

Pneumonia Detection from Lung X-ray Images using Local Search Aided Sine Cosine Algorithm based Deep Feature Selection Method

Soumitri Chattopadhyay

Department of Information Technology, Jadavpur University
 Plot No.8, Salt Lake Bypass, LB Block, Sector III, Salt Lake City, Kolkata -
 700106, INDIA
 Email: soumitri.chattopadhyay@gmail.com

Rohit Kundu

Department of Electrical Engineering, Jadavpur University
 188, Raja S.C. Mullick Road, Kolkata-700032, INDIA
 Email: rohitkunduju@gmail.com

Pawan Kumar Singh

Department of Information Technology, Jadavpur University
 Plot No.8, Salt Lake Bypass, LB Block, Sector III, Salt Lake City, Kolkata -
 700106, INDIA
 Email: pawansingh.ju@gmail.com

Seyedali Mirjalili*

Centre for Artificial Intelligence Research and Optimization, Torrens
 University, Australia
 Email: ali.mirjalili@gmail.com

Ram Sarkar

Department of Computer Science and Engineering, Jadavpur University
 188, Raja S.C. Mullick Road, Kolkata-700032, INDIA
 Email: ramjucse@gmail.com

*Corresponding author: Seyedali Mirjalili

*Corresponding author email: ali.mirjalili@gmail.com

Pneumonia Detection from Lung X-ray Images using Local Search Aided Sine Cosine Algorithm based Deep Feature Selection Method

Soumitri Chattopadhyay¹, Rohit Kundu² , Pawan Kumar Singh¹ , and Seyedali Mirjalili^{3,4,*}  Ram Sarkar⁵ 

¹ Department of Information Technology, Jadavpur University, Kolkata - 700106, India

² Department of Electrical Engineering, Jadavpur University, Kolkata - 700032, India
³ Centre for Artificial Intelligence Research and Optimization, Torrens University, Australia

⁴ Yonsei Frontier Lab, Yonsei University, Korean

⁵ Department of Computer Science & Engineering, Jadavpur University, Kolkata-700032, INDIA

Abstract. Pneumonia is a major cause of death among children below the age of five years, globally. It is especially prevalent in developing and under-developed nations where the risk factors for the disease such as unhygienic living conditions, high levels of pollution and overcrowding are higher. Radiological examination (usually X-ray scans) is conducted to detect pneumonia, yet it is prone to subjective variability and can lead to disagreements among different radiologists. To detect traces of pneumonia from X-ray images, a more robust method is therefore required, which can be achieved by using a Computer-Aided Diagnosis (CAD) system. In this research, we develop a two-stage framework, using the combination of deep learning and optimization algorithms, which is both accurate and time-efficient. In its first stage, the proposed framework extracts feature using a customized deep learning model called DenseNet-201 following the concept of transfer learning to cope with the scanty available data. In the second stage, we then reduce the feature dimension using an improved Sine Cosine Algorithm (SCA) equipped with Adaptive Beta Hill Climbing ($A\beta HC$) based local search algorithm. The optimized feature subset is utilized for the classification of “Pneumonia” and “Normal” X-ray images using a Support Vector Machines classifier (SVM). Upon an evaluation on a publicly available dataset, the proposed method demonstrates the highest accuracy of 98.36% and sensitivity of 98.79% with a feature reduction of 85.55% (74 features selected out of 512), using a 5-fold cross-validation scheme. Extensive additional experiments on continuous benchmark functions as well as the CEC-2017 test suite further showcase the superiority and suitability of our proposed approach in application to real-valued optimization problems. The relevant codes for the proposed method can be found in: [GitHub](#).

* Corresponding Author
Email: ali.mirjalili@gmail.com

Keywords: Pneumonia Detection · Lung X-Ray · Meta-heuristic · Local Search Optimization · Deep Learning

1 Introduction

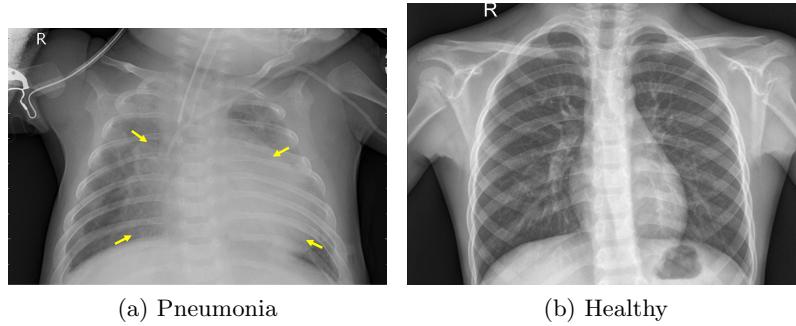


Fig. 1: Sample lung X-ray images of (a) a pneumonia infected individual, and (b) a healthy individual. The yellow arrows in the pneumonic X-ray show the white infiltrates that indicate the pneumonia infection.

Pneumonia is a respiratory disorder, where the alveoli or air sacs of the lungs get filled with fluids like water, puss, etc. The causative agent of pneumonia can be any microorganism- bacteria, viruses or fungi. Pediatric pneumonia is a common occurrence in developing and under-developed nations where along with the problem of overpopulation, necessary hygienic standards cannot be maintained. It accounts for more than 15% of the children (below 5 years of age) mortality [74]. Such nations also lack the medical resources for the regular screening of pneumonia affected individuals. Early diagnosis of the disease can cure the disease before it becomes fatal to life, thus decreasing the mortality rate significantly. The most common form of diagnosis of pneumonia is the examination of chest X-ray scans. Pneumonia appears as white infiltrates inside the lungs, as shown in Figure 1. However, detection of pneumonia from X-rays is highly dependent on the interpretative capability of the radiologists [60, 75], and can lead to early stages of pneumonia go unnoticed due to this subjective variability. Thus, to detect pneumonia traces in lung X-rays, an automation-assisted Computer-Aided Diagnosis (CAD) framework is needed to make accurate predictions.

Nowadays, deep learning has become an indispensable tool in Artificial Intelligence (AI) that uses Artificial Neural Networks (ANNs) for various pattern recognition and image classification problems [46, 67, 27, 47]. ANNs mimic the way humans learn through examples. Convolutional Neural Networks (CNNs) have been widely used for many complex image classification tasks since they

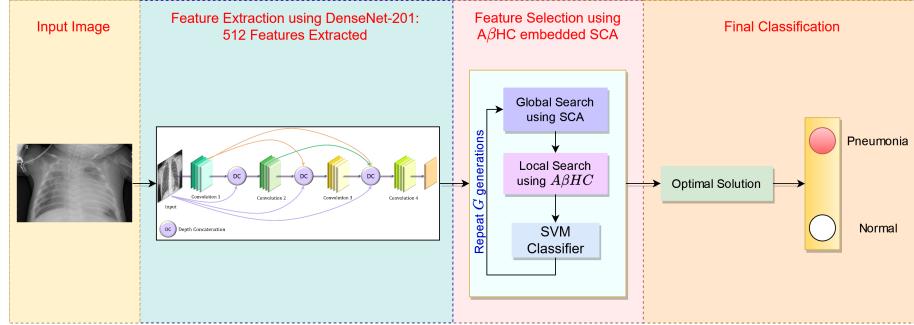


Fig. 2: The overview of the proposed framework for pneumonia detection using chest X-ray images: extracted feature set from DenseNet-201 is of dimensionality 512, and the feature selection steps using SCA global search which is explained in [Figure 4](#) and the local search algorithm A β HC is explained in [Figure 5](#).

help in reducing the number of trainable parameters through convolution and pooling operations. Moreover, they are translationally invariant and perform better than traditional Machine Learning models. However, such models require a large amount of data to perform optimally, acquiring which is challenging, especially in the biomedical domain where experts in the field need to manually annotate the images. Transfer learning is thus used in such cases, where the model weights trained on a large dataset are saved and then fine-tuned in the current problem having fewer samples. CNN models trained on the ImageNet dataset [20], containing 14 million images classified into 1000 classes, are used by many researchers in biomedical image classification problems. In the present research, we use the DenseNet-201 [36] CNN model using the concept of transfer learning for extracting features from the pneumonia chest X-ray images.

The number of features extracted from the DenseNet-201 model is quite large in dimension, and so we have applied a feature selection approach to reduce the dimensionality of the feature space to make the overall model computationally efficient. Selecting a subset of features from the original feature set using exhaustive search is an NP-hard problem, and therefore researchers rely on meta-heuristic algorithms to solve this problem instead [22, 29, 30, 31]. In this research, we use a local search embedded feature selection approach [73, 9], where a mathematics inspired meta-heuristic, called the Sine-Cosine Algorithm (SCA) [56], is used to perform a global search for the optimal feature set and the Adaptive Beta Hill Climbing (A β HC) algorithm [11] is used for performing a local search on the SCA output to enhance its performance. The final feature subset selected by this A β HC embedded SCA algorithm is fed to a Support Vector Machine (SVM) classifier to make the final predictions.

1.1 Motivation and Contributions

Bacterial or viral pneumonia affects a large fraction of the population, especially in underdeveloped and developing nations where living standard is poor and there are risk factors like overcrowding, bolstering the spread of the disease. Along with the associated risk factors, the insufficient medical infrastructure in such countries further exacerbates the proper diagnosis of pneumonia. Radiological examinations are performed as the preliminary diagnosis of pneumonia, but detecting traces in the early form of the disease is a daunting task even for experts, and thus this leads to subjective variability. Hence, a robust CAD framework needs to be developed for the early diagnosis of the disease. Most of the early methods for automatic pneumonia detection focused on performing end-to-end classification using CNN models. In this research, we consider the feature dimension reduction along with providing a more efficient framework for pneumonia detection following a local search embedded feature selection method. Before that, a pre-trained DenseNet-201 CNN model has been used for feature extraction to address the problem of scanty data.

According to the “No Free Lunch” theorem [76], there is no single meta-heuristic algorithm that can provide an optimal solution for all types of complex optimization problems. In the current study on pneumonia detection, the SCA embedded with the A β HC algorithm is used which gives the optimal results i.e., the optimal number of features has been selected from the original feature vector generated by a CNN model. Hence, it has been chosen as the optimization framework for this feature selection problem. The results justifying the same are shown in Section 4.3.

The key contributions of the present research are as follows:

1. A deep feature selection framework is proposed to reduce the dimensionality of the feature space, thus reducing the storage and time requirements significantly while enhancing the performance by eliminating the redundant features.
2. SCA [56] is used for performing the global search of the feature space formed by features generated by the DenseNet-201 CNN model from the lung X-ray scans.
3. To augment the performance of SCA, the A β HC algorithm [11] is used for performing a local search on the obtained feature subset. This reduces the feature space substantially by improving the exploitation capability of the SCA.
4. The proposed framework has been evaluated on a publicly available pneumonia lung X-ray dataset [43], achieving 98.36% accuracy and 98.79% sensitivity upon 5-fold cross-validation, outperforming state-of-the-art methods, while reducing the feature dimensionality by 85.55%.

2 Related Work

Detection of pneumonia from lung X-ray images using CAD systems [49, 38, 45] is a challenging task since the amount of publicly accessible labelled data is lim-

ited. To classify X-ray images into different classes, the expertise of radiologists is required which is an expensive and time-consuming process. Conventional Machine Learning methods have been used for pneumonia image classification, like, Kuo et al. [48] extracted 11 features from 185 schizophrenic patient data, like age, smoking habits, etc., and used classification and regression models like SVM, decision trees, logistic regression, etc. to provide a comparative analysis. The authors found that the decision trees classifier yielded the highest accuracy of 94.5%. Chandra et al. [16] performed segmentation of the lungs from X-ray scans and extracted 8 statistical features for classification using 5 different traditional classifiers, namely, logistic regression, sequential minimal optimization (SMO), classification via regression, random forest and multi-layer perceptron (MLP) models, achieving the highest accuracy of 95.39% on a 412-image dataset. Yue et al. [81] extracted 5 features from Computed Tomography (CT) scan images of 52 patients for pneumonia classification and obtained 97% as the best area under curve (AUC) value. The results obtained from these methods justify that they are not suitable for practical applications, and also, being evaluated on such small datasets make them hard to generalize.

Deep learning-based methods [63, 71, 39, 44], on the other hand, do not require handcrafted features to be extracted from the images for classification, instead the models learn the relevant informative features automatically. CNN based models are preferred for image classification problems since they efficiently extract translationally invariant features from the images using the convolution of the input image and the filters. They are computationally efficient for processing image data, where each pixel represents a feature since the trainable parameters are controlled through the pooling operations.

