

Proposed Self – Regulated Gray Wolf Optimizer based Extreme Learning Machine Neural Network Classifier for Lung Cancer Classification

I. Jasmine Selvakumari Jeya, M. Revathi, M. Uma Priya

Abstract-- The major issue in the development of pattern recognition towards lung cancer classification is the formation of feature extraction process and the proposed classifier model. In the proposed approach, a self-regulated gray wolf optimizer based extreme learning machine classifier is proposed to carry out lung cancer classification along with the statistical feature extraction methods. Simulation shows that the proposed approach works well and produces higher classification accuracy than the conventional classifier methods. The modeled Self-Regulated Gray Wolf Optimizer (SRGWO) and Extreme Learning Machine (ELM) along with feature and segmentation process shows highest improvement in comparison with the other existing literature studies in neural networks. In particular, the significant finding of this work employing ELM, SRGWO and feature analysis validates the correlation of Computed Tomography (CT) measures as well as classification pathological parameters. Thus, the proposed SRGWO and ELM classifier is developed in the present approach for lung cancer classification of CT images reducing the computational cost and time of all the earlier classifiers and as well increasing the classification accuracy. On performing trial runs for the proposed SRGWO – ELM to compute the classification results for the considered real time and Lung Image Database Consortium (LIDC) lung images, it has been noted that at certain trials, the extreme learning machine neuronal classifier is noted to get stuck up with the local minima problem and it is necessary to restart the generation process to achieve classification solutions.

I. INTRODUCTION

In the growing scenario, the potential method for improving the retention of survival of a patient is detection of the cancer at an early stage. Since the past decades, cancers such as breast cancer, cervical cancer, lung cancer, blood cancer, throat cancer, mouth cancer have been noted to occur. The present approach focuses on various analysis and classification of lung module of the human body. Considering various lung diseases like fibrosis, carcinoma and so on, lung cancer plays a major role for increasing the death rate in both men and women. From the current statistics it is noted that around 1.25 million people possess the lung cancer disease and almost around 1.18 million

people die because of this disease [1]. The patient's survival rate can be increased on detecting cancer at an early stage. Detecting lung cancer at the beginning stage is a hectic task. In general, several patients are detected with the occurrence of lung cancer in the middle or at the advanced stage of cancer. To identify the presence of lung cancer and to detect the abnormalities in the initial stage computer aided diagnosis helps in the diagnostic process. In the present approach, lung cancer classification method for lung CT images is developed employing a neuronal model – Extreme Learning Machine Neural Classifier (ELMNC). Several classifiers are developed in a manner to learn from the image data itself employing the training process, since the expert knowledge required for determining the classifier parameters are highly impracticable. The present approach contributes for medical image classification – specifically to lung cancer images, proposing a hybrid classifier model based on computational machine learning paradigm for carrying out the image classification in an effective and efficient manner. This proposed computational paradigm, hybridizes the exploration mechanism of self-regulated learning ability of Gray Wolf Optimizer (GWO) with that of the ELMNN for performing the cancer classification process. In the recent developments of the off-line learning architectures, Extreme Learning Machine Classifier is a single hidden layer feed forward neural network architecture, which has proven itself to be an excellent classifier architecture with more number of hidden layer neurons. In the present approach the proposed SRGWO is employed to compute the optimal set of parameters for the ELM neuron classifier thus reducing the number of hidden layers of neurons and optimizing the weights, so as to improve the classification accuracy.

Basically, the entire lung is segmented from the CT images and the segmented image aids to compute the necessary parameters. The various parameters employed in the current research for the classification process are sensitivity, specificity, classification accuracy and Mean Squared Error (MSE). The convergence of the proposed Self-Regulated Gray Wolf Optimizer based ELM Neural Classifier (SRGWO – ELMNC) is based on the minimization of the MSE and the increase in the classification accuracy rate of the image classification process.

Manuscript received September 16, 2019.

I. Jasmine Selvakumari Jeya, Associate Professor, Hindusthan College of Engineering and Technology, Coimbatore, T.N, India.(e-mail: wjasminejeya@gmail.com)

M. Revathi, Research Scholar & Assistant Professor, Hindusthan College of Engineering and Technology, Coimbatore, T.N, India.(e-mail: revathimuthusamy@gmail.com)

M. Uma Priya, Assistant Professor, Hindusthan College of Engineering and Technology, Coimbatore, T.N, India.(e-mail: umapriya.kce@gmail.com)

PROPOSED SELF – REGULATED GRAY WOLF OPTIMIZER BASED EXTREME LEARNING MACHINE NEURAL NETWORK CLASSIFIER FOR LUNG CANCER CLASSIFICATION

Most of the earlier literature reveals the applicability of Back Propagation Neural Network (BPNN) for carrying out the classification process of lung images [2]. Thus the paper focuses on developing the hybrid SRGWO – ELMNC for effective and efficient classification process and is as presented in the forth coming sections. The simulation proves that the proposed approach is able to achieve better classification accuracy, than the other classifiers. The proposed flow of lung cancer classification model is as shown in Figure 1.

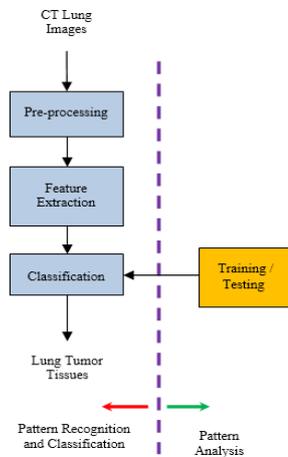


Figure 1: Proposed Flow of Lung Cancer Classification Model

II. FEATURE EXTRACTION PROCESS FOR LUNG CT IMAGES

Generally, feature extraction is a methodology wherein the input data are transformed into a set of features. Feature extraction is carried out to extract the prominent features from any of the given images and the prominent features are selected based on the expert opinion from the radiologists. Fundamentally, a feature is considered to be rich when the feature is characterized with an appropriate intensity, shape, location and texture. Originally, human expertise was

employed for identifying the useful features required for analysis. To complement the human expertise and the requirement for converting the raw data into a set of useful features, automatic feature construction methods are employed. There are various methods for extracting the features which include, statistical, structural, model based and transform based [3]. In the present approach, statistical method is employed to determine the spatial distribution of the intensity of the pixels and as well to determine the local features at each point in the image that is, a set of statistics is derived from the distributions of the local features. Statistical methods include the features derived from the histogram, auto-correlation method, gradient method, run length matrix method and gray level co-occurrence method [4].

