

Pleural Effusion Detection Using Optimized Pretrained Models

¹Rehab Fathi Ibrahim, ^{1,2}Nashwa Mohamed Elsayed, ²Alshaimaa Mostafa Mohammed and ³Ahmed Magdy Mohamed

¹Mathematics Department, Faculty of Science, Suez Canal University, Ismailia, Egypt

²Basic Science Department, Faculty of Informatics & Computer Science, British University in Egypt (BUE), El Shorouk City, Egypt

³Electrical Engineering Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt

Article history

Received: 15-01-2025

Revised: 15-03-2025

Accepted: 24-03-2025

Corresponding Author:

Rehab Fathi Ibrahim

Mathematics Department, Faculty of Science, Suez Canal University, Ismailia, Egypt

Email: rehabfathy9090@gmail.com

Abstract: Pleural effusion is an abnormal accumulation of fluid in the pleural cavity, posing a significant thoracic issue with various causes. Even with a variety of diagnostic techniques, accurate diagnosis remains challenging. Delays in diagnosis and treatment heighten the risks of the disease. In this paper, we introduce a novel approach by leveraging the Grey Wolf Optimizer (GWO) to systematically optimize pretrained models for detecting pleural effusion in chest X-rays. Our need for optimization stems from the obstacles encountered in recent studies on diagnosing pleural effusion. Previous studies have relied on trial-and-error parameter tuning, which has entailed high computational costs and made training more difficult. Proper parameter tuning improves model performance. Therefore, in this paper, we utilize the GWO model to fine-tune the parameters and enhance the performance of our pretrained model performance in detecting pleural effusion. However, due to the challenges of applying the optimization model to the entire dataset, we will use a subset, referred to as the "Small ChestX-Ray14" dataset. This not only reduces the computational demands of optimization but also addresses the dataset imbalance in ChestX-Ray14, which could affect performance. The GWO-ResNet18 model achieved a significant improvement in performance on the Small ChestX-Ray14 dataset, attaining an Area Under the Curve (AUC) of 95.1% and an accuracy of 89.07%. In comparison, the baseline ResNet18 model achieved an AUC of 93.3% and an accuracy of 86.59%. Building on this success, the optimized hyperparameters were transferred to the full ChestX-Ray14 dataset to evaluate the model's performance on a larger scale, resulting in an AUC of 88.48%, compared to 87.36% with unoptimized hyperparameters. This surpasses both ResNet18 with unoptimized hyperparameters and state-of-the-art methods. Our results indicate that optimizing pretrained models, specifically through the use of GWO, improves medical image interpretation and enhances disease diagnosis, such as that of pleural effusion.

Keywords: Grey Wolf Optimizer, Pleural Effusion, Pretrained Models

Introduction

Thoracic diseases, such as pneumonia, tuberculosis, cardiomegaly, and pleural effusion, are among the most common serious health issues, causing millions of deaths each year. Therefore, early diagnosis and treatment of these diseases are crucial in the medical field, as they can save many lives. (Mostafa *et al.*, 2022; Singh, 2024). Pleural effusion is an abnormal accumulation of fluid in the pleural cavity. It is one of the most serious thorax diseases, caused by a variety of reasons, and despite the

availability of several diagnostic modalities, its diagnosis is still challenging (Chen *et al.*, 2021; Liu *et al.*, 2022).

One of the most commonly used medical imaging techniques in diagnosing lung diseases is chest radiography, also known as Chest X-ray (CXR) (Usman *et al.*, 2022). CXRs have many advantages, such as their low cost and ease of use. Therefore, they are frequently utilized to diagnose lung diseases such as tuberculosis, pleural effusion, and interstitial lung disease. However, accurately interpreting and analyzing CXR images

remains a significant challenge for doctors, as distinguishing between similar diseases relies on the judgment of experienced specialists. Despite the high demand for medical image analysis, the number of qualified doctors capable of providing precise interpretations is limited. Moreover, human eyes can grow weary, so wrong interpretation and analysis of CXR images tend to occur, which can result in false diagnoses of lung diseases (Singh, 2024).

Researchers have been exploring how to utilize computer technology to enhance the accuracy of lung disease detection in medical images. Computer-aided detection systems have become a significant research focus due to the widespread use of CXRs and the difficulty of interpreting them. In this context, Deep Learning (DL) has emerged as a powerful approach, playing a crucial role in the medical field, particularly in improving the analysis of CXR images. By leveraging advanced DL models, the accuracy and efficiency of detecting chest diseases have significantly improved,

DL a key tool in modern medical diagnostics. (Sanida *et al.*, 2024; Wang *et al.*, 2021).

A key factor in the success of DL-based medical image analysis is the availability of large, well-annotated datasets. The National Institutes of Health (NIH) ChestX-Ray14 represents one of the most important and common CXR datasets, diagnosing chest and lung diseases (Wang *et al.*, 2017). Many recent studies have leveraged the rapid development of DL and the release of the ChestX-Ray14 dataset to classify and diagnose various chest diseases, such as cardiomegaly, pneumonia, mass, nodule, atelectasis, and pleural effusion. These advancements have contributed significantly to the detection of pleural effusion. Table 1 presents recent relevant studies that have contributed to the diagnosis of pleural effusion on the ChestX-Ray14 dataset. It includes the model used in each study for image classification, along with the main advantages and disadvantages of each model.

Table 1: Summary of Relevant Studies on Pleural Effusion Detection Using the ChestX-Ray14 Dataset

Ref	Model/Method	Merits	Demerits
Li <i>et al.</i> (2024)	MBRANet (Backbone Model: ResNet-50)	- Enhance focuses on disease regions and filters noise. - Classify diseases directly from feature maps, preserving spatial information. reduce the impact of rare disease cases.	- High computational cost - Requires careful tuning of attention parameters. - More complex model structure, leading to longer training time. - Struggles to generalize on datasets with different disease distributions.
Wang <i>et al.</i> , (2024)	CXR×MLAGCPL	- Improves multi-label classification by learning relationships between diseases. - Reduces bias toward frequent diseases. - Graphs improve interpretability by modeling co-occurrence dependencies.	- Requires more computation than Convolutional Neural Network (CNN) based methods. - Over-smoothing occurs when too many graph layers are used. - Long-tail label distribution can still affect results. - Harder to fine-tune due to multiple hyperparameters.
Jiang <i>et al.</i> (2024)	TransDD	- Improve Feature-Label Relationship Modeling - Make self-attention faster than standard Transformers. - Use attention mechanisms to improve classification.	- Requires more tuning than CNN-based models. - Slower than CNN models. - Affected by noisy labels in ChestXray14. - More complex training process than standard CNN models.
Kufel <i>et al.</i> (2023)	EfficientNet-B1	- Lightweight and require less GPU memory. - Improve feature extraction using pretrained ImageNet weights. - Improved Generalization with Transfer Learning.	- Sensitive to Hyperparameter Choices: small changes in hyperparameters can affect performance more than in standard CNNs. - Dataset label noise still affects model performance.
Liu <i>et al.</i> (2023)	ML-LGL (Backbone Model: DenseNet121)	- Uses label correlation to improve classification accuracy. - Works well for multi-label classification. - Mimics how radiologists learn diseases (from common to rare).	- No explicit label noise handling. - Computationally expensive (multiple training phases). - Relies on selection functions (correlation, similarity, frequencybased).
Mezina & Burget (2024)	InceptionV3 & ViT for features extraction (Fully Connected Layer for Classification)	- Transformer-based local feature extraction. - Combines local and global features effectively. - Lung segmentation preprocessing step.	- Requires high memory & computation due to ViT. - Extra preprocessing steps add complexity. - No explicit noise-handling mechanism.
Wu <i>et al.</i> (2023)	CTransCNN	- Capture both local and global features. - Combines Focal Loss and ASL, improving classification for rare diseases.	- High computational cost due to dual-model architecture. - Complex loss function requires fine-tuning, making training more difficult.
Zhu <i>et al.</i> (2022)	PCAN (Backbone Model: DenseNet-121)	- Pixel-wise classification improves detection of subtle abnormalities. - Adaptive weighting mechanism helps rare disease detection. - Attention mechanism helps focus on disease-relevant areas.	- Pixel-wise classification adds complexity without always improving global classification. - Does not explicitly handle noisy labels, which are common in ChestX-ray14.