Rajaraman et al. [63] proposed a novel visualization strategy for region of interest (RoI) localization from chest X-ray images and used a customized VGG-16 model to obtain a classification accuracy of 96.2% for pneumonia detection. The authors of [66] proposed a 10-layer CNN model and obtained a mean classification accuracy of 95.30% following a 5-fold cross-validation scheme. Both of these works used the lung X-ray dataset by Kermany et al. [43] for their experiments. Stephen et al. [71] and Sharma et al. [68] proposed CNN based models for end-to-end classification of chest X-ray images for pneumonia detection. To account for the scarcity of available data, they used image augmentation techniques. Stephen et al. got 93.73% accuracy and Sharma et al. got 90.68% accuracy upon an evaluation on the dataset by Kermany et al. [43]. The results are not satisfactorily high for use in the field. Data augmentation techniques such as image rotation, scaling, etc., add little information for the CNNs to learn from and thus do not provide a major boost in performance. Liang et al. [50] proposed a 49-layer residual CNN model and obtained 90.50% accuracy, while Rajpukar et al. [64] proposed a 121-layer CNN model and achieved only 76.8% f1-score for the pneumonia classification problem. Among other recent works, the authors of [54] extracted statistical as well as deep features and performed feature selection on the combined feature set which yielded a classification accuracy of 86.3% on the pneumonia detection task, and Dey et al. [21] leveraged ensemble feature

scheme that combined handcrafted features with deep features obtained using a customized VGG-19 model, which upon final classification using a Random Forest classifier yielded the highest accuracy of 97.94% on the dataset by Kermany et al. [43].

Transfer learning [82, 4] has been widely used in biomedical image classification to account for the data unavailability. Ayan et al. [13] proposed a transfer learning-based fine-tuning approach using Xception and VGG-16 models, obtaining 82% and 87% classification accuracies respectively on the lung X-ray dataset by Kermany et al. [43]. Rahman et al. [62], Zubair et al. [83] and Ibrahim et al. [37] used such methods in the recent past for pneumonia image classification. Rahman et al. used DenseNet-201 and obtained 98% accuracy, Zubair et al. used VGG-16 and obtained 96.6% accuracy, while Ibrahim et al. used AlexNet and obtained 94.43% accuracy on the dataset by Kermany et al. [43]. All of the aforementioned studies used pure transfer learning for the pneumonia classification task, the CNN models being pre-trained on the ImageNet dataset [20].

Meta-heuristic algorithms [35, 2] are computational intelligence paradigms mainly used for solving different complex optimization problems. Owing to their computational efficiency as well as superior performance in resource-constrained environments, meta-heuristic algorithms have been extensively used across the domains, including feature selection [5], neural architecture search [51, 8], task scheduling [70], handwritten script classification [15], image contrast enhancement [25, 6, 52], data clustering [40], multi-level image thresholding [34, 24] and solving class imbalance problem [69, 28] among others. Mostly, these algorithms are inspired from: (1) theory of evolution, such as Genetic Algorithm [33] and Differential Evolution [19]; (2) natural behaviour of organisms, such as the Whale Optimization Algorithm [57], Cuckoo Search [77] and Flower Pollination Algorithm [79]; (3) swarm intelligence, such as the Particle Swarm Optimization [42] and the Grey Wolf Optimizer [58]; and (4) physical or scientific phenomena, such as the Gravitational Search Algorithm [65] and the Multi-Verse Optimizer [59], to name a few. Among other recent works, Karami et al. [41] proposed the Flow Direction Algorithm that mimics the flow direction of a river to the outlet point in a drainage basin; Abdollahzadeh et al. [1] proposed a novel algorithm inspired from the social lives of gorilla troops; Hashim et al. [32] proposed a physics-inspired meta-heuristic based on the Archimedes' Principle; Abualigah et al. [3] proposed a mathematics-inspired algorithm that used basic arithmetic operations as means of exploration and exploitation of the search space for global optimization; and Banerjee et al. [14] used a nature-inspired meta-heuristic, called Red Deer Algorithm, for language-independent offline signature verification.

Researchers have also leveraged local search techniques [9] in combination with global optimizers to enhance exploitation of the search space and thereby achieve a better quality solution. Chatterjee et al. [17] proposed a novel local search embedded meta-heuristic framework for influence maximization in social networks. Yousri et al. [80] combined Flower Pollination Algorithm with a fractional-order calculus based local search method for image segmentation.

Guha et al. [31] proposed a hybrid feature selection model by incorporating co-operative game theory in Genetic Algorithm [33] for human action recognition, while the authors of [26] and [7] combined the A β HC [11] local search algorithm with popular nature-inspired meta-heuristics for feature selection.

3 Proposed Method

In this section, we provide a detailed explanation of each step of the proposed framework for pneumonia detection from lung X-ray images. Figure 2 shows the overall workflow of the proposed approach where the four stages are:

1. Feature extraction using DenseNet-201
2. Global search using SCA
3. Local search using A β HC
4. Classification on the optimal feature set using the Support Vector Machine (SVM) classifier

These stages are explained in the subsequent sections.

3.1 Feature Extraction using DenseNet-201

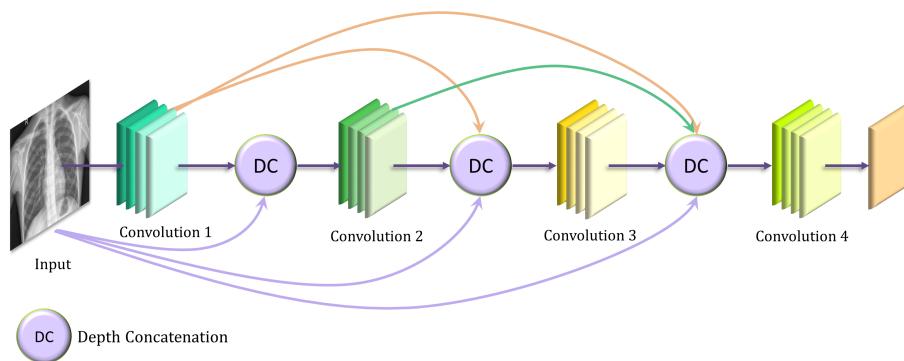


Fig. 3: Basic architecture of the DenseNet CNN model

In this research, a customized DenseNet-201 CNN model has been used for the extraction of features from the pneumonia chest X-ray images. The DenseNet architecture proposed by Huang et al. [36] provides a rich feature representation of the input images while being computationally efficient. This is because each dense layer in the DenseNet model is a concatenation of the current layer along with all of its preceding layers, thus preserving the important coarse level features. This also decreases the number of trainable parameters in the model by

controlling the new number of channels, thus making it computationally efficient. The basic DenseNet CNN model is shown in [Figure 3](#).

To capture the information effectively from the chest X-ray images using the pre-trained DenseNet-201 model, two fully connected (FC) layers have been added after flattening the final average pooling layer. The first customized FC layer is of size 4096 (FC-1) and the second is of size 512 (FC-2), following which is the final classification layer. Both the layers are associated with Leaky Rectified Linear Unit (LeakyReLU) activation function. The features have been extracted from this FC-2 layer, so we obtain a 512 size feature set to be optimized in the later stage. The flatten layer consists of 94080 units and directly mapping them to the final classification layer may lose important information. For this, we introduce these intermediate FC layers to cluster the important information before mapping to the classification layer.

3.2 Sine-Cosine Optimization Algorithm

SCA is a population-based meta-heuristic algorithm that initializes a set of random solutions and upon repeated evaluations based on an objective function through a set of constraints, the solutions evolve towards a global optimum. Since this process is stochastic in nature, a single run is not enough for the search process, however, a reasonable number of random solutions and iterations over the search space can increase the probability of finding the global optimal solution.

Two main stages in a population-based optimization algorithm are (a) exploration and (b) exploitation. For the exploration phase, the random solutions are combined by some means with a high rate of randomness to find suitable regions in the search space, while in the exploitation phase, there are gradual changes in the solutions with less randomness in the variations.

If P_i^t is the solution set at the current iteration (t) in the i^{th} dimension and x_i is the position of the destination point in the i^{th} dimension, then [Equation 1](#) is used to update the solutions in the $(t + 1)^{th}$ iteration, where r_1, r_2, r_3 and r_4 are random numbers and $r_4 \in [0, 1]$.

$$P_i^{t+1} = \begin{cases} P_i^t + r_1 \times \sin(r_2) \times |r_3 x_i^t - P_i^t|, & r_4 < 0.5 \\ P_i^t + r_1 \times \cos(r_2) \times |r_3 x_i^t - P_i^t|, & r_4 \geq 0.5 \end{cases} \quad (1)$$

As SCA has been proposed to solve continuous optimization problems where a solution comprises real values, to suit our feature selection purpose we need to map the continuous search space of SCA to a binary one. We have used the Sigmoid transfer function to do so as shown in [Equation 2](#).

$$T(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

The position of destination points x_i^{t+1} obtained from [Equation 1](#) will be updated according to [Equation 3](#).

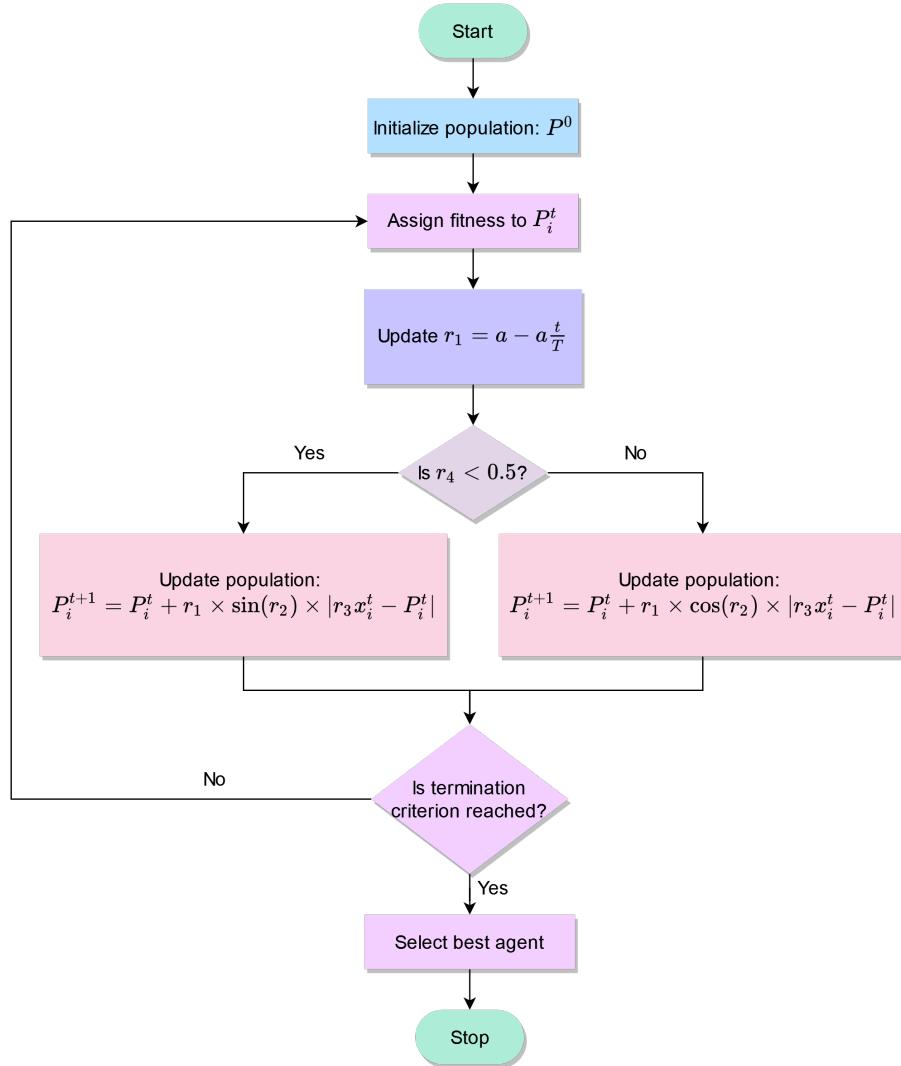


Fig. 4: Flowchart of the Sine-Cosine optimization algorithm: P_i^t represents the i^{th} individual of population P at iteration number t , T represents the maximum number of iterations and r_1, r_2, r_3 and r_4 represent random numbers.

$$x_i^{t+1} = \begin{cases} 1, & \text{rand}() < T(x_i^t) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $\text{rand}()$ yields a random number uniformly distributed between 0 and 1.

Feature selection is a multi-objective optimization problem, with the two objectives being: (a) to achieve the highest classification accuracy (i.e. maximization task), and (b) to select the lowest number of features (i.e. minimization task). It is evident that the objectives are contrasting in nature. To eliminate the contradiction, it is preferable to combine metrics of the two objectives into a single fitness function. We have used classification accuracy and feature reduction (i.e. the fraction of features reduced) for the same. This is because a higher feature reduction (i.e. less number of features selected) would imply a higher fitness value, so would imply a higher classification accuracy. The fitness function is shown in [Equation 4](#).

$$\uparrow \text{Fitness} = \alpha \times \eta + (1 - \alpha) \times \Delta \quad (4)$$

where η is the classification accuracy of the feature subset, Δ is the feature reduction given by [Equation 5](#), and $\alpha \in [0, 1]$ indicates the relative weight of the classification accuracy and the feature reduction. In our work, we have taken $\alpha = 0.99$ for all experimentation.

$$\Delta = \frac{(|D| - |S|)}{|D|} \quad (5)$$

where $|S|$ is the number of features selected, and $|D|$ is the original feature dimension. In our work, $|D| = 512$ (as specified in [Section 3.1](#)).