These statistical approaches are of first, second or higher order based on the number of points that define the local feature. The properties of images completely depend only on the individual pixel values in case of first order statistics and on the properties of pixel pairs for second order statistics. The first order statistics involve mean gray scale, standard deviation of the mean, the deviation of the pixel distribution given by skewness and the steepness of the pixel distribution given by kurtosis. The second order statistics employ gray level run-length and gray-level co-occurrence matrices. The higher order statistics tend to obtain higher discrimination indexes of the medical images. Basically, the statistical parameters are extracted from the Region of Interest (ROI) of the images. In the present work, ROI is the segmented single slices containing the two lungs. The first order statistical parameters considered in the paper are mean, standard deviation, skewness and kurtosis. The higher order statistics considered here includes entropy, homogeneity and angular second moment. The details on the statistical parameters for the proposed feature extraction process in lung images are as presented in Table

Table 1: Statistical Features for Evaluation

Statistical Features	Expression for Evaluation
First Order Statistics	
Mean	$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N I(i, j)$
Standard Deviation	$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - \mu)^2}$
Skewness	$S = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left[\frac{I(i, j) - \mu}{\sigma} \right]^3$
Kurtosis	$K = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left[\frac{I(i, j) - \mu}{\sigma} \right]^4$
Second Order Statistics	
Entropy	$E_p = \sum_{i=1}^M \sum_{j=1}^N \left(\frac{I(i,j)}{\sigma} \right) \log \left(\frac{I(i,j)}{\sigma} \right)$
Homogeneity	$H_n = \sum_{i=1}^M \sum_{j=1}^N \frac{I(i,j)}{1 + (i - j)^2} \text{ for } i \neq j$
Angular Second Moment	$ASM = \sum_{i=1}^M \sum_{j=1}^N \left(\frac{I(i,j)}{\sigma} \right)^2$
<i>I(i,j)</i> – intensity value of the pixel at (i,j) and <i>MxN</i> – size of the image	



III. PROPOSED SELF-REGULATED GRAY WOLF OPTIMIZER (SRGWO) FOR ELM TUNING

The rapid development of nature-inspired evolutionary optimization algorithms is been noted for the past two decades and they are noted to produce optimal solutions for the problems under consideration. In this category, a recently developed algorithm is the GWO algorithm, which is considered in the present approach with its variant for tuning the parameters of ELMNN. This section details the modeling and the fundamental flow process of the proposed SRGWO algorithm.

3.1 Gray Wolf Optimizer – Revisited

With the development of the bio-inspired meta-heuristic algorithms like Genetic Algorithm [5,6], Ant Colony Optimization [7], Particle Swarm Optimization [8], Bacteria Foraging Algorithm, Particle Swirl Algorithm and so on, it is studied that they are employed for finding optimal solutions to various problems by the scientists. The versatility of these algorithms is based on their flexibility, avoidance of local and global optima and faster convergence. Subsequent to these algorithms is the development of GWO inspired by the leadership and hunting process of gray wolves available in nature.

These Gray wolves live in groups and are available on the top of the food chain. Basically, 4 – 12 gray wolves live in a group and they by nature follow social dominant hierarchy. With respect to their social dominant behavior, they are classified into four hierarchies as alpha (α), beta (β), delta (δ) and omega (ω) wolves. The alpha wolves are the ones in the upper most hierarchy and are the leader wolves (may be a male or female). These alpha leader wolves actively involves in decision making on hunting process, the way in which the wolves have to travel, the place of sleep of the wolves, the time they have to wake up and other related decision making entities. All the remaining wolves in the hierarchy will strictly follow the decision taken by the alpha leader wolves. Alpha wolf being the dominant wolf is the only one that mates within the group. These wolves need not be the strongest wolf in the group but are the best leader of the group.

The second top hierarchies in the group are the beta wolves and they involve themselves in supporting the alpha wolves for decision making and in the absence of alpha wolves, beta wolf can act as an alpha wolf. The alpha wolf is respected by the beta wolf and these beta wolves dominate the other two remaining wolves' category. Beta wolves act on the words of alpha wolves and incorporate the alpha wolf commands within the group and communicate the result to the alpha wolf. Thus, beta wolf acts intermediate between the alpha wolf and the remaining delta and gamma wolves.

Next comes the delta wolves which obeys the alpha and beta wolves but dominate the omega wolf sector. The wolves coming under the delta category are, hunting wolves, elder wolves, care taker wolves and sentinels. The basic works of delta wolves include, acting as a security wolf by watching the boundary of their living place, providing warning messages when danger is noted to occur, protecting the group, and providing food for the group at the time of hunting and will take care of the sick wolves as doctors. The

last category in the hierarchy is the omega wolf and this submits itself to all the higher dominant wolves in the group. These wolves eat only at the end and when omega wolves are not there, the other three wolves – alpha, beta and delta wolves fights with each other and hence these omega wolves adjust in the group with all the other higher groups and completely satisfies the entire group. Based on the hierarchical nature of the wolves, the GWO algorithm is modeled as follows.

3.1.1 Modeling of GRAY Wolf Optimizer Algorithm

GWO algorithm is modeled based on their social hierarchy, prey encirclement behavior, their hunting behavior, searching nature (exploitation) and attacking behavior (exploration).

In the process of modeling GWO, the alpha is the ones with the fittest solution, beta are the ones with the second fittest solution and the third fittest solution are the delta parts. The other remaining solutions are considered as the omega wolves.

The optimization is carried out by alpha, beta and delta and the omega wolf follows these wolves. When the hunting process is initiated by the gray wolves, the fixed prey which is identified to be the target is rounded and will be encircled by the prey. The prey encirclement behavior is represented by the distance traversed and is given by Equation 1 and 2,

$$D = \left| C X_p(t) - X(t) \right| \quad (1)$$

$$\text{and } X(t+1) = X_p(t) - A \cdot D \quad (2)$$

The quantities in the Equations 1 and 2 are the vector quantity, which specifies the magnitude and the direction in which they move. In the given Equations 1 and 2, the position of the prey is given by X_p , the position of the gray wolf is given by X , coefficient vectors are given by C and A and the distance traversed during encirclement is given by D . The computation of quantities A and C is done as given below as Equations 3 and 4,

$$A = a(2r_1 - 1) \quad (3)$$

$$C = 2r_2 \quad (4)$$

In the above equations, r_1 and r_2 are the randomly generated vectors between 0 and 1 and the value of 'a' is initialized to 2 and this gets varied in decreasing order at the process of iteration.

The alpha wolves guide the hunting mechanism and at certain instances, the beta and delta wolves tend to participate in the hunting process.

In case of random search space, there lies the problem of identifying the prey location. Under such situation, it is noted that the hierarchy of wolves – alpha, beta and delta tend to be the best solutions and they will be aware of the location of prey.