Table 1: Continued

Ref	Model/Method	Merits	Demerits
Mao <i>et al.</i> (2022)	V-GCN-PPS (Backbone Model: VGGNet16BN)	- Uses GCN to model image relationships, allowing the model to learn disease connections. - Incorporates patient metadata, to improve prediction accuracy. - PPS helps detect rare diseases by linking similar cases.	- Higher computational cost due to graph-based processing. - Graph structures must be carefully tuned, making training more complex. - Requires additional metadata which may not always be available.
Yang (2025)	AFM-DViT (Backbone Model: DenseNet-ViT Hybrid Model)	- Hybrid feature extraction. - Federated Learning allows hospitals to train the model locally without sharing actual patient images. - AFM-DViT adapts to variations in X-ray images from different hospitals, ensuring consistent performance.	- High computational cost due to ViTbased model. - Slow Convergence: takes longer to train compared to purely CNN models. - Dependency on high quality data, limiting its effectiveness in data-scarce environments.
Huangsuwan <i>et al.</i> (2025)	FedDrip (Backbone Model: DenseNet-121)	- Uses synthetic images to improve federated learning in non-uniform data settings. - Reduces Data Scarcity Issues, improving classification when real data is limited. - Maintains Data Privacy: Training occurs locally, with only model weights shared.	- Dependence on Synthetic Data Quality. - Diffusion models require more processing power, increasing training time compared to standard CNN-based FL models. - Deviation in synthetic images can introduce bias, affecting classification accuracy.
Guan <i>et al.</i> (2023)	GMM-HM	- Better handling of noisy labels using GMM. - HM reduces bias from uncertain labels. - GMM filters uncertain samples, reducing overfitting to label noise and improving classification reliability.	- More complex than standard CNN due to noise detection steps. - Requires fine-tuning of noise detection thresholds for optimal performance. - May not generalize well to datasets without significant label noise.

From Table 1, we note that while deep learning models have advanced pleural effusion diagnosis, they face challenges such as high computational costs, complex structures, and impractical fine-tuning. Additionally, models trained on large-scale datasets like ChestX-Ray14 struggle with classification noise, dataset imbalance, and generalization issues. To address these issues, this paper proposes an optimized pretrained mode that systematically enhances hyperparameter selection, providing a more efficient and robust alternative. By leveraging hyperparameter optimization, the model reduces computational costs and enables fine-tuning without excessive manual adjustments. Unlike models with complex architectures, ResNet18 offers a lightweight yet effective solution, balancing performance and efficiency. Additionally, training first on the Small ChestX-ray14 helps mitigate dataset imbalance before applying the optimized hyperparameters to the full dataset, improving generalization and robustness while minimizing the impact of classification noise.

The role of optimization in improving model performance has been extensively recognized in recent research. Given the growing need for more efficient hyperparameter selection, various optimization techniques have proven effective across multiple research areas, including those in the medical field. For instance, Machine Learning (ML) classifiers have been applied to predict heart failure and its severity levels, with methods such as the Synthetic Minority Oversampling Technique (SMOTE) and Hyperband (HB) optimizer being used to identify optimal hyperparameters. This approach achieved its best results with Extra Trees combined with SMOTE, demonstrating

the potential of optimization for improving model accuracy (Abdellatif *et al.*, 2022).

Similarly, feature extraction with ResNet18 has been combined with meta-heuristic algorithms, such as Particle Swarm Optimization (PSO), Atom Search Optimization (ASO), and Equilibrium Optimizer (EO), to identify the most relevant features for breast cancer classification, with the EO and SVM combination achieving the highest accuracy (Atban *et al.*, 2023).

While these studies have focused on optimization for traditional ML models, their findings underscore the value of optimization in addressing computational challenges and improving predictive accuracy. Extending this concept to DL, optimization techniques have also been applied in other domains, such as spectrum recovery for compressive sensing, where the GWO demonstrated superior performance in solving minimization problems (Gamal *et al.*, 2022).

These examples highlight the versatility of optimization approaches and set the stage for their application to pretrained DL models, as explored in this paper. Building on these findings, we adopted GWO in this paper to fine-tune the hyperparameters of our pretrained ResNet18 model. GWO was chosen for its ability to balance exploration and exploitation during the optimization process, making it particularly effective in navigating complex solution spaces, such as those encountered in deep learning parameter tuning. Furthermore, GWO's hierarchical structure and adaptive mechanisms enable efficient convergence to optimal solutions, which is crucial when working with large datasets and computationally intensive tasks.

The ResNet18 model was chosen for its lightweight architecture, which offers reduced training time compared to other deep learning models while still achieving competitive performance in image classification tasks. This combination of GWO and ResNet18 provides a robust and efficient approach for optimizing and applying pretrained models to large medical datasets like ChestX-Ray14. However, optimizing the entire ChestX-Ray14 dataset is computationally intensive, requiring several months of processing time. Additionally, the dataset's class imbalance poses challenges. To address these issues, we constrain our model to a subset of the dataset, referred to as the Small ChestX-Ray14. The optimized hyperparameters derived from this subset will later be applied to the full ChestX-Ray14 dataset, thereby avoiding the need for manual parameter tuning and potentially yielding better results.

Materials and Methods

In this paper, we propose an optimized pretrained model using GWO. The ResNet18 model was selected as the base model. Its performance will be compared to that of the optimized GWO-ResNet18 model. Figure 1 illustrates the layers of the ResNet18 model.

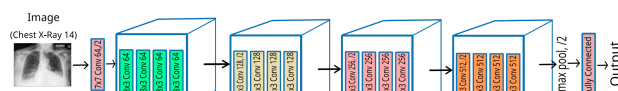


Fig. 1: Baseline ResNet-18 Architecture

GWO-ResNet18 model will be applied to a subset of the ChestX-Ray14 dataset, termed Small ChestXRay14, to improve computational efficiency and performance. By focusing on this smaller dataset, we aim to address issues related to dataset imbalance and lengthy processing times. The optimized hyperparameters derived from the GWO-ResNet18 model on the Small ChestX-Ray14 dataset will be used to enhance the performance of the model on the full ChestX-Ray14 dataset, eliminating the need for manual parameter tuning. An illustration of the research methodology is shown in Figure 2.

Dataset and Preprocessing

ChestX-Ray14 is one of the largest chest radiograph datasets, available at the NIH Clinical Center, utilized for diagnosing various thoracic diseases. It consists of 112,120 frontal-view X-ray images, including 51,708 that have been annotated with up to 14 pathologies: Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, and Pleural Thickening, while the remaining images have been labeled as No Finding (X. Wang *et al.*, 2017). The dataset contains 13,317 images labeled with effusion and 98,803 without effusion.

ChestX-Ray14 is considered an imbalanced dataset, which negatively impacts model performance during the classification process. Therefore, we selected only a subset to achieve dataset balance, leading to the creation of the Small ChestX-Ray14 dataset. Another important reason for this choice is that optimizing a pretrained model on the full ChestX-Ray14 dataset is extremely time-consuming given our computational capabilities, requiring several months of processing time. The Small ChestX-Ray14 dataset consists of 26,317 CXR images, 13,317 images with effusion and 13,000 images without effusion. We resized the images from (1024×1024) to (227×227) pixels. Table 2 provides a detailed breakdown of the total number of CXR images in Small ChestX-ray14, including the counts of images with and without effusion, as well as the numbers used for training and testing the model.

