Journal of Theoretical and Applied Information Technology
15th February 2023. Vol.101. No 3

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

# AN ADAPTIVE APPROACH TOWARDS PREDICTION AND DIAGNOSIS OF LUNG CANCER USING HEURISTIC GREY WOLF OPTIMIZATION APPROACH AND RANDOM FOREST CLASSIFIER–LUCAGO

**A.ANUPRIYA[1], ARUNKUMAR THANGAVELU[2 #]**

[1] Research scholar, Department of Computer Science, Bharathiyar University, Coimbatore,Tamilnadu,India.

[2]Professor,Department of Computer Science and Engineering,VIT University,Vellore,Tamilnadu,India.

E-mail:
[1]anupriya.mtech@gmail.com,[2]arunkumar.thangavelu@gmail.com
#Corresponding Author

**ABSTRACT**

Lung Cancer is considered to be a most disastrous health disease, where chances of recover is minimal as per World Health Organization. This research work supports on early prediction of lung cancer, which do help on chances of patients under risk to be diagnosed. Proposed work LUCAGO, focuses on understanding the chances of cancer attack and determining lifetime of patient whose prediction of lung cancer, which demands accuracy on prediction quality.LUCAGO frame work follows adaptive approach towards prediction of lung cancer and understanding the stochastic growth of cancer tumour cells using GWO for feature selection and random forest as classification approaches. LUCAGO approach adopts the knowledge of stochastic cancer symptoms as predictable patterns to suggest on prediction among detectable patient's using unstrained dataset.The performance evaluation of LUCAGO method demonstrates positive results, where GWO can be adopted effectively adopted by clinical research community oncologists to support on early identification of lung cancer. Work being verified over analytical metrics such as accuracy, error analysis and time taken for prediction shows that LUCAGO is far better in performance compared to GA, ACO approaches.
**Keywords** *: Lung Cancer, Early Prediction, GWO, Random Forest.*

## 1. INTRODUCTION

LUCAGO algorithm shows an analytical work of Grey Wolf animal's stocking of food nature and the results of the Artificial Bee Colony algorithm, Particle Swarm algorithm(PSO)[20] and Genetic Annealing (GA) is being compared. The Mean Absolute % Error (MAPE) [11] and accuracy is being measured which is the major metric for LUCAGO analysis. Mean Absolute Percent Error determines error size (%), which supports on large volume of data for determining the scale sensitivity. The percentage of error analyzed through MAPE. LUCAGO results shows a significant an improved accurate results in comparison to ACO, PSO algorithms.The major clue for lung cancer symptoms and suspiction is verifies from period of cough prolonged with stiffness in throat and history of chronic pulmonary diseases among close relatives. The laboratory test results suggest on improved diagnostic sensitivity for lung cancer. Prediction approaches depend on examining the level of gene diorders and patient's clinical history which consists of blood count rate, anti- CCP (Anti-cyclic citrullinatedprotein) [9], CRP (C-reactive protein), and related clinical data from tests [18].

This work adopts the following research objectives

1.To find significant predictor markers of Lung cancer from multiple cancer symptoms using classifier model for feature identification.

2.To propose on discrimination patterns between different subgroups of lung cancer as early prediction.

3.To evaluate the performance of the predominant lung cancer features using GWO over

normalised metrics to achieve optimal prediction.

LUCAGO involves the process of predicting an improved optimization model for predicting lung cancer which involved distinguishable stochastic patterns from datasets using an anomaly detection or outlier detection. LUCAGO approach follows regression based distance that uses a hybrid model of weight determination of lung cancer determinants or features with k-nearest Neighbor (kNN) [10] and Grey Wolf Optimization (GWO)[2] to predict on chances of lung cancer or a tumour or gene disorder prediction. Good analysis and measurement of cancer prediction / satisfaction rate which supports on net promoter score (NPS) [11] as major analytical understanding in any lung cancer prediction analysis is the primary motivation behind out this research work.

This research paper is arranged with section -1 focusing on introduction to analytical approach to predict lung cancer using Grey wolf optimization, while Section-2 supports on detailed survey and analysis of lung cancer and its prevalent challenges for early prediction design model over GWO and other related approaches. Section-3 focuses on LUCAGO model based design approach based on Grey Wolf Optimization and nearest neighbourhood prediction model which brings in possible optimization to early prediction. Detailed algorithm design and implementation is carried out in Section-4 while Section-5 focuses on experimental results, analysis and Section-6 concludes on outcome of work and future part of research work.

### 1.2 Gene Disorder Lung Cancer Prediction

Clinical research and study [4] [12] suggests that prediction of lung cancer primarily based on any genetic disorder which cloud be based on mutation of one gene (monogenic disorder), or based on multiple mutations of multiple genes (multifactorial inheritance disorder), or based on combination of mutations of multiple gene. Patient detected with lung cancer can also be due to environmental factors [19], or due to chromosomes damage[20] (changes in the number or structure of entire chromosomes),Clinical test and genetic test result confirm to support or rule out the chances of a genetic disorder suspect. Clinical condition help to determine patient's chance of developing or passing on a genetic disorder.

