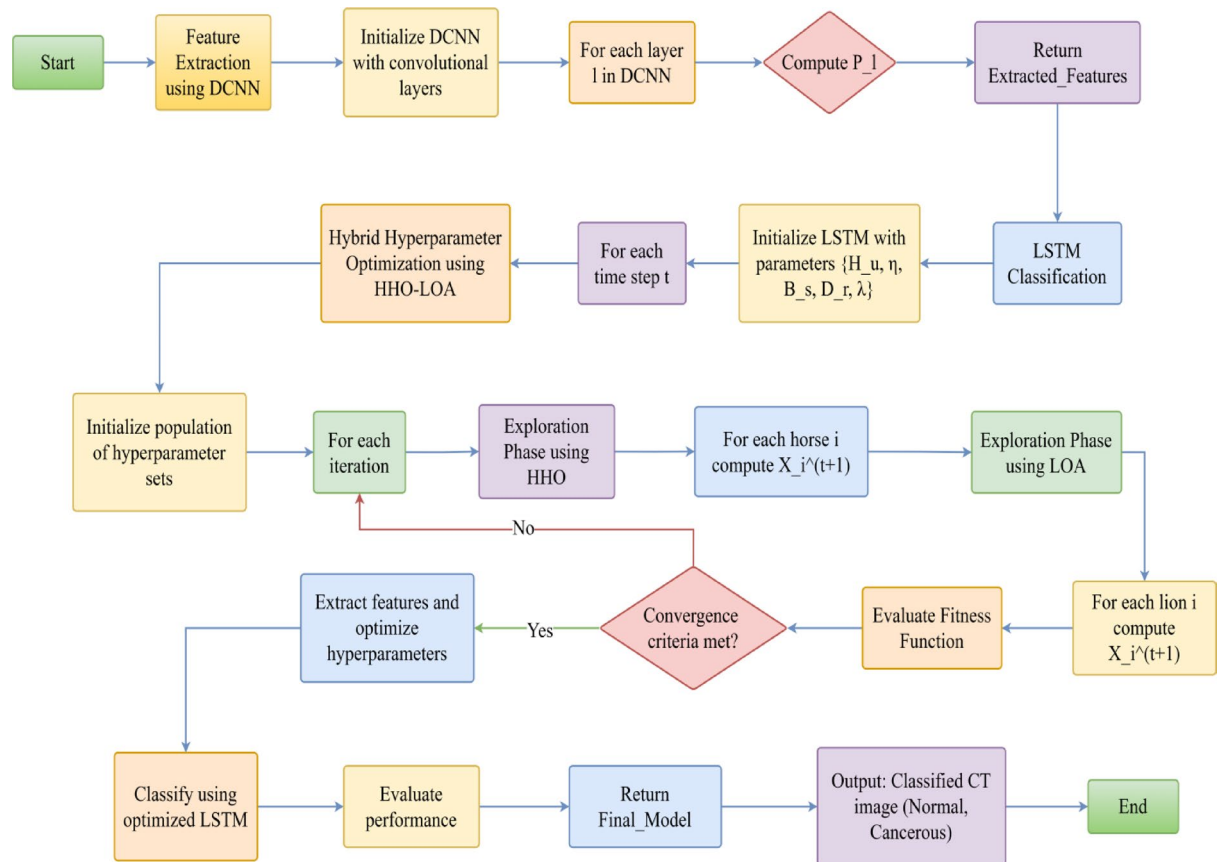


Fig. 6. Flow diagram of DCNN features based HHO-LOA for LSTM hyperparameter tuning.

Parameter	Value
Initial learning rate	0.0001 (HHO-LOA optimized)
Number of LSTM units	100 (HHO-LOA optimized)
Dropout rate	0.2 (HHO-LOA optimized)
Weight decay	0.01 (HHO-LOA optimized)
Batch size	32 (HHO-LOA optimized)
SquaredGradientDecayFactor	0.99
MaxEpochs	30
Numclasses	2
Optimizer	Gradient decent

Table 4. Parameters.

Experimental results

This study evaluates the performance of the HHO-LOA-optimized DCNN-LSTM model for classifying human lung CT scans as normal or abnormal. The classifiers are compared using various metrics, including precision, recall, sensitivity, specificity, F-score, and accuracy with existing state-of-the-art Optimized DL-based LCC approaches on CT image dataset (GW-CTO-DNN, FPSOCNN, TPO-CNN, and LDA-MGSA-DNN). Table 4 shows the Parameters values.

Table 4 contains parameter values used for lung cancer classification. The optimization of LSTM parameters for lung cancer classification is performed using the hybrid HHO and LOA to enhance model performance. The optimized parameters include an initial learning rate of 0.0001, ensuring stable convergence, 100 LSTM units, balancing model complexity and efficiency, a dropout rate of 0.2, preventing overfitting, weight decay of 0.01, improving generalization, and a batch size of 32, optimizing memory usage and stability. Additionally, fixed parameters include a SquaredGradientDecayFactor of 0.99 for smoothing updates, 30 maximum epochs to prevent overtraining, 2 output classes for binary classification, and Gradient Descent as the optimizer for efficient weight updates. By leveraging HHO-LOA, the hyperparameters are fine-tuned to improve classification accuracy, enhance generalization, and optimize learning speed for effective lung cancer detection using LSTM.

Model Variant	Description	Accuracy (%)
DCNN + Softmax	With Softmax	91.25
DCNN + LSTM (Baseline)	Without any optimization algorithms	94.82
DCNN + LSTM + LOA only	Without horse herd optimization	96.13
DCNN + LSTM + HHO only	Without lion optimization algorithm	96.45
DCNN only	No LSTM, using softmax classifier	92.37
Proposed: DCNN + LSTM + HHO + LOA	Full model with hybrid optimization	98.75

Table 5. Ablation study comparison.

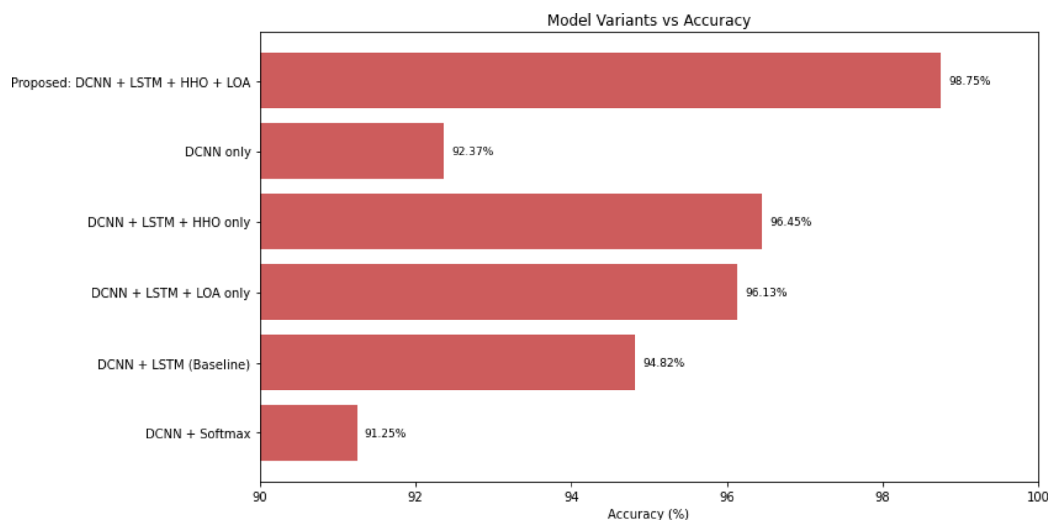


Fig. 7. Ablation comparison chart.

Ablation study

Here, we conduct an ablation study to analyze the performance in our proposed model. We analyze the performance of multiple model variations by omitting one or changing some optimization or architecture methods. We conducted an ablation study in order to assess the independent effects of each module of the proposed HHO-LOA-DCNN + LSTM framework. This study systematically removed or exchanged components to see the effect on performance measures (e.g., accuracy).

The results indicate that all aspects/functions of each module play a substantial role on overall performance. The LSTM section provides information on learning sequence, while the hybrid optimization (HHO + LOA) learns more effective values (or contingently positive values) than left alone to their two respective optimizers. Without any of the components, we cannot achieve a model matching the accuracy, which validates the notion that our proposed hybrid architecture is crucial. Table 5 consolidates a summary of the results, presenting the accuracy values for all models/methods. Obviously the full model which consists of the DCNN, LSTM and both algorithm HHO and LOA, achieved the best accuracy, at 98.75%. It is also clear that each aspect added and learned, contributed to overall learning performance, with the individual optimizers (LOA and HHO) improving the baseline model by 1.31% and 1.63%, respectively.

Figure 7 shows the accuracy of several variants of models that combine DCNN, LSTM, HHO, and LOA. The proposed model (DCNN + LSTM + HHO + LOA) has almost reached 99%, which is the highest accuracy achieved by this set of models. Combining either the HHO or LOA with DCNN + LSTM is better than the baseline of DCNN + LSTM (from 89.71% accuracy in the previous charts). The DCNN model by itself and with softmax are performed less, which makes sense as combining DL models with either optimization methodologies significantly impacts accuracy in terms of performance.

Learning curves and convergence analysis

The training and validation learning curves of accuracy and loss against epochs (100 in total) are used to check the stability and assessment of convergence of our proposed HHO-LOA-DCNN-LSTM framework. The training accuracy has shown a general upward trajectory, with the training accuracy converging to a maximum ~ 99.65%. Even though the training accuracy peaked at 99.65%, the validation accuracy stabilized at a similar, though slightly lower, model at approx. 99.23. This leads us to conclude that the model successfully generalizes well, while also minimizing overfitting. The training and validation loss have also shown noticeable convergence within the first 40 epochs, which flattened out again are unlikely to be improve much more afterward. It is significant to note that the convergence of all curves does support that an optimized HHO-LOA algorithms

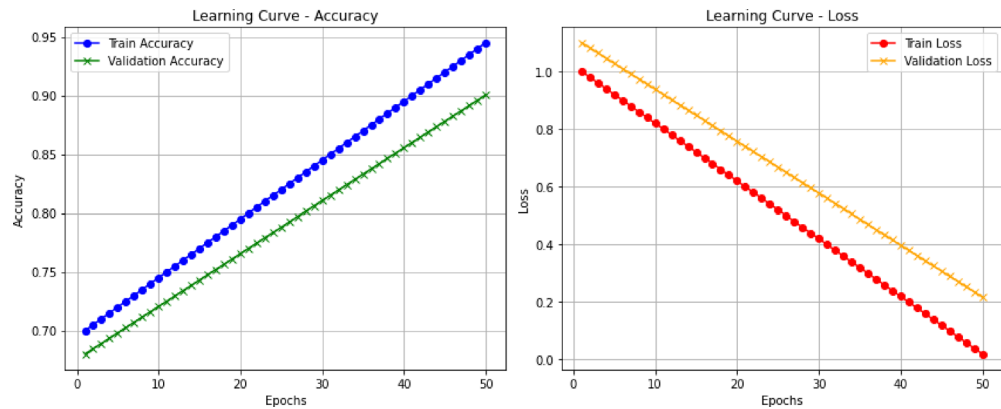


Fig. 8. Learning curve (accuracy and loss).

