ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN INTEGRATED DEEP LEARNING BASED ENHANCED GREY WOLF OPTIMIZATION FOR LUNG CANCER PREDICTION

T. DIVYA[1], Dr. J. VIJI GRIPSY[2]

[1]Research Scholar, PSGR Krishnammal College for Women, Coimbatore, India.

[2]Associate Professor, PSGR Krishnammal College for Women, Coimbatore, India.

E-mail: [1]divyathirumurthi@gmail.com, [2]vijigripsy@gmail.com

ABSTRACT

Lung cancer is an extremely harmful disease that represents the leading cause of death among both males and females within the nation. The survival spans for lung cancer patients within the 10%-20% range are limited to a duration of five years. Nevertheless, in the event that lung cancer is identified in its early stages and promptly treated, there is potential for a reduction in death rates. When lung cancer is identified at an early stage during the screening procedure, the clinical response to treatment may exhibit variability and provide very favourable outcomes. The implementation of a dependable and automated system might greatly facilitate the early identification of lung cancer, even in remote regions. This research presents a unique technique called Integrated Deep Learning-based Enhanced Grey Wolf Optimization for lung cancer prediction (IDL-EGWO). In order to address the issue of instability and convergence accuracy that occurs when using the Grey Wolf Optimizer (GWO) as a meta-heuristic algorithm with a robust capacity for optimum search, A weighted average GWO algorithm is suggested as a way to try to fix the problems with the GWO, such as the fact that it can get stuck in local optima and has a slow convergence rate in later stages. This technique incorporates an Artificial Neural Network (ANN) during the training phase. The research included a range of performance criteria, including precision, recall, f-measure, accuracy, execution time, and root mean squared error. According to the experiment, the IDL-EGWO algorithm demonstrated a higher accuracy rate of 97% compared to the previous methods.

Keywords: *Lung Cancer Prediction, Optimization, Deep Learning, GWO, ANN, MLP*.

1. INTRODUCTION

Lung cancer is a form of tumour originating from the lungs, exhibiting a malignant nature, and marked by the presence of genetic instability. Lung cancer is a significant contributor to the death rates in India related to cancer. By implementing regular evaluations of individuals, a substantial portion of these fatalities might potentially be prevented, thereby decreasing the likelihood of developing lung cancer. Chest radiography, CT scanning, and MRI are just a few of the imaging techniques that can help with the early detection of lung cancer. The timely identification of cancers at an early stage has been shown to enhance the likelihood of human survival in comparison to cases when malignancies are detected at an advanced stage [10]. Numerous researchers have conducted investigations into the use of machine learning techniques for cancer diagnosis. However, the success rate of early detection has not shown significant improvement. The incidence of lung cancer is significantly elevated in those who engage in smoking behaviour, with a risk that is at least 20-fold greater compared to those who do not smoke. The first stage in diagnosing lung cancer involves the identification of symptoms. The symptoms mostly show the impairment and functional decline of the lungs. The most prevalent signs of lung cancer are frequent chest discomfort and coughing [12]. Among lung cancer patients, additional often-seen symptoms include breathing difficulties, feeling weak, unexpected weight loss, bleeding, and fatigue. Too far, the scientific community has not yet devised a screening method capable of early-stage detection of lung cancer, thereby limiting the potential for enhanced survival rates [11]. Chest radiography is a widely accessible modality for screening purposes; nonetheless, its reliability remains inadequate. The creation of a screening tool is essential in light of the findings of several researchers, who have determined that the timely detection of early-stage malignancies significantly enhances the prospects for successful treatment.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The yearly screening of low-dose computed tomography (LDCT) is advised for those who currently smoke or have stopped smoking within the last 15 years [14]. According to the American Society of Clinical Oncology, those who have a smoking history of 30 years or more and fall within the age range of 55 to 74 have an increased susceptibility to developing lung cancer.

The use of deep learning techniques facilitates the assessment and understanding of complex medical information, hence providing support in the areas of diagnosis, management, and prognostication of treatment outcomes across many clinical scenarios [9]. The medical business stands to undergo a comprehensive transformation as a result of the integration of artificial intelligence. AI applications have made significant advancements, enabling their expansion into industries that were previously considered inaccessible without human expertise. This progress may be attributed to the abundance of digital data, advancements in machine learning, and the development of robust computer infrastructure [1]. In recent years, there has been significant progress in the fields of caption creation, photo identification, and voice recognition due to advancements in deep learning, an artificial intelligence approach. Additionally, Graphic Processing Units (GPUs) have made it easier to use parallel architectural deep learning methods, which has led to higher accuracy in a number of areas, such as predicting illness.

Motivation

In recent years, experts relied mostly on experiential knowledge and supplemented it with laboratory-tested data or clinical information in order to diagnose lung cancer. The test results examined in these labs show variations based on factors such as smoking habits, yellowing of fingers, anxiety levels, peer influence, the presence of chronic diseases, allergies, wheezing, alcohol consumption, coughing, experiencing shortness of breath, trouble swallowing, and chest discomfort. The primary objective of this study is to use deep learning and optimization techniques to accurately predict the occurrence of lung cancer based on a provided dataset.

Research Contributions

The main contributions of this research work are as follows:

- To develop a new model based on deep learning to predict lung cancer
- To propose an efficient algorithm for integration
- To integrate a deep learning-based optimization algorithm for accurate prediction

The next sections of the paper are structured as follows: Section 2 provides an overview of the background and existing research related to lung cancer. Section 3 outlines the approach used, which involves the utilization of neural networks for modeling lung cancer and optimizing the grey wolf optimizer throughout the training phase. In this study, Section 4 provides an analysis and interpretation of the experimental findings, and Section 5 offers a comprehensive conclusion of the work provided.