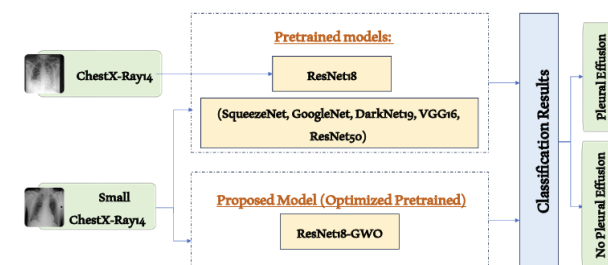


Fig. 2: Research Workflow, Highlighting Pretrained Model Evaluation and Optimization Process

Table 2: Details of Small ChestX-ray14 Dataset Split

Total No.	Effusion	No Effusion	Training	Testing
26317	13317	13000	21054	5263

Pretrained Models

In this research, we utilized several pretrained models to detect pleural effusion across our dataset. These models, previously trained on large-scale image datasets, were selected for their effectiveness in medical image analysis and their ability to enhance detection accuracy. Figure 3 illustrates the workflow for utilizing pretrained models in the detection of pleural effusion from CXR images. The pretrained models adopted in this paper are as follows:

- GoogleNet, also known as Inception v1, was developed by Google in 2014 to reduce computational complexity. It consists of 22 layers, including three convolutional layers, nine inception layers (each containing two convolutional layers), and one fully connected layer (Alom *et al.*, 2010; Sze *et al.*, 2017; Szegedy *et al.*, 2015).
- SqueezeNet is an 18-layer Convolutional Neural Network (CNN) introduced in 2016, maintaining AlexNet-level accuracy while requiring less memory and processing time for classification (Islam *et al.*, 2020; Nguyen *et al.*, 2018).
- DarkNet19 is a deep CNN that serves as basis for YOLOv2. It consists of 19 convolutional layers and

five max pooling layers. The DarkNet19 model typically employs 3×3 filters, doubling the number of channels after each pooling step (Redmon & Farhadi, 2017).

- VGG16 (Visual Geometry Group) is a widely used deep CNN architecture introduced in 2014 by Simonyan and Zisserman. The term "deep" refers to the number of layers, with VGG16 and VGG19 having 16 and 19 convolutional layers, respectively (Ghosh *et al.*, 2019). To reduce computational costs, larger filters (e.g., 5×5) are constructed from smaller ones (e.g., 3×3), minimizing the number of weights (Simonyan & Zisserman, 2014).
- ResNet-18, ResNet-50 (Residual Neural Network) is a type of CNN introduced in 2015 that has excelled in various computer vision tasks, including image classification and object detection. Unlike traditional CNNs, which require learning the entire feature map and are thus computationally expensive and slow, ResNets use skip connections. These connections allow ResNets to be significantly smaller than traditional CNNs while maintaining high performance (Shafiq & Gu, 2022).

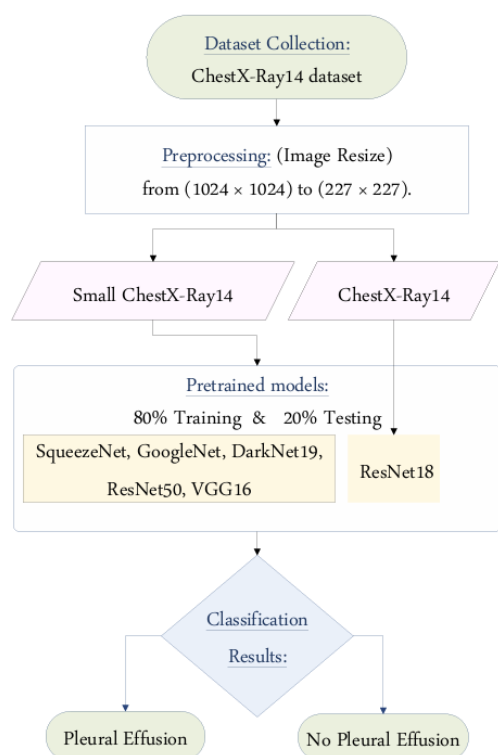


Fig. 3: Process Overview of Detecting Pleural Effusion Using Pretrained Models

The 18-layer version of the ResNet architecture, known as ResNet-18, consists of 17 convolutional layers and one fully connected layer (Džakmić *et al.*, 2020). ResNet-18 is widely used for various computer vision tasks, especially when computational resources are limited, following the general ResNet principles with residual connections to facilitate deep network training.

Similarly, ResNet-50, which has 50 layers with two or three convolutional layers per block, is widely used in image classification and object detection tasks, adhering to the same principles as ResNet-18 (Maeda-Gutiérrez *et al.*, 2020; Sarwinda *et al.*, 2021).

Proposed Method

In this paper, we introduce a novel model, GWOResNet18, designed to enhance the detection of pleural effusion in the Small ChestX-Ray14 dataset. Our approach involves optimizing the pretrained ResNet18 model using GWO. The integration of GWO with ResNet18 aims to fine-tune the model's hyperparameters effectively, thereby improving its performance and accuracy in identifying pleural effusion within ChestX-Ray14 images. Due to the extensive time required to apply this optimization directly to the full ChestX-Ray14 dataset, we first implemented it on the smaller subset, Small ChestX-Ray14. Once the hyperparameters are optimized on this subset, we will apply them to the full ChestX-Ray14 dataset, aiming to improve its performance without the need for direct optimization. Figure 4 illustrates the flowchart of the GWO-ResNet18 model.

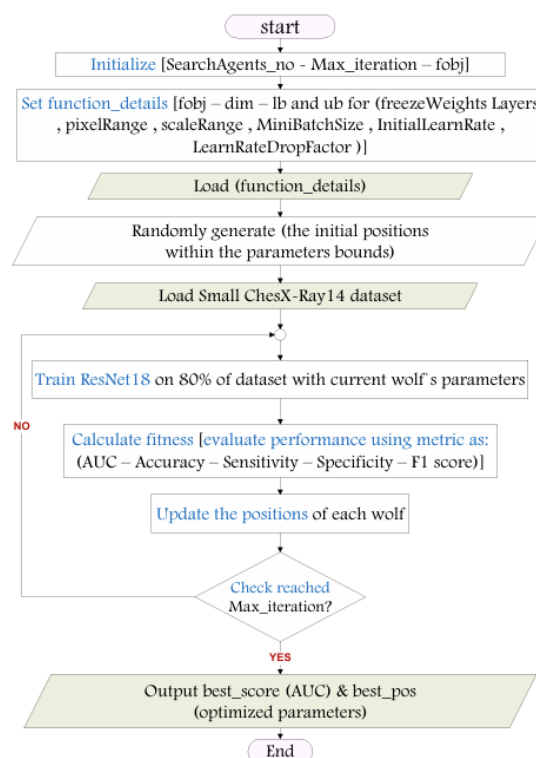


Fig. 4: Flowchart of our GWO-ResNet18 model

The Grey Wolf Optimizer (GWO) was introduced by S. Mirjalili in 2014. It is a swarm intelligence algorithm inspired by the social hierarchy and hunting behavior of grey wolves. GWO adopts a four-level social hierarchy, which includes α , β , δ , and ω wolves. The α (alpha) wolves are the leaders responsible for decision-making.