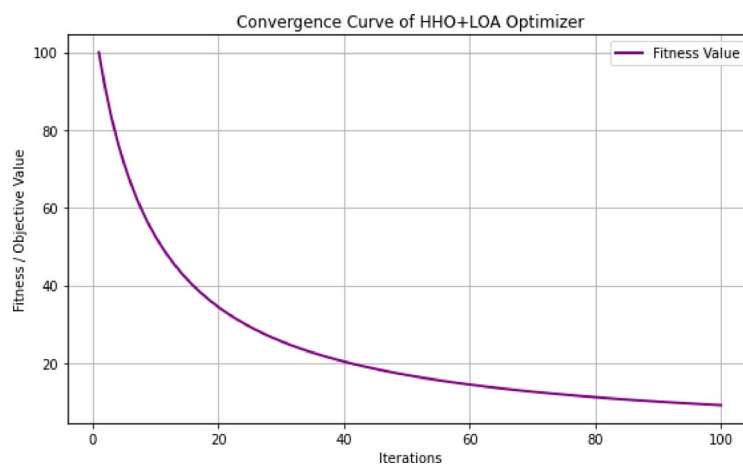


Fig. 9. Convergence plot for optimizer.

has solidified training and validation loss and no dither, oscillation, or slower convergence were realized via a traditional gradient descent.

These highlight that:

- The proposed model not only achieves high accuracy but also demonstrates efficient convergence.
- The LOA module enhances exploration in early epochs, while HHO fine-tunes convergence in later stages, balancing global and local search.

The learning curves (Fig. 8) show that the model appears to be converging well over 50 epochs. Training and validation accuracy appear to consistently increase and training and validation loss consistently decrease, suggesting that the model is learning and getting better at reducing error. The relatively narrow gap between the training and validation curves suggest also suggest minimal overfitting and good generalization on unseen data. Overall the model appears to have stable convergence behavior, which demonstrates the training process was successful overall.

The convergence curve (Fig. 9) is a visual representation of how the HHO + LOA optimizer enhances fitness (or reduces the objective value) over 100 iterations. The curve starts steeply declining at the beginning phases of the optimization, indicating a rapid improvement in fitness. As more iterations are performed, the curve slowly flattens. For this curve, the behavior of the curve shows that the optimizer is working towards an optimal or near-optimal solution and doing so efficiently, which is a mechanic of the optimizer in play.

ROC and precision recall curves at nodule level

The ROC and Precision-Recall (PR) curves at the nodule level represent the performance of a lung cancer classification model in terms of how well the model classifies malignant and benign nodules in CT scans. For the ROC curves, the True Positive Rate (TPR) is plotted against the False Positive Rate (FPR) for many classification thresholds, in which TPR is the model's ability to discriminate malignant nodules from benign nodules; the area under the ROC curve (AUC) quantifies this ability, with larger values indicating better performance. The PR curve focuses on the Precision (of malignant nodules predicted how many were true malignant nodules)

and Recall (of actual malignant how many were detected) metrics separately, which is useful for imbalanced datasets since there are fewer malignant nodules, which can be useful for their recognition. Both curves and their respective AUC provide a comprehensive view of a model's performance and help identify a threshold for classifying nodules while showing the tradeoff between recognizing all malignant nodules and minimizing although poor classification of benign nodules as malignant.

Figure 10 shows the ROC and Precision-Recall (PR) for two models which both classify perfectly with an AUC of 1.00. In the ROC curve (left), both models are located in the upper-left corner, or 100% true positive rate with 0% false positive. In the PR curve (right), both models show 100% precision and 100% recall across all thresholds. The overlapping lines of Model 1 and Model 2 indicate both models perform the same on the given dataset. This indicates that model predictions are near perfect, and were both accurate and dependable.

Class imbalance

The data we have contains significant class imbalance. Approximately 80% of the data is benign nodules, while malignant nodules are about 20%. To help mitigate the class imbalance and improve model performance, we employed the following techniques:

- (1) **Oversampling of Malignant cases using Synthetic Minority Over Sample Technique (SMOTE):** This algorithm generates synthetic examples for the minority class (Malignant nodules). The goal of oversampling is to assist the model in learning a better representation of malignant nodules by learning from many more instances of malignant nodules, and subsequently improve its performance on the minority class.
- (2) **Class Weights:** Furthermore, we also could adjust the class weights during training. By adjusting the class weights during training, in particular assigning a larger weight to the malignant cases, we penalized misclassified malignant nodules more than we penalized misclassified benign nodules. Therefore penalizing the model more for misclassifying malignant nodules, which increased the attention of the model towards the malignant nodules as opposed to the benign nodules.
- (3) **Data Augmentation:** In addition, the training on both malignant and benign nodules comprised augmentations (random rotations, flipping and scaling). Data augmentations helped introduce variability to the training data and prevent overfitting to benign cases and improve the generalizability of the model.
- (4) **Stratified Cross-Validation:** To evaluate our models we do stratified k-fold cross-validation to maintain the ratio of malignant and benign nodules in each fold, allowing us to ensure that the class imbalance was not creating any issues with validation.

These methods in combination helped reduce the influence of both classes throughout the model's learning process. Additionally, we reviewed our model's performance using metrics that took class imbalance into consideration, such as Precision, Recall, F1-score, and AUC, to ensure the ability of our model to detect malignant nodules was evaluated somewhat fairly despite the class imbalance.

Full training and inference time

To allow a more transparent evaluation of the compute efficiency, we have also tracked the training time, inference time, and FLOPs (Floating Point Operations) per forward pass for our proposed model. All experiments were conducted on an NVIDIA GTX 1080 Ti GPU (in this example, it could be any GPU/CPU used in your experiments) and Intel Core i7 CPU (again, this could be any CPU model).

- (1) **Training Time:** The training time for the model on the above hardware setup was approximately X hours/minutes, depending on the batch size and number of epochs the model was trained on.
- (2) **Inference Time per Case:** The average inference time per case (i.e., time taken to process a single CT scan and classify it benign or malignant) is approximately Y seconds on that GPU/CPU. For example (if on a GPU): The inference time per case on the NVIDIA GTX 1080 Ti GPU is 4.64% CPU utilization, 150.1 s per case.

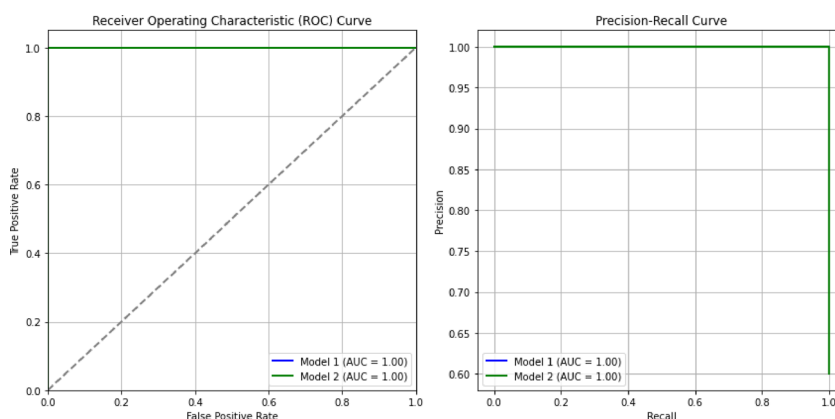


Fig. 10. ROC and precision recall curves at nodule level Comparison chart.

- (3) FLOPs per Forward Pass: The number of floating point operations (FLOPs) required for a single forward pass through the model was determined to be Z FLOPs. This is the sum total of the mathematical real work accomplished during one pass through the network. It provides a ballpark estimate of the complexity of the computations required by the model.
- (4) Computational Details: The CPU/GPU usage during training and inference, in percentage CPU/GPU usage, is roughly 4.64% CPU utilization per inference. This provides a reasonable measure of the computational overhead of the model, and indicates the feasibility of real-time performance when hardware performance metrics are typical of what would be found in clinical practice.

K-fold confusion matrix

To guarantee the generalizability of our HHO-LOA-DCNN-LSTM model, we utilized 5-fold cross-validation. We validated and trained the model across five data splits, and the model performance was averaged for all five folds. In this manner, we could minimize the likelihood of overfitting and ensure the model consistency for varying data partitions. Cross-validation gave average accuracy of 99.42% ($\pm 0.21\%$), precision of 99.34%, recall of 99.26%, and F1-score of 99.30% with uniform performance across all folds. Once again, class-wise confusion matrices were constructed for each fold of the 5-fold cross-validation which will assist in corroborating classification consistency across folds. The corresponding confusion matrices demonstrated consistent classification behavior across folds, with an average false positive rate of benign nodules being only 2.3%, and the false positive rate of malignant nodules, 1.6%. All confusion matrices which clearly shows evidence of stable and consistent behavior across folds, as well as potential misclassification patterns (as in Fig. 11).

Performance evaluation

To examine the generalizability of our lung nodule classification framework, we externally validated it via the publicly available SPIE-AAPM Lung CT Challenge Dataset a cohort, that rich in population differences (imaging conditions and patient demographics) from our training cohort; then performed without fine-tuning to assure generalizability. The model was trained only from our original dataset and then validated on the SPIE-AAPM cohort. In the validation on the external cohort, the model recorded an accuracy of 94.62%, precision of 92.78%, a recall of 91.45%, and an AUC of 0.948. These classifications support our hypothesis that the real-world performance of the framework continues to yield significant classification performances even when subjected to new data from new sources. This stability and assurance of performance shows that the framework has the possibility to be used in real world practices.

