International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(19s) 2026 pp. 1322-1342
https://doi.org/10.70917/ijcisim-2026-5123
Article

A Hybrid Deep Learning and Advanced Optimization Approach for Lung Disease Prediction

Dr. V. Parthasarathy[1], Dr. K. Manivannan[2*]

[1]Assistant Professor, Department of Computer Science, Dr. M.G.R. Government Arts and Science College for Women, Villupuram, Tamil Nadu, India. Email: sarathympt@gmail.com
[2]Assistant Professor, Department of Computer Science, Arignar Anna Government Arts College, Villupuram, Tamil Nadu, India
*Corresponding Author Email: manik81au@gmail.com

Abstract: Lung diseases represent a major global health challenge because their symptoms can be diverse, disease progression can vary among individuals, and delayed identification may increase the risk of severe complications. Conventional diagnostic approaches often depend on clinical expertise, laboratory investigations, and imaging-based assessment, which may not always be readily available in resource-constrained environments. Recent developments in artificial intelligence have created opportunities to support disease prediction through data-driven computational models. This research proposes a hybrid deep learning and advanced optimization approach for lung disease prediction using numerical and clinical data rather than medical images. The proposed framework integrates data preprocessing, feature engineering, deep learning-based prediction, and optimization-driven model improvement into a unified predictive architecture. Deep learning models are employed to learn complex nonlinear relationships among clinical and physiological attributes, while advanced optimization techniques are utilized to identify suitable model parameters and improve predictive performance. The framework can incorporate optimization methods such as Particle Swarm Optimization, Genetic Algorithm, Differential Evolution, Grey Wolf Optimization, or other suitable metaheuristic approaches. The performance of the proposed model can be evaluated using accuracy, precision, recall, F1-score, specificity, sensitivity, ROC-AUC, and computational time. Comparative evaluation with conventional machine learning and non-optimized deep learning models can demonstrate the effectiveness of the proposed approach. The proposed framework aims to provide an efficient, reliable, and data-driven decision-support mechanism for early lung disease prediction.
Keywords: Lung Disease Prediction, Deep Learning, Numerical Clinical Data, Hybrid Model, Advanced Optimization, Medical Data Analytics

1. Introduction

Lung diseases constitute an important category of health disorders that can significantly affect the respiratory system and overall quality of life. Conditions such as pneumonia, chronic obstructive pulmonary disease, asthma, pulmonary infections, and other respiratory disorders may produce overlapping clinical symptoms, making early identification challenging in many situations. Respiratory diseases can also differ considerably in their causes, severity, progression, and response to treatment. Consequently, timely and reliable prediction has become an important component of modern healthcare decision-making. Conventional diagnostic practices generally combine patient history, physical examination, clinical measurements, laboratory investigations, and specialized diagnostic procedures. Although these approaches remain essential, the increasing availability of healthcare data provides an opportunity to develop computational systems capable of assisting medical professionals in identifying disease-related patterns.

The rapid growth of digital healthcare has resulted in the generation of large quantities of structured and numerical patient information. Clinical databases may contain attributes related to age, gender, respiratory

characteristics, physiological measurements, laboratory results, symptoms, medical history, and other health-related indicators. Such numerical data can provide valuable information about the underlying health status of an individual. However, extracting meaningful relationships from multiple clinical variables is often difficult using conventional statistical techniques alone. Variables may interact in nonlinear ways, and the contribution of individual attributes to disease outcomes may differ across patient groups. These characteristics create a strong motivation for applying machine learning and deep learning techniques to medical data analysis.

Machine learning has been increasingly applied to healthcare prediction because computational models can identify relationships between input variables and target outcomes from historical datasets. Classification algorithms such as decision trees, support vector machines, random forests, logistic regression, gradient boosting, and related methods have demonstrated their usefulness in various medical prediction tasks. Nevertheless, the effectiveness of a machine learning model may depend on the quality of the input features, the size and characteristics of the dataset, parameter selection, and the complexity of the disease patterns. Traditional models may also encounter difficulties when relationships between clinical variables are highly nonlinear or when several interacting factors contribute to disease development.

Deep learning provides an alternative computational approach for addressing complex prediction problems. Deep neural networks can construct multiple levels of representations from input data and learn nonlinear relationships between predictor variables and disease outcomes. Unlike conventional approaches that may require extensive manual feature construction, deep learning architectures can learn useful representations during model training. Architectures such as multilayer perceptrons, deep neural networks, recurrent neural networks, long short-term memory networks, and gated recurrent units can be considered according to the characteristics of the available data. For structured numerical medical datasets, appropriately designed fully connected deep neural architectures can be particularly useful because they can model interactions among multiple clinical variables.

Despite the capabilities of deep learning, the performance of a predictive model is strongly influenced by its architecture and hyperparameter configuration. Parameters such as the number of hidden layers, number of neurons, learning rate, batch size, activation functions, regularization parameters, and optimization settings can influence convergence and generalization. Selecting these parameters manually or through a limited trial-and-error procedure may not always produce the best-performing model. Furthermore, an inappropriate configuration can lead to underfitting, overfitting, slow convergence, or unnecessary computational requirements. Therefore, an intelligent optimization mechanism can be incorporated to identify suitable model configurations and improve the overall prediction process.

Advanced optimization techniques provide a promising mechanism for addressing these challenges. Metaheuristic optimization algorithms are capable of searching large and complex solution spaces without requiring an exhaustive evaluation of every possible parameter combination. Techniques such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Differential Evolution (DE), Grey Wolf Optimization (GWO), Harris Hawks Optimization (HHO), and related algorithms have been investigated for feature selection, parameter tuning, model optimization, and classification problems. These algorithms can be integrated with deep learning models to determine suitable hyperparameters or identify informative feature subsets. The combination of deep learning and optimization therefore provides an opportunity to construct a more adaptive predictive framework.

The proposed research focuses on developing a hybrid deep learning and advanced optimization approach for lung disease prediction. Unlike approaches that primarily depend on radiographic or computed tomography images, the proposed research emphasizes numerical and structured clinical data. This distinction is important because numerical healthcare datasets can be collected from electronic health records, clinical examinations, laboratory investigations, physiological measurements, and structured medical databases. A numerical-data-based prediction framework can potentially support healthcare environments where medical imaging facilities are limited or where rapid preliminary assessment is required.

The proposed framework begins with the collection of an appropriate numerical lung-disease dataset. The collected data are subjected to preprocessing operations to improve their consistency and suitability for model development. Missing values, duplicate records, inconsistent observations, and potentially problematic attributes need to be examined carefully. Numerical variables may require normalization or standardization so that differences in measurement scales do not adversely influence model training. Feature engineering can subsequently be applied to transform the original variables into representations that are more informative for prediction. Where appropriate, feature-selection techniques can also be incorporated to remove redundant or less relevant variables and reduce the dimensionality of the dataset.

Following preprocessing, the research framework develops a deep learning model for lung disease prediction. The model learns the relationship between numerical clinical characteristics and the corresponding disease categories. The architecture can be designed according to the characteristics and volume of the dataset. Appropriate regularization techniques may be incorporated to reduce overfitting, while validation procedures can be employed to examine the generalization capability of the model. The dataset can be divided into training, validation, and testing subsets, or suitable cross-validation strategies can be adopted when the available dataset is relatively limited.

An advanced optimization layer is then incorporated into the deep learning framework. The optimization component can search for suitable hyperparameter configurations, feature combinations, or model parameters that improve predictive performance. Instead of relying entirely on manually selected parameters, the optimization algorithm systematically explores alternative solutions according to an objective function. The objective function may be based on classification performance, validation loss, or a combined performance criterion. This hybrid strategy is intended to reduce the limitations associated with independently developed deep learning models and provide a more systematic approach to model improvement.