The β (beta) wolves are second in command and the best candidates to become α ; they provide feedback to the α leader based on information gathered from other wolves. The δ (delta) wolves obey the α and β wolves but have authority over the ω wolves. Finally, the ω (omega) wolves are the lowestranking members of the pack, responsible for maintaining their safety. This hierarchical structure allows GWO to simulate the cooperative behavior of grey wolves during the optimization process. (Kohli & Arora, 2018; Mirjalili *et al.*, 2014). Mathematically, α , β , δ , and ω in GWO represent:

- Alpha (α): the best (fittest) solution.
- Beta (β): Second-best solution.
- Delta (δ): Third-best solution.
- Omega (ω): Remaining possible solutions.

Initialization is the first stage of the GWO algorithm, where the positions of wolves (X_i) in the search agent are generated. Each wolf's fitness is then evaluated using the optimization problem's objective function. The positions are updated in each iteration through three steps: encircling behavior, hunting behavior, and exploration behavior, inspired by the social behavior and hunting techniques of grey wolves (Magdy *et al.*, 2023; Sánchez *et al.*, 2017).

Encircling Behavior

The grey wolves encircle their prey (optimal solution) mathematically as (Mirjalili *et al.*, 2014):

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_P(t) - \vec{X}(t) \right| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}_P(t) - \vec{A} \cdot \vec{D} \quad (2)$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \quad (3)$$

$$\vec{C} = 2 \cdot \vec{r}_2 \quad (4)$$

Where:

- t : current iteration
- $\vec{X}_P(t)$: the position vector of the prey (optimal solution) at iteration t
- $\vec{X}(t)$: the position vector of the wolf at iteration t
- $\vec{C}$ and $\vec{A}$: coefficient vectors
- $\vec{a}$: linear decreased from 2 to 0 over iterations
- $\vec{r}_1, \vec{r}_2$: random vectors in $[0,1]$

Hunting Behavior

To mathematically simulate hunting behavior, let the first three best solutions (α , β and δ) guide the optimization, as follows (Mirjalili *et al.*, 2014):

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \left(\vec{D}_\alpha \right) \quad (5)$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \left(\vec{D}_\beta \right) \quad (6)$$

$$\vec{X}_3 = \vec{X}_\gamma - \vec{A}_3 \cdot \left(\vec{D}_\gamma \right) \quad (7)$$

Where:

$$\vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right| \quad (8)$$

$$\vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right| \quad (9)$$

$$\vec{D}_\gamma = \left| \vec{C}_3 \cdot \vec{X}_\gamma - \vec{X} \right| \quad (10)$$

The new position is computed as:

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (11)$$

Exploration Behavior

Exploration in GWO focuses on searching new areas of the solution space to find promising regions and avoid getting stuck in local optima. It is controlled by the parameter C , which generates random values in $[0, 2]$ to encourage diverse searches. The parameter A , influenced by a (which decreases from 2 to 0), promotes exploration when its value is high, allowing larger changes in the wolves' positions. These factors enable GWO to effectively explore the search space (Faris *et al.*, 2018).

System Implementation

The pretrained models as well as the proposed (GWO-Resnet18) model were implemented in the (MATLAB R2021a) environment. The system employed for training and testing these models, along with its detailed specifications, is provided below:

- Operating System: Windows 11 Pro
- Processor: 12th Gen Intel(R) Core (TM) i9-12900F 2.40 GHz
- RAM: 32.0 GB (31.8 GB usable).
- System Type: 64-bit operating system, x64-based processor.

Evaluation Metrics

In this subsection, we present the evaluation metrics used to assess model performance. The primary metric used is AUC, expressed as a percentage. AUC is a reliable evaluation measure ranging from 0 to 1, with 1 indicating perfect performance. Given the dataset imbalance, AUC is particularly valuable as it summarizes model performance across all classification thresholds. Expressing AUC as a percentage enables direct comparison with previous studies. Table 3 presents five additional evaluation metrics, which will be computed for a more comprehensive assessment. TP (True Positive) refers to the number of correctly identified pleural effusion cases.

TN (True Negative) refers to the number of CXR images correctly classified as not containing pleural effusion. FP (False Positive) indicates the number of CXR images incorrectly classified as containing pleural effusion. FN (False Negative) refers to the number of pleural effusion cases incorrectly classified as not containing pleural effusion.

Table 3: Formulas for the Utilized Evaluation Metrics

Metric	Equation	No.
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	(1)
F1 score	$\frac{2TP}{2TP+FP+FN}$	(2)
Precision	$\frac{TP}{TP+FP}$	(3)
Sensitivity (Recall)	$\frac{TP}{TP+FN}$	(4)
Specificity	$\frac{TN}{TN+FP}$	(5)

Experimental Settings

In this subsection, we outline the hyperparameter settings for both the pretrained models and the optimized GWO-ResNet18 model. Proper hyperparameter tuning, including learning rate and batch size, is critical for maximizing performance, ensuring accuracy, and improving generalization across the dataset. To evaluate pleural effusion detection, we first apply several pretrained models to the Small ChestX-Ray14 dataset and compare their performance against the optimized model trained exclusively on this dataset. For the ChestX-Ray14 dataset, we apply the baseline ResNet18 model, using hyperparameters adopted from (Rakshit *et al.*, 2019). We then assess its performance using the optimized hyperparameters derived from Small ChestX-Ray14.

For optimization, we utilized the GWO algorithm with the ResNet-18 model to detect pleural effusion in the Small ChestX-Ray14 dataset. Table 4 details the number of search agents, maximum iterations, and the lower and upper bounds (LB, UB) of the hyperparameters used in optimizing the GWOResNet18 model.

Table 4: Hyperparameters Settings for GWO-ResNet18 Model

Parameter	Parameter Setting	
search agents	5	
Maximum number of iterations	20	
	LB	UB
freezeWeights Layers	(1:5)	(1:10)
pixelRange	[-10 10]	[-40 40]
scaleRange	[0.95 1.05]	[0.85 1.15]
MiniBatchSize	64	128
InitialLearnRate	0.01	0.00001
LearnRateDropFactor	0.0001	0.001

The hyperparameters used for training the pretrained models, along with the optimized hyperparameters derived through GWO-ResNet18 on the Small ChestX-Ray14 dataset, will be presented in Table 6 once the optimization process is completed.

Results

This section presents the results of applying both pretrained and optimized models for pleural effusion detection. First, we report the performance of several pretrained models on the Small ChestX-Ray14 dataset.

Next, we evaluate the impact of optimization by assessing the performance of the optimized GWOResNet18 model on the same dataset, followed by a comparative analysis to measure the improvement achieved through optimization. Furthermore, we will compare the performance of the baseline ResNet18 model on the full ChestX-Ray14 dataset using two hyperparameter settings: unoptimized hyperparameters adopted from (Rakshit *et al.*, 2019) and those transferred from the optimized model on Small ChestX-Ray14. This evaluation highlights the impact of optimization on model performance.

Performance of Pretrained Models on the Small ChestX-Ray14 Dataset

In this subsection, we present the evaluation results of pretrained models utilized on the Small ChestX-Ray14 dataset for pleural effusion detection. The models were evaluated based on key metrics, such as AUC and accuracy, to ensure a thorough comparison of their classification performance. Table 5 summarizes the performance of the six pretrained CNN models, offering insights into their relative effectiveness in accurately detecting pleural effusion cases. Next, we illustrate the ROC curve and confusion matrix for the VGG16 model, which achieved the highest AUC on the Small ChestX-Ray14 dataset, in Figures 5 and 6, respectively. Similarly, we present the ROC curve and confusion matrix for the baseline ResNet18 model on the Small ChestX-Ray14 dataset in Figures 7 and 8, respectively.