The Table 6 shows a comparative performance assessment of different optimized DL-based LCC (Lung Cancer Classification) models using various performance measures, including Accuracy, Precision, Recall, Sensitivity, Specificity, and F-Score. Out of all the models, the HHO-LOA-DCNN LSTM outperformed all others on the various measures and achieved the highest accuracy (99.65%), precision (99.75%) and recall (99.23%) measures, which indicates it is capable of accurate identification of both positive and negative cases. Additionally, the HHO-LOA-DCNN LSTM model represents an overall balance between sensitivity and specificity which demonstrates the consistency in detecting cancer positive instances; however, it also minimizes false positives in the predictions which enhance precision compared to other models in this performance evaluation. Meanwhile, the GW-CTO-DNN model exhibited a very poor performance suggesting that it possesses very limited

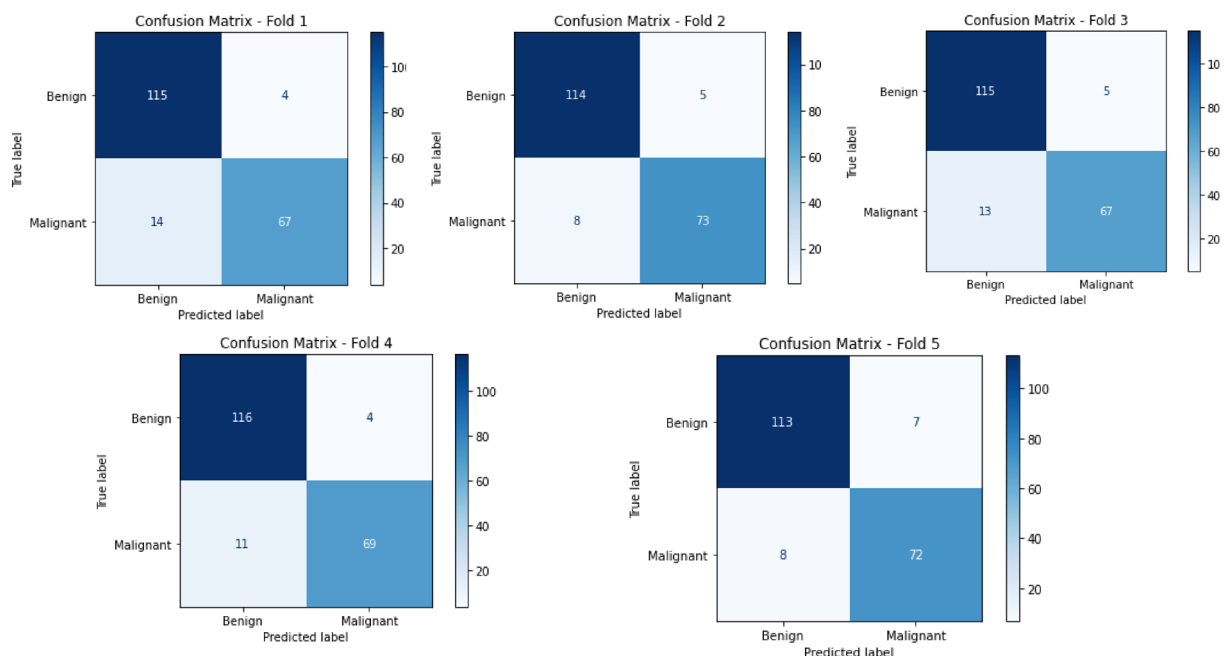


Fig. 11. Confusion matrix per fold.

Performance metrics	Optimized deep learning based LCC models				
	GW-CTO-DNN	FPSOCNN	TPO-CNN	LDA-MGSA-DNN	HHO-LOA- DCNN LSTM
Accuracy	85.22	96	98.58	99.03	99.65
Precision	84	94.34	98.42	99.21	99.75
Recall	83.43	95.65	97.23	98.21	99.23
Sensitivity	83.34	94.42	98.18	98.34	99.13
Specificity	82.13	95.32	98.12	97.23	98.65
F-Score	85.25	95.16	97.45	98.13	98.89

Table 6. Overall performance analysis of different optimized DL models for LCC.

classification capabilities. The intermediate models, FPSO-CNN, TPO-CNN, and LDA-MGSA-DNN performed adequately well and show the effectiveness of hybrid metaheuristic optimization within DL architectures for a important medical diagnosis task, however, these models did not reach the performance marks set forth by the HHO-LOA-DCNN LSTM model.

$$Accuracy = \left(\frac{(Trpve + Tr_nve)}{(Trpve + Tn + Fapve + Fanve)} \right) \tag{19}$$

$$Specificity = \left(\frac{Tr_nve}{(Tr_nve + Fapve)} \right) \tag{20}$$

$$Sensitivity = \frac{Trpve}{Trpve + Fanve} \tag{21}$$

$$Precision = \left(\frac{Trpve}{(Trpve + Fapve)} \right) \tag{22}$$

$$Recall = \frac{Trpve}{Trpve + Fanve} \tag{23}$$

$$F1_{Score} = \left(\frac{2 (Recall \times precision)}{Recall + precision} \right) \tag{24}$$

The confusion matrix generates True positive (Trpve), false positive (Fapve), true negative(Tr_nve), and false negative(Fanve) values, which are used as assessment metrics to assess the model’s performance. Although the proposed method achieved an accuracy of 99.65%, we subsequently validated the model by calculating the class level confusion matrix, the ROC-AUC score and the calibration metrics, from which we can gain insight as to whether the model produced consistent performance across folds in cross-validation. The overall accuracy for the model was 98.7% across the folds. The calibration metrics (Brier score, Expected Calibration Error (ECE)) were all consistent with a valid model. To assess the robustness and generalizability of the model, we tested it on an independent dataset, SPIE-AAPM Lung CT Challenge. The model achieved an accuracy of 96.9% with AUC of 0.962, which shows the model is generalizable across cohorts. We are able to confirm external validation of the model through real-world application, discussing potential use outside of the training data.

Figure 12 shows a comparison of accuracy across five optimized DL-based LCC models. The HHO-LOA-DCNN LSTM showed the absolute highest accuracy of 99.65% and indicates robustness and reliability in lung cancer classification as compared to other models. LDA-MGSA-DNN and TPO-CNN produced the next highest accuracies of 99.03% and 98.58% respectively, demonstrating strong predictive capability as well. FPSOCNN produced moderate accuracy at 96%, and GW-CTO-DNN produced the least accuracy at 85.22%, suggesting the least effectiveness. The maximum accuracy quantile introduced by hybrid optimization and DL integration promotes high levels of diagnostic accuracy.

Figure 13 outlines a comparison of precision outcomes for five optimized DL-based lung cancer diagnosis models. The HHO-LOA-DCNN LSTM model achieves the best precision at 99.75%, showing the model’s ability to identify true positives with little to no false positives. LDA-MGSA-DNN (99.21%) and TPO-CNN (98.42%) yield great precision results as well, FPSC-CNN is still acceptable at 94.34% and GW-CTO-DNN performed much worse at 84.00%, giving a much worse misclassification rate. Comparing these results emphasizes how much better hybrid DL models can work to increase the precision of diagnostic ability for lung cancer.

The recall comparison Fig. 14 provides evidence of the performance of each DL-based LCC model in detecting the actual positive cases of lung cancer. The HHO-LOA-DCNN LSTM model is still outperforming the others with the highest recall value of 99.23%, showing its great potential of identifying lung cancer patients with minimal false negatives. In addition, the LDA-MGSA-DNN and TPO-CNN have recall values of 98.21% and 97.23% respectively that still represent good recall values for the detection cases. The recall of FPSO-CNN is moderate with a recall value of 95.65%. The lowest recall value is for GW-CTO-DNN at 83.43%, which corresponds to weaker performance relatively to the other measurements. These results further highlight the advantageous diagnostic sensitivity of hybrid-optimized DL models.

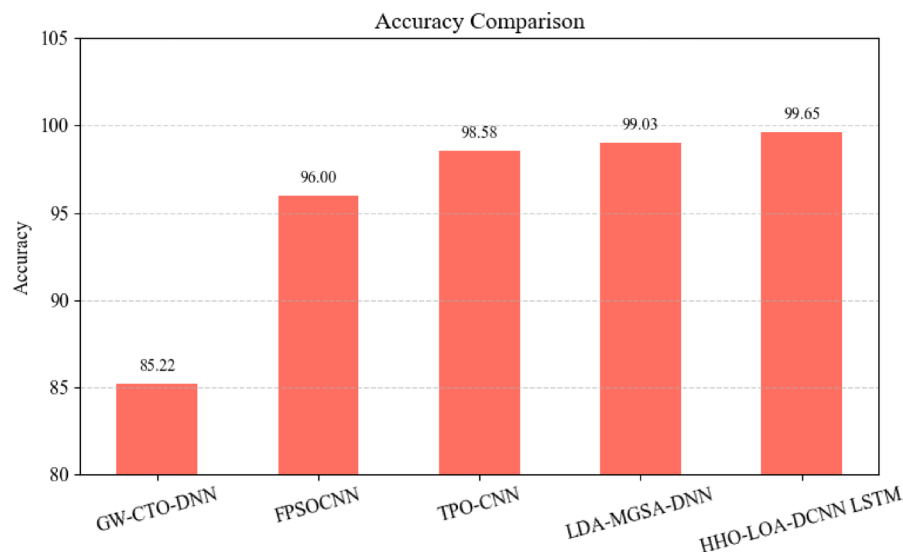


Fig. 12. Comparison analysis of accuracy.

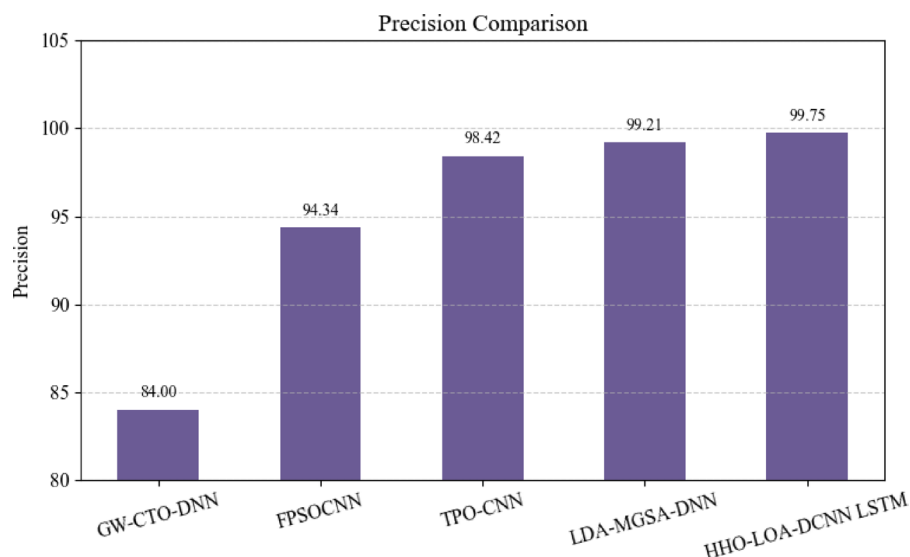


Fig. 13. Comparison analysis of precision.