2. LITERATURE REVIEW

Kannuswami et al. (2018) [6] proposed the use of an ANN to include texture and fractal information into a CAD system for the purpose of lung cancer detection. In this study, the use of fuzzy development was implemented as a means to enhance the rate of lung cancer diagnosis. Furthermore, the use of fractal and texture feature analysis enabled the identification of key characteristics. Ultimately, the detection strategies exhibited superior performance compared to the other detection mechanisms, resulting in improved detection accuracy.

In their study, Jiang et al. (2017) [5] proposed a methodology for the identification of lung nodules by recognizing several groups of patches within lung images. In this study, the Frangi filter was used to enhance the quality of the multi-group patches. Subsequently, an automated lung wall mending technique was implemented with the primary aim of preventing the omission of juxta-pleural nodules. Furthermore, this study developed four distinct CNN designs, each designed to address the four phases of nodule detection. The purpose of these architectures was to efficiently and accurately estimate the position of nodules. The simulation results also showed that the improved detection model worked better and had fewer false positives when it was used on the large datasets that were studied.

In their study, Saien et al. (2018) [17] proposed a hybridized classifier, namely a Random UnderSampling/boosting (RUSBoost) approach, to address the problem of imbalanced data in the lung

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

nodule dataset of the subjects. The RUSBoost classifier tells the difference between real nodules by looking at the features of both the nodule candidate population and the non-nodule population, which have properties that are not balanced. As a result, this strategy proved to be more successful in managing and segmenting the extensive clinical data sets, thereby achieving an increased convergence rate with fewer iterations.

Bouget et al. (2019) [2] did a study and came up with a 2D pipeline method that uses the U-Net technique to fix the problem of pixel-wise division and control information imbalances. The mask R-CNN algorithm improved the process of pixel-wise segmentation inside bounding boxes and improved instance recognition. In conclusion, a tracking technique was used to execute pixel-wise mark augmentation and 3D instance identification by measuring slices. In addition, the identified samples were characterized by a 3D mask that delineated the pixels, the degree of bouncing, and the centroid point. In this study, the SVM was used to diagnose lung cancer. This technique improved the feature selection process by taking into account both redundant and irrelevant information during training, resulting in reduced training time and computational complexity.

In their work, Palani and Venkatalakshmi (2019) [13] presented IoT-based lung cancer prediction modeling. Fuzzy cluster-enabled expansion and classification allowed frequent lung cancer surveillance. The research used remedial instructions to deliver healthcare functions. Fuzzy clustering was utilized to find transition zones for image splitting in this study. Additionally, fuzzy C-means clustering was performed to categorize the transitional zone in the lung cancer image. Otsu thresholding was also utilized to recover the transition zone from lung cancer images. The researchers also considered adding the right edge picture and morphological reducing techniques to improve segmentation. Object areas were extracted from edge lung cancer images using morphological clean-up and image area filling. Researchers used incremental classification with the optimization technique Association Rule Mining (ARM) to achieve success. They added CNN and temporal elements to the ordinary decision tree.

In their study, Haarburger et al. (2019) [4] proposed the use of a CNN as a means of assessing data from individuals diagnosed with cancer. This methodology successfully attained effective picture analysis, emphasizing the significance of processing high-dimensional data. Nevertheless, this approach is characterized by poor comprehension results.

Lai et al. (2020) [8] proposed a Deep Neural Network (DNN) that integrates biological phenomena and clinical information from several sources of knowledge. In this study, the principles of natural science were used to identify and validate predictive biomarkers. This approach demonstrated the highest hazard magnitude ratio with regard to both the training sets and the testing sets. Nevertheless, this methodology did not improve the accuracy of the categorization.

Kim et al. (2020) [7] were responsible for creating the Deep Learning Survival Prediction Model (DLPM). The data obtained was used as a tool to categorize surgical risk in cancer patients. This process yielded replicable analytical results. However, the cancer patients included in the verification set were of a small sample size, which therefore limited the statistical power of the applied mathematical analysis.

Mohamed et al. (2023) [18] introduced an innovative hybrid approach that enhances the precision of lung cancer classification by using a CNN model. The EOSA method was used to optimise the solution vector of the CNN architecture, which underwent training on separate 2D samples classified according to their anomalies. The EOSA-CNN algorithm did better than regular CNN and other hybrid algorithms based on metaheuristics. The fact that it had higher measures of specificity, sensitivity, recall, kappa, and accuracy demonstrated its superiority. The main achievement of this work is the effective use of the EOSA algorithm, a virus-based optimisation approach, to enhance the solution vector of the proposed CNN architecture.

Hussain et al. (2023) [19] suggest training the ML-CNN classifier on two categories: nodules (diseased, either malignant or normal) and non-nodules (non-diseased, either malignant or harmless, specifically normal). The ML-CNN with PSO model achieves accuracy, precision, sensitivity, specificity, and F-measure values of 98.45%, 98.89%, 98.45%, 98.62%, and 98.85%, respectively. The hybrid technique yields superior accuracy and achieves stronger convergence outcomes in comparison to other methods.

The authors (Oyelade and Ezugwu, 2021) [22] developed the Ebola optimization search algorithm (EOSA), drawing inspiration from the Ebola virus and its associated illness propagation model. The findings demonstrated that the suggested algorithm had similar performance to other cutting-edge optimisation methods in terms of scalability, convergence, and sensitivity analyses.

In their study, Shan and Rezaei (2021) [23] developed a feature selection technique using a novel optimisation approach known as Improved Thermal Exchange Optimization (ITEO). The objective of this method is to improve the efficiency and stability of the system. The segmentation of lung regions was achieved by the application of Kapur entropy maximisation and mathematical morphology. The authors obtained the 19 GLCM features from the segmented pictures for the final assessments. ITEO used a very effective artificial neural network, and the findings demonstrated that the suggested approach achieved an accuracy of 92.27%.

Priyadharshini and Zoraida (2021) [20] created bat-inspired metaheuristic convolutional neural network algorithms for the prediction of lung cancer using CAD technology. They decomposed the input picture using the discrete wavelet transform (DWT), resulting in a series of sub-bands. The low (LL) band refers to one of these sub-bands. The authors trained the lung cancer data using CNN, resulting in an accuracy of 97.43%.