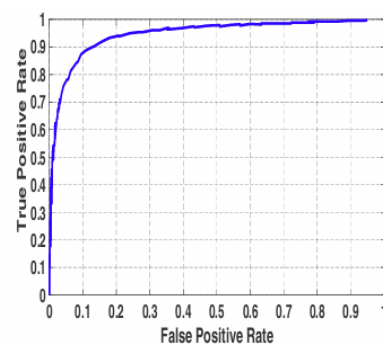


Fig. 5: The Roc Curve for the VGG-16 on Small ChestX-Ray14 Dataset

Confusion Matrix of VGG16 on Small ChestX-Ray14

	No Effusion	Sick with Effusion
No Effusion	2300	300
Sick with Effusion	292	2371
	No Effusion	Sick with Effusion
	Predicted Class	

Fig. 6: The Confusion Matrix for the VGG-16 Model on Small ChestX-Ray14 Dataset

Table 5: Pretrained Models Performance on Small ChestX-Ray14 (Best Results are Underlined)

Algorithm	AUC (%)	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1_score (%)	Elapsed time
SqueezeNet	92.5	85.10	86.40	83.74	86.50	83.74	00:16:40
GoogleNet	91.3	83.60	85.13	81.90	85.35	83.48	00:20:42
VGG16	<u>94.7</u>	<u>88.75</u>	<u>88.77</u>	89.03	<u>88.46</u>	<u>88.90</u>	00:45:37
DarkNet19	93	86.98	86.63	87.83	86.12	87.23	00:34:05
ResNet18	93.3	86.59	84.25	<u>90.39</u>	82.69	87.21	00:17:29
ResNet50	92.7	86.26	86.63	87.83	86.12	87.23	00:40:06

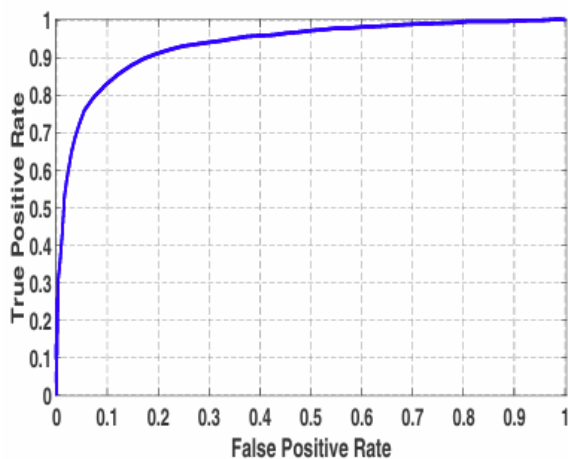


Fig. 7: The ROC Curve for ResNet18 Model on Small ChestX-Ray14 Dataset

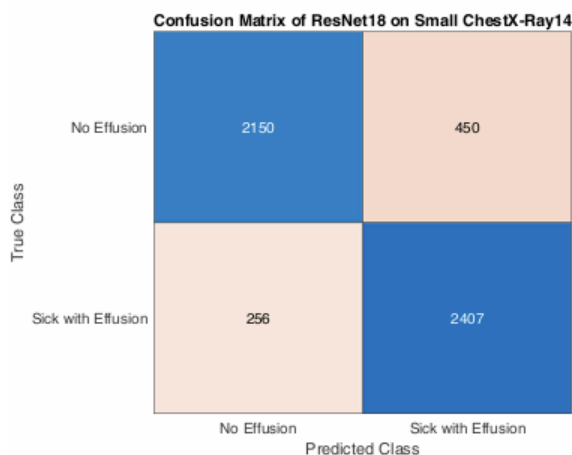


Fig. 8: The Confusion Matrix for the ResNet18 Model on Small ChestX-Ray14 Dataset

Performance of Optimized GWO-ResNet18 and Comparison with Pretrained Models on the Small ChestX-Ray14 Dataset

In this subsection, we present the optimized GWO-ResNet18 model results on the Small ChestX-Ray14 dataset. Table 6 presents the optimal hyperparameters derived through GWO-ResNet18 on the Small ChestX-Ray14 dataset, along with a comparison with those used with pretrained models.

Optimization led to improvements in key evaluation metrics, such as AUC and accuracy, enhancing the

model's ability to detect pleural effusion. The ROC curve and confusion matrix for the GWO-ResNet18 model are shown in Figures 9 and 10, respectively.

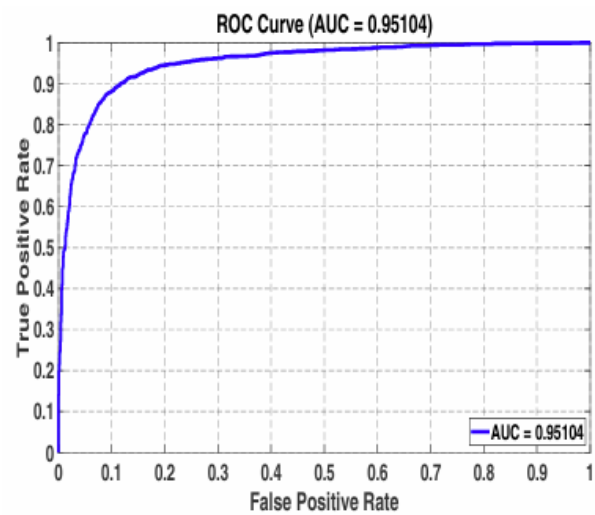


Fig. 9: ROC Curve of GWO-ResNet18 model on Small ChestX-Ray14 dataset

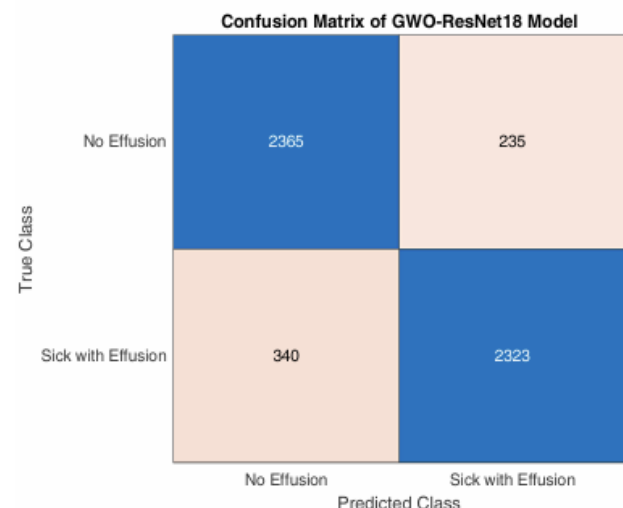


Fig. 10: Confusion Matrix of GWO-ResNet18 model on Small ChestX-Ray14 dataset

To demonstrate the impact of optimization, we compare the performance of the GWO-ResNet18 model with the baseline ResNet18 model, as presented in Table 7, and followed by a visual comparison in Figure 11. The results indicate a notable improvement in performance after optimization.