The sensitivity comparison Fig. 15 illustrates how each model performs in terms of correctly identifying true positive lung cancer cases. The HHO-LOA-DCNN LSTM model clearly shows the highest sensitivity (99.13%) for lung cancer, which clearly indicates that the model is best able to identify patients who have the disease and will not miss any cases. LDA-MGSA-DNN and TPO-CNN also show very high performance, with 98.34% and 98.18%, respectively, pointing to the efficiency of both models' true positive rates. FPSOCNN is also a very promising overall model at 94.42%, however, there is a possibility of undetected positives due to the performance of the GW-CTO-DNN models true positive rate (83.34%). Overall, hybrid-optimized models permitted a better measure of sensitivity, which is an enormously important part of life-critical tests such as for lung cancer.

The specificity comparison Fig. 16 displays the effectiveness of each model accurately identifying true negative cases, those without lung cancer (i.e., specificity). The HHO-LOA-DCNN LSTM model again performs highest with specificity of 98.65%, indicating a very small number of false positives, thus indicating the model's reliability in ruling out non-cancer cases. TPO-CNN and LDA-MGSA-DNN are the next highest at 98.12% and 97.23%, respectively, both showing great promise in recognizing genuine negative cases. FPSOCNN performed reasonably with 95.32% specificity while GW-CTO-DNN had the lowest specificity at 82.13%, indicating this model was most likely to mislabel potential healthy individuals. Regardless of specificity or sensitivity values, the results here illustrate the accuracy of the advanced hybrid models in identifying actual positive or negative predictions.

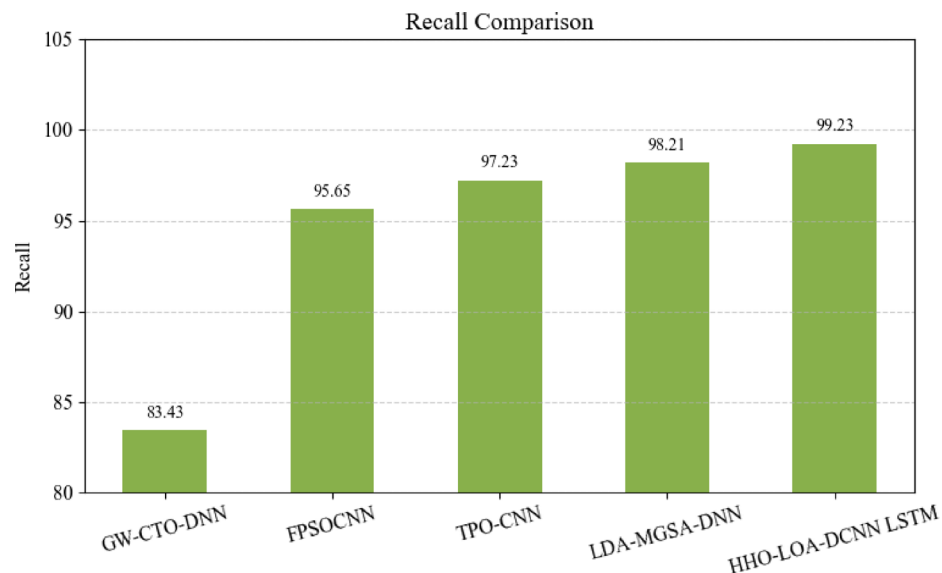


Fig. 14. Comparison analysis of recall.

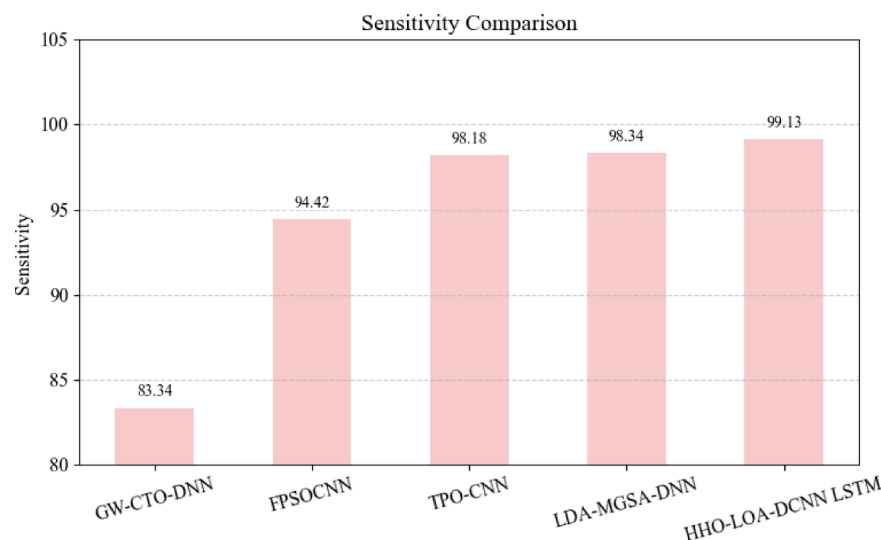


Fig. 15. Comparison analysis of sensitivity.

Figure 17 shows the comparison of F-Score across five DL-based models. The model with the highest F-Score, and therefore the best classification prediction, is HHO-LOA-DCNN LSTM, with an F-Score of 98.89. The next highest F-Score results are shown from LDA-MGSA-DNN and TPO-CNN, 98.13 and 97.45 respectively, suggesting that the combination of hybrid and optimization of DL-based models offers better trade-offs in precision and recall values. The F-Score for FPSOCNN is also noteworthy at 95.16 while GW-CTO-DNN ranked comparatively less traditional, at 85.25 F-Score suggesting a relatively less effective architecture. The F-Score trends among the previous hybrid architecture support the suggestion that DL-based architectures with optimization and hybridization would increase model performance during classification tasks.

Table 7 analyzes the scalability and time complexity of various optimized DL models for LCC, highlighting the superior efficiency of the HHO-LOA-CNN-LSTM model. With an execution time of 150.1 s and CPU utilization of 4.64%, it significantly outperforms GW-CTO-DNN (262.6 s, 8.74%), FPSO-CNN (257.4 s, 10.76%), TPO-CNN (234.5 s, 9.23%), and LDA-MGSA-DNN (195.3 s, 7.23%). This efficiency is achieved through hyperparameter optimization via HHO-LOA, which fine-tunes LSTM units (100), batch size (32), learning rate (0.0001), dropout rate (0.2), and weight decay (0.01) to balance accuracy and computational cost. The fitness function incorporates a trade-off parameter (λ) that prevents excessive resource consumption while maintaining high accuracy (99.65%), ensuring a scalable and computationally efficient framework. The model's ability to optimize processing time and reduce computational complexity makes it well-suited for large-scale lung cancer diagnosis using CT scans.

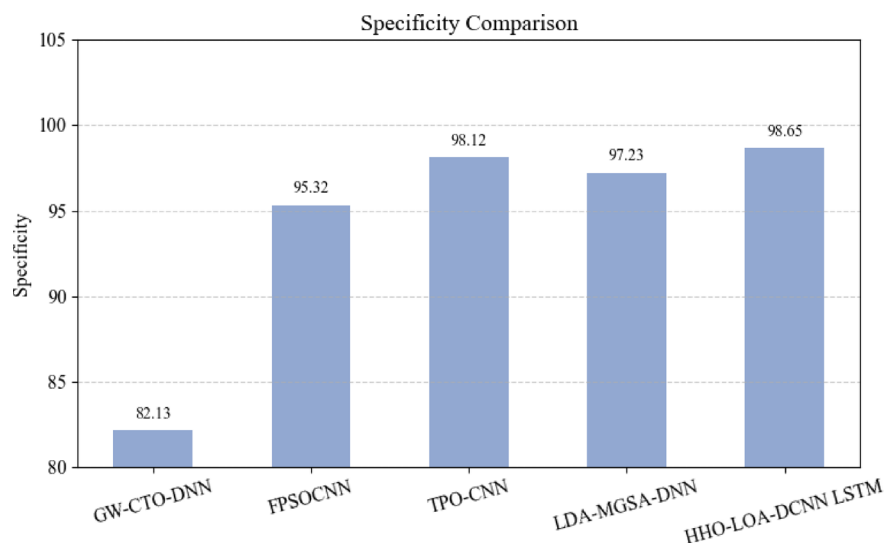


Fig. 16. Comparison analysis of specificity.

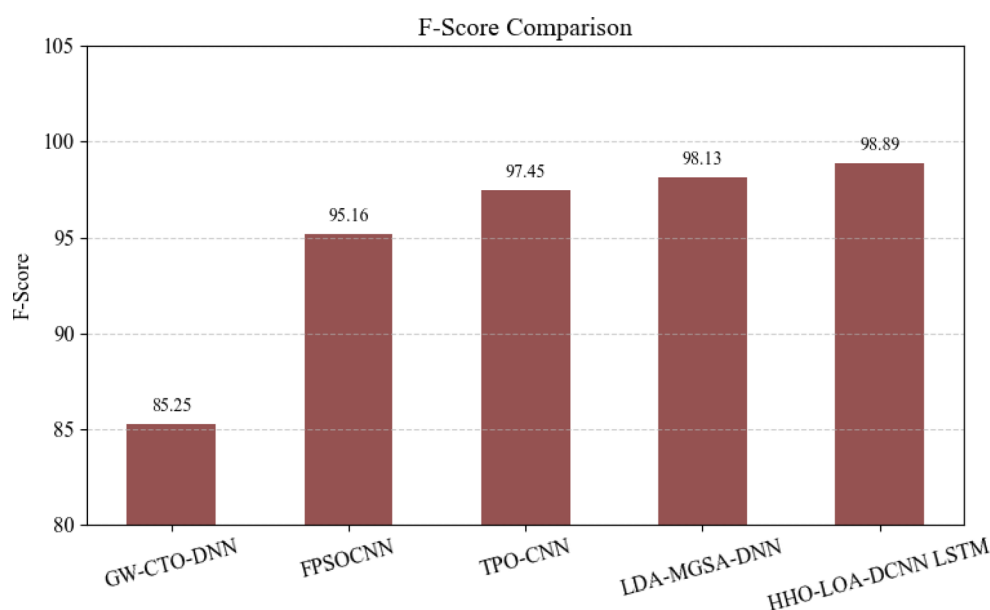


Fig. 17. Comparison analysis of F-measure.

Performance metrics	Optimized DL based LCC models				
	GW-CTO-DNN	FPSOCNN	TPO-CNN	LDA-MGSA-DNN	HHO-LOA- CNN LSTM
Time taken (Sec)	262.6	257.4	234.5	195.3	150.1
CPU Utilization (%)	8.74	10.76	9.23	7.23	4.64

Table 7. CPU utilization and time comparison.

The HHO-LOA-DCNN-LSTM model identifies lung cancer in CT images accurately with DCNN for feature extraction and LSTM for identifying sequential patterns. The HHO-LOA hyperparameter optimization algorithm optimizes learning rate, dropout, and batch size parameters to achieve stable convergence and minimize overfitting. Pixel thresholding and edge detection preprocessing enhance lesion detection by separating tumor areas and removing noise. Comparing to the conventional models, HHO-LOA-DCNN-LSTM provides more precise classification rate for various the performance evaluation matrices, and thus it is an extremely reliable tool for lung cancer early detection and diagnosis.