The four main parameters in [Equation 1](#): r_1, r_2, r_3, r_4 perform the following functions:

1. r_1 governs the direction of the next position's movement which can be the space between the solution P^t and destination x^t , or beyond it.
2. r_2 governs the amount of movement to be subtended towards or outwards from the destination.
3. r_3 allocates a random weight to the destination for stochastically emphasizing or de-emphasizing the effect of the destination x^t to define the distance, according to $r_3 > 1$ (emphasize) or $r_3 \leq 1$ (de-emphasize).
4. r_4 helps switch between the sine and cosine components.

To balance the exploration and exploitation of the SCA, the parameter r_1 is changed adaptively using [Equation 6](#) to change the range of sine and cosine components. a is a predefined constant, t is the current iteration and T is the maximum number of iterations.

$$r_1 = a \left(1 - \frac{t}{T} \right) \quad (6)$$

The flowchart for the SCA global search algorithm is shown in [Figure 4](#). The prime advantages of the SCA are as follows:

1. It has high exploration capability since it forms and improves a set of the random population (candidate solutions) and uses four randomized parameters to ensure the avoidance of the local optima.
2. If the absolute values returned by the sine and cosine functions are greater than 1 in the SCA, different regions of the search space are explored.
3. If the absolute values returned by the sine and cosine functions are less than 1, then the search space is exploited in the promising regions.
4. The SCA switches between exploration and exploitation phases smoothly according to the adaptive change in the parameter r_1 .
5. SCA tends to converge towards the best region in the search space since each candidate solution gets updated its positions around the best solution obtained so far.

3.3 Local Search: Adaptive Beta Hill Climbing

The A β HC [11] is an improved version of the β -Hill Climbing algorithm by Al-Betar et al. [10]. Suppose $\min\{g(s) \mid s \in S\}$ is the optimization problem, where $s = \{s_1, s_2, \dots, s_N\}$ is the starting provisional solution which will be moved to a neighbouring solution $s' = \{s'_1, s'_2, \dots, s'_N\}$, $g(s)$ is the objective function. Each decision variable $s_i \in S_i$, where $S = \{S_i \mid i = 1, 2, \dots, N\}$ is the possible range of values for the decision variable, i.e., $S_i \in [L_i, U_i]$, where L_i and U_i are the respective lower and upper bounds for that decision variable. N represents the number of decision variables.

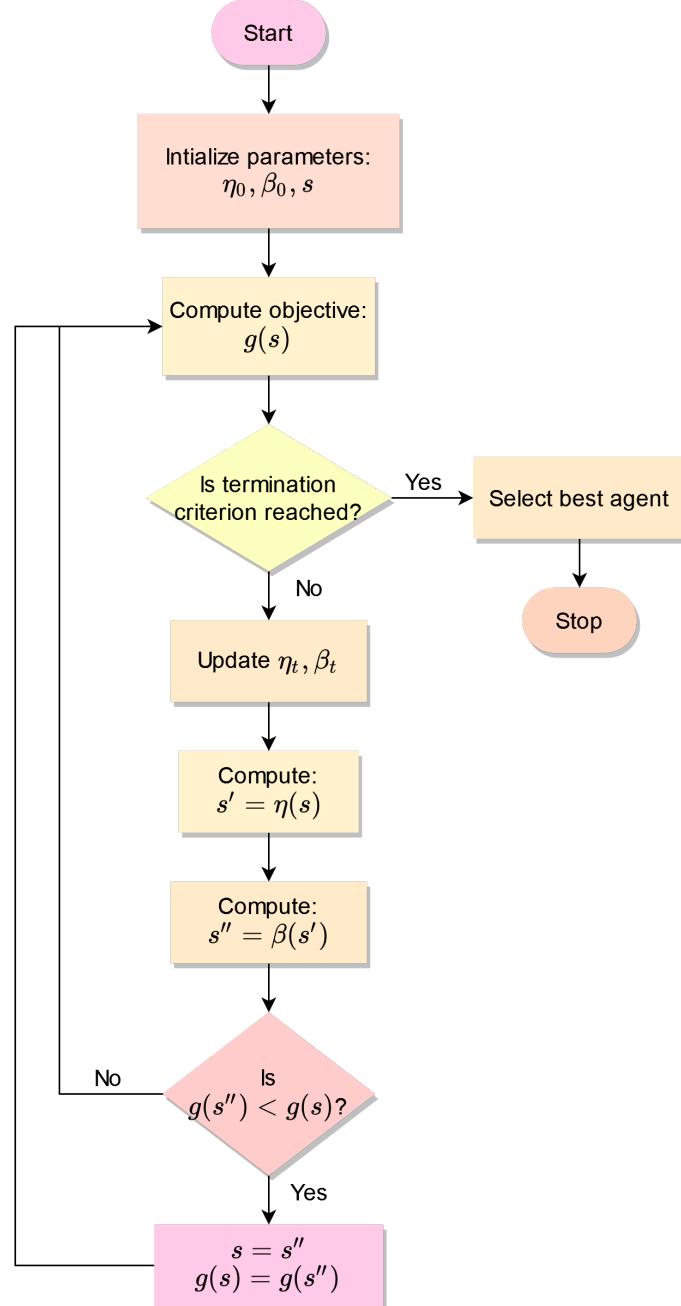
The solutions in each iteration are improved using the η -operator employing a concept called “random walk”, wherein for the solution s_i , a random neighbouring solution is visited according to [Equation 7](#), where i is randomly chosen from $\{1, 2, \dots, N\}$.

$$s'_i = s_i \pm U(0, 1) \times \eta, \quad \forall i \in [1, N] \quad (7)$$

The β operator is used as follows: for the current solution s' , the new solution s'' is calculated according to [Equation 8](#), where r is a uniform random number in $[0, 1]$, the range of β -parameter is $[0, 1]$, and $s_k \in S_i$ is randomly selected from the range of decision variable s'_i .

$$s''_i = \begin{cases} s_k, & r \leq \beta \\ s'_i, & \text{otherwise} \end{cases} \quad (8)$$

In the A β HC algorithm, the η and β parameters are updated in each iteration deterministically to control the exploration and exploitation respectively. A higher η value means a larger search space is covered, so it is reduced in each iteration starting from a value close to 1 in the first iteration according to [Equation 9](#), where t is the current iteration, T is the maximum number of iterations

Fig. 5: Flowchart of the Adaptive β -Hill Climbing algorithm

and P is a fixed number which degrades the value of η closer to 0 as the number of iterations increase.

$$\eta_t = 1 - \frac{t^{\frac{1}{P}}}{T^{\frac{1}{P}}} \quad (9)$$

The β parameter is updated within the specific range $[\beta_{min}, \beta_{max}]$ in each iteration t using [Equation 10](#). In our research, we have taken $\beta_{min} = 0.01$ and $\beta_{max} = 0.1$ for all experimentation.

$$\beta_t = \beta_{min} + \frac{t}{T} \times (\beta_{max} - \beta_{min}) \quad (10)$$

The flowchart of the A β HC algorithm is shown in [Figure 5](#).

3.4 Classification using SVM

SVM is a supervised learning model with an associated learning algorithm particularly effective for classification, numerical prediction and pattern recognition tasks. The SVM classifier was proposed by Vapnik et al. [\[18\]](#) for binary classification problems. The primary objective of the algorithm is to find the optimal hyperplane $f(W, Z) = W \cdot Z + b$ to distinguish between the two classes for a given dataset having features $Z \in \mathbb{R}^m$. In other words, for binary classification problems and given a set of training examples, an SVM training algorithm builds a model that assigns new examples to one of two classes. An SVM projects training samples to points in space to maximize the width of the gap between the two classes. New samples are then mapped into the same space and predicted a class based on which side of the separation they fall.

The parameter W is learned by the SVM using the equation mentioned in [Equation 11](#).

$$\min \frac{1}{k} \|W\|^2 + P \sum_{i=1}^k \max(0, 1 - y'_i (W^T Z_i + b)) \quad (11)$$

where, $\|W\|^2$ represents Manhattan (or $L1$) Norm, P represents penalty parameter, Z_i and y'_i denote the feature representation and the actual label of the i^{th} data item respectively, and the term $W^T Z_i + b$ represents the predictor function. The parameter k determines the trade-off between increasing the gap width and ensuring that the Z_i lie on the correct side of the hyperplane.

3.5 Analysis of Computational Complexity

We perform asymptotic analysis on the proposed A β HC-embedded SCA algorithm for feature selection. For a population size of P with the original feature dimension being D , the space complexity of the proposed algorithm is $O(P \times D)$. The time complexity in asymptotic notation is given by [Equation 12](#), where $GlobalIter$ is the number of iterations for the SCA based global searching,

$LocalIter$ is the number of iterations inside the A β HC based local searching, and $T_{fitness}$ is the time taken for fitness evaluation of an agent using [Equation 4](#).

$$\text{Complexity} = O(GlobalIter \times P \times T_{fitness} \times (D + (LocalIter \times D + T_{fitness}))) \quad (12)$$

4 Results and Discussion

In this section, we report the results obtained by the proposed framework on a publicly available pneumonia chest X-ray dataset by Kermany et al. [43] using a 5-fold cross-validation scheme. We also compare the proposed approach with some state-of-the-art methods, where the same dataset was used to justify the superiority of the proposed approach.

The pneumonia detection dataset by Kermany et al. [43] is a class-imbalanced dataset consisting of a total of 5856 chest X-ray scans out of which 1583 X-ray scans belong to class “Normal” (i.e., healthy cases), whereas 4273 X-ray scans belong to the class “Pneumonia”.

4.1 Evaluation Measures

Four commonly used evaluation measures have been considered in this study to evaluate the proposed framework on the said chest X-ray dataset, namely, Accuracy, Precision, Recall and F1-Score. We need to first define the terms “True Positives”, “False Positives”, “True Negatives” and “False Negatives” to define the said evaluation measures.

For a binary classification problem, let the two classes in the dataset be called “positive” and the “negative” classes. Then the said terms can be defined as follows:

- *True Positive* (TPo) refers to a sample belonging to the positive class being correctly classified by a model.
- *False Positive* (FPo) refers to a sample belonging to the negative class being incorrectly classified as belonging to the positive class.
- *True Negative* (TNe) refers to a sample belonging to the negative class being correctly classified by the model.
- *False Negative* (FNe) refers to a sample belonging to the positive class being incorrectly classified as belonging to the negative class.

Now the four evaluation metrics can be defined by following Equations 13, 14, 15 and 16.

$$Accuracy = \frac{TPo + TNe}{TPo + FPo + TNe + FNe} \quad (13)$$

$$Precision = \frac{TPo}{TPo + FPo} \quad (14)$$

$$\text{Recall (or Sensitivity)} = \frac{TPo}{TPo + FNe} \quad (15)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (16)$$

4.2 Implementation

Table 1: Results obtained by the proposed method on the pneumonia detection dataset using a 5-fold cross-validation scheme.

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	98.29	99.41	98.25	98.83
2	98.89	99.53	98.95	99.24
3	98.21	98.38	99.18	98.78
4	98.29	99.41	98.25	98.83
5	98.12	98.15	99.30	98.72
Avg. \pm Std. Dev.	98.36\pm0.30	98.98\pm0.66	98.79\pm0.51	98.88\pm0.21

Table 2: Accuracy and number of features selected in each stage of the proposed method in every fold of the 5-fold cross-validation scheme.

Folds	Transfer Learning: DenseNet-201		Global Search using SCA		SCA+A β HC	
	Accuracy (%)	Number of Features	Accuracy (%)	Number of Features	Accuracy (%)	Number of Features
1	97.87	512	98.12	87	98.29	76
2	98.38	512	98.72	84	98.89	70
3	96.15	512	97.86	104	98.21	72
4	96.84	512	97.78	102	98.29	78
5	97.27	512	98.12	93	98.12	76
Avg. \pm Std. Dev.	97.30\pm0.78	512\pm0	98.12\pm0.37	94\pm9	98.36\pm0.30	74\pm3

The proposed framework has been implemented using PyTorch [61] and Python3 on a 12 GB Nvidia K80 GPU. The CNN feature extractor was trained for 20 epochs using the Stochastic Gradient Descent with Momentum optimizer [72] with momentum = 0.9 and a learning rate of 0.001.