At this juncture, the initial three best solutions are noted and the remaining search agents will be advised to update their positions with respect to the position of the best search



PROPOSED SELF – REGULATED GRAY WOLF OPTIMIZER BASED EXTREME LEARNING MACHINE NEURAL NETWORK CLASSIFIER FOR LUNG CANCER CLASSIFICATION

agents meaning that the prey is located nearer to it. Initial first three best solutions are given by the Equations 5, 6 and 7,

$$D_\alpha = |C_1 \cdot X_\alpha - X|$$

$$X_1 = X_\alpha(1 - A_1) \tag{5}$$

$$D_\beta = |C_2 \cdot X_\beta - X|$$

$$X_2 = X_\beta(1 - A_2) \tag{6}$$

$$D_\delta = |C_3 \cdot X_\delta - X|$$

$$X_3 = X_\delta(1 - A_3) \tag{7}$$

The distance traversed by encirclement and the prey positions with that of the hunting positions of alpha, beta and delta are given by the Equations 5, 6 and 7 respectively. Finally, the position of the current search agent is given by the Equation 8,

$$X(t+1) = \frac{(X_1 + X_2 + X_3)}{3} \tag{8}$$

Generally, the optimization approach of the GWO starts with the random generation of population of gray wolves which are to be the candidate solutions. During simulation process, alpha, beta and delta wolf estimates the possible prey position.

The parameter in Equation 9 plays a vital role in the exploitation and exploration process and at the time of simulation they are decreased from 2 to 0. On $|A| > 1$, the solution gets diverged from the prey and $|A| < 1$, gets converged towards the prey.

The exploration and exploitation are dependent on the values of 'a' and 'A'. The main parameters to be adjusted includes 'a' and 'C' in GWO. The prey encirclement process in GWO forms a circle shaped neighborhood region and it defines a hyper sphere. The parameters A and C play a key role in defining the radii inside the circular search space.

3.1.2 Pseudo Code for GWO Algorithm Process

The GWO algorithm process is based on the positional movements of alpha, beta and delta wolves and the prey position which forms the entity in the random search space[9]. The pseudo code for the GWO algorithm is as presented in Table 2.

TABLE 2: Pseudo Code of GWO Algorithm

Start
Randomly initialize the population of gray wolves
Initialize the values of a, r ₁ and r ₂ and A and C
Calculate the fitness of each search agents
X _α - first best search agent (alpha wolf)
X _β - second best search agent (beta wolf)
X _δ - third best search agent (delta wolf)
While Maximum number of iterations not attained
For each search agent
Perform position update of the current search agent
End
Perform updation of a, A and C
Calculate the fitness of all search agents
Update the search agents positions X _α , X _β and X _δ .
Update the iteration count
End
Return the output value of the best search agent X _α .
Stop

The flowchart depicting the process of GWO is as shown in the Figure 2.

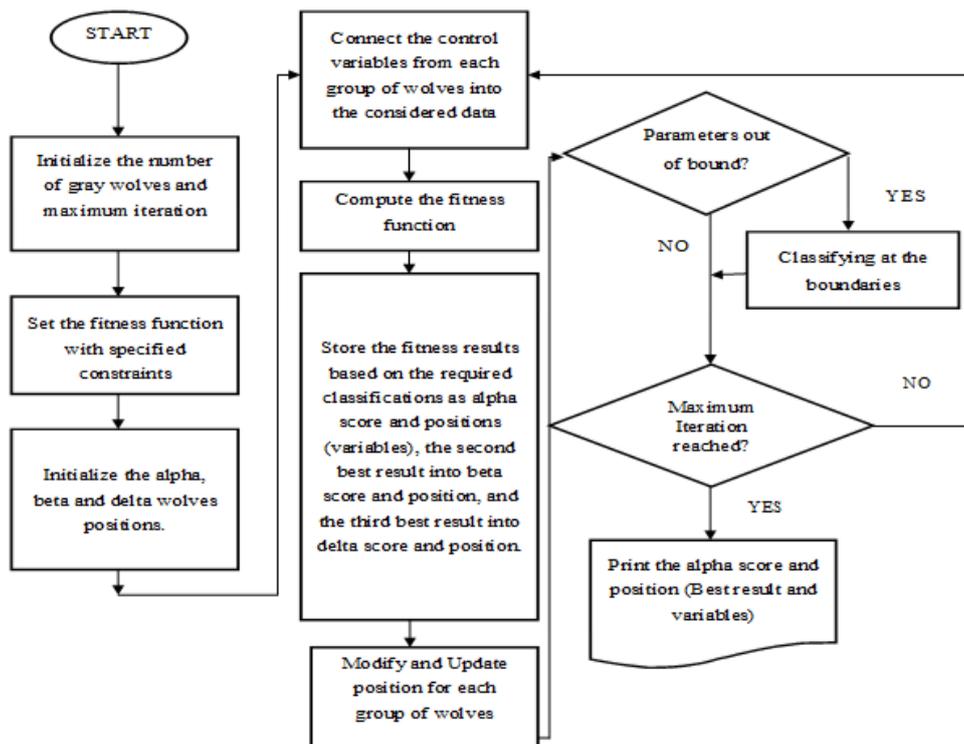


Figure 2: Process Flow of GWO Algorithm

3.2 Proposed Self-Regulated GWO Algorithm for ELM Tuning

Considering the traditional GWO algorithmic flow, based on the fitness computed the category of three

wolves are located as the best search agents and these best three wolves identified will progress in the iteration process and involve themselves to locate the optimal location of the presence of the prey. At the time of encirclement of the prey by the wolves, the distance traversed by the wolves is a key parameter which in turn makes the other wolves to update their positions randomly around the identified target prey location. From the traditional algorithm [10], it is noted that the GWO permits its search agents for updating their positions based on the positional location of alpha, beta and delta and it then starts attacking the prey. At this point, it can be inferred that there is a possibility for occurrence of local optima during the algorithmic flow. The encirclement of the prey with the distance traversed is noted to perform exploration process only to a certain extent and hence more operators are required to carry out the process of exploration.

Henceforth, this paper focuses on developing a Self-Regulated Gray Wolf Optimizer (SRGWO) to carry out an efficient exploration and exploitation process for search of the prey and achieve a better optimal solution. The following subsections will detail the flow of the proposed self-regulated GWO.

3.2.1 Mechanism for the Self-Regulated Learning GWO Algorithm

Let's revisit the distance traversed and prey position equations as given in the Equations 5 through 7. In all these equations, for the three wolves to be dominant in the hierarchy, the three best solutions are chosen from the fitness evaluated for all the generated search agents. The position of the current search agent is the average of the positions of the three wolves – alpha, beta and delta. The components C_1X_α , C_2X_β and C_3X_δ are the major components in the Equations 5 through 7 which decide the overall exploration and exploitation capacities of the proposed SRGWO algorithm. These C_1X_α , C_2X_β and C_3X_δ are the first three global best solutions in the current iteration. X_1 , X_2 and X_3 are the positions of the prey that takes the values contributed to be the best fitness values in the current iteration. The replacement is carried out in the distance traversed locations D_α , D_β and D_δ of the Equations 5 through 7 wherein along with the best fitness positions, the worst positions are also considered. Thus D_α , D_β and D_δ tend to possess two set of components based on self-regulated criterion of considering both the best and worst positions.