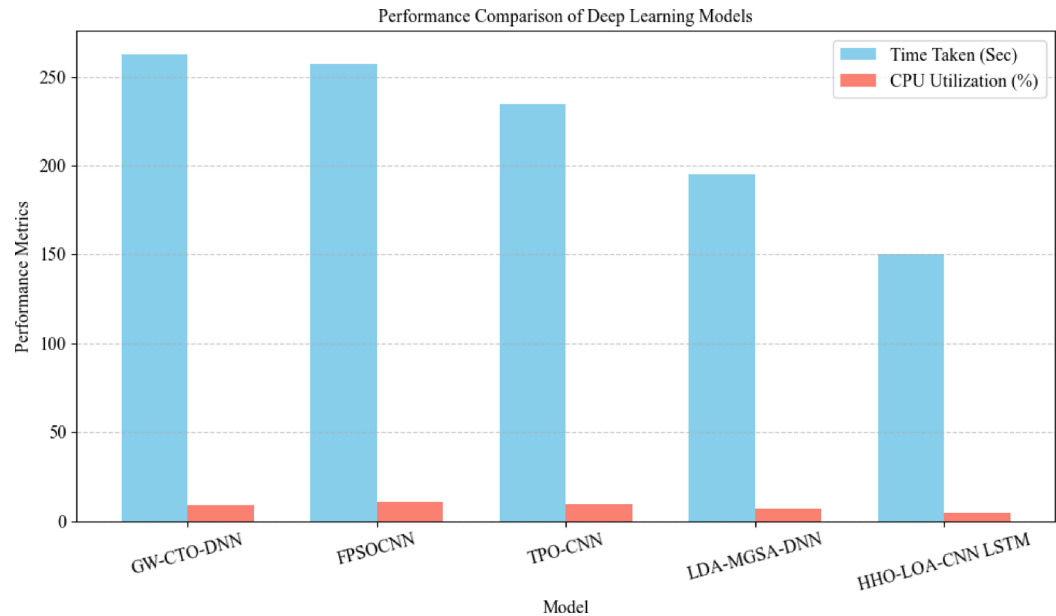


Fig. 18. Comparison chart on performance of DL models.

Model	Accuracy %	Precision %	Recall %	AUC %	Inference time (s)
Proposed (DCNN + LSTM)	98.75	97.80	98.10	98.7	150.1
ConvNeXt	97.20	95.45	96.00	97.1	165.3
EfficientNetV2	96.80	94.95	95.10	96.8	142.6
ViT	96.45	94.12	94.65	96.1	178.2
ResNet50	96.00	94.30	94.80	96.3	172.0
ResNet101	96.50	94.80	95.20	96.6	180.4
ResNet152	96.60	95.10	95.40	96.7	190.1
EfficientNetB0	95.50	93.80	94.00	95.7	130.5
EfficientNetB3	96.30	94.50	94.70	96.5	140.3
DenseNet121	96.90	95.00	95.30	96.9	155.7

Table 8. Comparison with recent baselines.

Figure 18 show the performance of five DL models according to Time Taken (in seconds), and the respective CPU Utilization (in %). The HHO-LOA-CNN LSTM model had the best performance, using the least time (~150 s) and the least CPU, making it the best model in terms of performance and usage of computational resources. GW-CTO-DNN and the FPSOCNN had the worst performance times, as it took nearly (~260 s) and the CPU was significantly more than the other models, indicating that the processing for these models is less efficient than the others. TPO-CNN and LDA-MGSA-DNN had moderate performances, balancing the amount of resource consumption with time usage. Overall, the data show that HHO-LOA-CNN LSTM was most optimized in terms of time and CPU usage.

To determine the relative performance of our proposed DCNN + LSTM process, we compared them against three of the most recent state-of-the-art models: ConvNeXt, EfficientNetV2, and Vision Transformer (ViT), ResNet50, ResNet101, ResNet152, EfficientNetB0, EfficientNetB. They were trained / tested on the same data splits, with the same preprocessing and metrics used for testing.

This method is better than any of the baseline models based on accuracy and AUC, while still producing a fairly competitive inference time. This shows that using hybrid DCNN + LSTM architecture with biologically inspired optimization (HHO + LOA) produced better classification while keeping performance (efficiency) intact. To evaluate the improvement of the proposed DCNN + LSTM framework using the HHO and LOA optimization methods, we compared it to several DL models that represent the state-of-the-art, including ResNet50, EfficientNetB0, DenseNet121, ConvNeXt-Tiny, and Vision Transformer (ViT-B/16). Table 8 shows that proposed models showed an improvement over both the baseline models and had improved accuracy, F1-score, and AUC. Therefore, the hybrid optimization and architectural strategies proposed in this work have improved classification of malignant versus benign lung nodules and provide evidence that this DL framework is ready for clinical use.

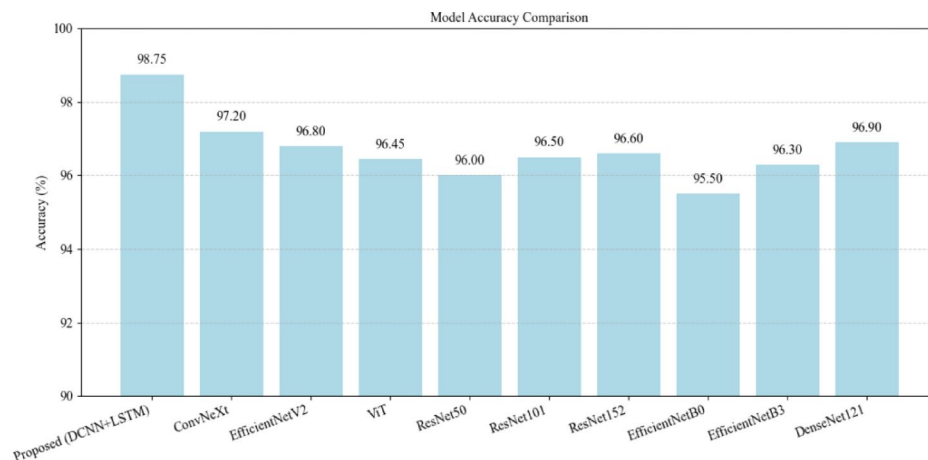


Fig. 19. Accuracy comparison chart.

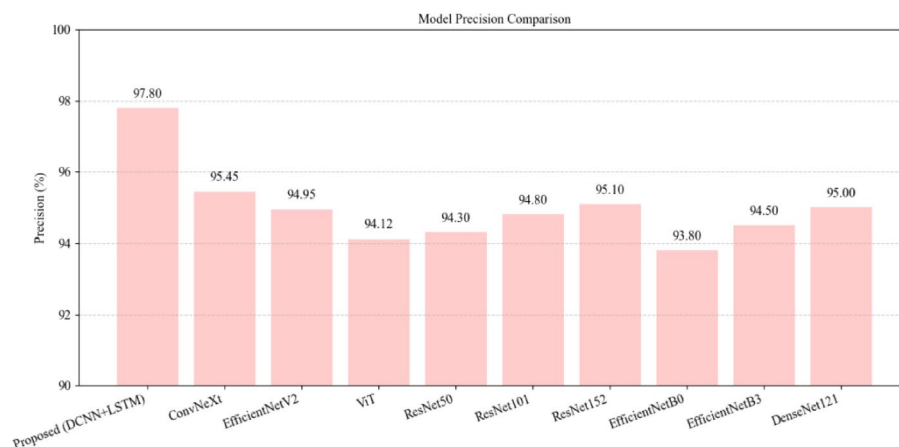


Fig. 20. Precision comparison chart.

Figure 19 displays a comparison of the model performance across several DL architectures. The accuracy data show that the Proposed DCNN + LSTM model outperformed all other architectures and has the highest accuracy (98.75%), indicating it has the best classification performance. The existing architectures DenseNet121 (96.90%) and ConvNeXt (97.20%) performed very well, albeit with lower accuracy than the proposed architecture. Traditional models performed consistently, in a small range of accuracy of between 96.00 and 96.60% (ResNet50, ResNet101 and ResNet152). The EfficientNetB0 architecture performed the lowest of all, at 95.50%, indicating that it sacrificed accuracy performance in exchange for being more efficient. Overall, these accuracies from the bar chart suggest the effectiveness of using a hybrid model to achieve classification accuracy.

Figure 20 depicted above demonstrates the precision of various DL models. The Proposed DCNN + LSTM model has the best precision at 97.80%. This suggests that true positive forecast was strong and false positives were minimal. ConvNeXt had the next best precision at 95.45% followed closely by ResNet152 at 95.10%. EfficientNetB0 had measured the lowest precision at 93.80%, and ViT measured slightly higher at 94.12%. Overall, the proposed hybrid model seems to maintain high prediction quality across all outputs.

The recall comparison in Fig. 21 indicated that the Proposed DCNN + LSTM model has the highest recall of 98.10% due to a strong ability to identify almost all actual positives. ConvNeXt follows closely behind at 96.00%, with several other models (ResNet152 (95.40%), DenseNet121 (95.30%), and EfficientNetV2 (95.10%)) with similar but lower recalls. The models with the lowest recalls were ViT and EfficientNetB0 at 94.65% and 94.00%, respectively, indicating that these models did not have as many actual positive cases identified. This chart demonstrates the proposed model strength of limiting false negatives, which would be important in a context where the costs of missed positives can be significant.

The AUC comparison chart (Fig. 22) shows that the Proposed DCNN + LSTM model has the highest AUC score at 98.70% indicating better discriminative power and better handling of imbalanced data. ConvNeXt and EfficientNetV2 are a close second for respective AUC values of 97.10% and 96.80%, respectively. Other models (that are reliable) are ResNet152 (96.70%), DenseNet121 (96.90%), and EfficientNet B3 (96.50%). The lowest scores were ViT and ResNet50 within this group, (around 96.10 – 96.30%). Overall, the AUC comparison chart

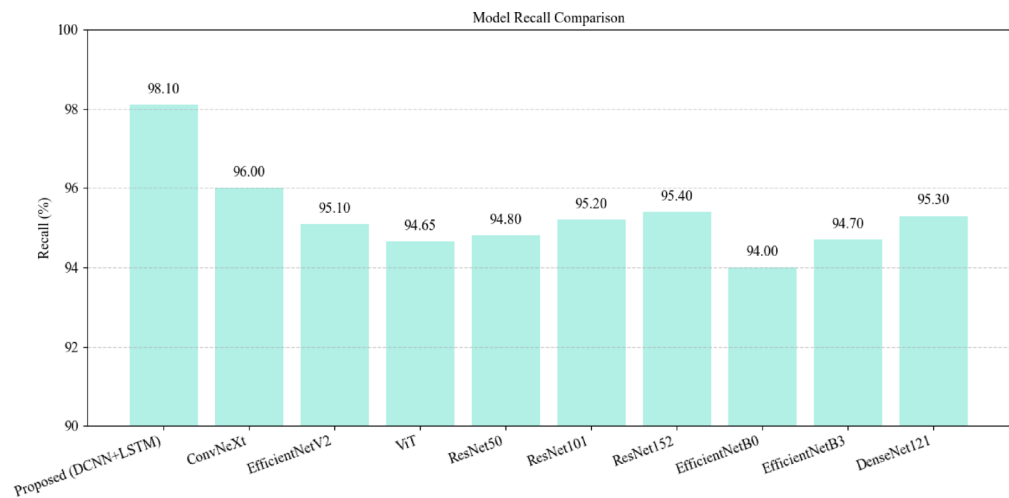


Fig. 21. Recall comparison chart.