Table 6: Comparison of Hyperparameters Configurations – Unoptimized vs. Optimized Hyperparameters on the Small ChestX-Ray14 Dataset

Hyperparameters	Hyperparameters Settings for Baseline (Pretrained Models)	Hyperparameters Settings for Optimized Model (GWO-ResNet18)
Input Size in pixel	227 × 227	227 × 227
Frozen Weights	First 10 layers	First 6 layers
pixelRange	[-30 30]	[-18 18]
scaleRange	[0.9 1.1]	[0.86 1.14]
optimizer	Sgdm	Sgdm
MiniBatchSize	64	97
MaxEpochs	5	5
InitialLearnRate	0.001	1e-2
LearnRateDrop Factor	Null	0.0004
LearnRateDrop Period	Null	Null

Table 7: Performance Comparison of GWO_ResNet18 vs. ResNet18 on Small ChestX-Ray14 (Best Results are Underlined)

Dataset	AUC (%)	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1_score (%)
ResNet18	93.3	86.59	84.25	<u>90.39</u>	82.69	87.21
GWO_ResNet18	<u>95.1</u>	<u>89.07</u>	<u>90.81</u>	87.23	<u>90.96</u>	<u>88.99</u>

Comparison of ResNet18 Performance in Detecting Pleural Effusion on the ChestX-Ray14 Dataset Using Transferred Optimized and Unoptimized Hyperparameters

In this subsection, we compare the performance of the ResNet18 model on the ChestX-Ray14 dataset using two different hyperparameter configurations. The first configuration is adopted from Rakshit *et al.* (2019). Whereas the second consists of hyperparameters transferred from the optimization performed on the Small Chest-XRay14 dataset. These configurations are detailed in Table 8.

The comparison evaluates the effectiveness of transferring optimized hyperparameters across datasets. The results are presented in Table 9, along with a visual

comparison in Figure 12, which further illustrates the impact of optimization. Finally, Table 10 provides a performance comparison against state-of-the-art approaches, using the AUC metric to highlight the competitive performance of our approach.

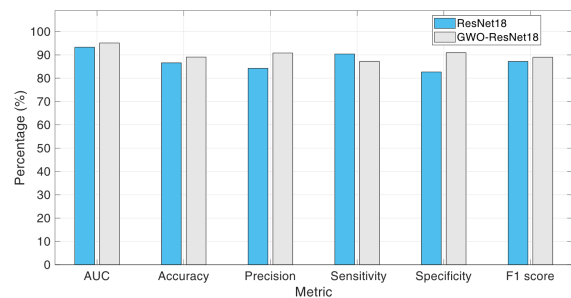


Fig. 11: GWO-ResNet18 vs. ResNet18 Performance on Small ChestX-Ray14

Table 8: Comparison of Hyperparameters Configurations – Baseline (Rakshit *et al.*, 2019) vs. Optimized (Transferred from Small ChestX-Ray14)

Hyperparameters	Baseline (Rakshit <i>et al.</i> , 2019)	Optimized Transferred from Small ChestX-Ray14
Input Size in pixel	224 × 224	227 × 227
Frozen Weights	First 10 layers	First 6 layers
pixelRange	[-30 30]	[-18 18]
scaleRange	-	[0.86 1.14]
optimizer	adam	Sgdm
MiniBatchSize	16	97
MaxEpochs	5	5
InitialLearnRate	1e-4	1e-2
LearnRateDrop Factor	0.5	0.0004
LearnRateDrop Period	5	Null

Table 9: Comparison of ResNet18 Performance on the ChestX-Ray14 Dataset Using Transferred Optimized Hyperparameters and Unoptimized Hyperparameters (Best Results are Underlined)

Configuration	AUC (%)	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1_score (%)
Unoptimized Hyperparameters	87.36	88.72	55.34	26.06	97.17	26.06
Optimized Hyperparameters	<u>88.48</u>	<u>88.97</u>	<u>57.42</u>	<u>27.6</u>	<u>97.42</u>	<u>37.28</u>

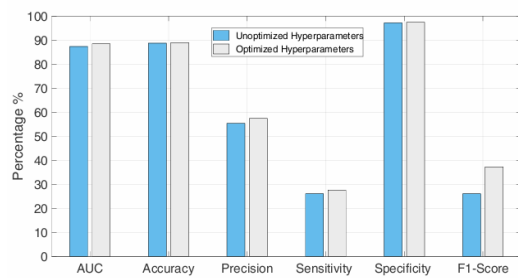


Fig. 12: Comparison of ResNet18 Performance on the ChestX-Ray14 Dataset using Transferred Optimized Hyperparameters and Unoptimized Hyperparameters

Table 10: Comparative Evaluation of ResNet18 performance Using Optimized Hyperparameters vs. State-of-the-Art Based on AUC Metric (Best Results are Underlined)

Ref	AUC (%)
Li <i>et al.</i> (2024)	88.1
Wang <i>et al.</i> (2024)	87.7
Jiang <i>et al.</i> (2024)	84.2
Kufel <i>et al.</i> (2023)	87.9
Liu <i>et al.</i> (2023)	84.7
Mezina & Burget (2024)	81.3
Wu <i>et al.</i> (2023)	83.7
Zhu <i>et al.</i> (2022)	84.1
Mao <i>et al.</i> (2022)	87.4
Yang (2025)	87.0
Huangsuwan <i>et al.</i> (2025)	77.8
Guan <i>et al.</i> (2023)	82.4
Ours	<u>88.48</u>

Discussion

In this paper, we explored the application of pretrained models for detecting pleural effusion in chest X-rays, with a particular focus on the impact of optimization in enhancing model performance. Several pretrained models, including SqueezeNet, GoogleNet, VGG16, DarkNet19, ResNet18, and ResNet50, were evaluated on the Small ChestX-Ray14 dataset, a balanced subset of the ChestX-Ray14 dataset. The creation of this smaller dataset addressed the class imbalance present in the original dataset and reduced the computational load typically associated with model optimization. The dataset was divided into 80% for training and 20% for testing, and model performance was assessed using AUC, accuracy, precision, sensitivity (recall), specificity, and F1 score metrics.

The integration of the GWO with the ResNet18 model demonstrated significant performance improvements, achieving an AUC of 95.1% on the Small Chest-XXray14 dataset, compared to an AUC of 93.3% with the baseline ResNet18 model. The GWOResNet18 model demonstrated overall performance improvements across multiple metrics, including AUC, accuracy, precision, specificity, and F1-score, compared to the baseline ResNet18. These improvements can be attributed to key hyperparameter optimizations. The

increased mini-batch size (64 → 97) likely contributed to better generalization by stabilizing weight updates. The adjusted pixel and scale ranges may have enhanced data augmentation, improving robustness to variations in the dataset. Additionally, increasing the initial learning rate (0.001 → 0.01) allowed the model to learn faster in the early stages, while the learning rate drop factor (0.0004) helped fine-tune the training by gradually reducing the learning rate. This adjustment likely sped up convergence and prevented the model from getting stuck in suboptimal solutions (local minima).

However, a trade-off was observed in sensitivity, which slightly declined. This is reflected in the confusion matrices, where the optimized model reduced false positives (450 → 235), enhancing specificity, but showed an increase in false negatives (256 → 340), leading to lower sensitivity. This suggests that while the optimization improved the model's confidence in correctly identifying healthy cases, it became slightly more conservative in detecting sick cases. Nonetheless, the overall gains in performance metrics indicate the effectiveness of the optimization in enhancing model reliability.

This result confirms our hypothesis from the Introduction, where we anticipated that optimization would enhance the model's performance. The GWO's strength lies in its balance between exploration and exploitation, allowing it to efficiently search the solution space. The hierarchical structure and adaptive search mechanism of GWO contributed to robust hyperparameter tuning, ultimately improving the detection accuracy for pleural effusion, a crucial task in medical image analysis.

Furthermore, the GWO-optimized ResNet18 model achieved a more balanced performance compared to the baseline ResNet18 model. While the baseline model exhibited a higher sensitivity (recall) of 90.38%, it suffered from lower specificity (82.67%), reflecting a tendency to favor positive cases at the expense of false positives. The optimization process shifted this balance, resulting in a sensitivity (recall) of 87.22% and an improved specificity of 90.96%. This demonstrates that the GWO optimization adjusted the model's decision boundary to reduce false positives, achieving a significant gain in specificity (+8.29%) while incurring a moderate reduction in sensitivity (recall) (-3.16%). This balance highlights the impact of optimization metrics, such as AUC, which consider both true positive and true negative rates. While the overall improvement underscores the potential of optimization techniques like GWO in enhancing pretrained models, the observed trade-off emphasizes the need for future optimization strategies to explicitly prioritize sensitivity (recall) alongside other metrics. This would ensure a more balanced improvement across all performance measures, addressing the critical requirements of medical diagnostic applications.