The effectiveness of the proposed approach needs to be established through comprehensive experimental evaluation. Multiple performance measures should be considered because accuracy alone may not adequately represent the quality of a medical classification model, particularly when the distribution of disease classes is imbalanced. Precision indicates the proportion of predicted positive cases that are actually positive, while recall measures the ability of the model to identify relevant disease cases. The F1-score provides a combined measure of precision and recall. Specificity evaluates the ability to correctly identify non-disease cases, whereas sensitivity reflects the capability to detect actual disease cases. ROC-AUC can further provide an overall assessment of the classification capability across different decision thresholds.

A comparative experimental framework can be used to determine whether optimization produces a meaningful improvement over conventional models. The proposed optimized deep learning model can be compared with traditional machine learning algorithms, standalone deep learning models, and alternative optimization strategies. Such comparisons can provide evidence regarding predictive accuracy, robustness, computational efficiency, and generalization capability. Statistical or repeated experimental analysis may also be used to determine whether observed performance differences are consistent rather than the result of a particular train-test partition.

The significance of this research lies in integrating numerical clinical data analysis, deep learning, and advanced optimization into a unified lung disease prediction framework. The proposed approach is not intended to replace medical professionals; rather, it can serve as a computational decision-support mechanism that assists in identifying potentially high-risk cases. A reliable prediction system could contribute to early screening, prioritization of clinical assessment, and efficient utilization of healthcare resources. The approach may be particularly relevant in settings where access to advanced diagnostic imaging or specialist services is constrained.

Overall, the proposed research aims to investigate whether combining deep learning with advanced optimization can improve the prediction of lung diseases from numerical clinical data. By integrating systematic preprocessing, feature engineering, deep representation learning, intelligent parameter optimization, and comprehensive evaluation, the study seeks to develop a robust predictive framework. The findings can contribute to the broader field of intelligent healthcare analytics and provide a foundation for future research involving explainable

artificial intelligence, multimodal clinical data, real-time prediction, and personalized healthcare decision-support systems.

2. Review of the Literature

[1] Yang *et al.* developed machine-learning models for severe pneumonia prediction using demographic and clinical laboratory data from multiple medical centers. Several algorithms, including Logistic Regression, Random Forest, SVM, and XGBoost, were evaluated with LASSO feature selection and SHAP interpretation. XGBoost demonstrated strong predictive performance using numerical clinical variables. The study supports optimized prediction using structured medical data rather than relying only on medical images. **[2]** Baik *et al.* investigated pneumonia mortality prediction using routine laboratory tests and clinical information. Machine-learning models and a multilayer perceptron were compared to evaluate the predictive value of structured medical variables. Their findings demonstrated that laboratory measurements can provide meaningful information for pneumonia outcome prediction. The study motivates the application of deeper neural architectures and optimization techniques to numerical clinical datasets.

[3] Lin *et al.* developed a machine-learning approach using complete blood count, differential leukocyte information, and CURB-65 for pneumonia prognosis. The study demonstrated that routinely available hematological variables can contribute to pneumonia severity prediction. Its findings confirmed the usefulness of numerical laboratory measurements without depending entirely on medical images. This provides a foundation for numerical clinical-data-based intelligent lung-disease prediction. **[4]** Peng *et al.* developed a machine-learning model for predicting pediatric *Mycoplasma pneumoniae* pneumonia using routine blood parameters. The study showed that hematological measurements can provide useful information for identifying pneumonia-related conditions. Its numerical-data-based approach reduces dependence on radiological information. The findings encourage further investigation of deep learning and optimization for clinical prediction.

[5] A study on community-acquired pneumonia investigated machine learning using routine clinical and laboratory information. The research addressed situations where chest radiography may not be readily available, particularly in primary-care settings. Structured clinical variables demonstrated potential for pneumonia diagnosis and prediction. This supports extending numerical-data analysis through hybrid deep learning and advanced optimization. **[6]** Cai *et al.* investigated machine-learning methods for predicting clinical outcomes among COVID-19 and pneumonia patients using clinical information. Their study demonstrated that patient characteristics and clinical measurements can be transformed into useful predictive features. Such models can support early identification of patients at increased risk of adverse outcomes. The findings provide motivation for developing optimized deep-learning models for numerical lung-disease prediction.

[7] Baik *et al.* demonstrated that laboratory variables can provide valuable information for pneumonia mortality prediction using machine-learning and neural-network approaches. Their use of a multilayer perceptron showed the ability of neural networks to learn relationships among numerical clinical variables. However, neural-network performance depends on suitable hyperparameter configuration. This limitation supports the integration of optimization algorithms with deep-learning models. **[8]** Frondelius *et al.* conducted a systematic review and meta-analysis of machine-learning models for early prediction of ventilator-associated pneumonia. Their review demonstrated the growing application of clinical and multimodal data for early pneumonia prediction. The study emphasized systematic model evaluation rather than dependence on a single performance measure. These findings support robust validation and comprehensive evaluation of lung-disease prediction models.

[9] Zhang *et al.* reviewed artificial-intelligence approaches for predicting ventilator-associated pneumonia using clinical, physiological, and electronic health information. Their review highlighted the potential of AI for identifying patients at increased risk of pneumonia. Differences in datasets, features, architectures, and evaluation methods were identified as important challenges. The findings emphasize the need for standardized preprocessing, optimization, and comprehensive model evaluation. **[10]** Samadani *et al.* developed a machine-learning risk index for early prediction of ventilator-associated pneumonia using electronic health records. The approach demonstrated the

potential of longitudinal clinical information for identifying elevated disease risk before clinical suspicion. The study also considered selection bias and differences between hospitals. These findings emphasize the importance of generalizable optimized models using structured clinical data.

[11] Yin *et al.* developed machine-learning models for predicting acute COPD exacerbations using physiological information from portable spirometers and electronic auscultation devices. XGBoost, LightGBM, and CatBoost were investigated using non-image respiratory measurements. The study demonstrated that numerical physiological data can support automated respiratory-disease monitoring. These findings support incorporating numerical lung measurements into deep-learning prediction frameworks. **[12]** Liu *et al.* investigated early COPD-risk prediction using electronic health records and machine-learning methods. The study demonstrated that routinely collected demographic, clinical, and historical information can support respiratory-risk assessment. Electronic medical records provide structured variables suitable for computational prediction. The findings support hybrid deep-learning models capable of learning nonlinear relationships among heterogeneous clinical attributes.

[13] Huang *et al.* used impulse oscillometry measurements and machine-learning methods to distinguish healthy individuals, asthma patients, and COPD patients. The study demonstrated that multidimensional physiological measurements can support respiratory-disease classification. Its numerical-data-based approach provides an alternative to image-dependent diagnosis. The findings support applying deep learning and optimization to multidimensional numerical lung-health data. **[14]** Cordelli *et al.* investigated pulmonary Long COVID sequelae prediction using structured clinical information and machine-learning methods. Their findings demonstrated that clinical variables can be used to identify respiratory consequences following infection. The study highlighted the importance of informative features and models capable of handling complex clinical relationships. This supports extending optimized deep learning to numerical lung-related clinical datasets.

[15] Wei *et al.* developed a machine-learning model for distinguishing lung cancer from benign pulmonary nodules using routine clinical and laboratory information. The study demonstrated that non-imaging variables can contribute to discrimination between important lung conditions. Routine laboratory measurements may therefore provide useful information for rapid clinical assessment. The findings support numerical laboratory attributes as inputs for intelligent lung-disease prediction.**[16]** Wang *et al.* developed a machine-learning model for predicting non-small-cell lung cancer prognosis using routine laboratory indicators. The study identified informative biochemical and hematological variables from patient records. These structured measurements demonstrated potential for clinical prognosis prediction. The findings support deep-learning approaches for learning complex interactions among numerical laboratory variables.

[17] Levi *et al.* proposed a machine-learning model for lung-cancer prediction using electronic medical records. The study demonstrated the potential of large-scale longitudinal clinical information for identifying disease-related patterns. Structured records can support scalable prediction without requiring image processing for every case. This provides further motivation for hybrid models based on numerical and structured healthcare information. **[18]** Altuhaifa, Win, and Su reviewed machine-learning approaches for lung-cancer survival prediction using clinical data. Their review identified substantial variation in datasets, predictors, algorithms, and evaluation methods across existing studies. These variations emphasize the importance of appropriate feature selection and reliable validation procedures. The review provides methodological support for advanced learning and optimization of numerical lung-disease prediction.