Lu et al. (2021) [21] developed a novel convolutional neural network that provides optimum lung cancer diagnosis. Marine predators are incorporated into the network as a metaheuristic technique. The MPA-based strategy demonstrated an accuracy of 93.4%, a sensitivity of 98.4%, and a specificity of 97.1%.

3. METHODOLOGY

3.1 Data Collection

This experiment utilized the lung cancer dataset from the Kaggle repository. The dataset comprises 309 instances, 15 attributes, and one class attribute. Table 1 describes the dataset.

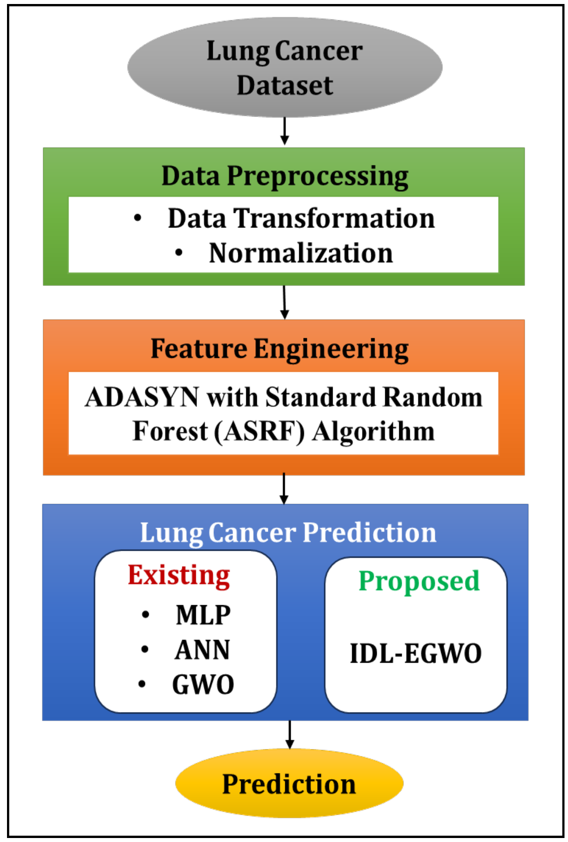

Figure 1: System Architecture

3.2 Data Preprocessing

Lung cancer detection begins with preprocessing, which fills gaps in the data and deletes unnecessary data.

Data Transformation

This dataset's GENDER and LUNG_CANCER attributes are objects. To use Sklearn's Label Encoder in Python to transform them into numbers.

Normalization

The utility class Label Encoder normalizes labels to 0–n_classes-1. It can also convert hash tables and comparable non-numerical labels to numerical ones.

Set all other properties to YES=1 and NO=0. Missing variables are imputed with three neighbours to increase model reliability.

Table 1: Dataset Description.

Attributes	Description
Gender	M(male), F(female)
Age	Age of the patient
Smoking	YES=2, NO=1
Yellow fingers	YES=2, NO=1
Anxiety	YES=2, NO=1
Peer_pressure	YES=2, NO=1
Chronic Disease	YES=2, NO=1
Fatigue	YES=2, NO=1
Allergy	YES=2, NO=1
Wheezing	YES=2, NO=1
Alcohol	YES=2, NO=1
Coughing	YES=2, NO=1
Shortness of Breath	YES=2, NO=1
Swallowing Difficulty	YES=2, NO=1
Chest pain	YES=2, NO=1
Lung Cancer	YES, NO

3.3 Feature Engineering

Several dataset characteristics must be extracted to simplify lung cancer diagnosis. It uses current features to generate new functionality using ADASYN with the Random Forest (ARF) algorithm. To merge ANXIETY and YELLOW_FINGERS into one feature like ANXYELFIN because the correlation matrix shows a higher than 50% correlation.

3.4. Integrated Deep Learning based Enhanced Grey Wolf Optimization for Lung Cancer Prediction (IDL-EGWO)

In the fundamental GWO method, the convergence factor undergoes a linear reduction from 2 to 0. However, in practical optimization problems, the algorithm's search process complexity results in a weakened search capability due to the linear variation of the convergence factor. Furthermore, it should be observed that the position update equation's first three wolf weight levels show equal values. The position of the wolf as the pack leader has a significant impact on its capacity to engage in hunting in its natural environment. This study suggests a framework for a nonlinear variation mode of convergence factor and the position updating the formula of weighted average while keeping in mind the previously mentioned limitations. In the position update equation, the coordinated algorithm's search capability simultaneously incorporates the beta distribution. The recommended method is utilized to accelerate the algorithm's overall convergence speed during the population initialization procedure, thereby ensuring its efficacy.

Scientists in the field of medicine like to use neural networks to represent unstructured problems because they can set up complex, non-linear links between input and output variables [6]. The fitness function is minimized by evaluating the root mean square error (RMSE) between the actual output and the output predicted by the ANN [3]. The difficulties of easily falling into local minima and exhibiting delayed convergence hinder the neural network approach. The neural network technique employs a method whereby distinct sets of weights and biases are generated throughout each iteration of the training phase. Consequently, every iteration will provide distinct prediction outcomes and rates of convergence. In order to address the limitations of the NN method, the GWO has been used to determine the most advantageous initial weights and biases for the NN algorithm. The GWO algorithm is designed to search for optimum solutions by exploring multiple pathways, with the aim of minimizing the likelihood of being caught in local minimums and enhancing the speed of convergence. The simulation results demonstrate a significant improvement in the algorithm's performance.

3.4.1 Grey Wolf Optimization (GWO) algorithm

It is one kind of swarm intelligence family of algorithms. This algorithm inherits the social behaviour of grey wolves, which is unique in its hierarchical administration and efficient group hunting of prey. Grey wolf (Canis lupus), which has another name as a timber wolf, is one of the largest members of the Canidae (dog) family. Mostly, this wild animal moves and hunts during the night [15]. Their favourite prey is large herbivores.

