International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Improving the Performance of Lung Cancer Detection at Earlier Stage and Prediction of Reoccurrence using the Neural Networks and Ant Lion Optimizer

S. Senthil, B. Ayshwarya, Shubha

Abstract---Lung cancer is considered to be the one among the most dreaded disease which will be the main reason for the death of individuals and having greater deterioration of death if it is not identified at primitive stage. Because of the fact that Lung cancer could be identified only after spreading to the parts of lungs to a greater extent and it is very tough to predict the presence of lung cancer at the earlier stage. Moreover, it involves greater error in the diagnosing the presence of Lung cancer by Radiologists and Expert Doctors. Therefor it is compulsory to design an intelligent and automated system for accurately predicting the cancer and stage at which the stage of cancer or enhancing the accuracy of prediction for detecting the cancer at earlier which will be much helpful in deciding the treatment type and depth of the treatment based on the extent of disease. Currently application of ANN strategies are the influential ways in supporting expert doctor for examining, complicated medical increase across a wider category of medical application. Back Propagation Network are ideal in recognizing lung cancer and there is no requirement involvement by expert doctors. Maximum number of applications of BPN in medical diagnosis will be utilized in the applications related to decision making of the presence or absence of disease; by which the performance will be reliant over the considered features and allocating the patient with minimum number of classes. Here this research paper establishes the idea of using BPN in the classification of the lung cancer and its stages and the predicting the possibility of recurrence. Along with the BPN, a nature inspired Meta Heuristics that is termed as Ant Lion Optimization Algorithm is used in optimizing the parameters and weights of Back Propagation Network. By using the Ant Lion Optimization Algorithm, the convergence mechanism is improved along with improving the accuracy of the proposed technique and it avoids the chance of getting caught within the clutches of local minima. By using this proposed method BPN network optimized with the help of antlion optimizer more accurate prediction of lung cancer is possible even at primitive stage and the predicting of chance of reoccurrence even after undergoing the appropriate treatment.

Keywords---BPN, ALO, Lung Cancer, Classification, Reoccurrence

Revised Manuscript Received on 30 July 2019.

***** Correspondence Author

Dr. S. Senthil,* is Professor and Director, School of Computer Science and Applications, REVA University, India.

Ayshwarya B, currently pursuing Ph.D. from REVA University, Bangalore

Dr. Shubha.A, Director and Professor-School of Commerce and Management Studies,

I. INTRODUCTION

UNRESTRAINED cell development in tissues of the lung is the cause of. Lung cancer, termed to be lung carcinoma, is a malevolent lung cancer.

The development might be distributed outside the lung with the procedure named as metastasis to adjacent tissues and additional portions in the body [1]. Division of Lung Cancer is as follows: small cell lung cancers (SCLC) and Non-Small Cell lung cancers (NSCLC). Reliant with the equipmental observation of tumor cells by means of Microscope, these segregations are performed. These categories of cancers develop, dispersed, and handled by diverse methods. Therefore, differentiation between these categories and stages of lung cancer status takes much significance in the diagnosing procedure.

Around 10%-15% of lung cancers will belong to the category of SCLC. Cancer in Lung parts will be more destructive in addition with quickly developing is SCLC and it contains 10% - 15 % proportion of lung cancers. The main cause of SCLC is because of deposition of toxins and Nicotine content in the lungs because of smoking of cigarettes. SCLCs develops swiftly in several parts of the body and can be identified only after spreading to the considerable extent.[2]. Stages of SCLC Category are presented as follows Limited stage: Cancer identified in either side of the chest, comprising only portion of the lung and adjacent lymph nodes. Extensive stage: Cancer distributed to various portions of the chest and the body.

Out of lung cancer category, that will be found in infected patients NSCLC is the most usual and takes about 85 % of portion in the entire cases of lung cancers. Depending on type of cells and their size that will find its place in tumor NSCLC is segregated into three types. They are:1. Adenocarcinomas 2. bronchioloalveolar carcinoma 3. Squamous cell carcinomas 4. Large Cell Carcinomas Majority of the adenocarcinomas will be identified in external, or marginal, regions in lungs. Sometimes they will prevail in lymph nodes and away from lymph nodes. A sub division of Adenocarcinoma that will be called as bronchioloalveolar carcinoma) which has the tendency of rapid progression in several portions of the lungs. This category of lung cancer appears to be like pneumonia while screening with chest X-ray.