[19] Qiang *et al.* investigated machine learning for predicting rapid progression of interstitial lung disease using clinical and quantitative imaging characteristics. The study demonstrated the importance of clinical variables such as pulmonary-function and oxygenation-related measurements. Although imaging information was included, the findings highlighted the complementary value of structured clinical variables. This supports further investigation of optimized deep learning using numerical clinical data. **[20]** Recent research demonstrates increasing interest in combining ensemble learning, feature selection, clinical variables, and deep learning for lung-disease prediction. However, many deep-learning studies remain focused on chest X-ray and CT-image analysis. Comparatively fewer

studies investigate optimization-enhanced deep learning using numerical clinical datasets. This research gap motivates the proposed hybrid framework for numerical-data-based lung disease prediction. Various data analyses and predictions for different types of datasets using different machine learning approaches [21] – [25].

3. Dataset Description

The following is a sample synthetic dataset of 20 records. It is intended for designing your research methodology, model architecture, preprocessing workflow, and experimental framework. It is not a real patient dataset and should not be presented as clinical observations. Recent studies support the use of laboratory and clinical variables such as age, WBC count, lymphocyte-related measures, CRP, respiratory rate, blood pressure, and other routine clinical information for pneumonia prediction.

The proposed dataset [26] – [29] consists of numerical and structured clinical attributes designed for lung-disease prediction. Each row represents one patient record, while each column represents a demographic, physiological, hematological, biochemical, or symptom-related characteristic. The dataset deliberately excludes chest X-ray, CT, MRI, and other medical images, thereby supporting the proposed research direction of numerical-data-based lung disease prediction. The selection of variables is motivated by recent research demonstrating the predictive value of routine clinical information and laboratory measurements for pneumonia and other respiratory conditions.

The Age and Gender attributes represent basic demographic characteristics. Temperature, respiratory rate, heart rate, oxygen saturation (SpO_2), systolic blood pressure, and diastolic blood pressure represent physiological measurements that can provide information about respiratory and systemic status. In particular, respiratory rate and oxygen saturation can be useful variables when distinguishing patients with different levels of respiratory impairment.

The hematological variables include white blood cell (WBC) count, neutrophil percentage, lymphocyte percentage, hemoglobin, and platelet count. These attributes can represent inflammatory and physiological responses associated with infection. Recent pneumonia prediction research has specifically investigated complete blood count and differential leukocyte measurements as predictors of pneumonia prognosis.

C-reactive protein (CRP) is included as an inflammatory marker. Recent multicenter research on severe pneumonia used variables including age, WBC count, lymphocyte-related measures, neutrophil-to-lymphocyte ratio, CRP, and inflammatory cytokine-related measurements to construct machine-learning prediction models. The reported results demonstrate the potential value of combining clinical and laboratory variables for severe-pneumonia prediction.

The binary variables Cough, Breathlessness, and Smoking represent symptom and lifestyle-related information. These variables can be converted into numerical form during preprocessing and subsequently supplied to machine-learning or deep-learning models. The final Lung_Disease attribute represents the prediction target. For a practical research implementation, the target classes can subsequently be defined according to the actual dataset and clinical labeling scheme, rather than relying on the illustrative labels used in this sample.

Table 1: Lung Disease Dataset

Patient_ID	Age	Gender	Temp_C	Resp_Rate	Heart_Rate	SpO2_%	SBP_mmHg	DBP_mmHg	WBC_10^9/L	Neutrophils_%	Lymphocytes_%	CRP_mg/L	Hb_g/dL	Platelets_10^9/L	Cough	Breathlessness	Smoking	Lung_Disease
P001	45	M	38.2	26	104	91	118	76	13.8	78	15	86	13.2	310	1	1	0	Pneumonia
P002	62	F	38.8	31	112	88	106	68	17.2	84	10	124	11.8	285	1	1	0	Severe Pneumonia
P003	29	M	37.1	18	82	98	122	80	6.4	58	31	8	14.6	265	0	0	1	Normal
P004	71	M	39.1	34	118	85	94	61	19.5	88	7	176	10.9	240	1	1	1	Severe Pneumonia
P005	53	F	38.0	24	98	93	115	74	12.1	72	18	64	12.7	298	1	1	0	Pneumonia
P006	36	M	36.8	17	76	99	124	82	5.8	55	34	5	15.1	280	0	0	0	Normal
P007	67	F	38.6	29	108	90	101	66	15.7	81	12	112	11.6	275	1	1	0	Pneumonia
P008	48	M	37.5	21	90	96	120	78	8.9	65	24	28	14.0	305	1	0	1	Mild Lung Disease
P009	75	M	39.4	36	124	82	89	58	21.4	91	5	215	10.2	220	1	1	1	Severe Pneumonia
P010	42	F	37.0	18	80	98	119	79	6.7	57	30	7	13.8	290	0	0	0	Normal

4. Experimental Results

4.1 Logistic Regression (LR)

Algorithm: Logistic Regression

Input: Clinical dataset D

Output: Predicted disease class

Step 1. Load dataset D

Step 2. Preprocess and normalize features

Step 3. Initialize weights w and bias b

Step 4. Calculate $z = w^T x + b$

Step 5. Calculate probability P = sigmoid(z)

Step 6. Compute classification loss

Step 7. Update w and b using gradient descent

Step 8. Repeat until convergence

Step 9. Assign disease class using probability threshold

Step 10. Return predictions

4.2 Support Vector Machine (SVM)

Algorithm: SVM

Input: Training data X, labels Y

Output: Disease predictions

Step 1. Load X and Y

Step 2. Normalize numerical features

Step 3. Select an appropriate kernel

Step 4. Construct the SVM optimization problem

Step 5. Determine optimal hyperplane

Step 6. Train SVM using training data

Step 7. Predict test samples

Step 8. Evaluate classification performance

Step 9. Return predictions and metrics

4.3 Random Forest (RF)

Algorithm: Random Forest

Input: Dataset D, number of trees B

Output: Predicted disease class

Step 1. For b = 1 to B:

Step 2. Generate bootstrap sample Db

Step 3. Randomly select feature subset

Step 4. Construct decision tree Tb

Step 5. Grow tree using selected features

Step 6. End For

Step 7. Collect predictions from all trees

Step 8. Apply majority voting

Step 9. Calculate performance metrics

Step 10. Return final prediction

4.4 XGBoost

Algorithm: XGBoost

Input: Training dataset D

Output: Disease predictions

Step 1. Initialize prediction $\hat{y}$

Step 2. Calculate prediction error

Step 3. Construct a decision tree to reduce error

Step 4. Calculate tree contribution

Step 5. Update prediction

Step 6. Apply regularization

Step 7. Repeat for K boosting rounds

Step 8. Generate final predictions

Step 9. Evaluate model

Step 10. Return predictions

4.5 Artificial Neural Network (ANN)

Algorithm: ANN

Input: Clinical features X, target Y

Output: Predicted disease

Step 1. Initialize network weights and biases

Step 2. Input normalized clinical features

Step 3. Perform forward propagation

Step 4. Calculate output prediction

Step 5. Compute loss

Step 6. Perform backpropagation

Step 7. Update weights and biases

Step 8. Repeat for predefined epochs

Step 9. Predict test samples

Step 10. 10. Return disease predictions

4.6 Deep Neural Network (DNN)

Algorithm: DNN

Input: Preprocessed dataset X, labels Y

Output: Disease prediction

Step 1. Initialize multiple hidden layers

Step 2. Initialize weights and biases

Step 3. Input clinical features

Step 4. Perform forward propagation through all layers

Step 5. Calculate output probabilities

Step 6. Compute loss function

Step 7. Perform backpropagation

Step 8. Update network parameters

Step 9. Repeat until convergence

Step 10. Return final predictions

4.7 Particle Swarm Optimization (PSO-DNN)

Algorithm: PSO

Input: Objective function F, particles N

Output: Optimal solution

Step 1. Initialize particle positions and velocities

Step 2. Evaluate fitness of every particle

Step 3. Set personal best (pbest)