Typically, Grey Wolves exhibit a preference for cooperative living in the form of a group structure. To maintain a social hierarchy, the population of wolves is classified into Alpha, Beta, Delta, and Omega. Alpha Wolf is in the top position in the pack. Alpha is unbiased by gender; it may be male or female. They are often referred to as the dictators of a pack, as they are in charge of taking decisions regarding the selection of a sleeping place, waking time, and hunting time. The next ranking wolf is Beta, which is subordinate to Alpha and can help Alpha take decisions. Beta is an advisor for Alpha Wolf and takes Superior's commands to the entire pack, reflecting the feedback to the top. Moreover, beta wolves are possibly the next choice to become an alpha when an existing alpha dies or becomes too old to lead. Although a beta wolf can command his lower-level wolves, it must obey the alpha leader [16].

The third hierarchy of wolves is delta, which can command the lower-ranking wolves (Omega) but must obey the alpha and beta wolves. Delta wolves comprise scouts, caretakers, sentinels, hunters, and elders in a pack. Scouts guard the borders and give warnings in possible dangerous situations. The responsibility of caretakers is to take care of the ill, weak, and wounded wolves in a pack. Sentinel wolves fight and protect the members of their group. Past alpha and beta wolves are the elders of the pack. Hunter wolves have the duty of assisting the alpha and beta wolves in providing food to the entire pack.

The lowest level in the hierarchy of a pack is Omega, which comprises any wolf that does not categorize under alpha, beta, or delta. These are normally called scapegoats. Omega wolves must obey all upper-lead wolves. These wolves are given the least importance either during a chance to get food or get protection in the pack. The process of optimization begins by initializing the parameters, generating the random population of search agents, evaluating the fitness value, and identifying the best three search agents as Alpha, Beta, and Delta wolves. The position of each agent is updated for each iteration.

Grey Wolves succeed in hunting by exploiting their unique group hunting behaviour.

i) Track the movement of the prey and move towards it.
ii) Encircling and raiding the prey till it ceases movement
iii) Invasion of the prey

The hunting nature of wolves as

1. Encircling the prey
2. Hunting
3. Attacking
4. Searching

Encircling the Prey

For the mathematical representation of grey wolves' social hierarchy and hunting behavior, the names of the wolves in the community are Alpha (A), Beta (B), Delta (D), and Omega (O). The prey's location is designated X_p, whereas the A, B, and D positions are X_A, X_B, and X_D, respectively.

$$D = |M.X_p(k) - X(k)| \quad (1)$$

$$X_{K+1} = X_p(k) - G.D \quad (2)$$

The coefficient vectors G and M are updated for each iteration, and index k indicates the index of the current iteration. The coefficients G and M can be initialized with

$$G = 2a.r_1 - a \quad (3)$$

$$M = 2.r_2 \quad (4)$$

The parameter 'a' takes the value between 0 and 2 in a decreasing manner for each iteration; r_1 and r_2 take random values between 0 and 1.

$$a = 2 - i\left(\frac{2}{Max}\right) \quad (5)$$

r_1 and r_2 take out any random values from 0 to 1, and Max is the maximum number of iterations.

Hunting

While searching, the top-ranked wolves regularly update their positions. The position update is defined using the following set of equations:

$$D_A = |M_1.X_A - X_K| \quad (6)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$D_B = |M_1.X_B - X_K| \quad (7)$$

$$D_D = |M_1.X_D - X_K| \quad (8)$$

Also, the updating of hunting agents is represented mathematically by

$$X_1 = X_A - G_1(D_A) \quad (9)$$

$$X_2 = X_B - G_2(D_B) \quad (10)$$

$$X_3 = X_D - G_3(D_D) \quad (11)$$

The final optimum solution for the hunter-wolf from a 2D perspective can be given as

$$X_{K+1} = \frac{X_1 + X_2 + X_3}{3} \quad (12)$$

Attacking the Prey

When the grey wolf stops moving, it means that it is satisfied with its hunting procedure and is ready to attack the prey. The limiting range of G is reduced as the value of a is reduced from [− 2a, 2a] to [-1, 1] in this stage of hunting. In the next movement, the wolf may take any forward step in between its present position and the position where the prey gets trapped.

Searching the prey

The searching behavior of grey wolves is determined by the positions of A, B, and D agents. Even though these three wolves branch off while searching for prey, they all congregate while attacking the prey. The value of |G| defines mathematically whether wolves are converging (|G| < 1) or diverging (|G| > 1).

3.4.2 Enhanced grey wolf optimization algorithm

The intervals of exploration and exploitation are equal in the GWO algorithm. This might potentially lead to an inefficient exploration procedure, resulting in extended periods of time spent looking. Furthermore, the individual wolf inside the GWO algorithm adapts its location by computing the average of the places occupied by the alpha, beta, and delta wolves. Nevertheless, due to the hierarchical structure of grey wolf packs, with A, B, and D in descending order, using a simple average may not be the most suitable or effective method for updating the location of a particular wolf.

To successfully reduce these concerns, three potential enhancements have been suggested:

1) The function of the convergence factor changes from being linear to being nonlinear during this process.
2) A weighted average is used to adjust the geographical coordinates of an individual wolf.
3) The act of a wolf attacking its prey occurs when the prey reaches a certain proximity to the wolf.