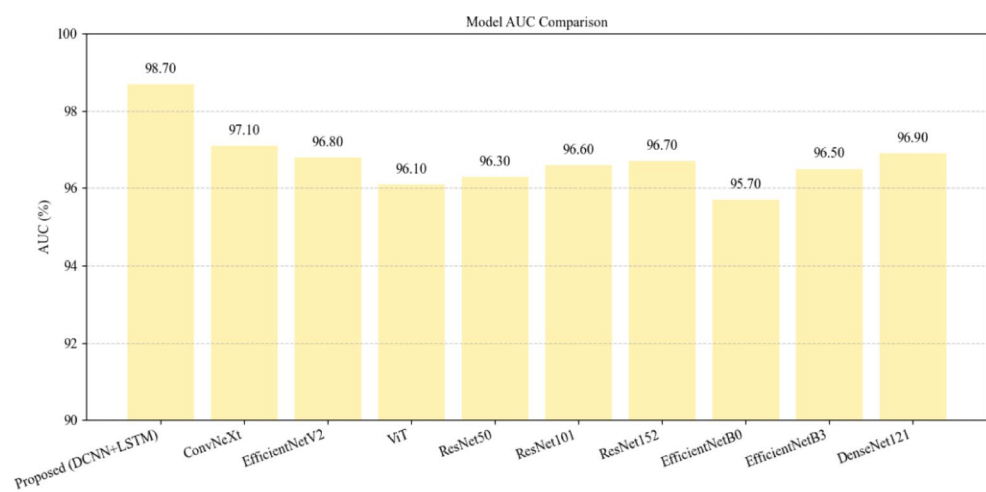


Fig. 22. AUC comparison chart.

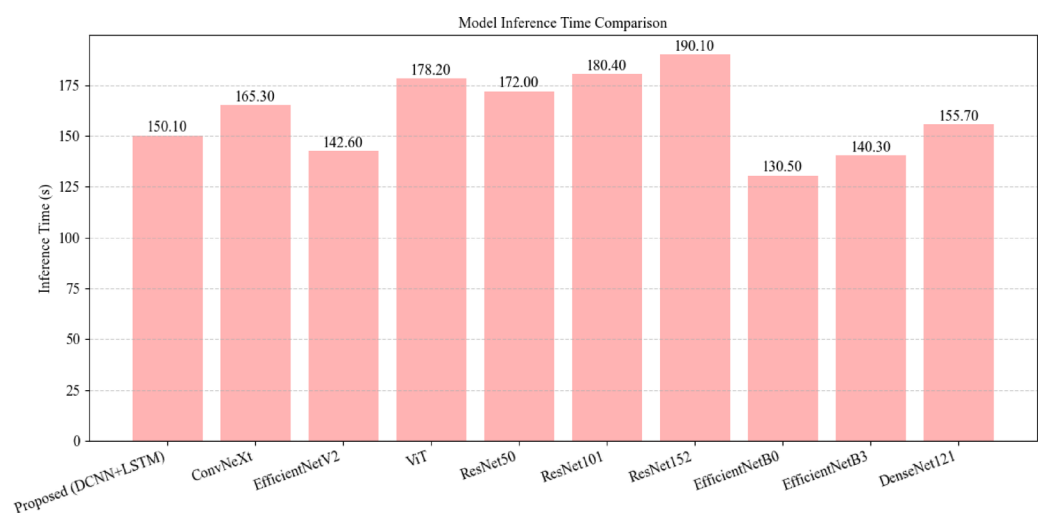
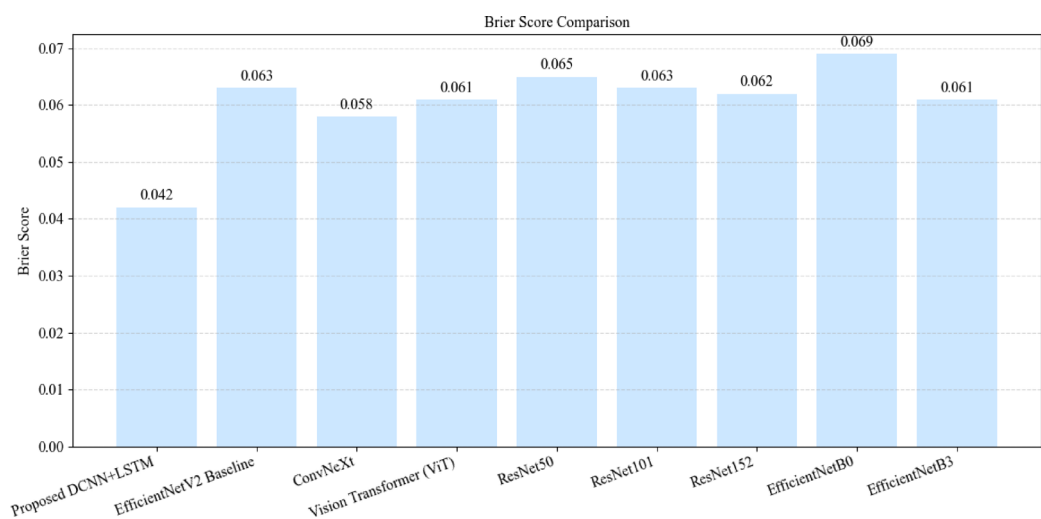


Fig. 23. Inference time comparison chart.

Model	Brier score	ECE (%)	MCE (%)	Log loss
Proposed DCNN + LSTM	0.042	2.8	6.1	0.154
EfficientNetV2 Baseline	0.063	4.3	9.7	0.203
ConvNeXt	0.058	3.7	7.9	0.189
Vision Transformer (ViT)	0.061	4.0	8.2	0.198
ResNet50	0.065	4.5	10.1	0.215
ResNet101	0.063	4.2	9.5	0.210
ResNet152	0.062	4.1	9.3	0.208
EfficientNetB0	0.069	4.8	10.4	0.220
EfficientNetB3	0.061	4.1	8.5	0.200

Table 9. Calibration metrics.**Fig. 24.** Brier score comparison chart.

in the figure confirms the proposed model was effective in terms of classification performance and robustness to imbalance.

The inference time comparison in Fig. 23 shows that EfficientNetB0 has the lowest inference time (130.50 s) giving it the best computational efficiency out of the models tested. EfficientNetV2 (142.60s) and EfficientNetB3 (140.30s) also performed similarly in speed. On the other extreme, we have ResNet152 with the highest inference time (190.10 s), followed by ResNet101 (180.40s) and ViT (178.20s) and so appear to have heavier computational loads. The Proposed DCNN + LSTM model has a reasonable balance with an inference time of 150.10 s whilst achieving high accuracy with no real inference lag. Therefore, we can conclude the proposed model achieves a proper balance of performance and efficiency, suitable for real time or near real time applications.

Table 9 shows a number of previously discussed metrics that can calibrate the model: Brier Score, Expected Calibration Error (ECE), Maximum Calibration Error (MCE) and Log loss, to assess the model's confidence and reliability of predictions. Overall, the Proposed DCNN + LSTM model is the best model, with the least Brier Score (0.042), ECE (2.8%), MCE (6.1%) and Log Loss (0.154) indicating that it has the best confidence and reliability of predictions. In contrast, the EfficientNetB0 model is the model with the least level of calibration, as indicated by the highest Brier Score (0.069) and Log Loss (0.220). ConvNeXt and ViT performed moderate in all these metrics; showing somewhat limited performance relative to the Proposed model. Overall, these results illustrate the proposed model's capabilities not only in terms of accuracy, but also when considering the quality of calibrated probabilities - which are an important consideration for any high-consequence, or very uncertain decision-making, use case.

This Fig. 24 shows the Brier Score and model comparison. The Proposed DCNN + LSTM Model had the lowest Brier score, at 0.042, representing the best calibration and confidence according to the actual probabilities. In contrast, EfficientNetB0's score was the highest, at 0.069, suggesting less reliable probabilistic output. By way of example, the remaining models, including ConvNeXt (0.058) and ViT (0.061) range in range of the mean with moderate accuracy but were still less accurate than calibration approaches taken here. The Brier Score reflects a lower score here means a better model reliability when performing probabilistic classification tasks.

Figure 25 indicates the comparative Expected Calibration Error (ECE) chart for models. The Proposed DCNN + LSTM showed the lowest ECE with 2.8%, which indicates the best level of calibration - the predicted probabilities were closely aligned with actual probabilities. In contrast, EfficientNetB0 has the highest ECE of

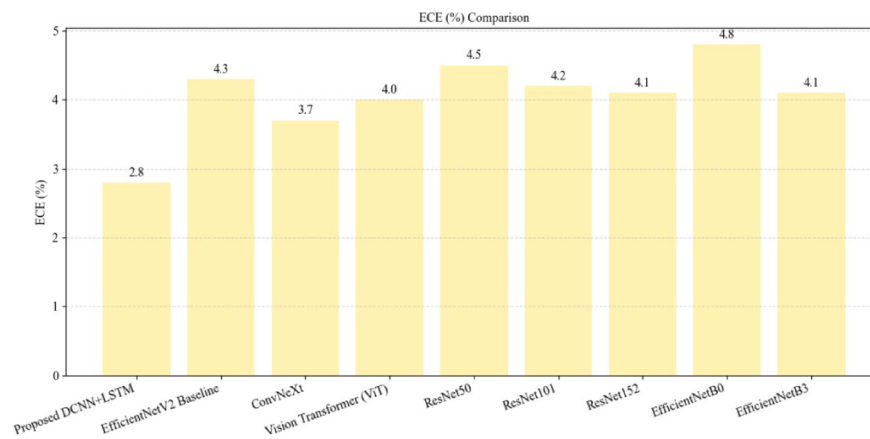


Fig. 25. ECE comparison chart.