To further explore the effect of optimization, we transferred the optimized hyperparameters obtained from the GWO-ResNet18 model on the Small ChestXRay14 dataset to the full ChestX-Ray14 dataset. We focused only on the ResNet18 model here, as the goal was to evaluate the effect of transferring the optimized hyperparameters. The results showed that transferring optimized hyperparameters we get from GWO-ResNet18 on the small dataset, improved model performance on the larger dataset, with the best AUC reaching 88.48%. This outperformed both the performance of ResNet18 with unoptimized hyperparameters, and state-of-the-art results. These findings suggest that hyperparameters optimized on a smaller, balanced dataset can effectively enhance performance on a larger, imbalanced dataset, though with some limitations.

Compared to state-of-the-art, our approach offers some advancements. While most prior research focused solely on AUC as the primary evaluation metric, we provided a more comprehensive evaluation using a broader set of metrics, including accuracy, precision, sensitivity (recall), specificity, and F1 score. Furthermore, previous studies did not apply optimization techniques nor address the problem of dataset imbalance. By introducing GWO-based optimization and tackling the dataset imbalance, our paper presents a novel approach to improving pleural effusion detection performance in medical image classification.

Despite these promising results, several limitations should be acknowledged. Transferring the optimized hyperparameters had a limited positive effect when applied to the entire ChestX-Ray14 dataset, possibly due to the complexity and size of the dataset. This was compounded by the time constraints that restricted us from applying optimization directly to the full dataset. Additionally, increasing the number of search agents or the maximum number of iterations in the optimization process might yield better results, but this would require a substantial computational investment. These constraints highlight the need for further exploration of optimization techniques in future research.

Future work should focus on applying optimization directly to larger datasets, such as the full ChestXRay14 dataset. Exploring other evolutionary algorithms may also provide further enhancements in model performance, potentially surpassing the improvements observed with GWO. Overcoming computational challenges and addressing the optimization demands in large datasets remains a critical avenue for future research.

Conclusion

Accurate and efficient detection of pleural effusion in chest X-rays is critical for improving diagnostic outcomes in medical imaging. In this paper, we demonstrated the effectiveness of utilizing an optimized

pretrained model for detecting pleural effusion, with a focus on the novel application of GWO to optimize the ResNet18 model. While standard pretrained models, such as VGG16 and ResNet18, performed well on the Small ChestX-Ray14 dataset, achieving AUCs of 94.7% and 93.3%, respectively, the optimized GWO-ResNet18 model achieved superior results with an AUC of 95.1%. This highlights the significant impact of optimization in enhancing pretrained model performance. Furthermore, a key contribution of this paper was the transfer of hyperparameters optimized on the Small ChestXRay14 dataset to the full ChestX-Ray14 dataset. This approach not only addressed computational constraints but also improved the performance of the ResNet18 model on the larger dataset, achieving an AUC of 88.48%, compared to 87.36% with unoptimized hyperparameters. These results underscore GWO as an effective optimization strategy for improving pretrained models in medical image analysis, particularly under resource constraints. The ability to fine-tune pretrained models efficiently offers practical benefits, such as reducing computational costs and training time—crucial factors in resource-limited medical environments. This work emphasizes the importance of integrating optimization algorithms like GWO into model development to enhance diagnostic performance. Future research could explore applying the optimization process directly to larger datasets, such as the full ChestX-Ray14 or CheXpert datasets, as computational resources become more available. Additionally, investigating other evolutionary optimization algorithms may yield further improvements and expand the applicability of this approach to broader tasks in medical image classification. Furthermore, integrating optimized models into real-time diagnostic systems could enhance their practical use in hospitals and clinical settings. Finally, future studies could expand this approach to other medical imaging tasks, such as detecting lung nodules or pneumonia, to assess its generalizability.

Authors's Contributions

Rehab Fathi Ibrahim: Contributed to the conception and design of the study, collected data, participated in data analysis and interpretation, drafted parts of the manuscript, and reviewed and approved the final version.

Nashwa Mohamed Elsayed: Participated in data analysis and interpretation, contributed to manuscript drafting, and reviewed and approved the final version of the manuscript.

Alshaimaa Mostafa Mohammed: Contributed to the conception and design of the study, participated in manuscript drafting, and reviewed and approved the final version.

Ahmed Magdy Mohamed: Identified and developed the research problem, contributed to the study's conception and design, participated in data analysis and interpretation, contributed to manuscript drafting, and reviewed and approved the final version.

Ethics

The authors declare that there are no ethical issues or conflicts of interest to report regarding the present.

References

- Abdellatif, A., Abdellatef, H., Kanesan, J., Chow, C.-O., Chuah, J. H., & Gheni, H. M. (2022). An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods. *IEEE Access*, 10, 79974–79985.
<https://doi.org/10.1109/access.2022.3191669>
- Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidiq, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S., & Asari, V. K. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics*, 8(3), 292–369.
<https://doi.org/10.3390/electronics8030292>
- Atban, F., Ekin, E., & Garip, Z. (2023). Traditional Machine Learning Algorithms for Breast Cancer Image Classification with Optimized Deep Features. *Biomedical Signal Processing and Control*, 81, 104534.
<https://doi.org/10.1016/j.bspc.2022.104534>
- Chen, Z., Chen, K., Lou, Y., Zhu, J., Mao, W., & Song, Z. (2021). Machine Learning Applied to Near-Infrared Spectra for Clinical Pleural Effusion Classification. *Scientific Reports*, 11(1), 9411.
<https://doi.org/10.1038/s41598-021-87736-4>
- Džakmić, Š. (2020). Evaluation of ResNet Network for Semantic Segmentation of Coral Reefs. *International Journal of Emerging Trends in Engineering Research*, 8(1.1), 54–61.
<https://doi.org/10.30534/ijeter/2020/0881.12020>
- Faris, H., Aljarah, I., Al-Betar, M. A., & Mirjalili, S. (2018). Grey Wolf Optimizer: A Review of Recent Variants and Applications. *Neural Computing and Applications*, 30(2), 413–435.
<https://doi.org/10.1007/s00521-017-3272-5>
- Gamal, G., Mohamed, F. A., Abdelazeem, A. A., & Ahmed, M. (2022). Enhancement of Spectrum Recovery in Cognitive Radio Using Artificial Intelligence based on Grey Wolf Optimizer. *Journal of System and Management Sciences*, 12(6), 439–453.
<https://doi.org/10.33168/jsms.2022.0626>
- Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., & De, D. (2020). Fundamental Concepts of Convolutional Neural Network. *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, 172, 519–567.
https://doi.org/10.1007/978-3-030-32644-9_36
- Guan, Q., Chen, Q., & Huang, Y. (2023). An Improved Heteroscedastic Modeling Method for Chest X-ray Image Classification with Noisy Labels. *Algorithms*, 16(5), 239.
<https://doi.org/10.3390/a16050239>
- Huangsuwan, K., Liu, T., See, S., Beng Ng, A., & Vateekul, P. (2025). FedDrip: Federated Learning With Diffusion-Generated Synthetic Image. *IEEE Access*, 13, 10111–10125.
<https://doi.org/10.1109/access.2025.3525806>
- Islam, Md. M., Tasnim, N., & Baek, J.-H. (2020). Human Gender Classification Using Transfer Learning Via Pareto Frontier CNN Networks. *Inventions*, 5(2), 16.
<https://doi.org/10.3390/inventions5020016>
- Jiang, X., Zhu, Y., Liu, Y., Cai, G., & Fang, H. (2024). Transdd: A Transformer-Based Dual-Path Decoder for Improving the Performance of Thoracic Diseases Classification Using Chest X-Ray. *Biomedical Signal Processing and Control*, 91, 105937. <https://doi.org/10.1016/j.bspc.2023.105937>
- Kohli, M., & Arora, S. (2018). Chaotic Grey Wolf Optimization Algorithm for Constrained Optimization Problems. *Journal of Computational Design and Engineering*, 5(4), 458–472.
<https://doi.org/10.1016/j.jcde.2017.02.005>
- Kufel, J., Bielówka, M., Rojek, M., Mitreǵa, A., Lewandowski, P., Cebula, M., Krawczyk, D., Bielówka, M., Kondoł, D., Bargiel-łaczek, K., Paszkiewicz, I., Czogalik, ł., Kaczyńska, D., Wocław, A., Gruszczyńska, K., & Nawrat, Z. (2023). Multi-Label Classification of Chest X-ray Abnormalities Using Transfer Learning Techniques. *Journal of Personalized Medicine*, 13(10), 1426.
<https://doi.org/10.3390/jpm13101426>
- Li, D., Huo, H., Jiao, S., Sun, X., & Chen, S. (2024). Automated Thorax Disease Diagnosis Using Multi-Branch Residual Attention Network. *Scientific Reports*, 14(1), 11865.
<https://doi.org/10.1038/s41598-024-62813-6>
- Liu, X.-T., Dong, X.-L., Zhang, Y., Fang, P., Shi, H.-Y., & Ming, Z.-J. (2022). Diagnostic Value and Safety of Medical Thoracoscopy for Pleural Effusion of Different Causes. *World Journal of Clinical Cases*, 10(10), 3088–3100.
<https://doi.org/10.12998/wjcc.v10.i10.3088>
- Liu, Z., Cheng, Y., & Tamura, S. (2023). Multi-Label Local to Global Learning: A Novel Learning Paradigm for Chest X-Ray Abnormality Classification. *IEEE Journal of Biomedical and Health Informatics*, 27(9), 4409–4420.
<https://doi.org/10.1109/jbhi.2023.3281466>
- Maeda-Gutiérrez, V., Galván-Tejada, C. E., Zanella-Calzada, L. A., Celaya-Padilla, J. M., Galván-Tejada, J. I., Gamboa-Rosales, H., Luna-García, H., Magallanes-Quintanar, R., Guerrero Méndez, C. A., & Olvera-Olvera, C. A. (2020). Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases. *Applied Sciences*, 10(4), 1245.
<https://doi.org/10.3390/app10041245>