## 2.0 RELATED LITERATURE ANALYSIS

The researchers conduct several experiments over lung cancer prediction and analysis where combining two different imbalanced data classification techniques or adapting ensemble methods or by using meta-heuristic algorithms aims for resolve the unbalanced data classification method.we discussed some of the related works here. The related works discussed in this research work focus on lung cancer prediction approaches using data mining approaches over swarm intelligence approaches and computation models followed while image analysis approaches and video streaming analysis is the limitations.

This section surveys on research works and primary challenges focusing on computational models for achieving prediction accuracy optimization and related works on swarm intelligence. The [17] experiments of combining two techniques, Smote + Tomek and Smote + ENN are proven effective for data sets with a smaller number of minority instances. The ensemble based imbalanced data classification solutions used to generate different representative training sets through resampling the imbalanced data distribution and then adopt the ensemble techniques to solve the problem. [18] Easy ensemble, balance cascade, [19] Under bagging, [20] RUS Boost are the popular ensemble based under sampling algorithms and [21] SMOTE Boost, [22][SMOTE Bagging] SMOTE Bagging are the popular ensemble based oversampling algorithms. The ensembles [23] BEV (Bagging Ensemble Variation), Data Boost [3], [4] Asymmetric Bagging, [15] Quasi Bagging are the well-known ensemble based imbalance learning algorithms which are popular for their resampling methods. [6] HyperSMURF algorithm creates random forests by using under sampling and smote techniques. The predictions of the random forests are combined through ensemble of ensemble approach. [7] Balanced Boost creates a representative training dataset by sampling the training data instances based on the weights of AdaBoost.M2 algorithm and balances the data distribution.

Cluster based under sampling [8], [9] partition the data distribution into K clusters and randomly choose instances in majority class to balance the distribution. [3] Cluster OSS is based on the concept of One Sided Selection (OSS). It uses k-means cluster to subset and selects the greater class instances which are near to the cluster centroid to create representative dataset. After this under

sampling procedure, Tomek links is used. [1] Diversified sensitivity under sampling algorithm groups the majority instances through clustering technique and under samples the instances using the stochastic sensitivity measure (SM). The selected instances are trained using a radial basis function neural network. [13] The sampling techniques and k-modes clustering approach using simple sampling technique to build representative training dataset to solve imbalanced data classification problem.

The ACO sampling method[14 ], adopts the popular Ant based Colony (ACO) algorithm is used to derive the potential subset of majority classes. Evolutionary under sampling algorithm [14 ] uses CHC - an elitist genetic algorithm and boosting technique to under sample the imbalanced data. The fitness function emphasize on diversity of the instance dataset learned and performance. The CBEUS [6 ] is an under sampling method which partitions the majority class instances into clusters by using k-means clustering and calculates the Euclid ean distance of the majority class instances with its cluster centroid. The genetic algorithm optimizes the threshold Euclid ean distance for instance selection and connection weight of ANN. LSHGWOBRNN framework [5] uses LSH-SMOTE in the pre-processing stage. It uses LSH algorithm to arrange instances into buckets. GWO is used for searching buckets and training bidirectional recurrent neural network BRNN. GWO-weighted ELM [16 ] solves the types of problem with unbalanced data by optimizing the regularization parameter of Extreme Learning Machine (ELM). Bacterial Foraging Optimized Weighted ELM [17 ] is proposed to increase the classification rate.

This research work proposes a random forest approach for classification modelling and meta-heuristic based grey wolf optimization algorithm for optimized feature selection approach.. In this section, a brief introduction to Grey Wolf Optimization (GWO) algorithm to better illustrate the proposed oversampling algorithm.

### 2.1 Need for Research work and Motivation

The research work carries on early forecast of lung cancer node, point at which induces the chances of risk among patients who are under risk of being diagnosed of lung cancer is major research discussion and primary motivation to implement this work. The proposed work LUCAGO, focuses on understanding the chances of cancer attack and analysing the lifetime of patient whose prediction of lung cancer, which suggests on accuracy quality.Proposed model is expected to work on minimum computation time along with demand for early prediction and need for cancer prediction as accuracy metric,where GWO model is found to be advantageous than other models.

On analysis of GWO as need for improved or modified version for research gap suggests on its design which considers patient under risk as a valuable prediction parameter is used for analysis which depends on the major parameters related to detection activity.An early prediction of lung tumour among patients demand to precise the cost and convergence to improve the exact error detection towards patient's cancer detection rates depends on applying the GWO with random forest methods to be applied as per survey and analysis [13] is determined.