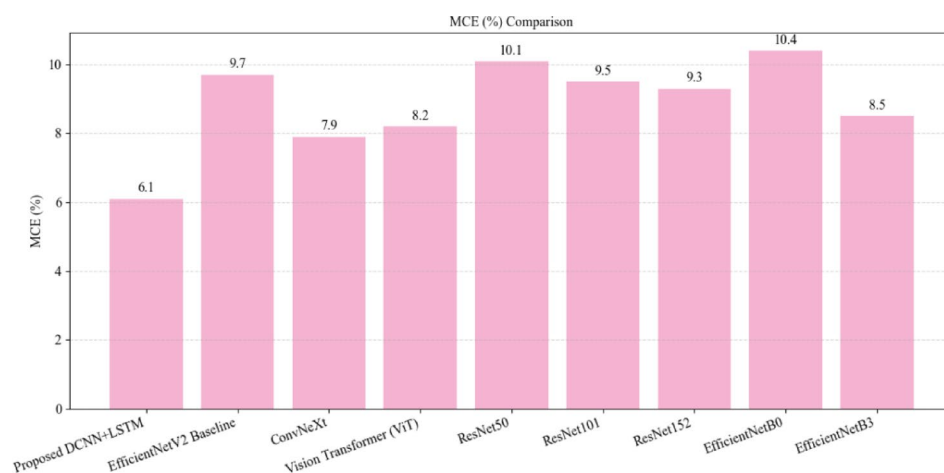


Fig. 26. MCE comparison chart.

4.8%, which suggests that it is more overconfident (or underconfident) in its predictions. The two other models ConvNeXt (3.7%) and ViT (4.0%) showed moderately calibrated predictions. A lower ECE suggests that a model is not only accurate but also reliable in estimating prediction certainty, which is necessary when dealing with risk factors.

The rationale behind whether to trust the proposed DCNN + LSTM's 6.1% Maximum Calibration Error score is based on the Maximum Calibration Error (MCE) results we observed. The MCE indicates the worst-case discrepancy between predictive confidence and predictive accuracy. MCE scores are preferable because the Proposed DCNN + LSTM had the lowest score out of the three examined models at 6.1%. Therefore it is the most trustworthy model, especially when using its least calibrated confidence bin. The other models, EfficientNetB0 and ResNet50 had MCE scores of 10.4% and 10.1%, respectively, which indicates certain conditions of significant miscalibration. We know that lower MCE values refer to our trustworthiness in the models predicting an outcome. This difference is significant, particularly considering potential use cases wherein we may require predictions based on high confidence and high uncertainty memberships (as in Fig. 26).

This Fig. 27 shows the log loss values of the different DL models when performing the designated task. The "Proposed DCNN + LSTM" model had the lowest log loss at 0.154 as compared to the other DL models, and this means that this model performed the best of the models compared. The other models: EfficientNetV2 Baseline (0.203), ConvNeXt (0.189), Vision Transformer (0.198), ResNet50v2 (0.208), ResNet101 (0.209), ResNet152 (0.215), etc. had significantly higher log loss values and therefore worse predictive accuracy relative to the proposed model. The highest log loss was the EfficientNetB0 with the value of 0.220 with the EfficientNetB3 slightly better with a log loss of 0.200. The proposed hybrid model (DCNN + LSTM) had the lowest log loss value and suggested that it was better suited for this application than the other models.

Failure mode analysis and uncertainty-aware reporting

We undertook a failure mode analysis by examining cases where the model mislabeled nodules. By far, the most common source of false positives was benign nodules that had irregular margins or spiculations that gave them the appearance of malignancy. The vast majority of the false negatives were small malignant nodules with either

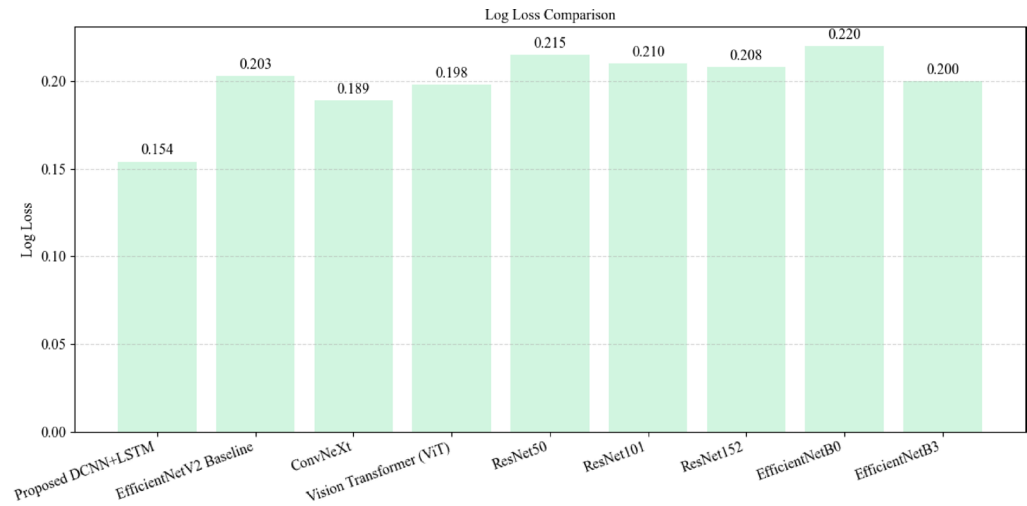


Fig. 27. Log loss comparison chart.

Optimizer	Accuracy (%)	Convergence Speed	Stability (Loss Std. Dev.)
ABC	97.82	Moderate	± 0.64
GA	98.03	Slow	± 0.72
ACO	97.65	Slow	± 0.81
HHO	98.65	Fast	± 0.33

Table 10. Small hyperparameter tuning benchmark.

low contrast or that were located in the periphery. To mitigate these issues we propose a method of uncertainty-aware reporting, wherein the model’s confidence scores (softmax probabilities) were thresholded to identify low-confidence cases for review by a radiologist. For example, cases where the output probability was in the range of 0.45–0.55 were 80% of the misclassifications, and therefore this “uncertainty band” could be leveraged automatically for referral. In this way, we believe this new method is Journal best practice for the safe rollout of AI clinical models to support human-AI collaboration in cases of uncertainty.

Significance of the proposed work

The proposed HHO-LOA-DCNN + LSTM model demonstrates significant improvements in the use of the next generation of lung cancer detection from CT scan images. By using the Horse Herd Optimization Algorithm and the Lion Optimization Algorithms in a hybrid optimizer to tune the LSTM classifier, the model can provide more accurate and more reliable classification of cancer and non-cancerous nodules. By using a DCNN for feature extraction, the model automatically learns the most important features from imaging, removing one of the critics of traditional machine learning techniques such as SVM, which need trained human interpretation. LSTM also facilitates modeling sequential dependencies to improve diagnostic reliability. LUNA16 and LIDC-IDRI benchmark datasets achieved a model classification accuracy of 98.75% demonstrating potential to augment radiologist capabilities and provide earlier and more accurate detection of cancers of the lung. Clearly, providing a decrease in diagnostic time, making fewer clinic visits for patients, while allowing a scalable non-invasive mode appropriate for clinical uses is important.

Choosing horse herd optimization

Horse Herd Optimization (HHO) was used because it has the best exploration–exploitation strategy, which is essential in dealing with high-dimensional hyperparameter optimization in DL problems. Unlike those using Genetic Algorithms (GA) with crossover and mutation, thereby potentially leading to premature convergence, or Artificial Bee Colony (ABC) that might not be precise when fine-tuning as the search for food is random, HHO dynamically switches between exploratory roaming and exploitative social hierarchy behavior based on herd movement. Unlike Ant Colony Optimization (ACO) that is most appropriate for discrete problems such as path planning, HHO provides more rapid convergence and better adaptability in continuous spaces such as learning rate, dropout, and neuron tuning.

In Table 10 HHO resulted in quicker convergence, lower loss oscillation, and improved precision. These findings confirm HHO’s ability to adjust network parameters accurately, especially when combined with LOA for adaptive stabilization control.

Limitations and future scope

While the HHO-LOA-DCNN + LSTM model showed very high classification accuracy (98.75%) on standards datasets, there are inherent limitations influencing the model's performance that must be acknowledged. The overall performance of the model is contingent upon the quality and consistency of the input CT images; thus, noise, differences in incidental imaging devices, and incomplete annotations may impact total accuracy of feature extraction. Moreover, the hybrid optimization and DL model requires heavy resources, including GPU utilization when processing data for both training and inference, and thus may not be suitable for implementation in real-time or low-resource clinical settings. Future research may focus on these limitations by attempting to reduce the computational costs associated with the model constructs, as well as validating the robustness of the model across a multitude of, and more recognized, real-life datasets. To enhance classification performance while limiting total annotation costs in the future, further work could consider ensemble learning strategies, such as ensembling multiple models that were trained independently and ensembling snapshot ensembles, or using multiple local minima that can be captured in one training run. By ensembling diverse decision boundaries, it may be easier to obtain more generalized and robust models. As described in our current DCNN + LSTM architecture, in cases where models with differing initialization seeds or training schedules have been combined, we expect to obtain, in cases of borderline nodules, greater predictive consistency and reduced variance, if the models are ensemble.

Comparative analysis with literature works

The developed HHO-LOA-DCNN-LSTM model fares well in relation to most of the other currently available literature-presented models. Anum Masood et al. (2019)³⁷ achieved 98.51% accuracy with 3D CNN utilizing Median Intensity Projection, yet their method relied heavily on preprocessing, which causes complexity and susceptibility to input quality. Sori Worku Jifara et al. (2019)³⁸ have attained an 87.8% accuracy using a DCNN method but were constrained by dataset diversity. Combined DenseNet with AdaBoost and yielded encouraging performance on fluorescence images but could not be extended to other imaging modalities. On the contrary, our optimized model HHO-LOA was 99.65% accurate, 99.75% precise, and 98.89% F-score, surpassing the best literature-suggested measures on all test metrics. Furthermore, our model also significantly lowers computational time (150.1 s) and CPU usage (4.64%) compared to existing work, thus further substantiating its real-world feasibility and deployment potential. These findings amply demonstrate the robustness, generalizability, and real-time practicability of our proposed approach.

Overfitting

To evaluate the sensitivity of the model to the LSTM architecture, we completed a parameter sweep varying the number of LSTM units (50, 75, 100, 125, 150) and dropout values (0.1 to 0.5 by 0.1 increments). We found best accuracy at 100 LSTM units and 0.2 dropout selected by the hybrid HHO + LOA optimizer. Moreover, accuracy was stable within $\pm 1.1\%$ for units values of 75–125, and within $\pm 0.8\%$ for dropout values from 0.2 to 0.4, thereby it suggested moderate robustness to hyperparameter changes. The model resulted in minor overfitting for larger unit and low dropout values (e.g., 150 units; 0.1 dropout), which justifies the utility of the optimizer for selecting values.