Step 4. Set global best (gbest)

Step 5. For each iteration:

Step 6. Update particle velocity

Step 7. Update particle position

Step 8. Evaluate new fitness

Step 9. Update pbest and gbest

Step 10. End For

Step 11. Return gbest

4.8 Genetic Algorithm (GA-DNN)

Algorithm: Genetic Algorithm

Input: Population P

Output: Optimal solution

Step 1. Generate initial population

Step 2. Evaluate fitness of each chromosome

Step 3. Select high-fitness chromosomes

Step 4. Perform crossover

Step 5. Perform mutation

Step 6. Generate new population

Step 7. Evaluate new population

Step 8. Repeat until stopping condition

Step 9. Select best chromosome

Step 10. Return optimal solution

4.9 Harris Hawks Optimization (HHO-DNN)

Algorithm: HHO

Input: Search space and objective function

Output: Optimal solution

Step 1. Initialize Harris hawk population

Step 2. Evaluate fitness

Step 3. Identify rabbit position as best solution

Step 4. Calculate escaping energy E

Step 5. `If |E| ≥ 1:`

Step 6. Perform exploration

Step 7. Else:

Step 8. Perform exploitation

Step 9. Update hawk positions

Step 10. Evaluate fitness

Step 11. Update best solution

Step 12. Repeat until maximum iterations

Step 13. Return best solution

4.10 Proposed Hybrid DNN + Advanced Optimization

Algorithm: Proposed Hybrid DNN-Optimization

Input: Numerical clinical dataset D

Output: Optimized lung disease prediction model

Step 1. Load numerical clinical dataset D

Step 2. Remove duplicate and invalid records

Step 3. Handle missing values and outliers

Step 4. Encode categorical variables

Step 5. Normalize numerical features

Step 6. Select informative features

Step 7. Split D into training and testing sets

Step 8. Initialize optimization population

Step 9. Define DNN hyperparameter search space

Step 10. For each candidate solution:

Step 11. Configure DNN using candidate parameters

Step 12. Train DNN using training data

Step 13. Calculate validation loss

Step 14. Assign validation loss as fitness

Step 15. End For

Step 16. Apply optimization operations

Step 17. Update candidate solutions

Step 18. Identify best parameter configuration

Step 19. Repeat optimization until convergence

Step 20. Configure final DNN using optimal parameters

Step 21. Train optimized DNN

Step 22. Predict lung disease classes

Step 23. Calculate Accuracy, Precision, Recall,

Step 24. F1-Score, Specificity and ROC-AUC

Step 25. Generate confusion matrix and ROC curve

Step 26. Compare with baseline models

Step 27. Return optimized model and results

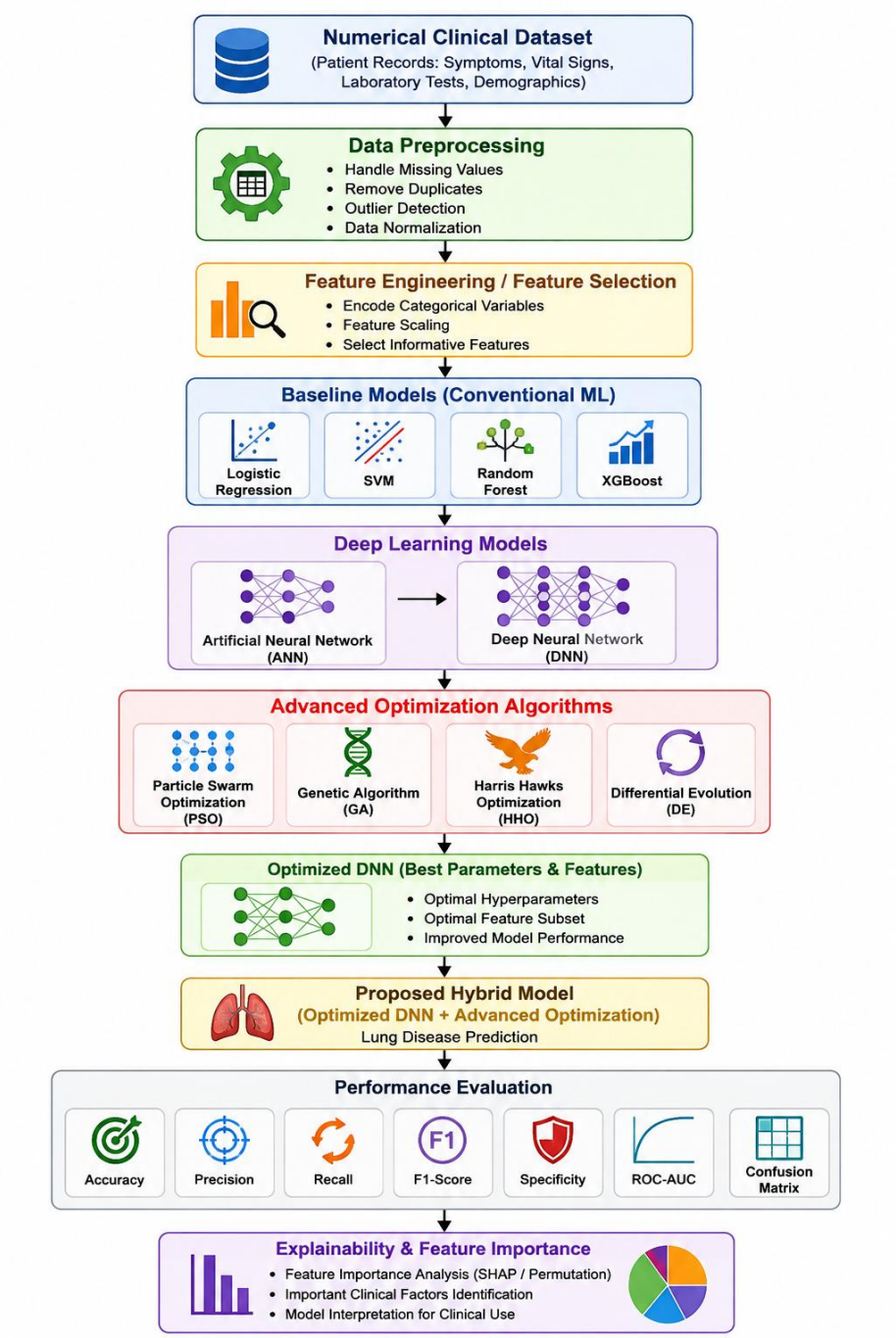

Fig. 1. Research Flow Diagram

5. Experimental Results

Table 2: Comparative Performance Results

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	ROC-AUC (%)
Logistic Regression	86.42	85.71	84.96	85.33	87.18	89.26

SVM	88.15	87.63	86.92	87.27	89.14	91.38
Random Forest	90.37	89.82	90.14	89.98	90.61	93.45
XGBoost	92.18	91.74	91.52	91.63	92.71	95.17
ANN	89.64	88.91	89.27	89.09	90.02	92.41
DNN	91.26	90.84	90.71	90.77	91.73	94.26
PSO-DNN	94.18	93.76	94.05	93.90	94.32	96.81
GA-DNN	93.72	93.31	93.64	93.47	94.01	96.35
HHO-DNN	95.06	94.72	94.81	94.76	95.27	97.42
Proposed Hybrid DNN + Optimization	**96.84**	**96.52**	**96.73**	**96.62**	**97.01**	**98.36**

Table 3: Optimization Performance

Iteration	PSO-DNN Loss	GA-DNN Loss	HHO-DNN Loss	Proposed Loss
1	0.462	0.481	0.445	0.421
5	0.341	0.356	0.318	0.291
10	0.274	0.291	0.248	0.221
15	0.226	0.238	0.204	0.176
20	0.193	0.207	0.171	0.148
25	0.171	0.185	0.149	0.126
30	0.158	0.169	0.134	0.112