Thereby during the initial stages of the iteration process, the majority of the components are influenced by the worst positions to establish the exploration and nearing the final iterations, the majority of components are influenced to establish exploitation. The self-regulating GWO tends to achieve faster convergence in comparison with that of the traditional GWO process due to the fact that the global best point takes the value from both the best position and worst position and each component position updation thus takes place based on the self-regulating principle. The self-regulating principle is represented by the Equations 9, 10 and 11,

$$D_\alpha = |C_1 \cdot X_\alpha - X| + |C_1 \cdot X_{\alpha(worst)} - X|$$

$$X_1 = X_\alpha(1 - A_1) \tag{9}$$

$$D_\beta = |C_2 \cdot X_\beta - X| + |C_2 \cdot X_{\beta(worst)} - X|$$

$$X_2 = X_\beta(1 - A_2) \tag{10}$$

$$D_\delta = |C_3 \cdot X_\delta - X| + |C_3 \cdot X_{\delta(worst)} - X|$$

$$X_3 = X_\delta(1 - A_3) \tag{11}$$

This self-regulating concept acts as a best diversifier in achieving the global mechanism for thoroughly exploring the solution space. Table 3 presents the pseudo code for the proposed SRGWO algorithmic process.

Table 3: Pseudo Code of the Proposed SRGWO Algorithm

Start
Initialize the population of gray wolves randomly
Initialize the values of a, r ₁ and r ₂ and A and C
Compute the fitness of each search agents
X _α - first best search agent (alpha wolf)
X _β - second best search agent (beta wolf)
X _δ - third best search agent (delta wolf)
X _{α(worst)} - first worst search agent (alpha wolf)
X _{β(worst)} - second worst search agent (beta wolf)
X _{δ(worst)} - third worst search agent (delta wolf)
When Maximum number of iterations not attained
For each search agent
Perform position update of the current search agent
End
Perform updation of a, A and C
Compute the fitness of all search agents
Update the search agents positions
Update the iteration count
End
Return the output value of the best search agent
Stop

IV. PROPOSED SRGWO BASED ELM NEURAL NETWORK CLASSIFIER FOR LUNG CANCER CLASSIFICATION

For the past decades it is well noted that Artificial Neural Networks (ANNs) are widely employed for pattern recognition and image classification applications [11]. This is due to their generalization and conditioning capability and as well on the requirement of minimal training points with faster convergence time.

ANN classifiers are observed to perform better and produce classification accuracy relatively higher than that of the traditional classifiers.

Numerous neural network architectures like multiple adaptive linear neuron network, radial basis function network, hop field neural network, probabilistic neural network, BPNN, support vector machines and its variants were used for pattern and image classification applications and the lack in all these cases are the incurred selection time because of the preprocessing speed delay.

Generally, to obtain better classification accuracy, more training data is to be employed than that of the testing data and the selection of suitable training algorithms is important to enable the considered application not to be stuck up with the local or global optima problem.



PROPOSED SELF – REGULATED GRAY WOLF OPTIMIZER BASED EXTREME LEARNING MACHINE NEURAL NETWORK CLASSIFIER FOR LUNG CANCER CLASSIFICATION

The said lacunae is to be overcome for improving the training performance and obtaining better classification accuracy is noted to be achieved via a neural network architecture, Extreme Learning Machine (ELM) Neuronal Classifier proposed in Zhu et al. wherein the training for single hidden layer feed forward neural networks is carried out [12].

4.1 Extreme Learning Machine Neuronal Classifier (ELMNC) –An Overview

ELMNN is developed in the year 1992 and is a supervised learning algorithm that aims to solve linear and non-linear classification problems [13]. On comparison with the other neural network architectures, ELMNN is a well-known single layer feed-forward neural net that possesses one hidden layer.

ELM can be applied for larger training samples and the variation in the number of hidden neurons with respect to the different ratios of the number of features of testing and training data are examined [14].

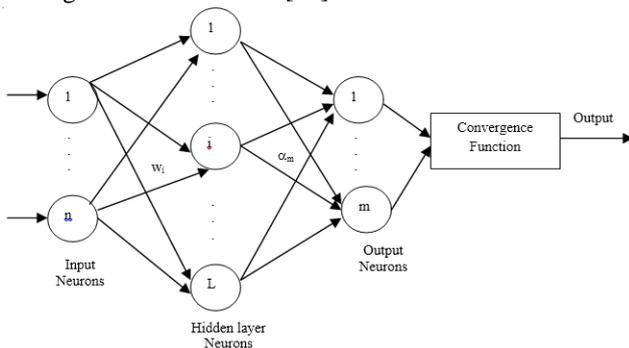


Figure 3: Architecture of the ELM Network (Single Layer Feed Forward Neural Network)

Basically, the ultimate aim of the Extreme Learning Machine neuronal classifier is to separate the presented classification data with a linear decision separating plane and then to maximize the margin in between the various categories of classes. This presents the convex quadratic programming problem. It can be noted that the training process of ELM involves solving a quadratic optimization problem which employs the applicability of optimization approaches from numerous recent mathematical or heuristic techniques. In general, ELM classifier is chosen for the considered application based on the following facts.

Even for difficult search spaces and for complex data sets this classifier tends to provide good solutions, because of the convergence characteristic ELM provides a competitive good solution. ELM approach is observed to reduce the computational burden and the enormous time taken in earlier classifier modules. ELM is noted to achieve good generalization performance in extremely fast learning speed. Figure 3 shows the basic architecture employed for image classification of ELM neural network [15,16]. The basic Extreme Learning Machine classifier algorithm [13] employed for classification application is as given below,

Consider N different samples $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}^m$, where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ represents the input vector and $y_i = [y_{i1}, y_{i2}, \dots, y_{im}]^T$ represents the target vector, the standard supervised learning feed forward network with M hidden nodes and activation function $g(x)$ are formulated as Equation 12:

$$\sum_{i=1}^M \lambda_i f_i(x_j) = \sum_{i=1}^M \lambda_i f(w_i \cdot x_j + b_i) = O_j, j = 1, \dots, N \quad (12)$$

Where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ represents the weight vector connecting the input nodes and the i-th hidden nodes, $\lambda_i = [\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{im}]^T$ represents the weight vector connecting the i-th hidden node and the output nodes, and b_i is the bias of the i-th hidden node. Assuming that the function approaches all N sample by zero error, i.e., there exist parameters (w_i, b_i) and λ_i such that Equation 13:

$$\sum_{i=1}^M \lambda_i f(w_i \cdot x_j + b_i) = y_j, j = 1, \dots, N \quad (13)$$