Persons getting infected by this category incline to consider an improved prediction on comparison with other lung cancer categories. Squamous cell cancers that will develop in central portion of chest near the bronchi can be called as epidermoid carcinomas. The Squamous Cell Lung Cancer will be found within the lung areas and dispersed to lymph nodes and has the tendency of developing bigger in size creating a hole. Large cell cancers occasionally mentioned as undistinguishable cancers, that belongs to the rarest category under NSCLC. Large Cell Category will be identified in lymph nodes and other farther places from lungs.

Stages of Non-Small Cell Lung Cancer is listed as Stage I: The cancer identified in the lung areas only Stage II: The cancer is identified in lung spread to adjacent lymph nodes. Stage III: Cancer spread to chest are apart from lungs and lymph nodes. This considered to be the advanced stage. Stage IV: This stage considered to be fully developed stage This condition prevails when the cancer reached to both lungs, and the cause for the collection of fluid inside the lungs, or spread to liver or other organs.[3]

Lung cancer takes the highest reason for death around the world due to cancer dominating the added proportion of the subsequence reasons of mortality because of cancer: colon, breast and prostate. During diagnosis almost 40% patients can be identified in the advanced stage. As a result of this the chance of survival with 5-year lies in the range 6% to 15% Different from the above-mentioned fact the chance of 5-year survival with the detection of lung cancer at stage I will be estimated as 70%. Therefore, it is recommended that if the disease is identified at an earlier stage, then chance for the survival is high the outcome in the life duration of persons infected by the disease might be immensely enhanced. Presently Spiral computed tomographic (CT) imaging modality will be perceived as the best screening procedure available in the detection of lung cancer at an early stage. It is uncertain that during the current circumstances by the utilization of CT screening, that the deaths because of lung cancer might be diminished by CT screening. Persons with greater danger of getting lung cancer are customary selections to detect the presence of lung cancer at an early stage such as CT screening. Consequently, there will be development in focus in the construction of techniques of specific danger forecast for lung cancer.[4]

The available and feasible remedial handling in primitive stage non-small cell lung cancer (NSCLC) patients, will be the comprehensive removal of the affected parts might be accomplished within hours and availability of growth of new tissues is contrived to be feasible. These opinions will be benefits of surgery which is not linked with additional procedures such as chemotherapy and radiotherapy. Nevertheless, there will be several circumstances which falls back in achieving the complete remedy after surgery. Around the range in between 30% to 55% of patients having NSCLC is having the chance of recurrence and pass away because of the disease in spite of undergoing remedial surgery. Consequently, several persons with lung cancer sooner or later expire because of recurrence subsequent to surgery. Additionally, surgery itself holds a particular quantity of danger. Nevertheless treatment-associated mortality is comparatively rare occurrence, but the existence of noteworthy danger in deaths, in addition with maximum number of persons will encounter a reduction in performance subsequent to surgery. Consequently, in spite of the developments in surgical process, curing with the help of surgical removal of affected portions will not be the perfect choice, while comprehensive removal of affected portion is feasible.[5].

Keeping the above-mentioned risks about the lung cancer in mind, the need of the hour is to develop an automated and intelligent prediction of the presence of lung cancer at an early stage with accurate detection as the primary motive and faster diagnostic procedure. Apart from the detection of presence of lung cancer.it is compulsory to predict the chance for recurrence once it was identified in an early stage and subsequent to treatment of the lung cancer. The motivation of the proposed work is the construction of the system that performs the computer-supported involuntary identification which will make Automated techniques to take part as routine in medicinal decision and arriving at the conclusion's procedures portion of radiologists' in addition with empowering the detection of disease at primitive phase which might permit the patient prognosis in an effective way. Along with the earlier detection of the presence of lung cancer, this paper highlights the importance of prediction of reoccurrence of the disease after detecting the disease and subsequent to the treatment that is provided by specialist doctors either by providing radiation therapy or through surgical procedure.