Enhanced Convergence Factor

The convergence factor demonstrates an initial rapid reduction during the early iterations, thereafter transitioning to a progressive decrease as the iterations progress towards completion. The objective is to minimize redundant investigation and decrease the duration of search by modifying the convergence factors α to a non-linear function.

$$\alpha = i\left(1 - \frac{2}{Max}\right)^2 \quad (13)$$

Weighted Average

The hunting behaviour of grey wolves is influenced by the spatial distribution of individuals occupying different hierarchical positions within the group, namely alphas, betas, and deltas. The weights assigned to the variables A, B, and D in the equation are equal, without considering any variations in their relative significance. Therefore, the value of (eq. 12) is adjusted to (Equation. 14).

$$X_{K+1} = \frac{w_A.X_1}{w} + \frac{w_B.X_2}{w} + \frac{w_D.X_3}{w} \quad (14)$$

Where $w = w_A, w_B, w_D$, and $w_A > w_B > w_D$. Thus, $w_A/w = 3/6$, and $w_B/w = 2/6$, $w_D/w = 1/6$. However, users are permitted to use their own discretion, provided that the weights assigned to alpha exceed those assigned to beta, which in turn exceed the weights assigned to delta.

Attacking prey

Grey wolves have been seen to exhibit pouncing behavior during their hunting activities, particularly when targeting smaller prey such as birds, hares, and other similar species. The act of attacking suggests that a wolf has the ability to initiate an attack on its victim while it is in close proximity, without needing explicit instruction from higher-ranking wolves.

$$X_{K+1} = X_{best} \pm r, \qquad when\ A = c, \quad (15)$$

Let r be a randomly generated number within the range of [0, 5]. Furthermore, the constant c is chosen from the range [0, 1], and X_{best} represents the present optimal location, denoted as X_A.

Algorithm 1. Enhanced Grey Wolf Optimization

Step 1: Denote the initial wolf population as X_i, where i represents the index ranging from 1 to m.
Step 2: Assigning initial values to the variables a, G, M, and Max.
Step 3: While (t<Maximum number of iterations)
Step 4: Calculate the fitness value of every search agent.
The first, second, and third best search agents are representing as X_A, X_B, X_D
Until the value of k exceeds the maximum number of iterations.
Step 5: For each search agent
// Using equations 6, 7, and 8, modify the search agent's location.
// Equations 9, 10, and 11 used to describe hunting movement
Step 6: end for
Step 7: Update a, G and M
// Attack
If A<c then
Utilizing the mathematical expression of 12, engaging in predatory behaviour against a target.
Step 8: Else
Step 9: Determine the fitness of each search agent within the given context.
Step 10: Update X_A, X_B and X_D
Step 11: End if
Step 12: k = k+1
Step 13: end while
Step 14: return X_A

3.4.3 Integrated Deep Learning with Enhanced Grey Wolf Optimization

The characteristics that have been chosen are used to train the ANN classifier in order to diagnose lung cancer. The ANN is composed of three distinct layers, including the input layer, the hidden layer, and the output layer. The input layer of neurons receives the properties of lung cancer data, which are represented as $(x) = x$. The hidden layer of an ANN is often characterized by the use of the tan-sigmoid activation function.

$$f(x) = \frac{2}{1+e^{-2x}} - 1 \quad (16)$$

The weight values of each input are denoted as $w1$, $w2$, …, wn. The adder function is responsible for calculating the weighted sum of the inputs.

$$u = \sum_{i=1}^{n} w_i x_i \quad (17)$$

The output layer of an ANN may be characterized as follows:

$$y = f(\sum_{i=1}^{n} w_i x_i + b_i) \quad (18)$$

Equation 18 represents the relationship between the output neuron value (denoted as y), the transfer function (denoted as (x)), the weight values (denoted as wi), and the chosen characteristics of lung cancer data (denoted as xi). The determination of lung cancer diagnosis is established by evaluating the values of the output neurons.

$$\varepsilon_0 = \frac{w_i x_i + b}{\varepsilon_H} \quad (19)$$

Equation 19 defines ε_o as the error rate seen on the output nodes, whereas ε_H represents the error rate observed at the hidden layer nodes.

The error rate that is seen on the output nodes is denoted by ε_o in Equation 19, while the error rate that is observed on the hidden layer nodes is denoted by ε_H.

In the first experiment, the ANN is trained via the EGWO algorithm. The EGWO algorithm is used for the purpose of determining the most favourable values for the starting weight and biases. The subsequent stage entails the training of the neural network through the use of the back-propagation method. The weights and biases were derived using the process of EGWO. This proposal would boost the efficacy of the back-propagation algorithm in the pursuit of global optima in modeling. The weights and biases are evaluated as a vector of variables inside the proposed model. The Root Mean Square Error (RMSE) metric is used to judge how appropriate each vector is. It measures the difference between what was observed and expected outcome. Equation 20 introduces the notion of RMSE, which is computed by using the target output (T_i) and the anticipated value (P_i) derived from an ANN. A reduction in the RMSE signifies an improvement in the model's performance.

Algorithm 3. Proposed IDL-EGWO

Input: Lung cancer dataset

Output: Lung Cancer Prediction

Step 1: Start the process

Step 2: Training ANN using EGWO

Step 2.1: Set number of pack (population)

Step 2.2: Set total number of iterations for optimization

Step 2.3: Generate ANN model based on back-propagation algorithm

Step 2.4: Run EGWO to find the best values of weights and biases (vector)

Step 2.5: Return the initial optimal weights and biases (a).

Step 3: Training using backpropagation

Step 3.1: Use EGWO results as initial weights and biases

Step 3.2 Return the trained ANN model

Step 4: End the process

The performance of the ANN-EGWO algorithm in accurately following the original and anticipated output for both the training and testing datasets. Upon analysis, it becomes evident that the performance of a normal ANN is prone to becoming ensnared in local minima and exhibits sluggish convergence. ANN-EGWO exhibits quicker convergence and achieves optimum values more effectively compared to ordinary ANN. The improved grey wolf optimizer we propose improves the back-propagation algorithm by addressing the limitations of ordinary artificial neural networks.