The given equations can be simplified as, $H\lambda = T$, where

Equation (3.14),

$$H(w_1, \dots, w_M, b_1, \dots, b_M, x_1, \dots, x_N) = \begin{bmatrix} f(w_1 \cdot x_1 + b_1) & \dots & f(w_M \cdot x_1 + b_M) \\ \vdots & & \vdots \\ f(w_1 \cdot x_N + b_1) & \dots & f(w_M \cdot x_N + b_M) \end{bmatrix}_{N \times M} \quad (14)$$

$$\lambda = \begin{bmatrix} \lambda_1^T, \dots, \lambda_M^T \end{bmatrix}_{M \times m}^T, T = \begin{bmatrix} T_1^T, \dots, T_N^T \end{bmatrix}_{N \times m}^T$$

The solution of the linear system is given by $\lambda = H^\dagger T$, where H^\dagger represents the Moore-Penrose generalized inverse of hidden layer output matrix H. The ELM algorithm can be given by the following three steps,

- Input: Consider a training data set $\{(x_j, y_j) | x_j \in \mathbb{R}^n, y_j \in \mathbb{R}^m, j = 1, \dots, N\}$, with an activation function g and that of the number of hidden nodes M ;
- Output: the weights of hidden layer to output layer;
- Step 1: Randomly generate the weights and bias (w_i, b_i) values, $i = 1, \dots, M$;
- Step 2: Compute the hidden layer output matrix H ;
- Step 3: Compute the weights matrix of output $\lambda = H^\dagger T$.

In comparison with that of the general Artificial Neural networks the extreme learning machine approach employs only a minimal number of parameters for tuning. The fundamental concept of ELM lies in the selection of a kernel function and their equivalent kernel parameter possesses control over the convergence speed and henceforth is the quality of final solution obtained.

4.2 Proposed SRGWO Based ELM for Lung Cancer Classification

The proposed approach contribution combines the concept of SRGWO for optimizing the weights in ELM Neural Network. This presented Self-Regulated Gray Wolf Optimizer with ELM enables the selection of input weights for increasing the generalization performance and as well for the conditioning of the single layer feed-forward neural network. The various steps involved for the proposed approach are as follows,

- Step 1: Initialize the search agents with a set of input weights and hidden biases:
 $[W_{11}, W_{12}, \dots, W_{1n}, \dots, W_{21}, W_{22}, \dots, W_{2n}, \dots, W_{H1}, W_{H2}, \dots, W_{Hn}, b_1, b_2, \dots, b_H]$. These presented weights and biases will be randomly initialized within the range of $[-1, 1]$ on D dimensions in the search space.
- Step 2: For each search agent in the group, the respective output final weights are computed at ELM as given in Equation 14.
- Step 3: Now Invoke SRGWO.
- Step 4: Then evaluate the fitness function which is the MSE of each search agent in the group, and is evaluated as Equation 15,

$$MSE = \frac{1}{N} \sum_{i=1}^N E_i^2 = \frac{1}{N} \sum_{i=1}^N (y_k^i - d_k^i)^2 \quad (15)$$

Where, N represents the number of training samples, and the term y_k and d_k specifies the error of the actual output and target output of the k^{th} output neuron of i^{th} sample.

Henceforth, the fitness function $f(x)$ is defined by the MSE. For avoiding the over-fitting of the single layer feed forward neural network, the fitness of each member is adopted as the MSE on only the validation set instead of the whole training set.

- Step 5: Find the Self-regulating acceptance criterion based on the fitness.
- Step 6: Update the distance traversed and prey position equations of each search agent's as given in Equations 9 through 11.
- Step 7: Terminating Criteria – The algorithm repeats steps 2 to 6 until certain criteria are met, along with the hard threshold value as the maximum number of iterations. When the training process stops; the algorithm reports the values of the optimal weights with minimal MSE as its solution.

Thus, the proposed SRGWO with ELM determines the best optimal weights W and bias b so that the fitness reaches the minimum value for achieving better generalization performance, with minimum number of hidden neurons, considering the advantages of both extreme learning machine classifier and self-regulated GWO.

During the process of the selection of optimal input weights, the SRGWO considers not only the MSE on validation set but the norm of the output weights as well. The proposed SRLGWO based ELM classifier combines the feature of SRLGWO into ELM classifier for computing the optimal weights and bias for making the MSE minimal.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In the present search, the proposed SRGWO – ELMNC algorithm is applied for lung cancer classification and the computed results are presented in the following subsections

5.1 Clinical Datasets

The details of lung benchmark dataset taken from the Lung Image Database Consortium (LIDC), normal healthy lung image dataset and real time clinical data from hospitals

and diagnostic centers employed in the simulation are given in Table 4.

The proposed algorithm is simulated in MATLAB R2012a environment for the medical image datasets and executed in a PC with Intel core i5 processor with 3.5 GHz speed and 2 GB RAM with 64 bit operating system.

Table 4: Clinical Datasets

Dataset	Availability	No of Cases	Tumor Types	Slice Thickness
Lung Image Database Consortium (LIDC)	Public	48	Healthy Lung Images	0.75 mm – 1.25 mm
Lung Image Database Consortium (LIDC)	Public	104	Cancer Images Well Circumscribed nodules – 22 Vascularized nodules – 46 Juxta – pleural nodules - 13 Pleural – tail nodules - 23	
Accura Diagnostics Centre (Coimbatore)	Private	90	Healthy lung Images	
PSG Institute of Medical Science & Research (PSGIMSR) (Coimbatore)	Private	51	Cancer Images Well Circumscribed nodules – 11 Vascularized nodules - 27 Pleural – tail nodules - 13	

From Table 4, it can be observed that 155 patient data are collected including both men and women affected by lung cancer and the healthy image data taken for training process is 138.

With the data samples from LIDC, hospital and diagnostic centre, it is noted that the average age of the patients considered include 64.7 years with the youngest patient of 18 years of age and the oldest patient of the age 85 years.

In case of lung, CT scan is carried out by giving a kilo voltage peak distribution of 120 – 140 KVp with a variation in the current from 25 to 40 mAs with respect to the age of the patients.

For the complete research the total number of lung nodules considered is 110 and of size greater than 3 mm and related to these lung nodules expert opinion is received from the radiologists.

Based on the size of nodules, the radiologists classify the cancer cells and present their opinion. There exists primary and secondary stage lung cancer nodules with four various kinds of nodules namely, well circumscribed nodules, vascularized nodules, juxta – pleural nodules and pleural – tail nodules.

Figure 4 represents the sample lung cancer image dataset and Figure 5 represents the sample healthy lung images which are employed for training and testing process.