The remaining paper is structured as follows; Section II introduces the Survey of various researches done by the various scholars in the area of Automatic prediction of lung cancer using classification Section III explains the proposed strategy, Evaluation Results of the proposed technique are described in Section IV. Finally, Section V is the Conclusion of the paper

II. LITERATURE SURVEY

Cancer takes the significant portion of the reasons for death of men and women because of health ailments. In assigning the proper treatment to the patients, identification of disease at primitive phase is the most compulsory requirement. Therefore, the demand for strategies in identification of the existence of cancer nodule in primary stage is growing. Error in finding the disease is the most worrying condition in the medical world. Lung cancer is the usually misidentified disease. Prior investigation of Lung Cancer protects massive quantity of human lives, not succeeding in the earlier diagnosis might result in critical issues triggering unexpected deadly termination of life. The rate of remedy is purely reliant over the primary identification and detection of the lung cancer. Identification of Expertise and data mining techniques are positioned itself in abundant applications in commercial and technical area. Fruitful information might be identified by applying the data mining and Artificial Intelligence Strategies in medical management framework.

Medical Management industry gathers vast quantities of medical information that, inopportunely, are not utilized to identify concealed knowledge. Identifications of concealed information and their associations frequently gets unnoticed. Finding of Lung Cancer Ailment might response complicated "what if" questions that conservative decision sustenance frameworks cannot.

Usual cancer category that affects the persons is lung cancer. Identification of lung cancer at primitive stage is the prominent requirement for selecting the method of treatment method and increasing the chance for survival. The Analysis by Dandıl, (2018) [6], projected unique strategy of Automated channel of Computed Tomography (CT) examinations with the purpose of identification of Lung Cancer at primitive stage by categorizations benign and malignant nodes. The projected channel containing four phases are improved for clarity in visual appearance is performed by extracting the lung volumes by a Unique Technique termed as LUVEM. By using Circular Hough Transform identification of affected nodes is performed. Using Self Organizing Maps, Lung nodules are segregated. For extraction of features, processing intensity, shape texture and combined features is performed and the extracted features are reduced by avoiding the redundant and irrelevant features using Principal Component Analysis. For the differentiation of benign and malign nodes classification is performed by Probabilistic Neural Network (PNN).

One of the most significant and critical feature of medical image processing lies in the technique of Efficient detection of presence lung cancer at a primitive phase itself. Numerous information mining techniques was utilized in detection of lung cancer at primitive phase. The research wok by Naresh and Shettar (2014) [7] recommends, the technique by utilizing CT Screening images that is in the format of DICOM helps in the identification of presence of Lung Cancer was established which will diagnose lung cancer during preliminary stage itself utilizing from the images taken patients. CT scan images that belongs to the format of DICOM (DCM). White Gaussian noise removal that is affecting the CT image that will be used in screening, is performed with the help of non-local mean filter in addition with segregating the lung using thresholding technique takes the significant consideration. To generate feature vector, the textural and structural characteristics are abstracted from the computed image The paper by Naresh and Shettar (2014) [7], applies three categories of classifiers namely SVM, ANN, and k-NN for the purpose that is involved in the identification of disease in addition with determining the extent of disease spread (stage I or stage II) in addition with analysis of performance using ANN, and KNN classifier with respect to diverse characteristics accuracy, sensitivity(recall), precision and specificity. It was observed that SVM accomplishes greater accuracy SVM procedure that attains highest accuracy supports patients and expert doctors in deciding corrective steps on time in addition with minimization of rate at which the death occurs because of this disease.