The EGWO method incorporates an ANN model, which is inspired by the hunting behaviour of individual grey wolves, to formulate the movement plan. Subsequently, the EGWO movement strategy chooses the candidate from the EGWO based on the excellence of their newly acquired positions. The collaboration between these two search techniques enhances the overall and localised search capabilities of the algorithm being suggested.EGWO combined with ANN preserves variety, improves the relationship between local and global search techniques, and prevents being trapped in local optima. The results of several experiments and statistical tests show that IDL-EGWO works better than other algorithms when used on benchmark functions with different properties. The IDL-EGWO method has the capability to address engineering design difficulties and the optimum power flow problem. Furthermore, the suggested approach may be used to address extensive, unconstrained global optimisation issues. The suggested approach may be modified to address more real-world and extensive optimisation challenges.

4. EXPERIMENTAL RESULTS

In this experiment, use Python to test how well the proposed IDL-EGWO lung cancer prediction method works and compare it to the MLP, ANN, and GWO methods by looking at their accuracy, precision, recall, and F-measure. The lung cancer dataset used for experimental purposes is sourced from the Kaggle repository. The dataset consists of 309 instances, 15 characteristics, and one class attribute.

4.1 Performance Metrics:

Precision:

Precision is determined from correctly classified lung cancer patients to totally classified lung cancer patients. It measures the following formula:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (20)$$

Recall:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Recall measures correctly identified the lung cancer-affected patients with the actual number of patients.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (21)$$

F-Measure:

F-measures are a combination of precision and recall. It is given as:

$$F - \text{Measure} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} * \text{Recall}} \quad (22)$$

Accuracy:

Accuracy is the percentage of correctly identified lung cancer patients in the total number of lung cancer patients. The following formula is used for calculating the accuracy score:

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP} \quad (23)$$

Execution Time:

Execution time is calculated in milli seconds.

Root Mean Squared Error:

Root Mean Square Error (RMSE) is the standard deviation of the residual or prediction errors. It measures how far the prediction varies from the ground truth value.

$$RMSE = \frac{1}{P}\sum_{k=1}^{P}(f(x_k) - y_k)^2 \quad (24)$$

Here $f(x_k)$ is the predicted value, y_k is the actual value and P is the number of samples.

Table 2 presents the performance metrics of the proposed IDL-EGWO algorithm, with a comparison with existing algorithms such as MLP, ANN, and GWO.

Table 2:. Performance Measures.

Algorithms	Precision	Recall	F-Measure	Accuracy
MLP	84	82	83	85
ANN	86	85	85	86
GWO	90	89	90	91
IDL-EGWO	**94**	**95**	**95**	**97**

From the analysis in Table 2, the proposed IDL-EGWO algorithm performs better than existing algorithms. The proposed IDL-EGWO algorithm achieves 10% higher precision, 13% higher recall, and 12% higher f-measure than MLP. The proposed IDL-EGWO algorithm achieves 8% higher precision, 10% higher recall, and 10% higher f-measure than ANN. The proposed IDL-EGWO algorithm achieves 4% higher precision, 6% higher recall, and 5% higher f-measure than GWO.

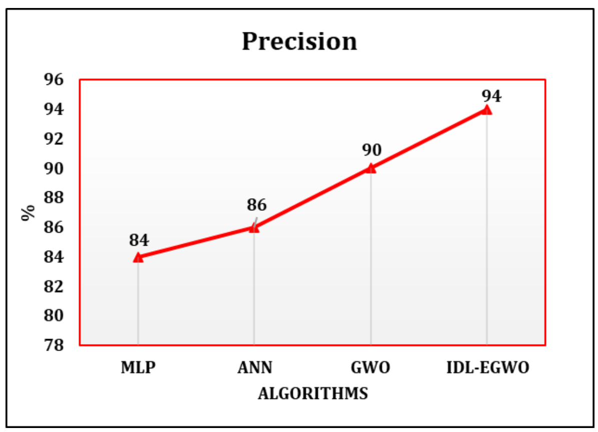

Figure 2 : Precision

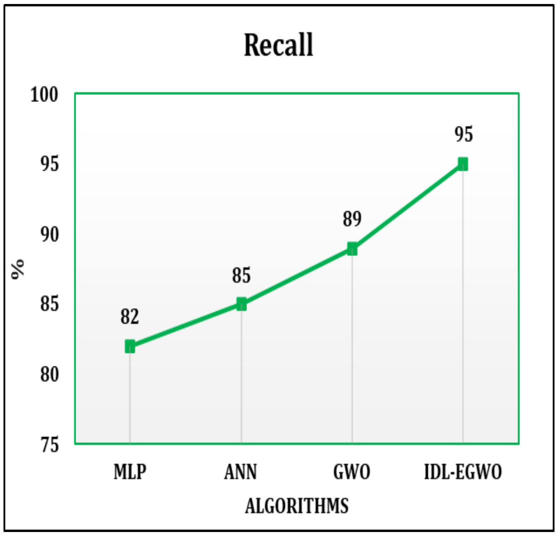

Figure 3 : Recall

Figure 2, 3, and 4 depicts the performance metrics respectively precision, recall, and f-measure pertaining to both the proposed and existing approaches. Figure 5 illustrates the accuracy rates of the algorithms used for the prediction of lung cancer.