PROPOSED SELF – REGULATED GRAY WOLF OPTIMIZER BASED EXTREME LEARNING MACHINE NEURAL NETWORK CLASSIFIER FOR LUNG CANCER CLASSIFICATION

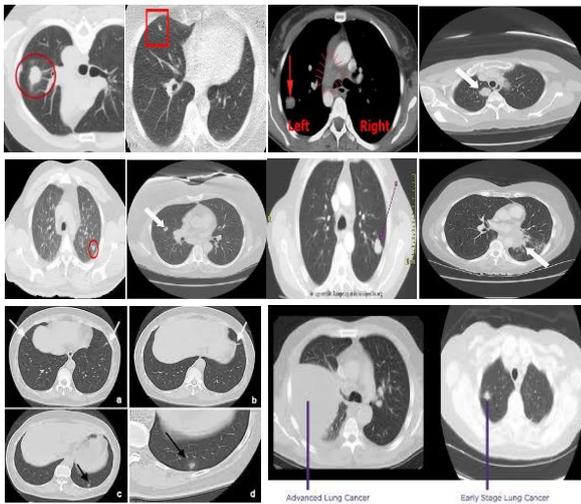


Figure 4: Samples of Lung Cancer Dataset

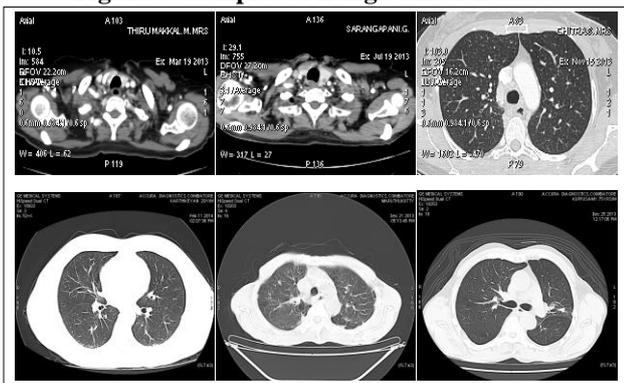


Figure 5: Samples of Healthy Lung Images

5.2 Metrics Considered for Image Classification

In the present approach, for implementing the proposed SRGWO – ELM algorithm for the clinical image datasets from the repositories and hospitals, the features are extracted and the parameters – classification accuracy, sensitivity and specificity are computed and tabulated.

Thus, in the present approach, the performance of the proposed methodology is computed in terms of sensitivity, specificity and classification accuracy [17]. Basically, sensitivity and specificity are the statistical measures employed basically to carry out the performance of the classification operations.

A perfect image classifier or a predictor will be described by 100 % sensitive and 100 % specific solutions. Since the considered application is of medical image data set involving complex images, the classification should be carried out in an accurate manner. Thus, sensitivity and specificity are selected to be the parametric indices for carrying out the medical image classification.

Sensitivity (true positive fraction) is the probability that a diagnostic test is positive as defined by Equation 16, given that the person has the cancer nodules.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{16}$$

Specificity (true negative fraction) is the probability that a diagnostic test is negative as defined by Equation 17, given that the person does not have the cancer nodules.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{17}$$

Classification Accuracy is the probability that a diagnostic test is correctly performed. Classification accuracy for the datasets is measured by the following Equation 18,

$$\text{Classification Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{18}$$

Where,

True Positives (TP) - Correctly classified positive cases - identifies the cancer nodules presence

True Negative (TN) - Correctly classified negative cases - identifies no cancer nodules are there

False Positives (FP) - Incorrectly classified negative cases - without cancer but identifies as cancer

False Negative (FN) - Incorrectly classified positive cases- with cancer but identified as no cancer

Thus, in the present approach, above mentioned three parameters are effectively employed to determine the classification accuracy of the image datasets.

Further to these three parameters, medical image classification problems generally involve ROC as a performance metric for performing effective image classification. ROC is a two dimensional measure for classification performance and is widely employed in image classification researches for assessing the performance of the considered image datasets.

ROC curve is a plot between the sensitivity and specificity with specificity on the x axis and sensitivity on the y- axis. On ROC curve being noted to rise rapidly towards the upper right hand corner of the plot or if the value of area under the curve is high, the image classification test is then said to be the one that performs well.

The image classification reliability is guaranteed when the area is close to 1.0, and when the area is noted to be close to 0.5, the image classification test is then said to be unreliable. Area under Receiver Operating Characteristics (AUROC) curve is used in the present approach for measuring the test accuracy. The area under ROC curve will be the part of the area of unit square. The region of convergence or AUROC for image classification is to satisfy the inequality condition as Equation 19,

$$0 \leq AUROC \leq 1 \tag{19}$$

On AUROC being close to 1 (area of unit square) this indicates reliable medical image classification. Fundamentally, integrating the area under the ROC curve, AUROC can be determined by evaluating the summation of the area of trapezoids of the curve or by Wilcoxon signed-rank test [18]. Henceforth, the approach contribution considers AUROC as a metric for performing image classification of the considered medical images from the database and the hospitals.

5.3 Feature Analysis and Segmentation of Lung Images

In medical processing applications, the small homogeneous portion region is manually selected and is referred to as ROI. For all the images considered from LIDC and from that of the hospitals and diagnostic centers, histogram equalization is performed. The features are extracted as per the first order and higher order statistical measures and are as tabulated in Table 5.



Table 5: Feature Measures for the Considered Medical Image Datasets

Image Datasets	Average Mean	Average Standard Deviation	Average Skewness	Average Kurtosis	Average Entropy	Average Homogeneity	Average Second Moment
LIDC cancer datasets	0.7654	0.7767	0.4563	0.7601	0.4457	0.9104	0.6572
Real time lung cancer datasets (PSGIMSR)	0.8156	0.8342	0.6154	0.7927	0.5129	0.8975	0.7128

Initially, morphological operations are employed to segment the lung from the CT images. First, the gray scale image is converted into binary image.

The input image pixels with an intensity greater than that of a threshold level is replaced with a value '1' and the remaining pixels with the intensity less than that of the threshold level is replaced with a value '0'.

The morphological operation is carried out to the binary image and the regularly employed structuring element based morphological operation is performed. The key advantage of employing morphological operation is its speed and its simplicity.

Figure 6 shows the sample input image and histogram equalization carried out on a sample cancer image. Figure 7 shows the lung cancer image with the final segmented cancer part.

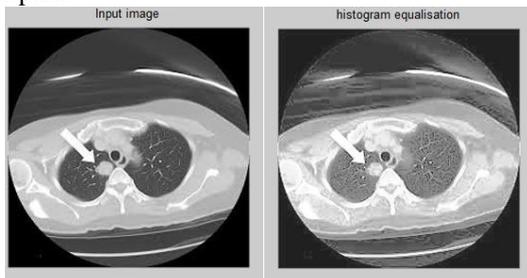


Figure 6: Sample Input Lung Cancer Image and its Histogram Equalization

5.4 Lung Cancer Classification with the Proposed SRGWO based ELM

The proposed Self – Regulated GWO based Extreme Learning Machine Neuronal Classifier is now invoked for the segmented output for efficient and accurate cancer classification. The complete details on the simulation process and the performance of the proposed SRGWO – ELM classifier for the dataset are presented in this section. The weights and bias for the input layer and hidden layer of the ELM neural architecture is optimized employing the proposed SRGWO.