A 5-fold cross-validation scheme has been used in the present research to robustly evaluate the performance of the proposed framework on the publicly available pneumonia dataset. The results obtained on every fold of the cross-validation, and the average and standard deviation over the 5 folds are tabulated in Table 1.

Table 2 shows the accuracy values and the final number of features selected (from a feature space of 512 features) by the proposed feature selection framework. As mentioned in the preceding section, the first stage of our framework involved transfer learning-based feature extraction using a customized pre-trained

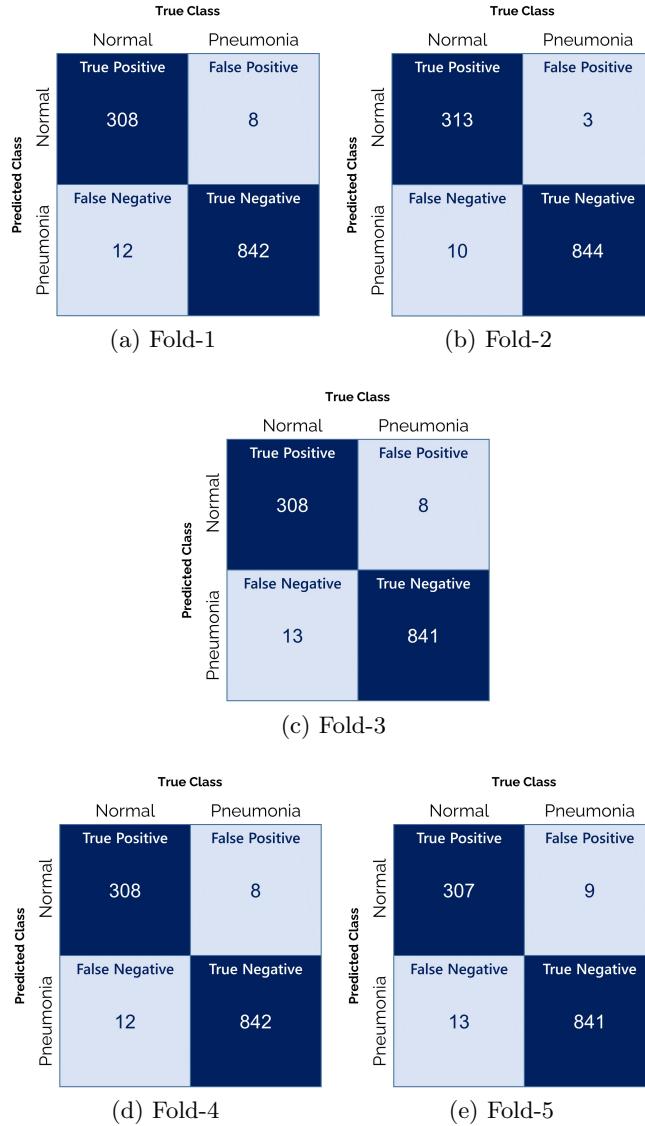


Fig. 6: Confusion matrices obtained by the proposed method using 5-fold cross-validation procedure on the pneumonia dataset.

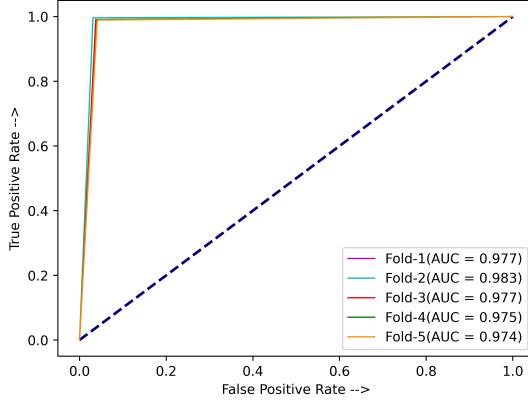


Fig. 7: Illustration of ROC curve obtained on each fold of cross-validation on the pneumonia dataset by the proposed framework and its corresponding AUC values.

DenseNet-201 model, using which a feature vector of 512 dimensions is obtained. The proposed feature selection method aims to improve upon the classification accuracy of the model as well as to reduce storage requirements by decreasing the feature space. The corresponding results obtained in the global search stage and A β HC embedded SCA have already been reported in [Table 2](#). It can be seen from [Table 2](#) that high accuracy of 98.36% and sensitivity (recall) of 98.79% have been obtained along with a feature dimension reduction of 85.55% (i.e., only 74 features are selected from the original 512 features). This justifies that the proposed framework gives a high-performance output while reducing the data storage requirements drastically.

The confusion matrices obtained by the proposed method on the five folds of cross-validation are shown in [Figure 6](#), which essentially depict the classification summary obtained upon each fold. From the figures, it is evident that the proposed approach has performed significantly well on the classification task under consideration, owing to the high True Positive (TPo) and True Negative (TNe) values obtained in each of the folds.

The Receiver Operating Characteristic (ROC) curves obtained by our proposed work on each of the folds of the cross-validation scheme are shown in [Figure 7](#). The corresponding Area Under the Curve (AUC) values is also reported for each fold. Based on these values, we can infer that the proposed method shows a high classification ability.

4.3 Comparison with state-of-the-art methods

[Table 3](#) tabulates the results obtained by comparing state-of-the-art methods with the proposed framework for pneumonia detection. The proposed method

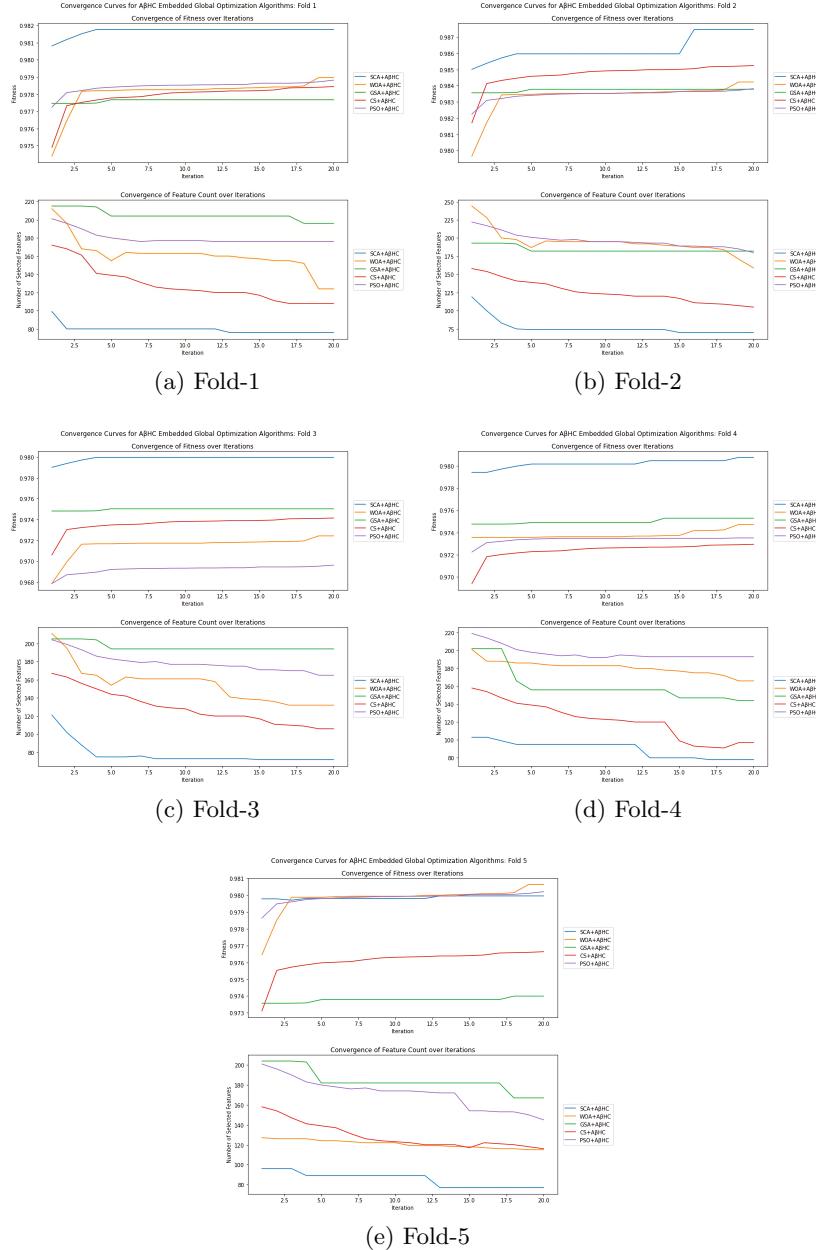


Fig. 8: Convergence plots obtained on the 5 folds of cross-validation on the pneumonia dataset by applying different algorithms.

Table 3: Comparison of the proposed feature selection framework with state-of-the-art methods on pneumonia chest X-ray dataset.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Mahmud et al. [53]	98.10	98.00	98.50	98.30
Dey et al. [21]	97.94	95.02	97.55	96.27
Zubair et al. [83]	96.60	97.20	98.10	97.65
Rajaraman et al. [63]	96.20	97.70	96.20	97.00
Saraiva et al. [66]	95.30	98.86	94.77	96.77
Ibrahim et al. [37]	94.43	-	98.19	-
Stephen et al. [71]	93.73	-	-	-
Sharma et al. [68]	90.68	-	-	-
Liang et al. [50]	90.50	89.10	96.70	92.70
Proposed Method	98.36	98.98	98.79	98.88

Table 4: Parameter settings considered for the comparative optimization algorithms for feature selection. Here, OA denotes optimization algorithm.

OA	Parameter(s)	Value(s)
PSO	Inertia weight (I) Constants (C_1, C_2)	I lies in [1 0] $C_1 = 2, C_2 = 2$
CS	Discovery rate of alien eggs (p_a)	$p_a = 0.25$
WOA	Encircling parameter (a) Shape of spiral (b)	a lies in [2 0] $b = 1$
GSA	Initial Gravitational constant ($G_{initial}$) Decay factor (α)	$G_{initial} = 6$ $\alpha = 20$

Table 5: Comparison among different global search optimization algorithms: the average over the 5-folds of cross-validation are reported. Here, OA denotes optimization algorithm and FS signifies the number of features selected.

OA	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FS
PSO [42]	97.81	99.07	97.95	98.50	200
CS [77]	97.76	98.37	98.49	98.42	176
WOA [57]	98.01	98.65	98.62	98.64	164
GSA [65]	97.84	98.62	98.10	98.37	221
SCA	98.12	98.45	99.00	98.72	94

Table 6: Comparison among different global search optimization algorithms after applying the A β HC local search optimization: the average over the 5-folds of cross-validation are reported. Here, OA denotes optimization algorithm and FS signifies the number of features selected.

OA+AβHC	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FS
PSO+A β HC	98.03	98.97	98.35	98.66	171
CS+A β HC	97.93	98.64	98.55	98.59	106
WOA+A β HC	98.08	98.51	98.76	98.63	140
GSA+A β HC	98.03	98.54	98.67	98.78	176
SCA+AβHC	98.36	98.98	98.79	98.88	74

Table 7: Results obtained on statistically analysing the SCA+A β HC algorithm with the other optimization algorithms embedded with the A β HC based local search using McNemar’s test. Here, OA denotes optimization algorithm

OA+AβHC	p-value
PSO+A β HC	1.30E-04
CS+A β HC	2.00E-02
WOA+A β HC	1.70E-04
GSA+A β HC	4.60E-03

outperforms all the previous methods in terms of all the four evaluation parameters used in this study. It is to be noted that all the methods used in this comparison perform end-to-end classification using a single CNN model, and none have used feature-dimensionality reduction methods to reduce the storage requirements. Hence, it can be inferred that our proposed framework is a highly effective and superior approach in detecting pneumonia from lung X-ray images.

The global search algorithm used in this study, i.e., SCA has been compared with the following algorithms, whose parameter settings have been described in [Table 4](#).

1. Particle Swarm Optimization (PSO) by Kennedy et al. [42]
2. Cuckoo Search (CS) algorithm by Yang et al. [77]
3. Whale Optimization Algorithm (WOA) by Mirjalili et al. [57]
4. Gravitational Search Algorithm (GSA) by Rashedi et al. [65]

The results of the aforementioned comparisons are presented in [Table 5](#). It is observed that the SCA shows the best performance in terms of both accuracy and the number of features selected. However, the WOA ranks second in performance and has greater precision than the former. Another observation is that although PSO yields relatively poor results in terms of accuracy and number of features selected than both WOA and SCA, it has the highest precision over all other global search optimization algorithms.

[Table 6](#) tabulates the results obtained after using the A β HC local search algorithm along with the aforementioned global search optimization algorithms.

It is evident from [Table 6](#) that embedding the local search algorithm enhances the performances of all the optimization algorithms in terms of both accuracy and number of features selected. However, the A β HC embedded SCA outperforms all of the other approaches in each of the evaluation metric values and also selects the lowest number of features. Further, it is also observed that the WOA+A β HC method ranks second in performance in terms of accuracy, while the PSO+A β HC method achieves a precision value closest to that of our proposed method. It is also worth noting that upon comparison among the optimization algorithm feature selection methods with and without embedding local search, the SCA shows the greatest feature reduction after A β HC has been added.