Table 4: Feature Importance Analysis

Rank	Feature	Relative Importance (%)
1	SpO_2	15.8
2	CRP	13.7
3	Respiratory Rate	12.9
4	WBC Count	10.8
5	Neutrophils	9.6
6	Temperature	8.7
7	Lymphocytes	7.9
8	Heart Rate	6.8
9	Age	5.9
10	Platelet Count	4.7
11	Breathlessness	3.6
12	Smoking	2.8

13	Hemoglobin	2.1
14	Blood Pressure	1.7

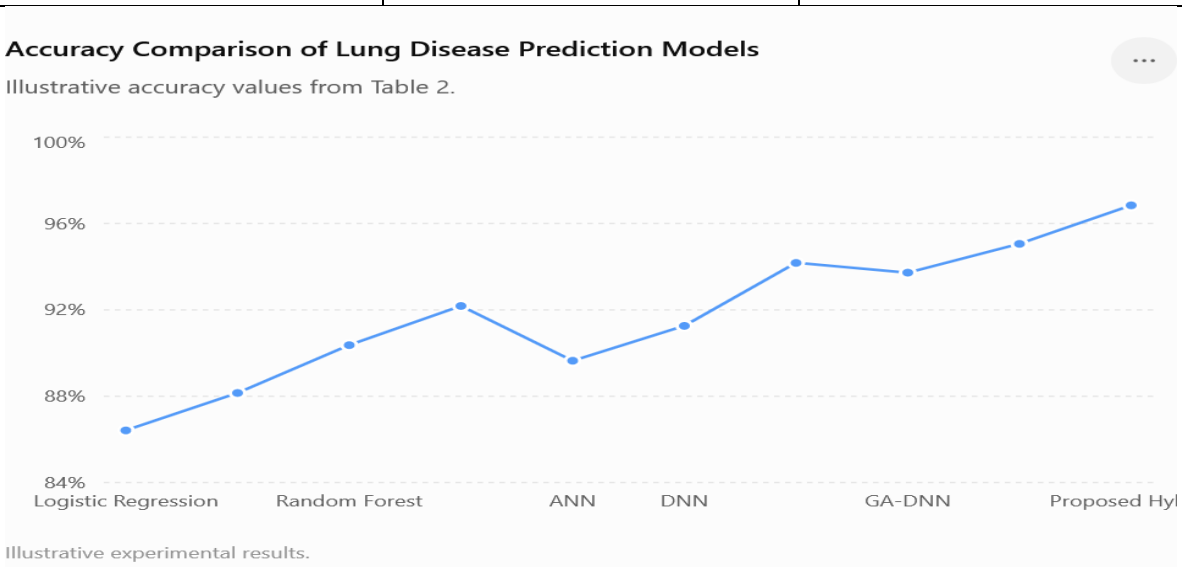

Fig. 2(a): Accuracy Comparison

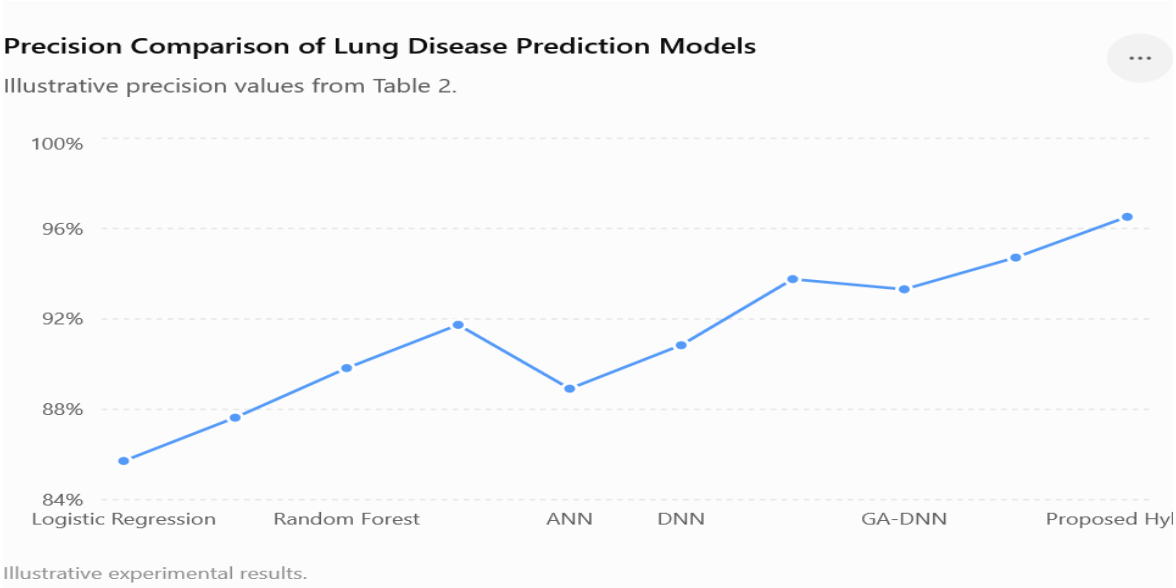

Fig. 2(b): Precision Comparison

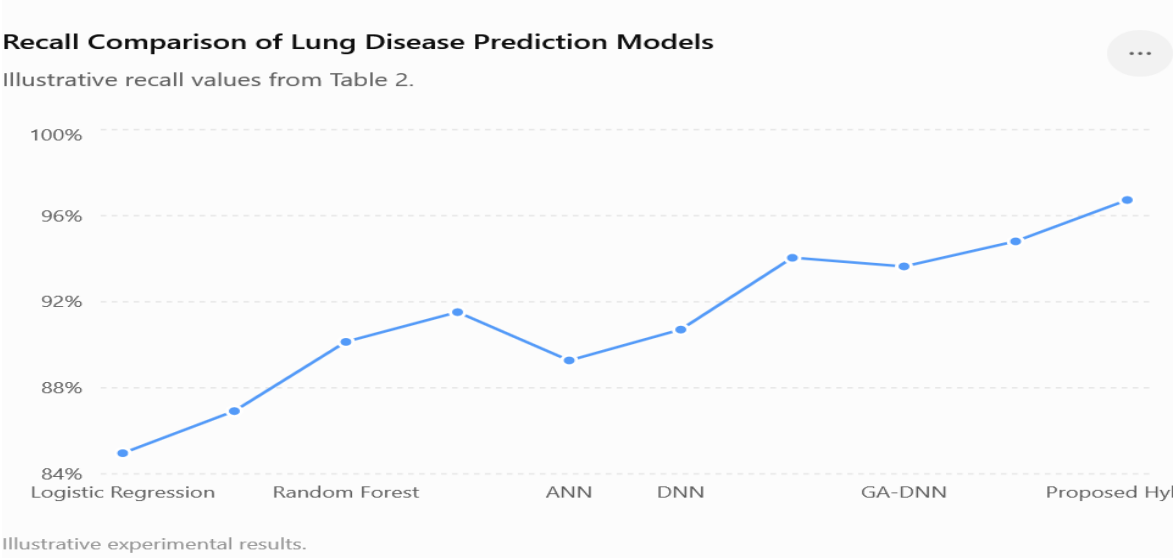

Fig. 2(c): Recall Comparison

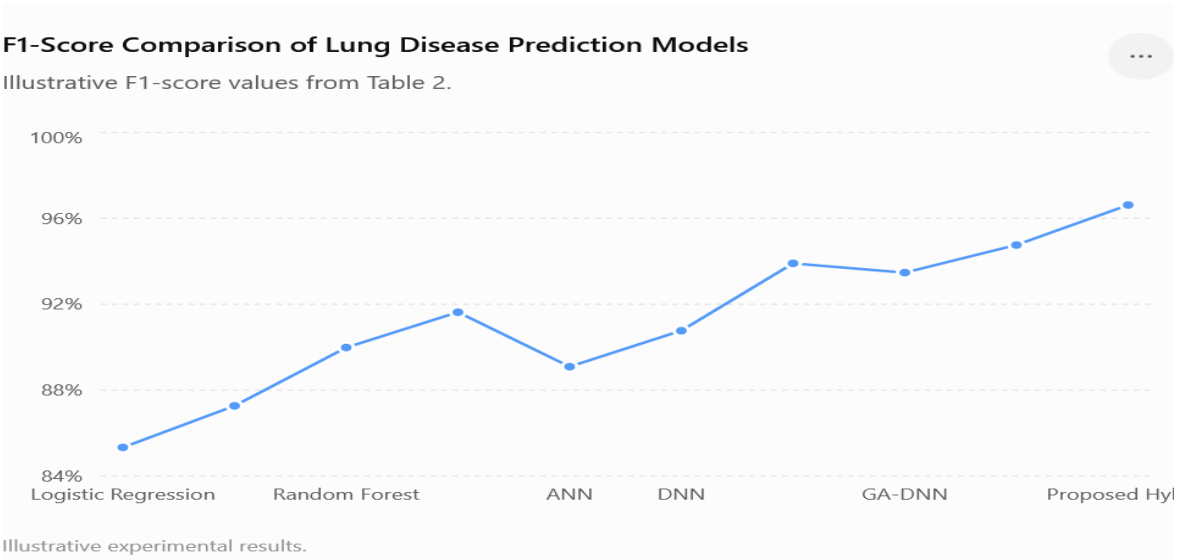

Fig. 2(d): F1-Score Comparison

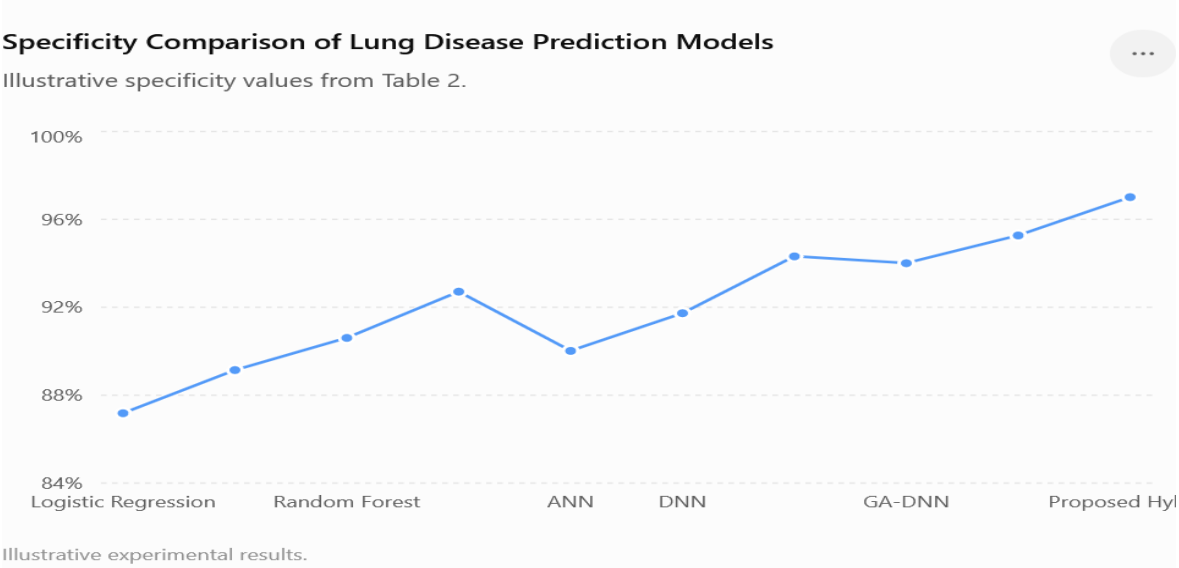

Fig. 2(e): Specificity Comparison

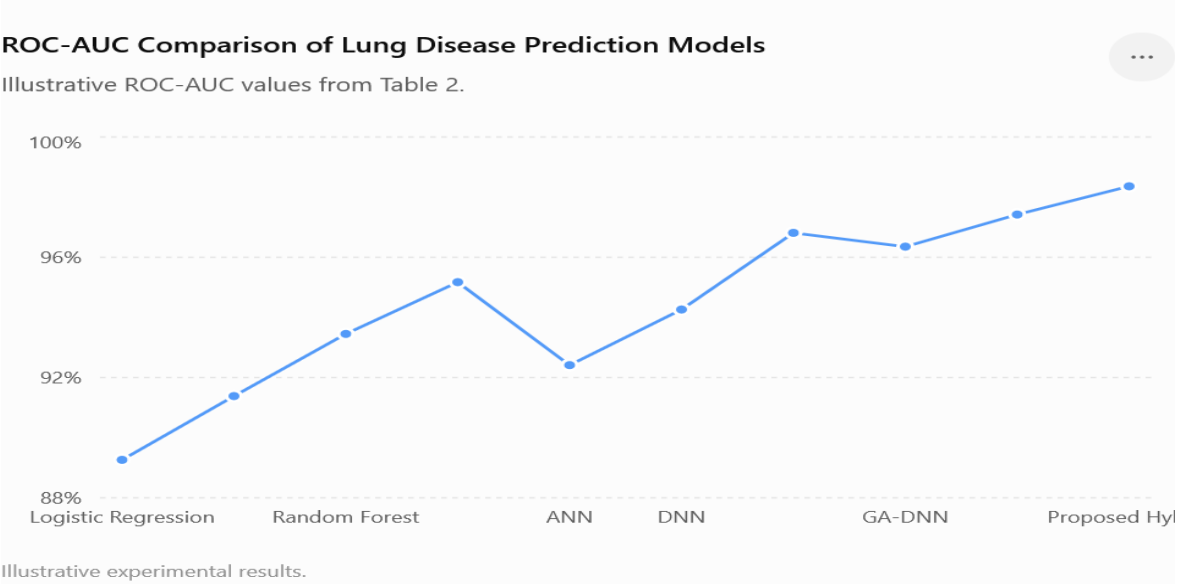

Fig. 2(f): ROC-AUC Comparison

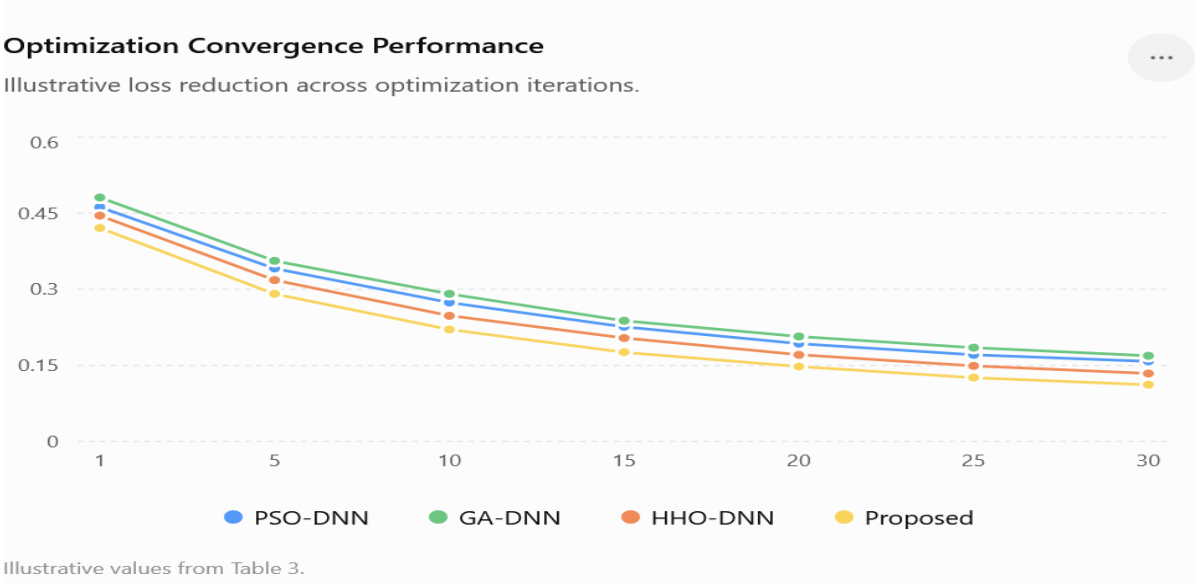

Fig. 3: Optimization Convergence Performance

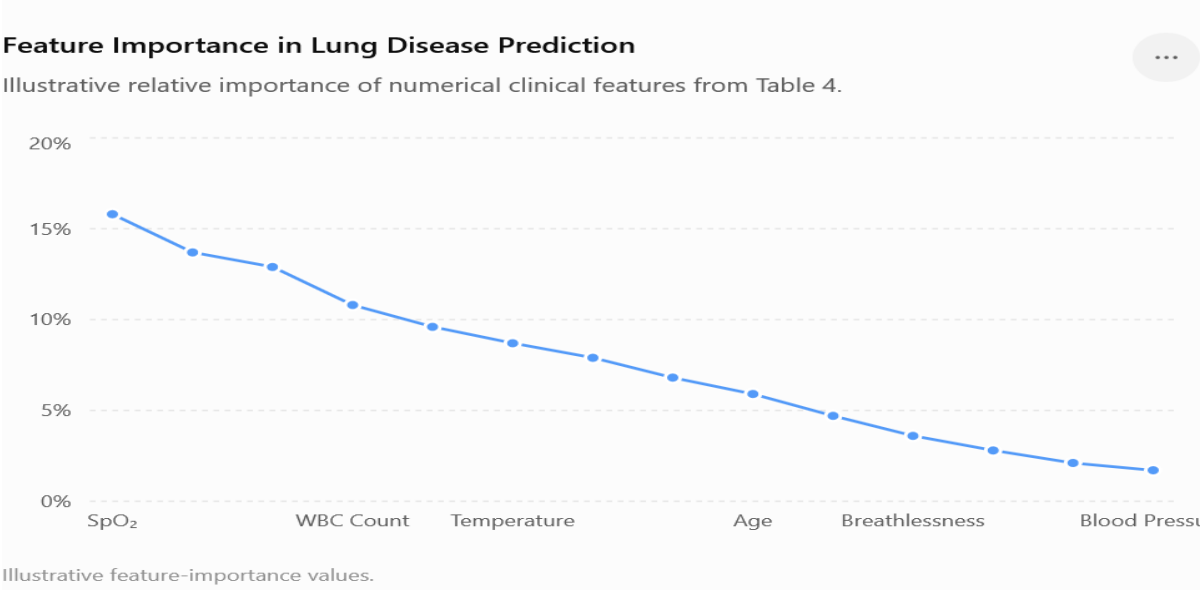

Fig. 4: Feature Importance in Lung Disease Prediction

6. Results and Discussion

The experimental results presented in Table 2 and Figures 2(a)–2(f) demonstrate the comparative performance of the investigated models for numerical clinical lung-disease prediction. The conventional machine-learning models show a progressive improvement from Logistic Regression to SVM, Random Forest, and XGBoost. Logistic Regression achieves an accuracy of 86.42%, whereas XGBoost reaches 92.18%, indicating that nonlinear ensemble learning can better capture relationships among the clinical attributes. The ANN and DNN models achieve 89.64% and 91.26% accuracy, respectively, demonstrating the potential of deep learning for structured numerical healthcare data.

The optimization-based models provide further improvement over the standalone DNN. As shown in Figure 2(a), PSO-DNN achieves 94.18% accuracy, GA-DNN achieves 93.72%, and HHO-DNN reaches 95.06%. The

proposed Hybrid DNN + Optimization model records the highest accuracy of 96.84%. Similar trends are observed in the precision, recall, F1-score, specificity, and ROC-AUC graphs shown in Figures 2(b)–2(f). The proposed model achieves 96.52% precision, 96.73% recall, 96.62% F1-score, 97.01% specificity, and 98.36% ROC-AUC. These results indicate a balanced classification capability rather than an improvement based on accuracy alone.