The inputs to the ELM network are 7 and the hidden neurons (*H*) is determined as 5 with the help of best parameters selected with SRGWO. The number of outputs specify each of the lung nodules namely, well circumscribed nodules, vascularized nodules, juxta – pleural nodules, pleural – tail nodules and healthy lung without cancer tissues. Consequently, the structure of the single layer feed forward ELM neural network is 7-5-5. Table 6 presents the algorithmic parameters for the proposed self – regulated GWO and ELM.

Table 6: Algorithmic Parameters of the Proposed SRGWO – ELM

Parameters	SRGWO	Parameters	ELM
Population size	20	Initial weights and Bias between networks	Random between 0 and 1
No. of generations	100	Learning rate	1
a	2	No. of Neurons in input and output layers	Based on data sets considered
r1 and r2	0.5	Maximum Iteration	100

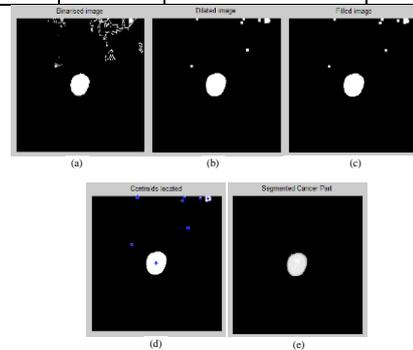


Figure 7: (a) Binarised Image (b) Dilated Image (c) Filled Image (d) Centroids Located (e) Segmented Output for a Sample Lung Cancer Image

Table 7 compares the proposed SRGWO – ELM simulation results with the existing tested classifiers. The SRGWO – ELM classification performance is noted to be appreciably good. Hence, it can be noted that the Extreme Learning Machine classifier outperforms the standard binary linear Support Vector Machines (SVM) and BPNN for medical image classifier and proves better classification of healthy and lung cancer nodules. Further, it is noted that there exists a significant difference in the computation time amongst the approaches.

Table 7: Result of the Comparison of Lung Cancer Nodule Classification to State of Art Techniques

Image Datasets	BPN	Linear SVM	GA-ELM	GWO-ELM	Proposed SRGWO - ELM
LIDC cancer datasets	0.84	0.90	0.93	0.95	0.97
Real time lung cancer datasets (PSGIMSR)	0.87	0.89	0.90	0.92	0.96

The proposed SRGWO – ELM classifier showed the highest overlap with the manual segmentation for all the tissues. Table 8 depicts the average value based on the cancer segmentation and the classification methods based on feature extraction model and compared to the state of art techniques [2,19,20]. The proposed SRGWO – ELM model proves better and shows improved values on sensitivity, specificity, classification accuracy and AUROC for the lung cancer datasets. The algorithm segmented lung cancer part with its histogram is as shown in Figure 8.



PROPOSED SELF – REGULATED GRAY WOLF OPTIMIZER BASED EXTREME LEARNING MACHINE NEURAL NETWORK CLASSIFIER FOR LUNG CANCER CLASSIFICATION

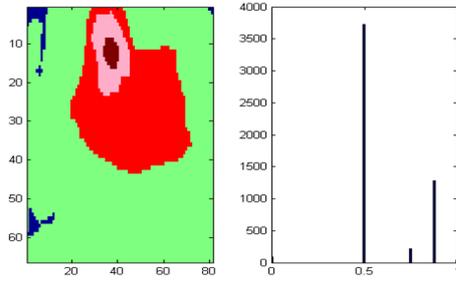


Figure 8: Lung Cancer Classification Using the Proposed SRGWO – ELM Algorithm

From Table 8, it can be inferred that the proposed SRGWO – ELM approach is noted to achieve highest classification accuracy for both the cancer datasets from the database consortium and as well the datasets from the hospitals and diagnostic centre.

Also, it can be noted that the area under the curve value tends towards 1, proving the reliability nature of the proposed algorithm.

Table 8: Comparison of Various Lung Cancer Classification Methods

Image Datasets	Metrics for Image Classification	Hybrid Logistic Regression – Artificial Neural Network Approach [17]	Hopfield Neural Network & Fuzzy Clustering Approach [18]	Back propagation Neural Network Approach [2]	Proposed SRGWO – ELM Approach
LIDC cancer datasets	Sensitivity	68.75	70.37	92.12	98.77
	Specificity	96.45	97.18	98.67	99.63
	Classification Accuracy	88.12	89.07	91.11	96.72
	AUROC	0.905	0.901	0.900	0.980
Real time lung cancer datasets (PSGIMSR)	Sensitivity	72.39	81.37	90.65	96.57
	Specificity	85.47	89.01	95.47	98.21
	Classification Accuracy	84.69	87.65	90.65	95.69
	AUROC	0.723	0.872	0.870	0.976

5.5 Lung Cancer Classification Analysis

The aim of the proposed work is to correctly classify the normal and cancer images. In the analyses of the images, each image is classified into one of the two classes, either normal or cancer affected images. Based on the features that have been extracted, the proposed classifier separates the same. Combining the publically available dataset (LIDC with healthy lung and with cancer affected) as well the real time datasets (from Accura diagnostic centre and PSGIMSR), dataset of 152 and 141 patients are analyzed respectively with the features extracted and the segments obtained; that are noted to be well separated for the two classes (affected and not affected) i.e., the patients were carefully selected with respect to the average feature value of the measures.

Table 9 presents the classification accuracy computed between various classifiers with that of the proposed classification model. The considered images were then classified by the proposed SRGWO – ELM classifier with

leave-one-out cross-validation that is the classifier was trained with 29 patients and then the one patient not used in training was classified. This process is rotated in a manner such that all the patients are used as a test set. In Table 9, it is noted that the samples chosen are 152 cases of LIDC data and 141 cases for real time data. The samples are converted into their respective pixel values and are input to the ELM model. The considered samples were loaded in a sequential manner; and delay process was well within the permissible extent. Table 10 presents few other metrics observed during the run time of the proposed SRGWO – ELM algorithm along with the existing classifiers for LIDC dataset. From Table 10, it can be observed that the proposed classifier model is noted to achieve the best solution for the various other network training parameters considered. Figure 9 shows the plot of the training efficiency of the proposed approach for LIDC dataset with respect to the number of iterations. From the plot, it can be noted that the training efficiency is 98.83%.