Detection of persons suffering from Non-Small Cell Lung Cancer (NSCLC) at a primitive phase with greater danger of reappearance might support in the identification of persons in such a way that they might obtain added advantage from ancillary remedial action. The research by Wu et al., (2017) [8] recommended the utilization of computational histomorphometry image classifier utilizing nuclear arrangement, texture, shape, along with the tumor framework in forecasting the chance of reappearance of lung cancer at primitive phase of NSCLC with the help of using the digitized H&E Tissue Microarray (TMA) slides. Utilizing the reviewing reasons of primitive phase of patients suffering from the category NSCLC (Cohort #1, n = 70), they developed a supervised classification prototype encompassing the maximum analytical characteristics related with sickness reappearance. The protype is authenticated on two individual groups of primitive phase NSCLC patients, Cohort #2 (n = 119) and Cohort #3 (n = 116). A multivariable Cox comparative threat prototype of Cohort #2, combining gender along with conservative analytical variables like node status in addition with which the extent of disease specified that the automatically mined histomorphometry score is autonomous factor for analysis.

Cancer rate of occurrence is swelling with an alarming rate globally. Though admitting the fact that the cancer is preventable and curable, it is only possible to identify the presence of cancer in the infected patients at mature stage. In addition, even after the treatment in most cases, there is chance of recurrent of the disease. Hence, forecasting of cancer recurrent takes much significance in order to decide on particular handlings might be required. Nevertheless, conservative techniques in forecasting the reappearance of cancer depends only over histopathology and the outcomes obtained were not trustworthy. The microarray gene expression methodologies are fascinating technique which might forecast reappearance of cancer with the help of investigating the gene expression of illustration cells. The microarray technique permits investigators in examining the expression of thousands of genes concurrently. The work by Win et al., (2014) [9] explains the most advanced machine learning technique reliant strategy termed averaged one-dependence estimators with sub Sumption resolution that helps in handling the issue of forecasting, using DNA microarray gene expression data, if the specific tumor might revisit inside particular duration, that will be commonly up to 5 years. Entropy Reliant Gene Opting Technique is utilized for minimizing the processing for picking of appropriate analytical genes which is straightaway accountable to identify the chance of reappearance

Remedial problems will be usually identified in each solitary member. Tumor takes the position as extraordinary out of the greatest insecure diseases a person be infected from. The process of differentiating at primitive stage is s extremely complex as the affected person can observe the impacts only at later stages. As a result, the prediction of swelling of cells in the lung area at primitive stage of is the most significant one since it supports the increased feasibility of deciding over treatment method. The strategy provides the greater complication in enhancing the chance of survival of infected patients.

A Recurrent neural network with Levenberg–Marquardt model that falls under the category of Higher order Neural Networks along with utilizing glowworm swarm optimization algorithm is suggested by Selvanambi, et al., (2018) [10] to handle the multimodal sickness. Implementation with suggested techniques will be tried with the help of standard available dataset, subsequent results establish that the higher-order recurrent neural framework by using optimization technique named as glowworm swarm provides greater accuracy of 98% while investigating in comparison with the normal optimized neural network.

Forthcoming activity forecast depending on accessible time sequence dimensions is an appropriate investigation field particularly for medical framework, such as analysis and evaluation in involvement of applications. For identifying and analysis of sicknesses, brain dynamics, electroencephalogram time sequence is considered and investigated periodically for accessing the expertise in current in addition with impending, mental states. Because of its chaotic nature, electroencephalogram time sequence necessitates the utilization of dedicated methodologies for competent identification. Motivation of research by Kose 2018) [11] is for establishing hybrid framework constructed by artificial intelligence methodologies in handling with electroencephalogram periodical sequence. Artificial neural networks in combination with the ant-lion optimizer, that are latest intellectual optimization methodology, was implanted for understanding the associated framework and function certain forecasting applications on electroencephalogram time sequence. As per the accomplished results, the framework might competently forecast the forthcoming conditions of target time sequence and still outclasses on comparison with additional hybrid neural network-dependent framework and substitute periodical sequence forecasting techniques.