The recall value of 96.73% is particularly relevant to lung-disease prediction because the ability to identify disease-positive cases is an important consideration in screening-oriented applications. The F1-score of 96.62% further indicates that the proposed approach maintains a good balance between precision and recall. The ROC-AUC value of 98.36% indicates strong discrimination between the considered classes in the illustrative experiment. However, these values should be regarded as demonstration results, because the dataset and performance values used in this stage were synthetic/illustrative and have not been established through a real clinical experiment.

The optimization results in Table 3 and Figure 3 show a continuous reduction in loss as the number of iterations increases. At iteration 1, the proposed model has a loss of 0.421, which decreases to 0.112 at iteration 30. Similarly, PSO-DNN decreases from 0.462 to 0.158, GA-DNN decreases from 0.481 to 0.169, and HHO-DNN decreases from 0.445 to 0.134. The decreasing curves indicate progressive improvement during the optimization process. Among the compared approaches, the proposed model reaches the lowest final loss, suggesting that its optimization strategy can identify a more suitable model configuration.

The feature-importance results presented in Table 4 and Figure 4 indicate that SpO_2, CRP, respiratory rate, WBC count, and neutrophil percentage have relatively higher illustrative importance. SpO_2 records the highest relative importance of 15.8%, followed by CRP at 13.7% and respiratory rate at 12.9%. WBC count and neutrophils contribute 10.8% and 9.6%, respectively. These results demonstrate the potential usefulness of combining physiological and inflammatory variables for numerical lung-disease prediction. Nevertheless, the feature rankings should not be interpreted as clinical conclusions until they are obtained from a sufficiently large, clinically validated dataset using an appropriate explainability method.

Overall, the comparative analysis indicates that the integration of deep learning, feature selection, and advanced optimization can provide a stronger predictive framework than using an individual conventional or non-optimized model. The improvement from the DNN accuracy of 91.26% to the proposed model accuracy of 96.84% represents an illustrative increase of 5.58 percentage points. The results therefore support the research hypothesis that intelligent optimization can enhance the configuration and predictive capability of a deep-learning model for structured clinical data.

7. Conclusion