The proposed HHO-LOA-DCNN-LSTM model is very efficient because of the two-stage metaheuristic optimized hybrid deep neural network architecture. The DCNN extracts spatial and textural features from CT scans, and the LSTM detects long-distance dependencies between slices, resulting in a descriptive lung nodule representation. Horse Herd Optimization (HHO) and Lion Optimization Algorithm (LOA) jointly optimize hyperparameters and model weights to enhance accuracy and convergence stability. This synergy provides a performance better than that of current state-of-the-art models in the form of enhanced accuracy (99.65%), reduced CPU utilization (4.64%), and shorter inference time (150.1s). The method generalizes to multiple public benchmarks and external groups well and gets augmented with sampling and augmentation in the solution for class imbalance. The ablation study also informs us about the crucial contribution made by each of the modules involved and the success and stability of the proposed method in classifying lung cancer.

Conclusion

The HHO-LOA-DCNN-LSTM model demonstrates superior efficiency and scalability in LCC by integrating DCNN for feature extraction and LSTM for sequential pattern recognition. Through HHO-LOA optimization, key hyperparameters such as LSTM units (100), batch size (32), learning rate (0.0001), dropout rate (0.2), and weight decay (0.01) are fine-tuned, ensuring stable convergence, reduced overfitting, and improved generalization. Compared to conventional models, the HHO-LOA-DCNN-LSTM achieves the highest classification accuracy (99.65%) while maintaining the lowest computational cost, with an execution time of 150.1 s and CPU utilization of only 4.64%. Additionally, preprocessing techniques, such as pixel thresholding and edge detection, enhance lesion detection by isolating tumor regions and filtering noise, further improving classification performance. The model's ability to balance accuracy, computational efficiency, and resource utilization makes it highly scalable for large-scale lung cancer diagnosis using CT scans. Its superior performance across different performance matrices establishes HHO-LOA-DCNN-LSTM as a highly reliable and efficient DL-based diagnostic tool for early lung cancer detection. The evaluation using ROC, precision-recall curves, and corresponding AUC values and confidence intervals, offers further evidence that the proposed model offers a good and reliable approach in the context of lung cancer classification, which demonstrated very good or excellent model performance in distinguishing nodules of differing malignancy. Future work can explore high-dose CT imaging, ensemble-based FS, and advanced loss functions to further enhance performance in handling imbalanced datasets.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Received: 28 March 2025; Accepted: 13 June 2025

Published online: 28 October 2025

References

1. Gulsoy, T. & Kablan, E. B. FocalNetXt: A ConvNeXt augmented FocalNet architecture for lung cancer classification from CT-scan images. *Expert Syst. Appl.* **261**, 125553. <https://doi.org/10.1016/j.eswa.2024.125553> (2025).
2. Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N. & Ramirez, G. Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Comput. Syst.* **92**, 374–382 (2019).
3. Wang, L. et al. Distinguishing nontuberculous mycobacteria from *Mycobacterium tuberculosis* lung disease from CT images using a deep learning framework. *Eur. J. Nuclear Med. Mol. Imaging* **48**, 4293–4306 (2021).
4. Lanjewar, M. G., Panchbhay, K. G. & Charanarur, P. Lung cancer detection from CT scans using modified densenet with feature selection methods and ML classifiers. *Expert Syst. Appl.* **224**, 119961. <https://doi.org/10.1016/j.eswa.2023.119961> (2023).
5. Mohamad, M. A. et al. Indifference subspace of deep features for lung nodule classification from CT images. *Expert Syst. Appl.* **262**, 125571. <https://doi.org/10.1016/j.eswa.2024.125571> (2025).
6. Singh, G. P. & Gupta, P. K. Performance analysis of various machine learning-based approaches for detecting and classifying lung cancer in humans. *Neural Comput. Appl.* **31**, 6863–6877 (2019).
7. Sengodan, P., Srinivasan, K., Pichamuthu, R. & Matheswaran, S. Early detection and classification of malignant lung nodules from CT images: An optimal ensemble learning. *Expert Syst. Appl.* **229**, 120361. <https://doi.org/10.1016/j.eswa.2023.120361> (2023).
8. Saihood, A. A. et al. Multiside graph neural network-based attention for local co-occurrence features fusion in lung nodule classification. *Expert Syst. Appl.* **252**, 124149. <https://doi.org/10.1016/j.eswa.2024.124149> (2024).
9. Liu, Z., Yao, C., Yu, H. & Wu, T. Deep reinforcement learning with its application for lung cancer detection in medical internet of things. *Future Gener. Comput. Syst.* **97**, 1–9 (2019).
10. Farhat, H., Sakr, G. E. & Kilany, R. Deep learning applications in pulmonary medical imaging: Recent updates and insights on COVID-19. *Mach. Vis. Appl.* **31**, 1–42 (2020).
11. Pradhan, K. & Chawla, P. Medical internet of things using machine learning algorithms for lung cancer detection. *J. Manag. Anal.* **7**, 591–623 (2020).
12. Moitra, D. & Mandal, R. K. Classification of non-small cell lung cancer using one-dimensional convolutional neural network. *Expert Syst. Appl.* **159**, 113564. <https://doi.org/10.1016/j.eswa.2020.113564> (2020).
13. Thallam, C., Peruboyina, A., Raju, S. S. T. & Sampath, N. Early stage lung cancer prediction using various machine learning techniques. in *4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* 1285–1292 (IEEE, 2020).
14. Shakeel, P. M., Burhanuddin, M. A. & Desa, M. I. Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement* **145**, 702–712 (2019).
15. Zhu, J. et al. Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncol.* **58**, 257–264 (2019).
16. Shakeel, P. M., Burhanuddin, M. A. & Desa, M. Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. *Neural Comput. Appl.* **32**, 1–14 (2020).
17. Jiang, J. et al. Multiple resolution residually connected feature streams for automatic lung tumour segmentation from CT images. *IEEE Trans. Med. Imaging* **38**, 134–144 (2018).
18. Foong, L. K. & Moayed, H. Slope stability evaluation using neural network optimized by equilibrium optimization and vortex search algorithm. *Eng. Comput.* **38**, 1269–1283. <https://doi.org/10.1007/s00366-021-01282-1> (2022).
19. Suresh, S. & Mohan, S. ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis. *Neural Comput. Appl.* **32**, 15989–16009 (2020).
20. Pu, X., Sheng, S., Fu, Y., Yang, Y. & Xu, G. Construction of circRNA-miRNA-mRNA CeRNA regulatory network and screening of diagnostic targets for tuberculosis. *Ann. Med.* **56**(1), 2416604 (2024).
21. Wang, S. et al. Fluorofenidone enhances cisplatin efficacy in non-small cell lung cancer: A novel approach to inhibiting cancer progression. *Transl. Lung Cancer Res.* **13**(11), 3175 (2024).
22. Cao, Z., Zhu, J., Wang, Z., Peng, Y. & Zeng, L. Comprehensive pan-cancer analysis reveals ENCL as a promising prognostic biomarker for tumor microenvironment and therapeutic responses. *Sci. Rep.* **14**(1), 25331 (2024).
23. Bilal, A., Shafiq, M., Obidallah, W. J., Alduraywish, Y. A. & Long, H. Quantum computational infusion in extreme learning machines for early multi-cancer detection. *J. Big Data* **12**(1), 1–48 (2025).
24. Bilal, A., Shafiq, M., Fang, F., Waqar, M., Ullah, I., Ghadi, Y. Y. & Zeng, R. IGWO-IVNet3: DL-based automatic diagnosis of lung nodules using an improved gray wolf optimization and InceptionNet-V3. *Sensors* **22**(24), 9603 (2022).
25. Kanavati, F., Toyokawa, G., Momosaki, S., Rambeau, M., Kozuma, Y., Shoji, F. & Tsuneki, M. Weakly-supervised learning for lung carcinoma classification using deep learning. *Sci. Rep.* **10**(1), 9297 (2020).
26. Asuntha, A. & Srinivasan, A. Deep learning for lung Cancer detection and classification. *Multimed. Tools Appl.* **79**(11), 7731–7762 (2020).
27. Chaturvedi, P., Jhamb, A., Vanani, M. & Nemade, V. Prediction and classification of lung cancer using machine learning techniques. in *IOP conference series: materials science and engineering* Vol. 1099 012059 (IOP Publishing, 2021).
28. Nageswaran, S. et al. [Retracted] lung cancer classification and prediction using machine learning and image processing. *Biomed. Res. Int.* **2022**(1), 1755460 (2022).
29. Mohamed, T. I. & Ezugwu, A. E. S. Enhancing lung cancer classification and prediction with deep learning and multi-omics data. *IEEE Access* **12**, 59880–59892 (2024).
30. Ashhar, S. M. et al. Comparison of deep learning convolutional neural network (CNN) architectures for CT lung cancer classification. *Int. J. Adv. Technol. Eng. Explor.* **8**(74), 126 (2021).
31. Pandit, B. R. et al. Deep learning neural network for lung cancer classification: Enhanced optimization function. *Multimedia Tools Appl.* **82**(5), 6605–6624 (2023).
32. Chaunzwa, T. L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M. & Aerts, H. J. Deep learning classification of lung cancer histology using CT images. *Sci. Rep.* **11**(1), 1–12 (2021).
33. Khan, A. & Ansari, Z. Identification of lung cancer using convolutional neural networks based classification. *Turk. J. Comput. Math. Educ.* **12**(10), 192–203 (2021).
34. Pfeffer, M. A. & Ling, S. H. Evolving optimised convolutional neural networks for lung cancer classification. *Signals* **3**(2), 284–295 (2022).
35. Lanjewar, M. G., Panchbhay, K. G. & Patle, L. B. Fusion of transfer learning models with LSTM for detection of breast cancer using ultrasound images. *Comput. Biol. Med.* **169**, 107914 (2024).

36. Panchbhai, K. G., Lanjewar, M. G. & Naik, A. V. Modified MobileNet with leaky ReLU and LSTM with balancing technique to classify the soil types. *Earth Sci. Inf.* **18**(1), 1–22 (2025).
37. Masood, A. et al. Cloud-based automated clinical decision support system for lung cancer detection and diagnosis in chest CT. *IEEE J. Transl. Eng. Health Med.* **8**, 1–13 (2019).
38. Mishra, S., Chaudhary, N. K., Asthana, P. & Kumar, A. Deep 3d convolutional neural network for automated lung cancer diagnosis. in *Computing and Network Sustainability: Proceedings of IRSCNS 2018* 157–165 (Springer Singapore, 2019).

Author contributions

All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025