[Figure 8](#) shows the convergence curves for the local search embedded optimization algorithms positioned on the same objective space. This is done to compare their convergence behaviours in terms of both fitness and number of features selected over iterations, in each fold of the 5-fold cross-validation scheme. From all the plots shown in [Figure 8](#), it can be said that our proposed SCA+A β HC method shows a superior convergence behaviour having a greater convergence rate and also avoiding premature convergence, as compared to the other optimization algorithms.

The McNemar's statistical hypothesis test [23] is performed to statistically analyse the results obtained by the different global optimization methods embedded with the A β HC based local search. McNemar's test is a non-parametric statistical test for paired nominal distribution. [Table 7](#) shows the results obtained on performing the test. The " p -value" is the indicator for whether the null hypothesis, that the two models are statistically similar, can be rejected. A " p -value" of less than 5% indicates that the null hypothesis can be rejected, and it can be concluded that the two models that were compared were statistically different. Since in [Table 7](#), for every case $p - value < 0.05$, we can say that the framework with SCA+A β HC statistically different from the other frameworks having different global optimization algorithms.

From both [Table 5](#) and [Table 6](#), it is evident that our proposed feature selection approach outperforms the other optimization algorithms both in terms of accuracy and the final number of features selected. Moreover, [Table 7](#) justifies the fact that the proposed model is statistically dissimilar to the other frameworks.

4.4 Additional Tests: Evaluation on Continuous Optimization Problems

To further examine the robustness and applicability to other domains of engineering optimization, we extensively evaluate our proposed local search embedded approach on a variety of standard real-valued numerical optimization functions as well as compare them with other popular optimization algorithms in the literature.

Classical Benchmark Functions We apply the proposed SCA embedded with A β HC local search on 23 benchmark test functions shown in [Table 8](#). Each

Table 8: Benchmark test functions used to test the proposed SCA+ α HC algorithm. Functions F1-F7 are unimodal test functions with variable dimensions, F8-F13 are multimodal test functions with variable dimensions and F14-F23 are multimodal test functions with fixed dimensions. For variable dimension problems, the dimension has been set to 30 in this study.

Function	Description	Dimension	Range	f_{min}
F1	$f_1(x) = \sum_{i=1}^n x_i^2$	30	[-100, 100]	0
F2	$f_2(x) = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	30	[-10, 10]	0
F3	$f_3(x) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j \right)^2$	30	[-100, 100]	0
F4	$f_4(x) = \max_i \{ x_i , 1 \leq i \leq n \}$	30	[-100, 100]	0
F5	$f_5(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	30	[-30, 30]	0
F6	$f_6(x) = \sum_{i=1}^n [(x_i + 0.5)^2]$	30	[-100, 100]	0
F7	$f_7(x) = \sum_{i=1}^n tx_i^4 + \text{random}(0, 1)$	30	[-128, 128]	0
F8	$f_8(x) = \sum_{i=1}^n -x_i \sin(\sqrt{ x_i })$	30	[-500, 500]	-418.9829 $\times n$
F9	$f_9(x) = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10]$	30	[-5.12, 5.12]	0
F10	$f_{10}(x) = -20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e$	30	[-32, 32]	0
F11	$f_{11}(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	30	[-600, 600]	0
F12	$f_{12}(x) = \frac{\pi}{n} \left\{ 10 \sin(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 [1 + 10 \sin^2(\pi y_{i+1})] + (y_n - 1)^2 \right\} + \sum_{i=1}^n u(x_i, 10, 100, 4)$ $y_i = 1 + \frac{x_i + 1}{4} u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 - a & x_i < a \\ k(-x_i - a)^m & x_i \leq -a \end{cases}$	30	[-50, 50]	0
F13	$f_{13}(x) = 0.1 \left\{ \sin^2(3\pi x_1) + \sum_{i=1}^n (x_i - 1)^2 [1 + \sin^2(3\pi x_i + 1)] \right. \\ \left. + (x_n - 1)^2 [1 + \sin^2(2\pi x_n)] \right\} + \sum_{i=1}^n u(x_i, 5, 100, 4)$	30	[-50, 50]	0
F14	$f_{14}(x) = \left(\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^j (x_i - a_i)^6} \right)^{-1}$	2	[-65, 65]	1
F15	$f_{15}(x) = \sum_{i=1}^{11} \left[a_i - \frac{x_i (b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4} \right]^2$	4	[-5, 5]	1
F16	$f_{16}(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^3 + x_1 x_2 - 4x_2^2 + 4x_2^4$	2	[-5, 5]	-1.0316
F17	$f_{17}(x) = (x_2 - \frac{5}{4}x_1^2 x_1^2 + \frac{5}{8}x_1 - 6)^2 + 10 \left(1 - \frac{1}{8x_1} \right) \cos x_1 + 10$	2	[-5, 5]	0.398
F18	$f_{18}(x) = [1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1 x_2 + 3x_2^2)] \\ \times [30 + (2x_1 - 3x_2)^2 \times (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1 x_2 + 27x_2^2)]$	2	[-2, 2]	3
F19	$f_{19}(x) = -\sum_{i=1}^4 c_i \exp\left(-\sum_{j=1}^3 a_{ij} (x_j - p_{ij})^2\right)$	3	[1, 3]	-3.86
F20	$f_{20}(x) = -\sum_{i=1}^4 c_i \exp\left(-\sum_{j=1}^3 a_{ij} (x_j - p_{ij})^2\right)$	6	[0, 1]	-3.32
F21	$f_{21}(x) = -\sum_{i=1}^5 [(X - a_i)(X - a_i)^T + c_i]^{-1}$	4	[0, 10]	-10.1532
F22	$f_{22}(x) = -\sum_{i=1}^7 [(X - a_i)(X - a_i)^T + c_i]^{-1}$	4	[0, 10]	-10.4028
F23	$f_{23}(x) = -\sum_{i=1}^{10} [(X - a_i)(X - a_i)^T + c_i]^{-1}$	4	[0, 10]	-10.5363

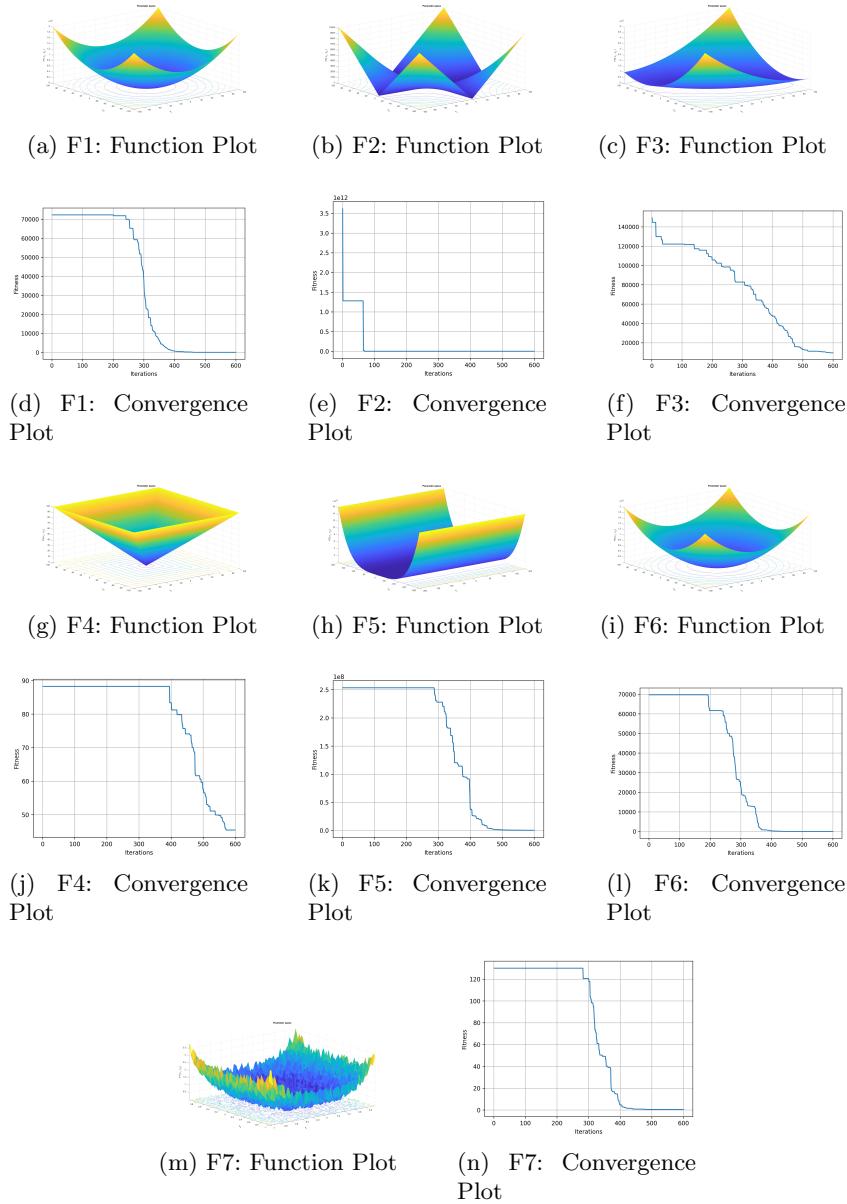


Fig. 9: Function and convergence plots obtained by the proposed SCA+A β HC algorithm framework on unimodal benchmark functions F1-F7

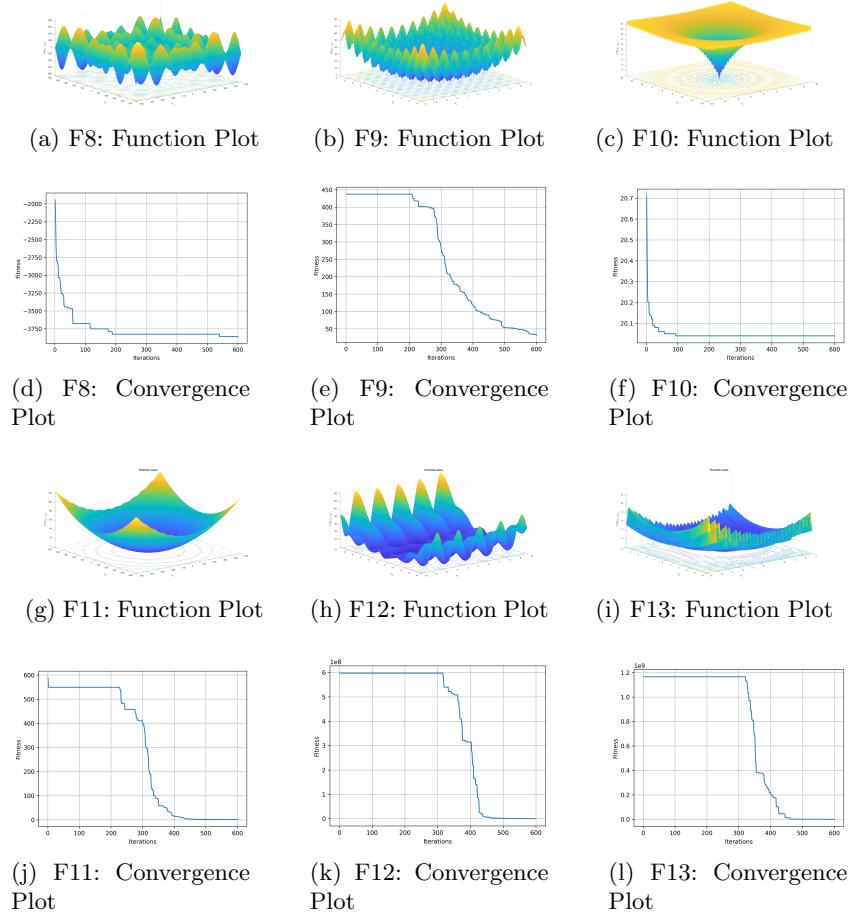


Fig. 10: Function and convergence plots obtained by the proposed SCA+A β HC algorithm framework on multimodal benchmark functions with variable dimensions F8-F13