## 3.0 THE FUNCTIONAL ARCHITECTURE OF LUCAGO USING GWO & RANDOM FOREST APPROACH

The survey of patient's cancer tumour from clinical dataset implies on application of data analytics and related aspects of computational intelligence over data mining. Research in various domain of applied statistics[21] over clinical analysis attempts to predict cancer tumour among patients at Stage-3 or final growth of cancer over variable time period. Survey suggest on the challenges over detecting cancer among patients who are suffering from cancer based on their primary instincts or symptoms This gives an idea of which observes modelling approaches for improving optimization based on approaches are combined together. This approach suggests on formulating a new optimization algorithm, based on this applied domains, what ever bear on good study to find the absolute result within stipulated time. Few researchers [8][22] were find solution to the localization problem through mixed breed algorithm.

### 3.1 Design of LUCAGO Model

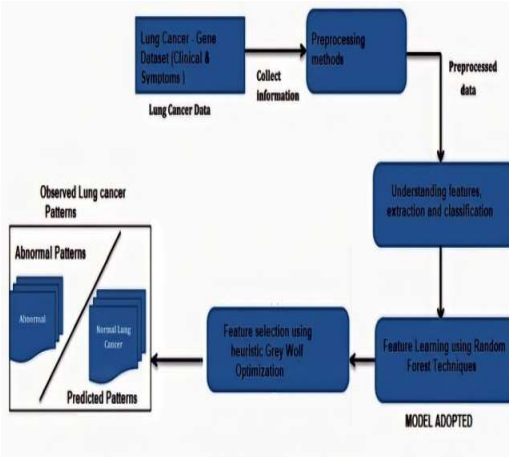


*Fig-1 : LUCAGO – Functional architecture*

LUCAGO frame work (Figure-1) supports on prediction of lung cancer and understanding the stochastic nature of cancer tumour cells using machine learning and random forest as classification approaches. Clinical dataset obtained from patient who are suffering from lung cancer predicted and patient symptom dataset. To predict on lung cancer patterns from dataset, modified heuristic Grey Wolf Optimization and random forest technique for feature selection approach is followed. The analysis adopts clinical dataset,which consists of patient's clinical parameters and cancer symptoms as shown in Table-1.

### 3.2 Proposed LUCAGO Functional Modules

Many research works concern on the predictor modelling approach towards cancer disease diagnosis and prediction.According to prediction accuracy metrics it is a very important that misleads in the direction of feature extraction and modelling. The proposed LUCAGO comprises Random Forest classifer with GWO to support on feature selection and classification methods. This approach adopts random forest works which adopts towards the advantage of selecting vital features by using wrapper-based selection approaches. This approach suggests on the variable support for determining local minima or global optima for differing data sets. Similarly, GWO finds an advantage for its nature of rejecting unwanted features but also get affected in its global searchability. The need for proposed approach is to defne on assignment of primary weight for supported features to suggest using random forest approach with GWO. The hybrid approach suggests on global search being combined with the local search capability of random forest and GWO. Though the hybrid model combines the best effects of swarm intelligence with classification model. Random forest exhibits the robustness to suggest on early classification over control parameters, which is computationally efficient with GWO. Survey finds that SVM approach has minimal adjustable parameters hence it takes longer time to converge or adaptiveness to predict on the convergence factor. Suvery suggests that hybrid model is always required to support as an efficient way of finding the global optima based on lighter computational effort.

The overall architecture of the proposed system is discussed in section 3.0 and its framework is shown in Figure-2. Considering the property of the non-cancerous cell data whose features are identified by combining GWO and Random Forest approach for faster convergence over large dataset. Here GWO approach is used as a classifier to differentiate the Non-cancerous cells and cancer symptoms verified. The detailed implementation steps of the proposed method are concisely given as follows:

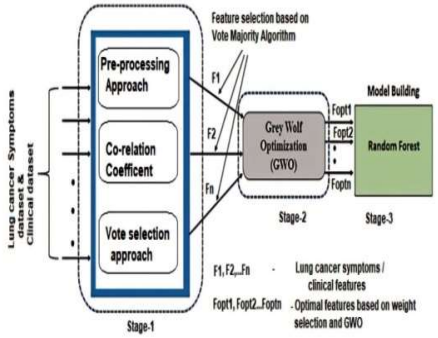


*Fig-2: LUCAGO Framework*

GWO approach adopts a cross study method for considering primary higher weightage parameters of the active lung cancer cells detected, This model helps to consolidate on the inclination plunge approach for improving the initial forerunner deteting parameters and using Least Square Estimator (LSE) for minimzing the chance of errors. In this research work, patient's detected cancer clinical datsset countsarelinked forsmoothning the startingmetrics of itsclassifier and the symptoms are being mapped.LUCAGO adopts the following steps to define on wolf population:

**Step 1:** Initialize on the position of Grey Wolf with its location parameters, including its deflection rate, angle of prey location, memory consideration rate, and prey hit (PHA) adjustment rate.

**Step 2:** Initialize and randomize the population of patients using equation 2 & 3 for cancer cell as target to be detected. The population is initialized with random solutions and fitness of the population.is evaluated appropriately

**Step 3:** Analyze on the target by arbitrarily organising and reorganizing the target arrangement vectors inside its vector space limits. The wolf and their arrangement defined over vectors settings to the detect the target parameters of the GWO.