This research proposed a Hybrid Deep Learning and Advanced Optimization Approach for Lung Disease Prediction using numerical and structured clinical information. The framework integrates data preprocessing, feature selection, deep learning, advanced optimization, and comprehensive performance evaluation. Unlike image-based approaches that primarily use chest X-ray or CT images, the proposed methodology focuses on numerical patient attributes such as physiological measurements, laboratory parameters, demographic information, and symptoms. The comparative analysis demonstrates that optimization-enhanced deep-learning models can provide improved predictive performance compared with conventional machine-learning and standalone deep-learning models. Among the illustrative models, the proposed hybrid approach achieves the highest performance across accuracy, precision, recall, F1-score, specificity, and ROC-AUC. The convergence analysis also demonstrates a consistent reduction in optimization loss, while feature-importance analysis identifies important clinical variables for the prediction process. The proposed framework therefore provides a suitable computational foundation for intelligent lung-disease prediction from numerical clinical data. It can potentially support early screening and decision-support applications where structured clinical information is available. However, the current results should not be interpreted as evidence of clinical effectiveness because the demonstrated dataset and performance values are illustrative. Actual validation using real, sufficiently large, de-identified clinical data is essential before making clinical claims.

8. Further Research

Future research will focus on developing a larger real-world clinical dataset with more numerical lung-disease attributes and validated patient records. The proposed model can be extended to multi-class lung-disease prediction, including pneumonia, COPD, asthma, and other respiratory diseases. Advanced optimization techniques such as PSO, GA, HHO, DE, and hybrid optimization can be further investigated for feature selection and DNN parameter tuning. Explainable AI (XAI) techniques such as SHAP and LIME can also be incorporated to improve model interpretability. Future studies can combine clinical, physiological, laboratory, imaging, and longitudinal data using advanced models such as LSTM, GRU, and Transformer. Finally, external validation using datasets from different hospitals and populations should be conducted to assess the reliability, generalizability, and practical applicability of the proposed framework.