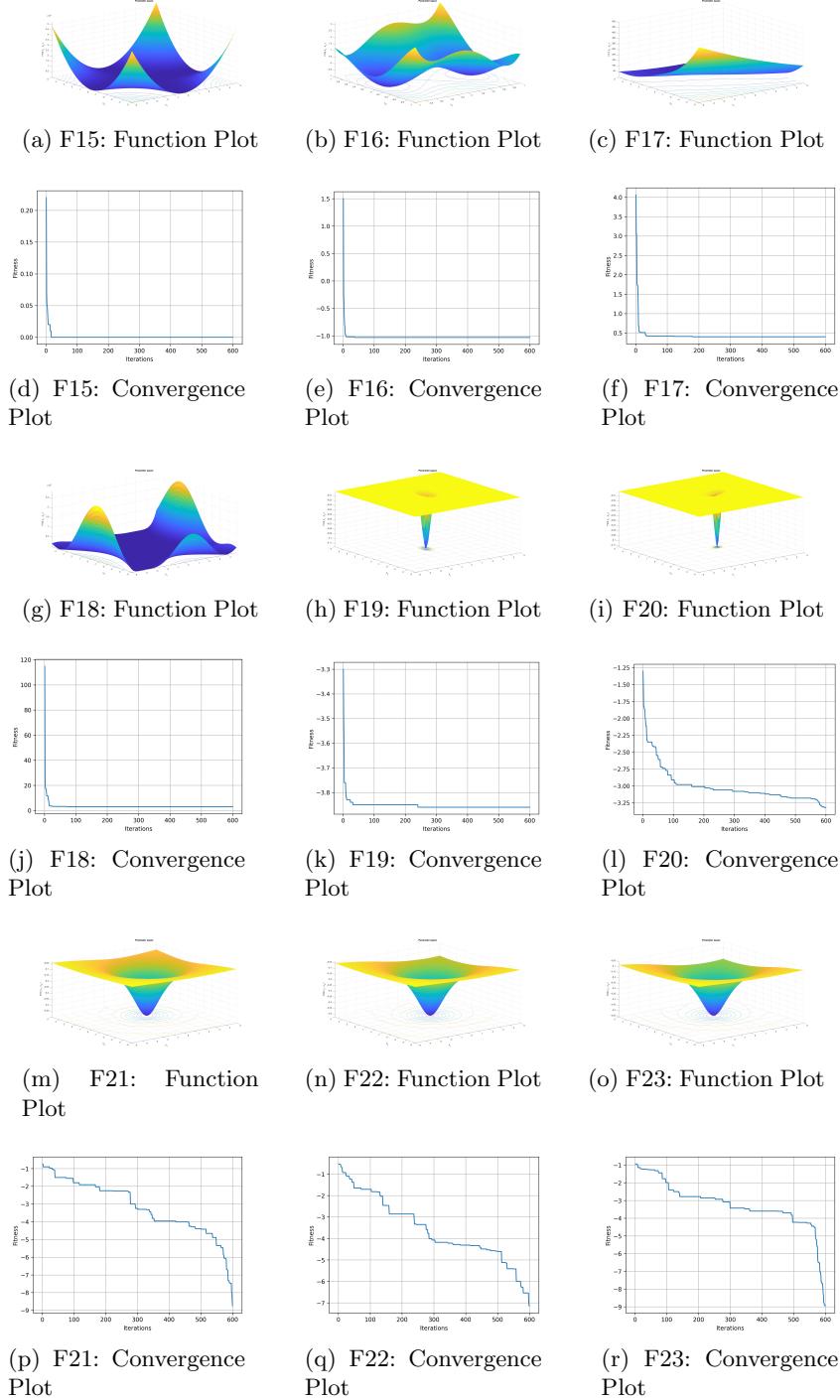


Fig. 11: Function and convergence plots obtained by the proposed SCA+A β HC algorithm framework on multimodal benchmark functions with fixed dimensions F15-F23

Table 9: Parameter settings for the comparative meta-heuristic optimization algorithms on continuous benchmark functions. Here, OA denotes optimization algorithm.

OA	Parameter(s)	Value(s)
GA	Selection strategy	Roulette wheel
	Crossover probability	0.8
	Mutation probability	0.05
PSO	Inertia weight (I)	I lies in [1 0]
	Constants (C_1, C_2)	$C_1 = 2, C_2 = 2$
GWO	Convergence operator (a)	a lies in [2 0]
BAT	Initial loudness (A)	$A = 0.5$
	Pulse rate (r)	$r = 0.5$
	Minimum frequency (Q_{min})	$Q_{min} = 0$
	Maxmimum frequency (Q_{max})	$Q_{max} = 2$
GSA	Initial Gravitational constant ($G_{initial}$)	$G_{initial} = 100$
	Decay factor (α)	$\alpha = 20$
MFO	Convergence parameter (a)	a lies in [-2 -1]
	Shape of logarithmic spiral (b)	$b = 1$
CS	Discovery rate of alien eggs (p_a)	$p_a = 0.25$
MVO	Minimum wormhole existence probability (WEP_{min})	$WEP_{min} = 0.2$
	Maximum wormhole existence probability (WEP_{max})	$WEP_{max} = 1$
FPA	Switching probability (p)	$p = 0.8$

function that has been used to perform the evaluation is taken from one of the following categories:

1. Unimodal test functions with variable dimensions (F1-F7)
2. Multimodal test functions with variable dimensions (F8-F13)
3. Multimodal test functions with fixed dimensions (F14-F23)

The population size has been set to 30 for all experiments. For the variable dimension problems, the dimension used for the tests has been set to 30. The proposed method is compared with the following popular optimization algorithms, the parameter settings for which have been tabulated in [Table 9](#):

1. Genetic Algorithm (GA) [\[33\]](#)
2. Particle Swarm Optimization (PSO) [\[42\]](#)
3. Grey Wolf Optimizer (GWO) [\[58\]](#)
4. Bat Algorithm (BAT) [\[78\]](#)
5. Gravitational Search Algorithm (GSA) [\[65\]](#)
6. Moth Flame Optimization (MFO) [\[55\]](#)
7. Cuckoo Search (CS) [\[77\]](#)
8. Multi-Verse Optimizer (MVO) [\[59\]](#)
9. Flower Pollination Algorithm (FPA) [\[79\]](#)

Table 10: Comparison of the proposed SCA+A β H C approach with some standard optimization algorithms in the literature on the benchmark test functions used in this study. For functions F1-F13 (variable dimension problems) the dimension is 30. Details of the benchmark functions are provided in Table 8.

Function	Proposed	PSO	GWO	BAT	GSA	MFO	CS	MVO	FPA
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg
F1	2.05E+00	3.02E+00	1.93E+03	5.70E+02	1.83E+04	3.01E+03	1.18E-27	1.47E-27	6.50E+04
F2	8.91E-04	5.96E-04	2.47E+01	5.68E+00	3.58E+02	1.35E-03	9.71E-17	2.71E+08	3.05E+00
F3	8.60E+03	5.78E+03	2.65E+04	3.44E+03	4.05E+04	8.21E+03	5.12E-05	2.03E+04	1.38E+05
F4	4.55E+01	1.32E+00	5.17E+01	1.05E+01	4.39E+01	3.64E+00	1.24E+00	1.94E+06	8.51E+01
F5	3.70E+02	9.64E-04	1.95E+01	1.31E-01	1.06E+01	6.35E+00	2.70E+01	7.78E+01	2.10E+07
F6	5.70E+00	7.77E+01	9.01E+02	2.81E+02	1.87E+01	2.92E+03	8.44E-01	3.18E+01	6.609E+04
F7	4.30E-01	2.00E+01	1.91E+01	1.50E+01	1.07E+01	3.05E+00	1.70E+01	4.57E+03	7.98E+02
F8	-3.15E+03	2.06E+02	-1.20E+04	4.50E+01	3.80E+03	2.40E+03	5.91E+02	-2.35E+03	8.48E+02
F9	4.88E+00	1.00E+01	9.03E+00	4.58E+00	2.55E+02	1.35E+01	2.15E+00	5.00E+00	1.52E+02
F10	1.60E+01	8.94E+00	1.30E+01	1.51E+00	1.75E+01	3.67E+01	1.05E+03	1.70E+03	1.92E+10
F11	1.85E+00	1.09E+01	1.01E+01	2.43E+00	1.70E+02	3.17E+01	4.76E+03	6.01E+02	4.59E+00
F12	3.41E+05	8.94E+01	4.77E+00	1.50E+01	1.51E+07	9.88E+06	4.84E-02	2.12E-02	4.71E+08
F13	1.05E+05	6.24E-08	1.52E+01	4.52E+01	5.75E+07	2.08E+07	5.50E-01	2.22E+01	9.403E+08
F14	1.00E+00	4.97E-07	9.95E+01	4.52E+16	1.39E+00	4.60E+01	4.17E+00	1.27E+01	6.71E+03
F15	4.75E-05	6.57E-06	3.33E+02	2.70E-02	1.61E-03	4.60E+01	1.25E+02	3.00E+02	4.10E-03
F16	4.09E+00	6.92E-08	1.54E+02	4.26E+01	1.75E+02	7.56E+01	1.00E+02	1.25E+02	4.80E+00
F17	3.98E-01	3.05E-07	3.05E+02	3.05E+01	3.05E+02	3.05E+01	3.05E+02	3.05E+02	1.00E+00
F18	3.00E+00	1.66E-06	5.31E+01	7.93E+01	9.22E+01	1.37E+02	5.68E+01	6.23E+01	8.52E+01
F19	-3.80E+00	1.90E+01	-3.42E+00	3.08E-01	-3.84E+00	3.14E-01	-3.84E+00	-3.84E+00	1.44E+03
F20	3.32E+00	2.09E-01	1.61E+00	4.60E-01	3.11E-01	2.94E-02	3.20E-00	6.13E-02	3.25E-00
F21	-8.75E+00	4.17E-10	-6.60E+00	3.78E+00	-4.16E+00	9.20E-01	-8.64E-01	2.56E+04	-4.27E+04
F22	-7.14E+00	3.05E-15	-5.58E+00	2.61E+01	-1.04E-01	6.78E-04	-5.61E+00	-8.82E+00	3.08E+00
F23	-8.95E+00	6.01E-11	-4.70E+00	3.20E+01	-4.72E+00	1.74E+00	-1.01E+01	6.75E+01	-3.01E+01

The average fitness, as well as standard deviation values of the results obtained on the test functions by the optimization framework across 30 independent runs, are shown in [Table 10](#). It can be observed that our proposed method outperforms BAT for all the test functions considered here. Further, our proposed method performs superior to PSO for all test functions except for F4 and F8, where PSO performs marginally better than the former. In comparison with GWO, our approach performs marginally inferior for exactly 8 cases and performs significantly better on the other 15 functions. The proposed optimization framework also outperforms FPA by significant margins on all functions except for F12 and F13. It may also be noted that all algorithms except GA and GSA perform equally on F19, while our method shows a highly comparable performance against MVO on functions F21-F23, outperforming the latter on other benchmark functions. The results obtained by the proposed approach are competitive and comparable to state-of-the-art, thus justifying the reliability of this framework.

The function plots and the convergence graphs obtained through experiments on unimodal, multimodal with variable dimensions, and multimodal functions with fixed dimensions have been shown in [Figure 9](#), [Figure 10](#) and [Figure 11](#) respectively. We have set the number of iterations per run as 600 for all experimentation. The convergence curves show that the proposed method effectively optimizes the test functions over iterations. Furthermore, the promising convergence behaviour on simple unimodal functions shown in [Figure 9](#) signifies the enhanced exploitation capability of the proposed approach, while [Figure 10](#) illustrates that in the case of multimodal functions having several local optima, our proposed approach results in a steady convergence rate avoiding premature convergence, which thereby depicts a satisfactory convergence behaviour. In a nutshell, the results indicate that our proposed framework can be effectively extended to solve continuous optimization tasks also.

CEC-2017 Test Suite Further, we test our proposed $\text{A}\beta\text{HC}$ embedded SCA approach on the CEC'17 test suite [\[12\]](#) comprising real-parameter continuous optimization problems. A total of 15 functions have been considered, each belonging to one of the following categories:

1. Unimodal functions (C1 and C3)
2. Multimodal functions (C4-C8)
3. Hybrid functions (C11-C14)
4. Composition functions (C21-C24)

The relevant CEC'17 test functions are described in [Table 11](#). Following the specifications given in [\[12\]](#), for each function the search range is taken as $[-100, 100]$, while the dimension is set as 10.

The proposed method has been compared with the meta-heuristic algorithms mentioned in [Table 9](#) with population size set as 30 for each algorithm. We report the mean and standard deviation of fitness values obtained by each algorithm over 30 independent runs on the aforementioned test functions, the results are tabulated in [Table 12](#). The convergence plots obtained by the proposed

Table 11: Description of the CEC-2017 test functions used to test the proposed SCA+ $A\beta$ HC algorithm.