## 4.0 GREY WOLF OPTIMIZATION

The research challenge on an logical methods of initial predictions of cancer tumour using GWO approach on irrelevant exceptions is supported by LUCAGO, which is designed as a easy and adaptive method. This approach adopts the knowledge of stochastic cancer symptoms to find patient's symptoms using trained dataset.

Ali Asghar et al., [2] defined and proposed GWO approach, whose design is primarily activated by grey wolves devour hunting behavior. The algorithm follows different behavioral aspects of wolves and hunting format of hierarchical leadership strategy surrounding the prey or target. Inside a limited space, the wolves live together and follow a cooperated hunting behaviour as a whole which defines on high-end carnivores in a group size of 10–15 wolves. Wolves are categories as alpha, beta, delta based on their hunting strategies [5] adopted where a wolf and its group of wolves involve together to acquire various hunting methods towards co-ordinator wolf, can take the further action to be done and directs supporter wolves, spies of victim wolves that helps in hunting. GWO hunting follows the below steps to achieve the goals:

(i). To identify the target prey using chasing and tracking model.

(ii). Circling the target prey continuously until the prey's mobility is stopped.

(iii). Attacking the prey from all directions.

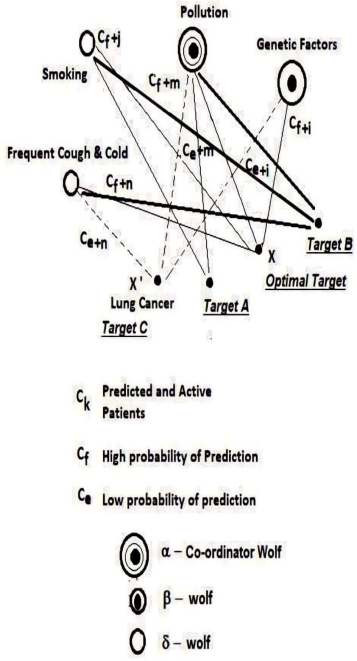


*Fig-3 : Lung cancer prediction analysis implemented using GWO*

Grey wolf optimizer analyze on patterns of lung cancer over iterative datasets [Refer-Fig:3] This approach also works on support to continue with corresponding values in cancer detection feature class labels of dataset. Instances of detection cancer cells on related samples of cell appearance and its neighbourhood cell determination indicates on its maximum relevance criterion or intensity. This approach decides on the intensity of cancer cells which use neighbourhood detection models such as K-Means or mutually away from other related tumour objects, to support on minimum redundancy.Deciding on determination of cancer cells and understanding notion of clinical trial outcome,GWO method recommends on lung cancer prediction based on related cell features being verified.

The procedure to choose the best target to feed on GWO approaches works basis on the phenomenon of encircling iteratively and hunting the prey by different utilization levels (a) to recognize the target prey location and encircle them, (b) to traverse all possible ways for the best final result in search vector space. The role of identifying the encircling the prey involves position vector of target as X, X' and Y, Y'.

To understand the notion of analysis of cancerous tumours, its clinical analysis or symptoms related to cancer, this work adopts Grey Wolf Optimization approach over random forest. This approach suggest on analysis of patient's possessing cancer over different time spans as a important parameter to be verified. Every patient is considered as a single person who may be possessive of a disastrous disease or identified to be active at any specific time period.

LUCAGO model discusses on three prediction processes being followed at individual level[10]:

(i). observed time for a patient being involved for test and analysis over a cancer tissue prediction

(ii). time taken for patient detected with cancer tumour tissue to suggest on clinical process after observation.

(iii). Time taken among group of patients whose presence on cancer tissue which indicate chance of tissue.

$$\vec{A} = |\vec{Y} \cdot \vec{Xp}(k) - \vec{X}(k)|$$
$$\vec{A}(k+1) = \vec{Xp}(k) - \vec{X} \cdot \vec{Y} \quad \text{-----(1)}$$

The equation represents, 'k' - current iteration of target prediction,

X and Y - coefficient distance vectors between the targets and search wolf

X(p) - Position vectors of target cells

A – the search wolf involved

(k+1) – defines the next optimal target new position within vector space which is computed by

Mean of all attack wolf positions.

Prediction of new target positions and its location of next near by target is an iterative operation where the wolves change their location with respect to their earlier nearby prey prediction, hence Equation (2) is as follows:

$$\vec{A_1} = \vec{A_\alpha} - (\vec{X} * \vec{m})$$
$$\vec{A_2} = \vec{A_\beta} - (\vec{Y} * \vec{n})$$
$$\vec{A_3} = \vec{A_\rho} - (\vec{Z} * \vec{i}) \quad \text{-------(2)}$$

The initial position of wolves are considered as$A_1$, $A_2$, and $A_3$ recognized for the grey wolves α, β, and ρ. The new relocated distance as calculated between wolf and prey target X, Y and Z indicate on co-efficient vectors considered as m, n, and i.