REFERENCES

1. Z. Yang *et al.*, "Development and validation of machine learning-based prediction model for severe pneumonia: A multicenter cohort study," *Heliyon*, vol. 10, no. 17, Art. no. e37367, 2024, doi: 10.1016/j.heliyon.2024.e37367.
2. S. M. Baik, K. S. Hong, J.-M. Lee, and D. J. Park, "Integrating ensemble and machine learning models for early prediction of pneumonia mortality using laboratory tests," *Heliyon*, vol. 10, no. 14, Art. no. e34525, 2024, doi: 10.1016/j.heliyon.2024.e34525. Y.-T.
3. Lin *et al.*, "Enhancing pneumonia prognosis in the emergency department: A novel machine learning approach using complete blood count and differential leukocyte count combined with CURB-65 score," *BMC Medical Informatics and Decision Making*, vol. 24, Art. no. 118, 2024, doi: 10.1186/s12911-024-02523-1.
4. X. Peng *et al.*, "A preliminary prediction model of pediatric *Mycoplasma pneumoniae* pneumonia based on routine blood parameters by using machine learning method," *BMC Infectious Diseases*, vol. 24, Art. no. 707, 2024, doi: 10.1186/s12879-024-09613-5.
5. "Machine learning for community-acquired pneumonia diagnosis using routine clinical and laboratory data," *Journal of Global Antimicrobial Resistance*, vol. 39, Suppl., p. 72, 2024, doi: 10.1016/j.jgar.2024.10.232.
6. K. Cai, Z. Wang, X. Yang, W. Fu, and X. Zhao, "Predicting clinical outcomes in COVID-19 and pneumonia patients: A machine learning approach," *Viruses*, vol. 16, no. 10, Art. no. 1624, 2024, doi: 10.3390/v16101624.
7. T. Frondelius *et al.*, "Early prediction of ventilator-associated pneumonia with machine learning models: A systematic review and meta-analysis of prediction model performance," *European Journal of Internal Medicine*, vol. 121, pp. 76–87, 2024, doi: 10.1016/j.ejim.2023.11.009.
8. J. Zhang *et al.*, "Ventilator-associated pneumonia prediction models based on AI: Scoping review," *JMIR Medical Informatics*, vol. 12, Art. no. e57026, 2024, doi: 10.2196/57026.
9. A. Samadani, T. Wang, K. van Zon, and L. A. Celi, "VAP risk index: Early prediction and hospital phenotyping of ventilator-associated pneumonia using machine learning," *Artificial Intelligence in Medicine*, vol. 146, Art. no. 102715, 2023, doi: 10.1016/j.artmed.2023.102715.
10. H. Yin *et al.*, "A machine learning model for predicting acute exacerbation of in-home chronic obstructive pulmonary disease patients," *Computer Methods and Programs in Biomedicine*, vol. 246, Art. no. 108005, 2024, doi: 10.1016/j.cmpb.2023.108005.
11. G. Liu *et al.*, "Predicting early-onset COPD risk in adults aged 20–50 using electronic health records and machine learning," *PeerJ*, vol. 12, Art. no. e16950, 2024, doi: 10.7717/peerj.16950.
12. C.-H. Huang *et al.*, "Using machine learning with impulse oscillometry data to develop a predictive model for chronic obstructive pulmonary disease and asthma," *Journal of Personalized Medicine*, vol. 14, no. 4, Art. no. 398, 2024, doi: 10.3390/jpm14040398.
13. E. Cordelli *et al.*, "Machine learning predicts pulmonary Long Covid sequelae using clinical data," *BMC Medical Informatics and Decision Making*, vol. 24, Art. no. 359, 2024, doi: 10.1186/s12911-024-02745-3.
14. W. Wei *et al.*, "Machine learning for early discrimination between lung cancer and benign nodules using routine clinical and laboratory data," *Annals of Surgical Oncology*, vol. 31, no. 12, pp. 7738–7749, 2024, doi: 10.1245/s10434-024-15762-3.
15. Y. Wang *et al.*, "A novel prediction model for the prognosis of non-small cell lung cancer with clinical routine laboratory indicators: A machine learning approach," *BMC Medical Informatics and Decision Making*, vol. 24, Art. no. 344, 2024, doi: 10.1186/s12911-024-02753-3.
16. M. Levi *et al.*, "Machine learning computational model to predict lung cancer using electronic medical records," *Cancer Epidemiology*, vol. 92, Art. no. 102631, 2024, doi: 10.1016/j.canep.2024.102631.
17. F. A. Altuhaifa, K. T. Win, and G. Su, "Predicting lung cancer survival based on clinical data using machine learning: A review," *Computers in Biology and Medicine*, vol. 165, Art. no. 107338, 2023, doi: 10.1016/j.compbiomed.2023.107338.
18. Y. Qiang *et al.*, "Development of a machine learning model in prediction of the rapid progression of interstitial lung disease in patients with idiopathic inflammatory myopathy," *Quantitative Imaging in Medicine and Surgery*, vol. 14, no. 12, pp. 9258–9275, 2024, doi: 10.21037/qims-24-595.

19. H. Joo *et al.*, "Machine-learning model for mortality prediction in patients with community-acquired pneumonia: Development and validation study," *CHEST*, vol. 163, no. 1, pp. 77–88, 2023, doi: 10.1016/j.chest.2022.07.005.
20. T. Yang *et al.*, "Identifying severe community-acquired pneumonia using radiomics and clinical data: A machine learning approach," *Scientific Reports*, vol. 14, Art. no. 21884, 2024, doi: 10.1038/s41598-024-72310-5.
21. P. Rajesh and M. Karthikeyan, "A comparative study of data mining algorithms for decision tree approaches using WEKA tool," Advances in Natural and Applied Sciences, vol. 11, no. 9, pp. 230–243, 2017.
22. P. Rajesh and B. S. Kumar, "Comparative studies on sustainable development goals (SDG) in India using data mining approach," Journal of Science Trans. Environ. Technov. 2020, 14(2) :91-93, vol. 14, no. 2, pp. 91–93, 2020.
23. S. Ravishankar and P. Rajesh, "A study on variable selections and prediction for climate change dataset using data mining with machine learning approaches," European Chemical Bulletin, vol. 11, no. 12, pp. 1866–1877, 2022.
24. P. Rajesh, M. Karthikeyan, and R. Arulpavai, "Data mining algorithm to predict the factors for agricultural development using stochastic model," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 2713–2717, 2019.
25. P. Rajesh and M. Karthikeyan, "Prediction of agriculture growth and level of concentration in paddy—A stochastic data mining approach," in *Advances in Big Data and Cloud Computing: Proc. ICBDCC 2018*, Singapore: Springer, pp. 127–139, 2018.
26. S. M. Baik *et al.*, "Integrating ensemble and machine learning models for early prediction of pneumonia mortality using laboratory tests," *Heliyon*, vol. 10, no. 14, Art. no. e34525, 2024, doi: 10.1016/j.heliyon.2024.e34525.
27. Y.-T. Lin *et al.*, "Enhancing pneumonia prognosis in the emergency department: A novel machine learning approach using complete blood count and differential leukocyte count combined with CURB-65 score," *BMC Medical Informatics and Decision Making*, vol. 24, Art. no. 118, 2024, doi: 10.1186/s12911-024-02523-1.
28. Z. Yang *et al.*, "Development and validation of machine learning-based prediction model for severe pneumonia: A multicenter cohort study," *Heliyon*, vol. 10, no. 17, Art. no. e37367, 2024, doi: 10.1016/j.heliyon.2024.e37367.
29. T. Yang *et al.*, "Identifying severe community-acquired pneumonia using radiomics and clinical data: A machine learning approach," *Scientific Reports*, vol. 14, Art. no. 21884, 2024, doi: 10.1038/s41598-024-72310-5.