Function	Category	Description
C1	Unimodal functions	Shifted and Rotated Bent Cigar Function
C3		Shifted and Rotated Zakharov Function
C4	Multimodal functions	Shifted and Rotated Rosenbrock's Function
C5		Shifted and Rotated Rastrigin's Function
C6		Shifted and Rotated Expanded Scaffer's C6 Function
C7		Shifted and Rotated Lunacek Bi-Rastrigin Function
C8		Shifted and Rotated Non-Continuous Rastrigin's Function
C11	Hybrid functions	Hybrid Function 1 (N=3)
C12		Hybrid Function 2 (N=3)
C13		Hybrid Function 3 (N=3)
C14		Hybrid Function 4 (N=4)
C21	Composition functions	Composition Function 1 (N=3)
C22		Composition Function 2 (N=3)
C23		Composition Function 3 (N=4)
C24		Composition Function 4 (N=4)

SCA+ $A\beta$ HC algorithm on the CEC'17 test suite are shown in [Figure 12](#). For each of the experiments on the CEC'17 test suite, the optimization algorithms have been run for 600 iterations. As can be seen from [Figure 12](#), the proposed method shows steady and satisfactory convergence behaviour on most of the test functions in terms of convergence rate, although for C13 and C14 the graphs do somewhat indicate premature convergence. Apart from that, from [Table 12](#), it can be observed that the proposed algorithm outperforms PSO, GWO, GSA, MFO and FPA in all the test functions except C12. Further, the BAT algorithm performs significantly inferior to the proposed method in all the test functions. This justifies that the proposed SCA+ $A\beta$ HC algorithm produces competitive and robust results on the CEC'17 test suite, often outperforming popularly used meta-heuristics.

5 Conclusion and Future Work

Pneumonia has been a cause of worry to developing as well as under-developed nations alike, and given the rise in pollution levels and improper civic hygiene, its accountability for human life would continue its ascent. However, accurate diagnosis in the early stages and subsequent treatment can cure patients before the disease becomes lethal. As such, CAD systems have been preferred to radiological interpretations owing to the subjective variability of the latter. Researchers of various fields continue to investigate and explore methods in this regard. The primary objective of our study is to contribute to this research field. In this work, we have proposed a two-stage framework in which we have leveraged the combined effectiveness of deep learning and meta-heuristic based optimization algorithms to detect traces of pneumonia from X-ray scan images obtained from a publicly available dataset. Transfer learning is performed us-

Table 12: Comparison of the proposed SCA+ $\alpha\beta$ HIC approach with standard optimization algorithms in the literature on CEC'17 test functions used in this study. The parameter settings of each comparative algorithm are tabulated in Table 9, while the specifications of the test functions are provided in Table 11.

Function	Proposed	GA	GWO	BAT	GSA	MFO	CS	MVO	FPA
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg
C1	9.00E-09	4.15E-07	3.21E-07	0.240E+09	0.93.06E+11	0.160E+10	0.100E+10	0.11.192E+09	0.6.77E+10
C3	1.27E-03	1.33E-02	6.60E-03	0.3.73E+02	0.1.34E+02	0.1.00E+04	0.04.1.16E+03	0.03.3.56E+04	0.04.8.68E+04
C4	4.36E-02	2.25E-01	4.62E-02	6.87E-01	1.33E-02	0.2.08E+02	0.1.00E+01	6.19E-03	5.41E-01
C5	5.30E-02	8.00E-02	5.30E-02	0.3.46E+01	0.6.70E+02	0.1.00E+00	0.1.00E+02	0.2.630E+01	0.1.6.72E+01
C6	6.00E-02	8.14E-01	6.16E-01	2.24E+01	0.2.25E+03	4.97E+02	0.3.66E+02	0.3.574E+03	0.3.100E+03
C7	7.09E-02	4.30E-02	7.36E-02	7.16E-02	1.88E+02	0.1.99E+02	0.2.24E+01	0.2.99E+03	0.3.54E+03
C8	1.45E-02	6.22E-02	8.20E-02	8.18E-02	0.1.88E+01	0.1.55E+01	0.1.48E+02	0.1.64E+01	0.1.23E+04
C11	1.10E-03	7.39E-02	1.34E-03	0.3.12E+02	0.3.00E+06	1.10E+06	1.41E+04	1.17E+04	1.27E+04
C12	4.07E-07	1.81E-05	7.35E-06	5.37E-01	1.11E+05	3.84E+06	1.84E+05	8.58E+06	7.77E+07
C13	1.71E-05	2.13E-03	8.08E-05	0.3.00E+07	0.1.300E+04	0.6.393E+06	0.1.91E+08	0.2.21E+05	0.6.50E+05
C14	1.67E-03	5.94E-02	2.41E-03	7.46E-02	6.393E+02	0.2.1.0E+08	0.2.500E+08	0.3.45E+05	0.9.45E+05
C21	2.23E-03	5.78E-02	2.22E-03	6.75E+02	0.2.2.1E+03	0.1.1E+04	0.1.3.43E+04	0.1.578E+02	0.2.74E+04
C22	2.38E-03	6.37E-01	2.22E-03	5.21E-02	5.39E+04	0.1.20E+02	0.1.60E+02	0.1.40E+02	0.1.24E+02
C23	2.64E-03	1.88E-02	2.64E-03	5.49E-02	5.05E+03	0.2.62E+02	0.2.51E+02	0.1.63E+04	0.1.46E+04
C24	2.72E-03	2.08E-02	2.89E-02	2.91E-03	0.1.3.68E-03	0.6.531E-01	0.1.20E+03	0.1.550E+01	0.1.30E+01

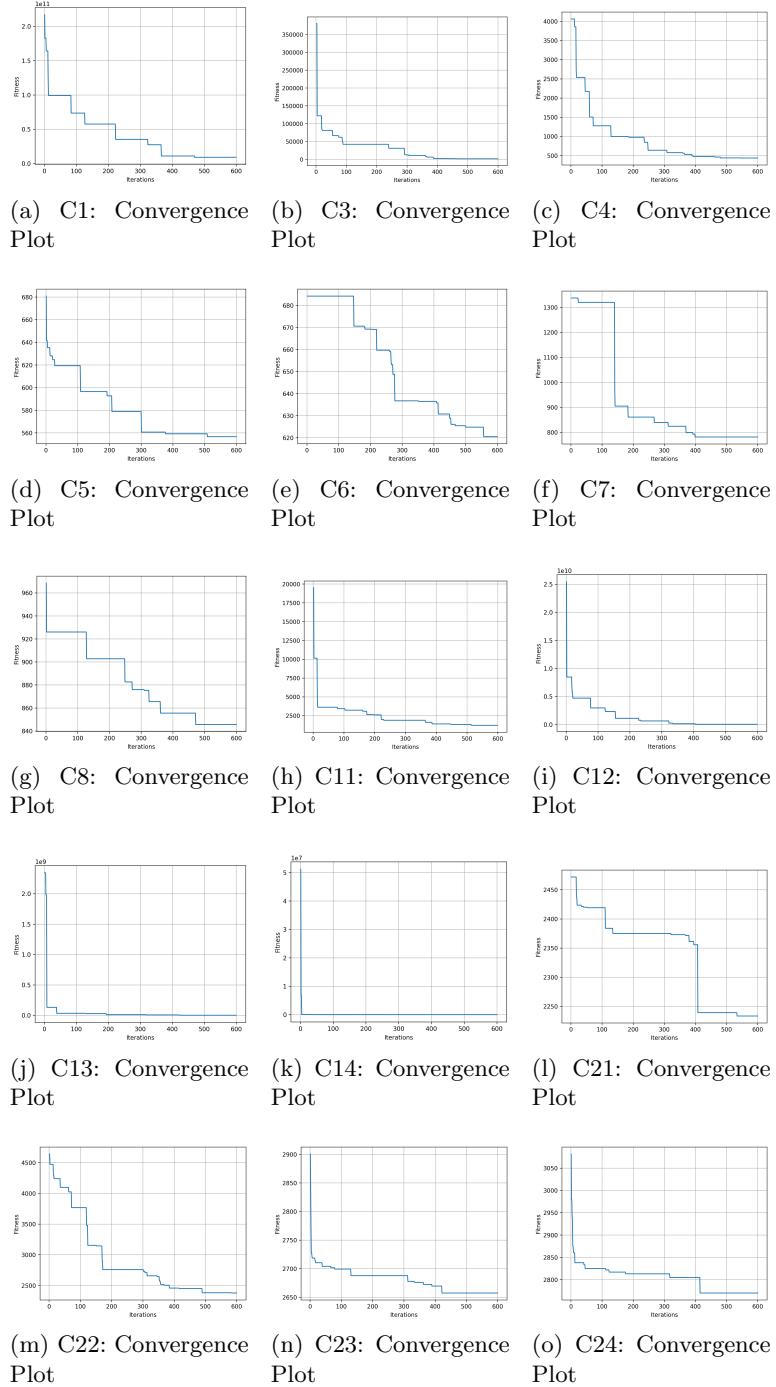


Fig. 12: Convergence plots obtained by the proposed SCA+A β HC algorithm framework on the CEC'17 test functions used in this study. C1 and C3 are unimodal, C4-C8 multimodal, C11-C14 hybrid and C21-C24 are composition functions.

ing a pre-trained DenseNet-201 model for feature extraction purposes. Then, the most effective features are selected from the extracted features using the meta-heuristic algorithm SCA, whose exploitation ability is improved by embedding the A β HC local search strategy. The final classification is performed by the SVM classifier, and it results in an accuracy of 98.36% and a sensitivity of 98.79% upon 5-fold cross-validation, which outperforms state-of-the-art methods on the same dataset. Further, our proposed architecture yields a high reduction in feature dimensionality of 85.55%, which considerably reduces storage requirements and thereby decreases computational needs. Additional analysis of the proposed method on continuous optimization problems reveals the superiority of the method compared to existing optimization algorithms on a variety of test suites, and thus the method can be applied on other domains of image analysis and optimization, thereby increasing its credibility. The only limitation, however, is the increased time complexity of the method caused by the embedding of the A β HC local search algorithm at the benefit of enhancing the exploitation ability of SCA for superior optimization performance.

Since the proposed method provides an end-to-end pipeline, it is well-suited for practical use, and therefore can effectively assist radiologists in their diagnosis. We expect our study can be used to provide further insights into the diagnosis of pneumonia to the research community. To contribute to this field in future, we intend to work with other available pneumonia datasets and continue studies using deep learning and different FS algorithms, as well as investigate alternate local search strategies. We may also try using generative models (such as Variational Autoencoders and Generative Adversarial Networks) to create synthetic X-ray images to mitigate the data scarcity problem and improve generalization. Last but not the least, we may also try performing segmentation on the lung X-ray images as a pre-processing step before classification, since RoI localization may enhance classification performance by removing redundant pixels as well as help radiologists in detecting anomalies more easily.

Acknowledgements

The authors would like to thank the Centre for Microprocessor Applications for Training, Education and Research (*CMATER*) research laboratory of the Computer Science and Engineering Department, Jadavpur University, Kolkata, India for providing the infrastructural support.

Conflict of Interest

All the authors declare that there are no conflicts of interest in the publication of this paper.

Bibliography

- [1] Abdollahzadeh B, Soleimanian Gharehchopogh F, Mirjalili S (2021) Artificial gorilla troops optimizer: A new nature-inspired metaheuristic algorithm for global optimization problems. *International Journal of Intelligent Systems*
- [2] Abualigah L, Diabat A (2021) Advances in sine cosine algorithm: a comprehensive survey. *Artificial Intelligence Review* pp 1–42
- [3] Abualigah L, Diabat A, Mirjalili S, Abd Elaziz M, Gandomi AH (2021) The arithmetic optimization algorithm. *Computer methods in applied mechanics and engineering* 376:113609
- [4] Agarwal N, Sondhi A, Chopra K, Singh G (2021) Transfer learning: Survey and classification. In: *Smart Innovations in Communication and Computational Sciences*, Springer, pp 145–155
- [5] Agrawal P, Abutarboush HF, Ganesh T, Mohamed AW (2021) Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019). *IEEE Access* 9:26766–26791
- [6] Ahmed S, Ghosh KK, Bera SK, Schwenker F, Sarkar R (2020) Gray level image contrast enhancement using barnacles mating optimizer. *IEEE Access* 8:169196–169214
- [7] Ahmed S, Ghosh KK, Garcia-Hernandez L, Abraham A, Sarkar R (2021) Improved coral reefs optimization with adaptive β -hill climbing for feature selection. *Neural Computing and Applications* 33(12):6467–6486
- [8] Akay B, Karaboga D, Akay R (2021) A comprehensive survey on optimizing deep learning models by metaheuristics. *Artificial Intelligence Review* pp 1–66
- [9] Al-Behadili H (2021) Stochastic local search algorithms for feature selection: A review. *Iraqi J Electr Electron Eng* 17(1):1–10
- [10] Al-Betar MA (2017) β -hill climbing: an exploratory local search. *Neural Computing and Applications* 28(1):153–168
- [11] Al-Betar MA, Aljarrah I, Awadallah MA, Faris H, Mirjalili S (2019) Adaptive β -hill climbing for optimization. *Soft Computing* 23(24):13489–13512
- [12] Awad N, Ali M, Liang J, Qu B, Suganthan P (2017) Problem definitions and evaluation criteria for the cec 2017 special session and competition on single objective real-parameter numerical optimization. Tech. rep.
- [13] Ayan E, Ünver HM (2019) Diagnosis of pneumonia from chest x-ray images using deep learning. In: *2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT)*, IEEE, pp 1–5
- [14] Banerjee D, Chatterjee B, Bhowal P, Bhattacharyya T, Malakar S, Sarkar R (2021) A new wrapper feature selection method for language-invariant offline signature verification. *Expert Systems with Applications* p 115756
- [15] Chakraborty S, Mondal R, Singh PK, Sarkar R, Nasipuri M (2022) Genetic dropout: An application to handwritten indic script classification. In:

- Advanced Techniques for IoT Applications, Springer Singapore, Springer Singapore
- [16] Chandra TB, Verma K (2020) Pneumonia detection on chest x-ray using machine learning paradigm. In: Proceedings of 3rd International Conference on Computer Vision and Image Processing, Springer, pp 21–33
 - [17] Chatterjee B, Bhattacharyya T, Ghosh KK, Chatterjee A, Sarkar R (2021) A novel meta-heuristic approach for influence maximization in social networks. *Expert Systems* p e12676
 - [18] Cortes C, Vapnik V (1995) Support-vector networks. *Machine Learning* 20:273–297
 - [19] Das S, Suganthan PN (2010) Differential evolution: A survey of the state-of-the-art. *IEEE transactions on evolutionary computation* 15(1):4–31
 - [20] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, Ieee, pp 248–255
 - [21] Dey N, Zhang YD, Rajinikanth V, Pugalenthhi R, Raja NSM (2021) Customized vgg19 architecture for pneumonia detection in chest x-rays. *Pattern Recognition Letters* 143:67–74
 - [22] Dhangupta S, Ghosh M, Mirjalili S, Sarkar R (2020) Selective opposition based grey wolf optimization. *Expert Systems with Applications* 151:113389
 - [23] Dietterich TG (1998) Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation* 10(7):1895–1923
 - [24] Dinkar SK, Deep K, Mirjalili S, Thapliyal S (2021) Opposition-based laplacian equilibrium optimizer with application in image segmentation using multilevel thresholding. *Expert Systems with Applications* 174:114766
 - [25] Draa A, Bouaziz A (2014) An artificial bee colony algorithm for image contrast enhancement. *Swarm and Evolutionary computation* 16:69–84
 - [26] Ghosh KK, Ahmed S, Singh PK, Geem ZW, Sarkar R (2020) Improved binary sailfish optimizer based on adaptive β -hill climbing for feature selection. *IEEE Access* 8:83548–83560
 - [27] Ghosh M, Sen S, Sarkar R, Maulik U (2021) Quantum squirrel inspired algorithm for gene selection in methylation and expression data of prostate cancer. *Applied Soft Computing* 105:107221
 - [28] Gillala R, Vuyyuru KR, Jatoh C, Fiore U (2021) An efficient chaotic salp swarm optimization approach based on ensemble algorithm for class imbalance problems. *Soft Computing* pp 1–11
 - [29] Guha R, Ghosh M, Chakrabarti A, Sarkar R, Mirjalili S (2020) Introducing clustering based population in binary gravitational search algorithm for feature selection. *Applied Soft Computing* 93:106341
 - [30] Guha R, Ghosh M, Mutsuddi S, Sarkar R, Mirjalili S (2020) Embedded chaotic whale survival algorithm for filter-wrapping feature selection. *Soft Computing* 24(17):12821–12843
 - [31] Guha R, Khan AH, Singh PK, Sarkar R, Bhattacharjee D (2021) Cga: A new feature selection model for visual human action recognition. *Neural Computing and Applications* 33(10):5267–5286

- [32] Hashim FA, Hussain K, Houssein EH, Mabrouk MS, Al-Atabany W (2021) Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems. *Applied Intelligence* 51(3):1531–1551
- [33] Holland JH (1992) Genetic algorithms. *Scientific american* 267(1):66–73
- [34] Houssein EH, Helmy BED, Oliva D, Elngar AA, Shaban H (2021) A novel black widow optimization algorithm for multilevel thresholding image segmentation. *Expert Systems with Applications* 167:114159
- [35] Houssein EH, Mahdy MA, Shebl D, Mohamed WM (2021) A survey of metaheuristic algorithms for solving optimization problems. In: *Metaheuristics in Machine Learning: Theory and Applications*, Springer, pp 515–543
- [36] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. IEEE, pp 4700–4708
- [37] Ibrahim AU, Ozsoz M, Serte S, Al-Turjman F, Yakoi PS (2021) Pneumonia classification using deep learning from chest x-ray images during covid-19. *Cognitive Computation* pp 1–13
- [38] Jain A, Mehra N, Khare P (2021) A survey on pneumonia detection methods using computer-aided diagnosis. *International Journal* 9(7)
- [39] Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJ (2019) Identifying pneumonia in chest x-rays: a deep learning approach. *Measurement* 145:511–518
- [40] José-García A, Gómez-Flores W (2016) Automatic clustering using nature-inspired metaheuristics: A survey. *Applied Soft Computing* 41:192–213
- [41] Karami H, Anaraki MV, Farzin S, Mirjalili S (2021) Flow direction algorithm (fda): A novel optimization approach for solving optimization problems. *Computers & Industrial Engineering* 156:107224
- [42] Kennedy J, Eberhart R (1995) Particle swarm optimization. In: *Proceedings of ICNN'95-international conference on neural networks*, IEEE, vol 4, pp 1942–1948
- [43] Kermany D, Zhang K, Goldbaum M (2018) Labeled optical coherence tomography (oct) and chest x-ray images for classification. <http://doi.org/10.17632/rscbjbr9sj.2>, <https://doi.org/10.17632/rscbjbr9sj.2>
- [44] Khamkar C, Shah M, Kalyani S, Bhowmick K (2021) Pneumonia detection using x-ray images and deep learning. In: *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*, Springer, pp 141–152
- [45] Khan W, Zaki N, Ali L (2021) Intelligent pneumonia identification from chest x-rays: A systematic literature review. *IEEE Access*
- [46] Kundu R, Basak H, Singh PK, Ahmadian A, Ferrara M, Sarkar R (2021) Fuzzy rank-based fusion of cnn models using gompertz function for screening covid-19 ct-scans. *Scientific Reports* 11(1):1–12
- [47] Kundu R, Das R, Geem ZW, Han GT, Sarkar R (2021) Pneumonia detection in chest x-ray images using an ensemble of deep learning models. *PLoS One*
- [48] Kuo KM, Talley PC, Huang CH, Cheng LC (2019) Predicting hospital-acquired pneumonia among schizophrenic patients: a machine learning approach. *BMC medical informatics and decision making* 19(1):1–8

- [49] Li Y, Zhang Z, Dai C, Dong Q, Badrigilan S (2020) Accuracy of deep learning for automated detection of pneumonia using chest x-ray images: a systematic review and meta-analysis. *Computers in Biology and Medicine* p 103898
- [50] Liang G, Zheng L (2020) A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer methods and programs in biomedicine* 187:104964
- [51] Liu Y, Sun Y, Xue B, Zhang M, Yen GG, Tan KC (2021) A survey on evolutionary neural architecture search. *IEEE Transactions on Neural Networks and Learning Systems* pp 1–21
- [52] Luque-Chang A, Aranguren I, Pérez-Cisneros M, Valdivia A (2021) A novel metaheuristic approach for image contrast enhancement based on gray-scale mapping. In: *Metaheuristics in Machine Learning: Theory and Applications*, Springer, pp 609–634
- [53] Mahmud T, Rahman MA, Fattah SA (2020) Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization. *Computers in biology and medicine* 122:103869
- [54] Masud M, Bairagi AK, Nahid AA, Sikder N, Rubaiee S, Ahmed A, Anand D (2021) A pneumonia diagnosis scheme based on hybrid features extracted from chest radiographs using an ensemble learning algorithm. *Journal of Healthcare Engineering* 2021
- [55] Mirjalili S (2015) Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-based systems* 89:228–249
- [56] Mirjalili S (2016) Sca: a sine cosine algorithm for solving optimization problems. *Knowledge-based systems* 96:120–133
- [57] Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Advances in engineering software* 95:51–67
- [58] Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Advances in engineering software* 69:46–61
- [59] Mirjalili S, Mirjalili SM, Hatamlou A (2016) Multi-verse optimizer: a nature-inspired algorithm for global optimization. *Neural Computing and Applications* 27(2):495–513
- [60] Neuman MI, Lee EY, Bixby S, Diperna S, Hellinger J, Markowitz R, Servaes S, Monuteaux MC, Shah SS (2012) Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. *Journal of hospital medicine* 7(4):294–298
- [61] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. (2019) Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* 32:8026–8037
- [62] Rahman T, Chowdhury ME, Khandakar A, Islam KR, Islam KF, Mahbub ZB, Kadir MA, Kashem S (2020) Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray. *Applied Sciences* 10(9):3233

- [63] Rajaraman S, Sema C, Incheol K, George T, Antani S (2018) Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Applied Sciences* 8(10)
- [64] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, et al. (2017) Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:171105225
- [65] Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) Gsa: a gravitational search algorithm. *Information sciences* 179(13):2232–2248
- [66] Saraiva AA, Ferreira NMF, de Sousa LL, Costa NJC, Sousa JVM, Santos D, Valente A, Soares S (2019) Classification of images of childhood pneumonia using convolutional neural networks. In: BIOIMAGING, BIOSTEC, pp 112–119
- [67] Sen S, Saha S, Chatterjee S, Mirjalili S, Sarkar R (2021) A bi-stage feature selection approach for covid-19 prediction using chest ct images. *Applied Intelligence* pp 1–16
- [68] Sharma H, Jain JS, Bansal P, Gupta S (2020) Feature extraction and classification of chest x-ray images using cnn to detect pneumonia. In: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, pp 227–231
- [69] Shaw SS, Ahmed S, Malakar S, Garcia-Hernandez L, Abraham A, Sarkar R (2021) Hybridization of ring theory-based evolutionary algorithm and particle swarm optimization to solve class imbalance problem. *Complex & Intelligent Systems* pp 1–23
- [70] Shukri SE, Al-Sayyed R, Hudaib A, Mirjalili S (2021) Enhanced multi-verse optimizer for task scheduling in cloud computing environments. *Expert Systems with Applications* 168:114230
- [71] Stephen O, Sain M, Maduh UJ, Jeong DU (2019) An efficient deep learning approach to pneumonia classification in healthcare. *Journal of healthcare engineering* 2019
- [72] Sutskever I, Martens J, Dahl G, Hinton G (2013) On the importance of initialization and momentum in deep learning. In: International conference on machine learning, PMLR, pp 1139–1147
- [73] Tubishat M, Idris N, Shuib L, Abushariah MA, Mirjalili S (2020) Improved salp swarm algorithm based on opposition based learning and novel local search algorithm for feature selection. *Expert Systems with Applications* 145:113122
- [74] WHO (2019) Pneumonia. World Health Organization URL <https://www.who.int/news-room/fact-sheets/detail/pneumonia>
- [75] Williams GJ, Macaskill P, Kerr M, Fitzgerald DA, Isaacs D, Codarini M, McCaskill M, Prelog K, Craig JC (2013) Variability and accuracy in interpretation of consolidation on chest radiography for diagnosing pneumonia in children under 5 years of age. *Pediatric Pulmonology* 48(12):1195–1200
- [76] Wolpert DH, Macready WG, et al. (1995) No free lunch theorems for search. Tech. rep., Technical Report SFI-TR-95-02-010, Santa Fe Institute, Santa Fe Institute

- [77] Yang XS, Deb S (2009) Cuckoo search via lévy flights. In: 2009 World congress on nature & biologically inspired computing (NaBIC), Ieee, pp 210–214
- [78] Yang XS, Gandomi AH (2012) Bat algorithm: a novel approach for global engineering optimization. *Engineering computations*
- [79] Yang XS, Karamanoglu M, He X (2014) Flower pollination algorithm: a novel approach for multiobjective optimization. *Engineering optimization* 46(9):1222–1237
- [80] Yousri D, Abd Elaziz M, Mirjalili S (2020) Fractional-order calculus-based flower pollination algorithm with local search for global optimization and image segmentation. *Knowledge-Based Systems* 197:105889
- [81] Yue H, Yu Q, Liu C, Huang Y, Jiang Z, Shao C, Zhang H, Ma B, Wang Y, Xie G, et al. (2020) Machine learning-based ct radiomics method for predicting hospital stay in patients with pneumonia associated with sars-cov-2 infection: a multicenter study. *Annals of translational medicine* 8(14)
- [82] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q (2020) A comprehensive survey on transfer learning. *Proceedings of the IEEE* 109(1):43–76
- [83] Zubair S (2020) An efficient method to predict pneumonia from chest x-rays using deep learning approach. *The Importance of Health Informatics in Public Health during a Pandemic* 272:457