Total number of active wolves involved in search and knowledge of target prediction is defined as N, and maximum number of target location available for prediction defined as M. The objective of predicting on number of targets involved for search, F(x) as shown in Equation (3).

$$F(x) = \sum_{i=1}^{M} \left( \vec{A}_i . \frac{\sum_{j=1}^{N} \vec{X} j * \vec{d}}{M} \right) \text{------(3)}$$

The target or prey location is determined by co-ordinator grey wolves which is always on state to attack. The performance of attack is suggested by wolves encircling the optimal prey based on earlier predicting solution (training set of data) within the search space ( discussed in Equation (1) and (2)). Determining the location of prey is suggested by position vector [Fig-3] based on group of grey wolves prediction approach and attacking the targets at an instance of time 't'. . The flowchart [Fig-4] elaborates the steps of surrounding and attacking the prey. At time 't+1', group of other wolves 'Wi', change or adapt their position based on vector space movement for prey prediction whose next step to get best solution for further analysis.

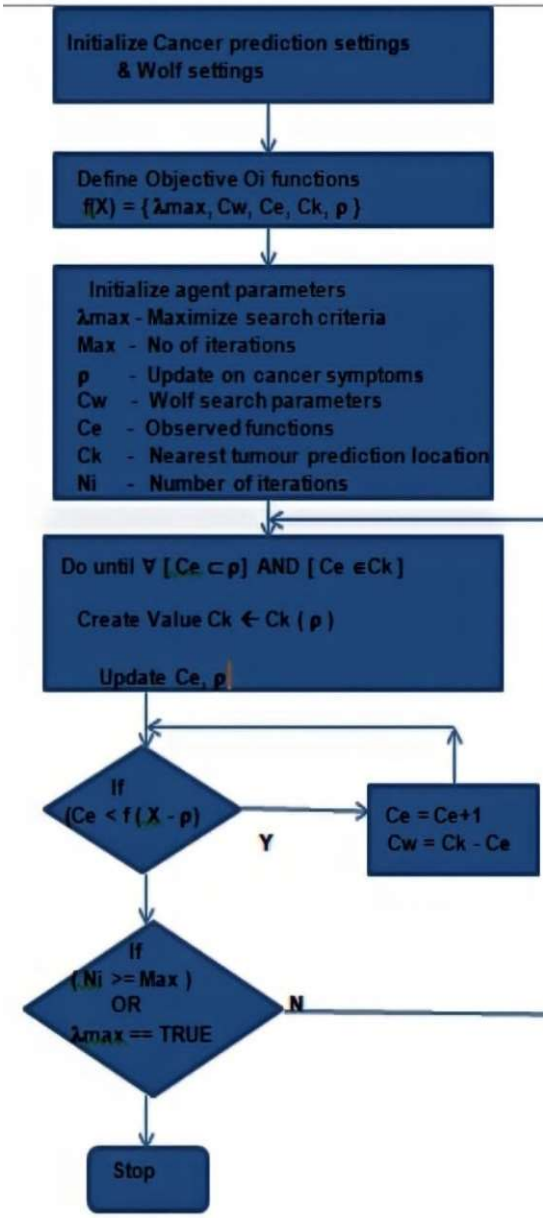


*Fig-4 : Flowchart Indicating GWO Over Lung Cancer Prediction*

The main aim is to find best food hunting & prey as target limit using LUCAGO algorithm. The objective function (F(x)) is discussed in Section 4.Fig-4represents the flowchart of working of LUCAGO approach over Grey Wolf Optimization with random forest approach for the target neighbourhood selection approach. Adopting grey wolf algorithm demands faster execution time as an iteration of $\alpha_{max}$, defined as variable limit which defines the maximum expected outcome of accuracy of cancerous cells analyzed. LUCAGO updates on lung cancer symptoms and its location, which adopts vital role as feature analysed towards cancer tumour prediction.

### 4.1 LUCAGO Algorithmic Design

A detailed analysis and focus on design of Grey Wolf Optimization approach for prediction of lung cancer is discussed in this section. Food hunting behaviour of Grey Wolf as a team for its prey is the primary interest of this work and hence food foraging behaviour as a group work is major concern for analysis. Predicted the lung tumour is depicted as food for grey wolf and the process of hunting is depicted as predicting the related symptoms and cancer tumour as an object.

The objective statement is defined as Fx (x=1,2...m) *// to define fitness of cancer tissue prediction, which is criteria for search*

**1:** Wolf population initialized as Wi (i=1,2,...n), n being set of decision variables where pulse frequency is Cp at Wi, Initialize pulse rate Ri, such that Cp is dependent on Ri

**2:***// set of observed variables*

Cf : Clinical values reference to cancer

Ce : Observed cancer symptoms

Ck : next nearest cancer tumour

Cw : worst vector observed, based on wolf based behaviour chances of missing a cancer tumour

Gk : cancer prediction and time taken for prediction relation matrix ($n_{row}$ x $n_{col}$) whose weight is an eigen vector

**3:** Generate Cf;

**4:**Forx = 1...$n_{row}$ do

for y = 1 …$m_{col}$ do

Cf →Gk[x][y]