Table 9: Accuracy Results Amongst Various Classifiers on Combining the Healthy and Cancer affected Lung Images

Datasets (Combined for Classification)	Image Types	BPN	SVM	ELM	GWO – ELM	Proposed SRGWO - ELM
Lung Image Database Consortium (LIDC) (152 cases – 48 healthy and 104 cancer affected)	Healthy	87.8 %	90.6%	81.2%	93.2 %	96.1%
	Well Circumscribed nodules	84.7%	86.1%	80.4%	92.8%	94.9%
	Vascularized nodules	80.8%	85.1%	79.4%	91.6%	95.3%
	Juxta – pleural nodules	84.9%	88.3%	75.4%	89.8%	91.2%
	Pleural – tail nodules	78.2%	75.4%	70.8%	86.5%	89.2%
Accura Diagnostics Centre & PSGIMSR (141 cases – 90 healthy and 51 cancer affected)	Healthy	90.6%	94.5%	87.9%	96.5%	98.7%
	Well Circumscribed nodules	86.4%	90.8%	84.2%	93.7%	97.4%
	Vascularized nodules	91.8%	93.2%	89.0%	95.1%	98.0%
	Juxta – pleural nodules	91.6%	92.4%	88.0%	94.9%	96.2%
	Pleural – tail nodules	86.4%	89.7%	81.6%	94.3%	97.8%



Table 10: Results of Comparison Network Training Parameters for the LIDC Dataset

Parametric Measures	Hybrid Logistic Regression – Artificial Neural Network Approach [17]	Hopfield Neural Network & Fuzzy Clustering Approach [18]	Back Propagation Neural Network Approach [2]	Proposed SRGWO – ELM Approach
Norm	378.67	231.98	196.72	121.47
MSE Error	0.9231	0.9745	0.1120	0.0097
Training Efficiency % Mean	86.25	81.26	94.57	98.83
Training Efficiency % STD	6.71	6.47	5.64	6.05
Testing Efficiency % Mean	90.35	88.61	95.84	97.12
Testing Efficiency % STD	6.13	6.09	5.45	6.03
Hidden neurons	19	7	12	5
Accuracy%	92.80	89.07	97.14	98.64
Time min	3.27	2.96	3.50	2.01

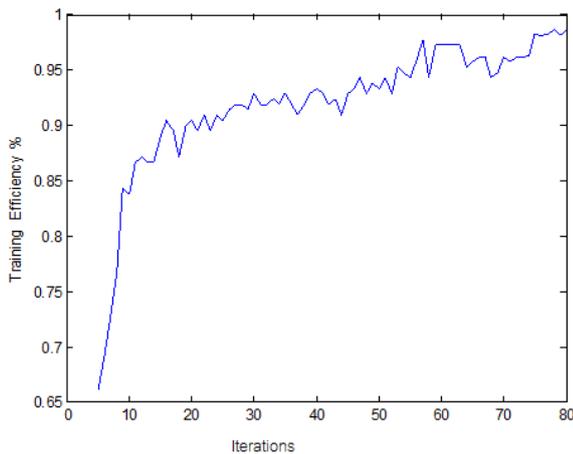


Figure 9: Training Plot of the Proposed SRGWO – ELM Approach

VI. CONCLUSION

A Self-Regulated Gray Wolf Optimizer based Extreme Learning Machine classifier is proposed to carry out lung cancer classification along with the statistical feature extraction methods. Simulation shows that the proposed approach works well and produces higher classification accuracy than the conventional classifier methods. The modeled SRGWO and ELM along with feature and segmentation process shows the highest improvement than all the methods found in the existing literature studies in neural networks.

REFERENCES

1. D. M. Parkin, "Global cancer statistics in the year 2000," *The Lancet Oncology*, vol. 2, no. 9, pp. 533-543, 2001.
2. J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 202-209, 2014.
3. G. Castellano, L. Bonilha, L. M. Li and F. Cendes, "Texture analysis of medical images," *Clinical Radiology*, Elsevier, vol. 59, no. 12, pp. 1061-1069, 2004.

4. R. M. Haralick, K. Shanmugam and I. H. Dinstein, "Textural features for image classification," *IEEE Transaction on Systems, Man and Cybernetics*, vol. 3, pp. 610-621, 1973.
5. E. Bonabeau, M. Dorigo and G. Theraulaz, "Swarm intelligence: from natural to artificial systems," Oxford University Press, New York, USA, 1999.
6. I Jasmine SelvakumariJeya, J Suganthi, "Watermarking Relational Databases Using Genetic Algorithm with Ring Crossover Technique", *International Review on Computers and Software*, vol. 9, Iss. 3, pp. 456-467, 2014.
7. M. Dorigo, M. Birattari and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28-39, 2006.
8. J. Kennedy and R. Eberhart, "Particle swarm optimization in neural networks," *Proceedings of IEEE International Conference*, pp. 1942-1948, 1995.
9. H. Turabieh, "A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease", *American Journal of Operations Research*, vol. 6, no. 2, pp.136, 2016.
10. S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46-61, 2014.
11. S. N. Sivanandam and S. N. Deepa, "Principles of soft computing," Wiley India Ltd, 2007.
12. Q. Y. Zhu, A. K. Qin, P. N. Suganthan and G. B. Huang, "Evolutionary extreme learning machine," *Pattern Recognition*, vol. 38, no. 10, pp. 1759-1763, 2005.
13. G. B. Huang, Q. Y. Zhu and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489-501, 2006.
14. D. Chyzyk, A. Savio and M. Grana, "Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI," *Neurocomputing*, Elsevier, vol. 128, pp. 73-80, 2014.
15. J. Zhai, Q. Shao and X. Wang, "Architecture Selection of ELM Networks Based on Sensitivity of Hidden Nodes", *Neural Processing Letters*, pp. 1-19, 2015.
16. I Jasmine SelvakumariJeya, Deepa S N, "Lung Cancer Classification Employing Proposed Real Coded Genetic Algorithm Based Radial Basis Function Neural Network Classifier", *Computational and Mathematical Methods in Medicine*, vol. 2016, pp. 1-15, 2016.

**PROPOSED SELF – REGULATED GRAY WOLF OPTIMIZER BASED EXTREME LEARNING MACHINE
NEURAL NETWORK CLASSIFIER FOR LUNG CANCER CLASSIFICATION**

18. Z. Wang, Y. Hu and L. T. Chia, "Image-to-class distance metric learning for image classification", Proceedings of European Conference on Computer Vision, Springer Berlin Heidelberg, pp. 706-719, 2010.
19. F. Wilcoxon, "Individual comparisons by ranking methods," Biometrics bulletin, vol. 1, no. 6, pp. 80-83, 1945.
20. T. Tunc, "A new hybrid method logistic regression and feedforward neural network for lung cancer data," Mathematical Problems in Engineering, vol. 2012, pp. 1-10, 2012.
21. F. Taher, N. Werghi, H. Al-Ahmad and R. Sammouda, "Lung cancer detection by using artificial neural network and fuzzy clustering methods," American Journal of Biomedical Engineering, vol. 2, no. 3, pp. 136-142, 201