The research work by Yamany et al., (2015) [12], Ant Lion Optimizer (ALO) is established in training Multi-Layer Perceptron (MLP). ALO is utilized in determining the weights and biases of the MLP for accomplishing least error in addition with improved classification accuracy. Four typical classification datasets are utilized to standardize the functionality of the suggested technique. Further the functionality of the projected technique is analyzed in comparison with optimization algorithms, such as, Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO). Investigational outcomes exhibited that the ALO technique in combination with the MLP is competent to solve the issue of local and accomplished the greater rate of accuracy.

Support Vector Machine (SVM) is fascinating and considered to be one among the best available classifiers. Classification accuracy is considerably high for SVM characteristics with kernel featrues and penalty parameter (C). The work by Tharwat, & Hassanien, (2018) [13], an innovative Chaotic Antlion Optimization (CALO) algorithm was established to optimize the characteristics of SVM classifier, in order to decrease classification error. This research considered six typical datasets that are listed in UCI machine learning data repository. With the purpose of assessing the suggested framework (CALO-SVM), To atuhenticate, the outcomes of the CALO-SVM algorithm is analyzed with grid exploration, that is conservative technique in exploring characteristic values, usual Ant Lion Optimization (ALO) SVM, and optimization algorithms such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). Investigational outcomes assured that the projected framework is having the capacity of determining the best standards of the SVM characteristics and circumvents the issue of getting caught in the local optima. Outcomes established least rate of error while analyzing with GA and PSO techniques.

Presently, differential analysis of lung cancer, precise classification of cancer types (adenocarcinoma, squamous cell carcinoma, and small cell carcinoma) is the most necessary one. Nevertheless, enhancing the accurateness in addition with steadiness of analysis is complicated problem. The analysis by Teramoto, et al., (2017) [14] established a computerized classification arrangement for the identification lung cancers accessible in microscopic images utilizing a Deep Convolutional Neural Network (DCNN), that will be important deep learning technique. The DCNN utilized for classification encompasses three convolutional layers, three sharing layers, and two completely linked layers. Assessment of investigations are carried out conducted, by providing the training to utilizing original database with computation of graphics unit [GUI] Microscopic images will be initially cut and resampled for getting images with resolution of 256 × 256 pixels and, for prevention of overfitting challenge, gathered images were enlarged via rotation, tossing, and filtering. The Chances of occurrences all categories of cancers will identified utilizing the designed arrangement and its classification correctness will be appraised utilizing threefold cross validation. Thus, the newly designed framework is helpful for classification of lung cancers from microscopic images.

III. PROBLEM STATEMENT

Most of the lung cancer types can be detected at matured stage after the cancer has been spread to considerable extent by using traditional techniques that is adopted by Physicians Radiologist worldwide. Though by detecting Lung cancer at that above-mentioned stage, even by providing the most sophisticated treatment, chance of survival of the patient is very low. Apart from the above the mentioned problem, the problem of misdiagnosis is another main cause of worry. Some times a benign category might be identified as malignant and vice versa by Doctors. This also will put the life of the patients in very high-risk situation.

One method to ovecome this concern is by considering computer supported analysis technique as a tool to support radiologists and physicians. Provided with an input CT and conceivable added appropriate infected persons metadata, such procedures focused in delivering a measurable outcome linked to the risk of lung cancer. It is stated to take into consideration of the objective of such framework as twofold. Initially in minimizing the inconsistency in evaluating and observing danger of the lung cancer in between inferring by the diverse physicians. Undeniably,

computer supported strategies was appeared to be enhancing reliability between diverse physicians in a several of medical background, involving detection of node affected by cancer (4) and screening with the help of CT (5) and it may be anticipated that such decision aiding tackles can deliver the similar advantage in classifying the nodule Second, it can provide enhanced functionality in classification that helps the least knowledgeable or not-skillful physicians in evaluation of the danger of a specific node to be malignant.

This research uses Artificial Neural Network in classification of Lung Cancer. Along with this Artificial Network, the parameters of ANN are optimized with the nature inspired meta heuristic technique called as Ant-Lion Optimization Technique. By using the Ant-Lion Optimization technique along with the ANN, the classification accuracy of ANN is enhanced to a considerable extent in the prediction of presence of Lung cancer.