Initialize Fx as (Ck, Cf)

for each Cf value

Generate Ce, where ∀ Ce⊂Cf

Generate Cw, where ∀Cw⊂Cf

endfor

**5:**

forx = 1...$n_{row}$ do

for y= 1 …$m_{col}$ do

if (Cw != NULL ) OR Cp (Ck, Cf) <Gk[x][y]) then *// continue this process until all active wolves are checked*

for all Cx do

*// Check on the fitness of wolf for observed cancer tumour*

Wi'∈ ( $W_{i1}$, $W_{i2}$, …. $W_{in}$ }

Gk[x][y] ∈Wi'

endfor

Cw ←Cf- Ce *// an update on unexpected symptom observed*

Ce = Ce +1

endfor

**6:**Ck = Cf +1 *// Update wolf behaviour in range with nearest pattern detected*

Update Cw, Ce, Cf

The process of detecting various abnormal cancer symptoms with reference to distinguishable patterns in lung cancer clinical datasets can be considered as anomaly detection or outlier detection which is the major discussion of this research.

## 5.0 RESULTS AND ANALYSIS

### a)Dataset Adopted :

Dataset used for analysis of proposed approach LUCAGO, considers all possible observed symptoms of cancer for analysis. Dataset considered for analysis engages the deterministic features of cancer for more than 2 years of verification process, observed consistently from the data collected from a clinical institute. The records of both in-patient and out-patient observed as dataset collected over period of Jan-2020 to Nov-2021, manages more than 12730 records for 117 different related cancer symptoms . The variables used for analysis over limited prediction time as a differential parameter as shown in Table-1. GWO's control parameter values can be set and modified based on execution of Ck recommendations.

*Table -1 : Dataset property*

| | |
|---|---|
| Density of attribute | 211 |
| Recordset | 12730 |
| % of Missing Values | 23% |
| Numerical Attribute size | 9436 |
| Categorical Attributes | 259 |

Analysis observes that more than 39% of lung cancers are diagnosed based on image report analysis and using tumor localization which suggests on the location of tumour size and site observed. This indicates that the cancer site which suggests on tumour where it started, which commonly referred to as stage I or initial stage of cancer. Survey only suggests on the challenge where patients with lung cancer whose bodily symptoms are determined only at final stages of the disease, this stage indicates on suspicion of lung cancer but discovered where the cancer has grows and spread more than 70% over the suspected patient .

To determine classification accuracy for cancer forecast, total recall and F1-score analysis for correct number of predictions is detected from the total forecast. The observed values are determined from variables which is conditional based on observed experiment outcome. The model performance is being assessed using range of cancer datasets which verifies on assess of cancer prediction accuracy. Different cancer datasets were considered such that available number of features are controlled for analysis. LUCAGO adopts random forest classifier in conjunction with GWO heuristic bio-inspired algorithm to result on enhanced classification accuracy.

*Table-2 : Analytical Classification And Performance Metrics of LUCAGO With Other Approaches*

| Method | Balance Accuracy | F1 Score | Sensitivity | Specificity | MCC |
|---|---|---|---|---|---|
| GWO-SVM | 0.8757 | 0.9622 | 0.9494 | 0.782 | 0.6801 |
| GA | 0.8390 | 0.8854 | 0.9321 | 0.7458 | 0.605 |
| ACO | 0.8659 | 0.8626 | 0.9352 | 0.7665 | 0.6621 |
| LUCAGO | 0.9262 | 0.9801 | 0.9846 | 0.8554 | 0.8102 |

such as error analysis, F1-score, accuracy, and specificity which assess on the overall effectiveness of LUCAGO hybrid heuristic classifier model.Three different lung cancer datasets [ ] are adopted as trained model as well tested and analysed for this research work . Results show that LUCAGO, the heuristic-based classification approach suggests on a reasonably consistent classification result.

### 5.1 Performance Analysis

The performance analysis of LUCAGO over computational models such as GA and ACO is discussed.The observed F1-score and related accuracy metrics for an early prediction rate of cancer tumours along with balanced accuracy rate, observed error rate and time complexity involved in prediction is studied for variable datasets.The study, also adopts the analytical approach of F1-score, precision, and recall measures on classification performance model for lung cancer as disease classes.

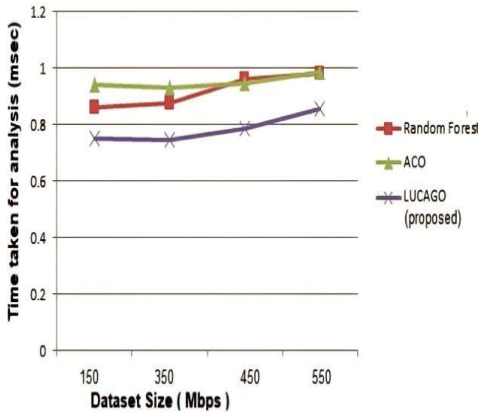


*Fig-5:Observed accuracy based on analytical metrics*

Observed cancer observance accuracy rate using LUCAGO is shown in Fig-5, which involves supporting prediction as life time of a patient as major parameter for analysis. The percentage of observed prediction rate from LUCAGO as forecast analysis shows an average variable rate of 74.81% of prediction rate in comparison to GA whose average rate is 67.29% of high inconsistent rate compared over ACO which demonstrates average rate of 62.73% as analysed.