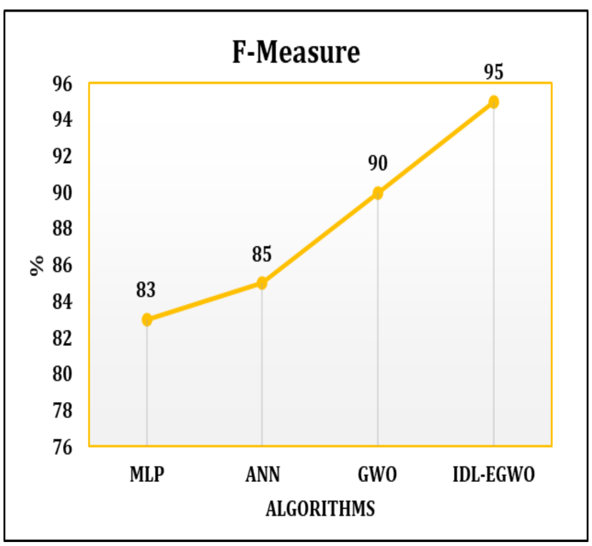

Figure 4: F-Measure

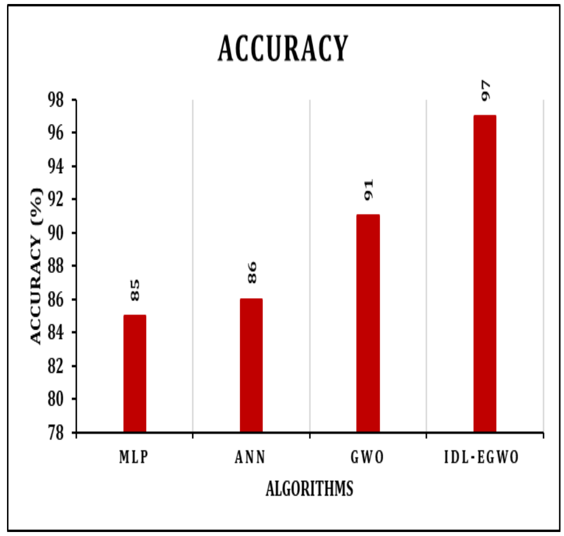

Figure 5 : Accuracy

Table 3 presents the observed durations for the execution of both the proposed algorithm and existing methods. The execution time is calculated in milli seconds. From the table, it is obvious that the proposed IDL-EGWO algorithm

Table 3: Execution Time.

Algorithms	**Execution Time (Milli seconds)**
MLP	920
ANN	910
GWO	840
IDL-EGWO	**760**

Here, the proposed algorithm executes with less time than existing algorithms. The proposed IDL-EGWO algorithm executed 160 milliseconds less than MLP, 150 milli seconds less than ANN, and 80 milli seconds less than GWO.

Figure 6: Execution Time

Figure 6 illustrates the amount of time needed to complete lung cancer prediction algorithms. The proposed IDL-EGWO algorithm performs with minimum time to predict the lung cancer.

Table 4 describes the root mean squared error for proposed and existing algorithms. Lower the values of RMSE represents the higher the performance of the working model. From the table 4, it is obvious that the proposed IDL-EGWO algorithm obtains minimum RMSE of 4% than other algorithms.

Table 4: Root Mean Squared Error for Proposed IDL-EGWO Algorithm.

Algorithms	RMSE (%)
MLP	16
ANN	15
GWO	11
IDL-EGWO	**4**

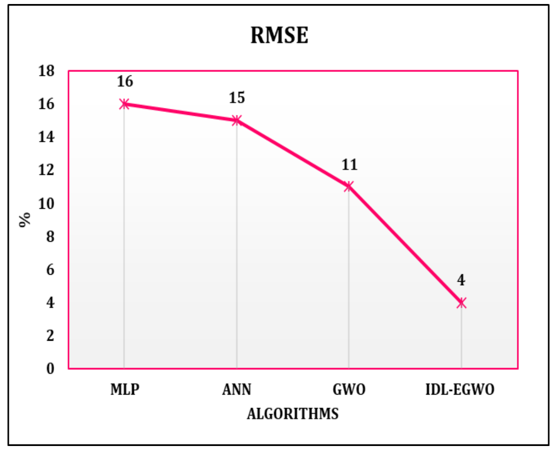

Figure 7: RMSE value for Proposed IDL-EGWO Algorithm

Figure 7 shows the RMSE value for the proposed IDL-EGWO algorithm. Table 3 describes how the proposed IDL-EGWO algorithm performs with a 4% RMSE value. The proposed IDL-EGWO algorithm performs a 12% minimum RMSE value compared to MLP, 11% minimum compared to ANN, and 7% minimum compared to the GWO algorithm.

Discussion

This section examines the primary factors that contribute to the excellence of the IDL-EGWO algorithm compared to other algorithms. The primary factor contributing to the effectiveness of the proposed method in both exploration and convergence is the acquisition of knowledge from neighbouring dimensions. Utilising this learning technique enables wolves to avoid local optima, leading to a thorough exploration of the search area. In addition, the neighbourhood structure used in IDL-EGWO is determined by a concept that facilitates both diversity and intensification during the optimisation process. Taking into account the distance, a greater distance corresponds to a higher variety of visiting wolves, meeting these criteria. In contrast, as the distance decreases, the number of neighbouring entities decreases. Furthermore, it preserves the variety necessary to address challenges in such intricate activities. The primary rationale is to use the advantages of both EGWO and ANN, since they complement each other in improving the equilibrium between exploration and exploitation as well as avoiding local optima.

5. CONCLUSION

This research introduces a novel approach that combines deep learning techniques with an upgraded version of the GWO algorithm for the purpose of predicting lung cancer. The outcomes of this research indicate that the use of IDL-EGWO makes it easier for ANN to perform its functions to identify the most favourable starting weights and biases. Consequently, it leads to accelerated convergence rates and a decrease in the RMSE. The investigation revealed that the IDL-EGWO method had a much higher accuracy rate of 97% and a lower RMSE value of 4% compared to existing techniques. However, we maintain that more investigation is necessary to explore the optimal arrangement of the neural network architecture, encompassing variables, selection of transfer functions, and choice of learning functions.

REFERENCES:

[1] Bhoj Raj Pandit, Abeer Alsadoon, P.W.C. Prasad, Sarmad Al Aloussi, Tarik A. Rashid, Omar Hisham Alsadoon, Oday D. Jerew, "Deep Learning Neural Network for Lung Cancer Classification: Enhanced Optimization Function", Multimedia Tools and Applications. DOI: https://doi.org/10.1007/s11042-022-13566-9,2022.

[2] Bouget D., Jorgensen A., Kiss G., Leira H. O., and Lango T, "Semantic Segmentation and Detection of Mediastinal Lymph Nodes and Anatomical Structures in CT Data for Lung Cancer Staging, International Journal of Computer Assisted Radiology and Surgery", 14(6), 977–986, 2019.