- Magdy, A., Hussein, H., Abdel-Kader, R. F., & Salam, K. A. E. (2023). Performance Enhancement of Skin Cancer Classification Using Computer Vision. *IEEE Access*, 11, 72120–72133.
<https://doi.org/10.1109/access.2023.3294974>
- Mao, C., Yao, L., & Luo, Y. (2022). ImageGCN: Multi-Relational Image Graph Convolutional Networks for Disease Identification With Chest X-Rays. *IEEE Transactions on Medical Imaging*, 41(8), 1990–2003.
<https://doi.org/10.1109/tmi.2022.3153322>
- Mezina, A., & Burget, R. (2024). Detection of Post-Covid-19-Related Pulmonary Diseases in X-Ray Images Using Vision Transformer-Based Neural Network. *Biomedical Signal Processing and Control*, 87, 105380.
<https://doi.org/10.1016/j.bspc.2023.105380>
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46–61.
<https://doi.org/10.1016/j.advengsoft.2013.12.007>
- Mostafa, F. A., Elrefaei, L. A., Fouda, M. M., & Hossam, A. (2022). A Survey on AI Techniques for Thoracic Diseases Diagnosis Using Medical Images. *Diagnostics*, 12(12), 3034.
<https://doi.org/10.3390/diagnostics12123034>
- Nguyen, T. H. B., Park, E., Cui, X., Nguyen, V. H., & Kim, H. (2018). fPADnet: Small and Efficient Convolutional Neural Network for Presentation Attack Detection. *Sensors*, 18(8), 2532.
<https://doi.org/10.3390/s18082532>
- Rakshit, S., Saha, I., Wlasnowolski, M., Maulik, U., & Plewczynski, D. (2019). Deep Learning for Detection and Localization of Thoracic Diseases Using Chest X-Ray Imagery. *Artificial Intelligence and Soft Computing*, 11509, 271–282.
https://doi.org/10.1007/978-3-030-20915-5_25
- Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517–6525.
<https://doi.org/10.1109/cvpr.2017.690>
- Sánchez, D., Melin, P., & Castillo, O. (2017). A Grey Wolf Optimizer for Modular Granular Neural Networks for Human Recognition. *Computational Intelligence and Neuroscience*, 2017(1), 1–26.
<https://doi.org/10.1155/2017/4180510>
- Sanida, M. V., Sanida, T., Sideris, A., & Dasygenis, M. (2024). An Advanced Deep Learning Framework for Multi-Class Diagnosis from Chest X-ray Images. *J*, 7(1), 48–71.
<https://doi.org/10.3390/j7010003>
- Sarwinda, D., Paradisa, R. H., Bustamam, A., & Anggia, P. (2021). Deep Learning in Image Classification Using Residual Network (ResNet) Variants for Detection of Colorectal Cancer. *Procedia Computer Science*, 179, 423–431.
<https://doi.org/10.1016/j.procs.2021.01.025>
- Shafiq, M., & Gu, Z. (2022). Deep Residual Learning for Image Recognition: A Survey. *Applied Sciences*, 12(18), 8972.
<https://doi.org/10.3390/app12188972>
- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations*, 1–14.
<https://doi.org/10.48550/arXiv.1409.1556>
- Singh, S. (2024). *Multimodal Machine Learning for Medical Imaging published dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy*.
<https://doi.org/10.25949/24882099.V1>
- Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE*, 105(12), 2295–2329.
<https://doi.org/10.1109/jproc.2017.2761740>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).
<https://doi.org/10.1109/cvpr.2015.7298594>
- Usman, M., Zia, T., & Tariq, A. (2022). Analyzing Transfer Learning of Vision Transformers for Interpreting Chest Radiography. *Journal of Digital Imaging*, 35(6), 1445–1462.
<https://doi.org/10.1007/s10278-022-00666-z>
- Wang, G., Wang, P., & Wei, B. (2024). Multi-Label Local Awareness and Global Co-Occurrence Prior Learning Improve Chest X-Ray Classification. *Multimedia Systems*, 30(3), 132.
<https://doi.org/10.1007/s00530-024-01321-z>
- Wang, J., Zhu, H., Wang, S.-H., & Zhang, Y.-D. (2021). A Review of Deep Learning on Medical Image Analysis. *Mobile Networks and Applications*, 26(1), 351–380.
<https://doi.org/10.1007/s11036-020-01672-7>
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3462–3471.
<https://doi.org/10.1109/cvpr.2017.369>
- Wu, X., Feng, Y., Xu, H., Lin, Z., Chen, T., Li, S., Qiu, S., Liu, Q., Ma, Y., & Zhang, S. (2023). TransCNN: Combining Transformer and CNN in Multilabel Medical Image Classification. *Knowledge-Based Systems*, 281, 111030.
<https://doi.org/10.1016/j.knosys.2023.111030>
- Yang, J. (2025). AFM-DViT: A framework for IoT-driven medical image analysis. *Alexandria Engineering Journal*, 113, 294–305.
<https://doi.org/10.1016/j.aej.2024.10.118>