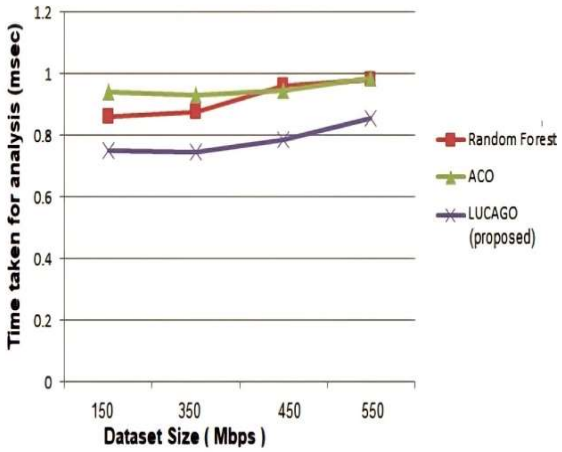


*Fig-6: Observed Time taken for analysis for variable datasets*

Error analysis on different models is discussed in Fig-7, which shows that LUCAGO shows an average of 0.437 % of error in comparison with ACO and SA approaches which demonstrates as 0.0583 % of error for ACO and 0.0687% for GA approach.

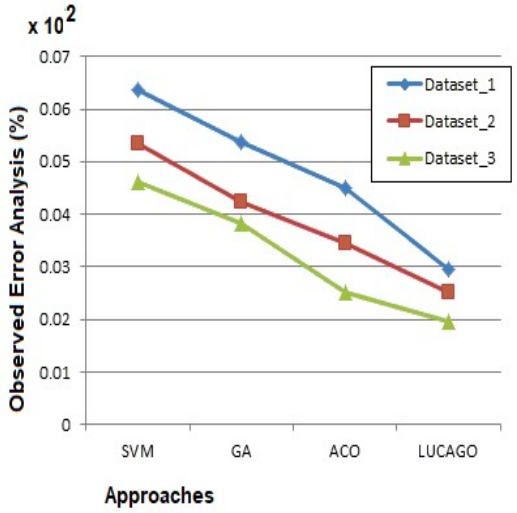


*Fig-7 : Observed error analysis for LUCAGO*

The proposed LUCAGO work shows an improved outcome whose observed accuracy is 74.81 % compared to analytical models such as ACO and GA. LUCAGO also shows an better performance in terms of error analysis and observed time for differing datasets.

Performance of LUCAGO is differs over variable large datasets and streaming analysis. Machine learning models need to be incorporated to as future research work which is major limitation.

## 6.0 CONCLUSION & FUTURE WORK

Predicting Lung cancer is considered as a major and challenging medical challenges to discuss in this research based on the gene disorder complexity of cancer cells or tumours. Medical survey and analysis suggests that there are types of cancer to be concerned related to lung cancer. If treatment for lung cancer is delayed, there is a significant increase in the risk of death. Detection of lung cancer at an early stage and treating at an early stage and curing will minimize risks.

LUCAGO works on Random Forest as classifier model and modified Grey Wolf Optimization to identify on features related to lung cancer , GWO approach supports on prediction towards the patterns for development of lung cancer. Though LUCAGO supports on the objective of this system is to predict on patient's cancer detection using tumour and detection of gene disorder as an early warning, which suggests on higher accuracy and minimal error rate with minimal time for prediction compared to other heuristic approaches. The performance evaluation of LUCAGO method produces positive results, which demonstrates that GWO can be adopted effectively by clinical research community oncologists to support on early identification of lung cancer.

Based on the research probems discussed in Section-1, the objectives are achieved . (a).The classifier model for feature identification using significant predictor markers of Lung cancer based on cancer symptoms is designed and analysed.

(b).An earlier lung cancer prediction approach based on cancer clinical discrimination patterns is analysed.

(c).The performance of predominant lung cancer features is evaluated using GWO over random forest and its metrics are analysed.

Research community can extend this work towards focusing on deep learning approaches for multi-objective problem and bringing in adaptive prediction of cancer among differing type and age group of people.

## REFERENCES

[1.] Abdullah D. M., Abdulazeez A. M., Sallow A. B. Lung cancer prediction and classification based on correlation selection method using machine learning techniques. Qubahan Academic Journal, 1(2):141–149, 2021.

[2.] Ali Asghar, Heidari Parham, Pahlavani, An efficient modified grey wolf optimizer with Lévy flight for optimization tasks, Applied Soft ComputingVolume 60, November 2017, Pages 115-134.

[3.] Ashwini S. S., Kurian M. Z., Nagaraja M. Emerging Research in Computing, Information, Communication and Applications . Singapore: Springer; Lung cancer detection and prediction using customized selective segmentation technique with SVM classifier; pp. 37–44, 2022.