[3] F. Leena vinmalar and Dr.A.KumarKombaiya –Oxidation Communication-Book2,ISSN:0209-4541,Published Paper on "Lung Cancer Diagnosis Using Hybrid Dragonfly Optimization and Radial Basis Neural Network Classification"- June 2019.

[4] Haarburger C., Weitz P., Rippel O., and Merhof D., "Image-Based Survival Prediction for Lung Cancer Patients Using CNNS", In IEEE 16*th* International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, 1197–1201.

[5] Jiang H., Ma H., Qian W., Gao M., and Li Y., "An Automatic Detection System of Lung Nodule Based on Multigroup Patch-Based Deep Learning Network", IEEE Journal of Biomedical and Health Informatics, 22(4), 1227–1237,2017.

[6] Kannuswami P., Poobal S., and Ramya M, "Artificial Neural Network Based Lung Cancer Detection for PET/CT Images", Indian J Sci Technol, 10(42), 1–13,2018.

[7] Kim H., Goo J. M., Lee K. H., Kim Y. T., and Park C. M, "Preoperative CT-Based Deep Learning Model for Predicting Disease-Free Survival in Patients with Lung Adenocarcinomas", Radiology, 296(1), 216–224,2020.

[8] Lai Y. H., Chen W. N., Hsu T. C., Lin C., Tsao Y., and Wu S, "Overall Survival Prediction of Non-Small Cell Lung Cancer by Integrating Microarray and Clinical Data with Deep Learning, Scientific Reports", 10(1), 1–11,2020.

[9] Lu X., Nanehkaran Y. and Karimi Fard M., "A method for optimal detection of lung cancer based on deep learning optimized by marine predators' algorithm," Computational Intelligence and Neuroscience, doi: 10.1155/2021/3694723,2021.

[10] Mapanga W., Norris S. A., Chen W. C., Blanchard C., Graham A., Baldwin-Ragaven L., et al. "Consensus study on the health system and patient-related barriers for lung cancer management in South Africa," Plos one, 16(2), p.e0246716, 2021. doi: 10.1371/journal.pone.0246716.

[11] Maxwell O., Kayode A. A., Olusola Y. O., Joshua A. I., and Chinedu O. V, "Lung Cancer: A Chronic Disease Epidemiology; Prevalence Study", Asian Journal of Advanced Research and Reports, 1–7, 2019.

[12] Mohamed TIA, Oyelade ON, Ezugwu AE, "Automatic detection and classification of lung cancer CT scans based on deep learning and EBOLA optimization search algorithm", PLoS One, Aug 17;18(8): e0285796. doi: 10.1371/journal.pone.0285796. PMID: 37590282; PMCID: PMC10434933,2023.

[13] Palani D. and Venkatalakshmi K., "An IoT Based Predictive Modelling for Predicting Lung Cancer Using Fuzzy Cluster Based Segmentation and Classification", Journal of Medical Systems, 43(2), 1–12, 2019.

[14] Phillips I., Ajaz M., Ezhil V., Prakash V., Alobaidli S., McQuaid S. J., South C., Scuffham J., Nisbet A., and Evans P, "Clinical Applications of Textural Analysis in Non-Small Cell Lung Cancer", The British Journal of Radiology, 91(1081), 20170267,2018.

[15] Prabukumar M., Agilandeeswari L., and Ganesan K., "An Intelligent Lung Cancer Diagnosis System Using Cuckoo Search

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Optimization and Support Vector Machine Classifier", Journal of Ambient Intelligence and Humanized Computing, 10(1), 267–293, 2019..

[16] Qiang Li, Huiling Chen, Hui Huang, Xuehua Zhao, ZhenNao Cai, Changfei Tong, Wenbin Liu, Xin Tian, "An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis", Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 9512741, 15 pages, 2017. https://doi.org/10.1155/2017/9512741.

[17] Saien S., Moghaddam H. A., and Fathian M., "A Unified Methodology Based on Sparse Field Level Sets and Boosting Algorithms for False Positives Reduction in Lung Nodules Detection", International Journal of Computer Assisted Radiology and Surgery, 13(3), 397–409, 2018.

[18] Mohamed TIA, Oyelade ON, Ezugwu AE, "Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm", PLoS One. Aug 17;18(8):e0285796. doi: 10.1371/journal.pone.0285796. PMID: 37590282; PMCID: PMC10434933, 2023.

[19] Hussain Ali, Yossra, Varghese Sabu Chooralil, Karthikeyan Balasubramanian, Rajasekhar Reddy Manyam, Sekar Kidambi Raju, Ahmed T. Sadiq, and Alaa K. Farhan. "Optimization System Based on Convolutional Neural Network and Internet of Medical Things for Early Diagnosis of Lung Cancer", Bioengineering 10, no. 3: 320.https://doi.org/10.3390/bioengineering10030320, 2023.

[20] Priyadharshini P. and Zoraida B. S. E., "Bat-inspired metaheuristic convolutional neural network algorithms for CAD-based lung cancer prediction", Journal of Applied Science and Engineering 24, no. 1, pp.65–71, 2021.

[21] Lu X., Nanehkaran Y. and Karimi Fard M., "A method for optimal detection of lung cancer based on deep learning optimized by marine predators algorithm", Computational Intelligence and Neuroscience, 2021. doi: 10.1155/2021/3694723

[22] Oyelade O. N. and Ezugwu A. E., "A Bioinspired Neural Architecture Search Based Convolutional Neural Network for Breast Cancer Detection Using Histopathology Images", Scientific Reports, vol. 11, no. 1, pp. 1–28, 2021.

[23] Shan R. and Rezaei T., "Lung cancer diagnosis based on an ann optimized by improved teo algorithm", Computational Intelligence and Neuroscience, https://10.1155/2021/6078524\t"_blank" 10.1155/2021/6078524,2021.