[4.] Brocken P., Kiers B. A., Looijen-Salamon M. G. Timeliness of lung cancer diagnosis and treatment in a rapid outpatient diagnostic program with combined 18FDG-PET and contrast enhanced CT scanning. Lung Cancer . 2012;75(3):336–341.

[5.] Chen, R.-L., Zhao, J., Zhang, X.-C., Lou, N.-N., Chen, H.-J., Yang, X., Su, J. Xie, Z, Zhou, Q, Tu, H.Y.,Crizotinib in advanced non-small-cell lung cancer with concomitant ALK rearrangement and c-Metoverexpression. BMC Cancer 2018, 18, 1171.

[6.] D'Cruz J., Jadhav A., Dighe A., Chavan V., Chaudhari J. Detection of lung cancer using backpropagation neural networks and genetic algorithm. Computing Technologies and Applications . 2016;6:823–827.

[7.] DinaElreedyAmir F.Atiya,A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance, Information Sciences, Volume 505, December 2019, Pages 32-64.

[8.] Fakih, M., O'Neil, B., Price, T.J., Falchook, G.S., Desai, J., Kuo, J., Govindan, R.; Rasmussen, E.,Ngang, J, Phase 1 study evaluating the safety, tolerability, pharmacokinetics (PK), and efficacy of AMG 510, A novel small molecule KRASG12C inhibitor, in advanced solid tumors. J. Clinical. Oncology,37, 3003, 2019.

[9.]Gainor, J.F., Lee, D.H., Curigliano, G., Doebele, R.C., Kim, D.-W., Baik, Clinical activity and tolerability of BLU-667, A highly potent and selective RET inhibitor in patients (pts) with advanced RET-fusion+ non-small cell lung cancer (NSCLC), J. Clin. Oncol. 37, 9008, 2019.

[10.] Hongyi Zhou and Jeffrey Skolnick, A knowledge-based approach for predicting gene–disease associations, Bioinformatics, 32(18), 2831–2838, 2016.

[11.] A. G. Hussien, Crow Search Algorithm: Theory, Recent Advances, and Applications, IEEE Access, vol. 8, pp. 173548-173565, 2020

[12.] ES Krishna, Arunkumar Thangavelu, Attack detection in IoT devices using hybrid metaheuristic lion optimization algorithm and firefly optimization algorithm, International Journal of System Assurance Engineering and Management, 1-14, 2021.

[13.] Landi, L., Chiari, R., Tiseo, M., D'Incà, F., Dazzi, C. Chella, A., Delmonte, A., Bonanno, L., Crizotinib, MET-Deregulated or ROS1-Rearranged Pretreated Non–Small Cell LungCancer (METROS): A Phase II, Prospective, Multicenter, Two-Arms Trial. Clin. Cancer Res. 25, 7312–7319, 2019.

[14.] Maclin, D. Opitz, Popular Ensemble Methods: An Empirical Study, Journal Of Artificial Intelligence Research, Vols. 11:169-198, 2011.

[15.] Miller H. A., Yin X., Smith S. A., Evaluation of disease staging and chemotherapeutic response in non-small cell lung cancer from patient tumor-derived metabolomic data. Lung Cancer . 2021;156:20–30.

[16.] Y. Qiu, S. Zhang, X. Zhang and L. Chen, Detecting disease associated modules and prioritizing active genes based on high throughput data, BMC Bioinformatics, vol. 11:26, 2010.

[17.] Ravi Salgia, Isa Mambetsariev, Tingting Tan, Amanda Schwer, Dary P. Pearlstein, Complex Oncological Decision-Making UtilizingFast-and-Frugal Trees in a Community Setting—Role of Academic and Hybrid Modeling, Journal of Clinical Medicine, 9, 1884, 2020.

[18.] Shantakumar B.Patil, Y.S.Kumaraswamy, Intelligent and Effective Heart Attack Prediction System Using Data Mining and ArtificialNeural Network, European Journal of Scientific Research , Vol.31 No.4, pp.642-656, 2009.

[19.] Seyed Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, Grey Wolf Optimizer, Advances in Engineering Software, Volume 69, Pages 46-6, March 2014.

[20.] Vijayan Sugumaran, Arun Kumar Sangaiah, Arunkumar Thangavelu, Computational Intelligence Applications in Business Intelligence and Big Data Analytics, CRC Press, June 2017

[21.] X. Wang, Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis, IEEE Transactions on Cybernetics, vol. 50, no. 9, pp. 3950-3962, Sept. 2020

[22.] Yongqian Qiang, Youmin Guo, Xue Li, Qiuping Wang, Hao Chen, Duwu Cuic,.The Diagnostic Rules of Peripheral Lung cancer Preliminary study based on Data Mining Technique. Journal of Nanjing Medical University, 21(3):190-195, 2007

[23.] Zhou,H. Comprehensive prediction of drug-protein interactions and side effects for the human proteome. Scientific Rep., 5, 11090, 2015