IV. PROPOSED METHODOLOGY

Lung cancer is the most important reason of death out of all cancers existing least chance for survival as per the several online cancer communities. The former medical analysis demonstrated that the detection of lung cancer at a primitive stage by an appropriate examination pattern is the only solution to reduce the rate of death due to lung cancer. Radiological screening consumes lot of time in addition with the involvement of several medical experts. Because of this human error can also creep in. Subsequently, the exactness of screening outcomes prominently relies over the knowledge of doctors. Computer Aided Detection (CAD) in radiology can deliver a purposeful and beneficial manner in which doctors focusing at enhancing correctness and supporting in identification of cancer at primitive stage saving the time and cost of expenditure incurred in various screening procedure in investigation of cancers

The Lung database are taken from the internet source and provided as an source for processing by the proposed system. Proposed strategy has following Pahses. Phase one will be Pre-Processing Second one will be the Feature Extraction Phase, Third Phase will be the Feature Selection and Final Phase will be Classification and optimization of the Classification Process.

Preprocessing of the vague data is performed for extracting and selecting the features from the data set. Then with the selected features trained and tested by back propagation neural network in which parameter is optimized by naturally inspired meta heuristics Ant Lion Algorithm. This back propagation neural network along with the Ant Lion Optimization technique identifies the presence of cancer in a precise and quicker manner

A. Data Set Used

Datasets belong to real-world circumstances so that they are fascinating and appropriate, even though least amount is quite sufficient for us to examine.

i) Requirements of Dataset for Experiments:

Well-Understood: The knowledge about the data set in such a way what is the information content encompasses, the reason for gathering, about the issue which has to be solved in order to format the analysis.

Baseline: It takes much significance to develop the knowledge about the techniques are recognized to be function well and the ranks accomplished by them in order to obtain the fruitful information about the analysis. This takes much significance while initializing and training in order to obtain the faster response for analyzing the performance.

Plentiful: Requirement of several datasets to pick by fulfilling the characteristics that will be useful in examining.

A collection database for machine learning oriented is available with the UCI Machine Learning Repository which will be available from internet and it is open source. The data set are accommodated and preserved in the Center for Machine Learning and Intelligent Systems in the University of California, Irvine. Each dataset contains separate webpage that presents the entire facts about its inclusion and any pertinent research that is examining it. The datasets from the internet will take up the format of ASCII files, frequently the helpful CSV arrangement [15]

Benefits of the Repository

1. All datasets are obtained from the field, referring that they contain real-world potentials.
2. Datasets encompass broader variety of topics.
3. Particulars contained in the datasets will be abridged with the help of concepts such as types of attributes, amount of illustrations, quantity of features and year of establishment that might be arranged and examined.
4. Datasets are analyzed that refers which are recognized according to characteristics that are influencing anticipated to generate better Outcomes. This will deliver a helpful foundation for analysis.
5. Maximum number of datasets are minor in size stating that it can be viewed in a text editor that you can readily load them in a text editor or MS Excel and analyze them, and could be speedily simulated.

Lung Cancer Data set from UCI Machine learning repository [16] is utilized for the employment of the designed technique.

B. Data Pre-Processing

The initial step of the proposed techniques is data preprocessing by which raw data can be converted into the suitable format for which the providing the data set for training and testing and for further Analyzing.

C. Dimension Reduction using Principal Component Analysis

It is stated that in the process of classification accuracy and execution time of classification largely reliant on dimension of the data. So, for quicker classification and enhancing the accuracy of the classification process it is compulsory to select the appropriate features by avoiding redundant and irrelevant features.

PCA is utilized for minimizing the dimension of the data, but here, it is utilized to convert the feature space into principal components feature space and arranged principal components in downward direction.

The purpose of utilizing the PCA is to minimize the dimension of the data whereas keeping modifications that exists in original dataset. It delivers the method for recognizing patterns of data, and articulating the data in order to highpoint the resemblances and differentiations. PCA utilized for the conversion of input vectors to the new exploration space transform the input vectors to the new search space. By utilizing PCA algorithm 7 features are chosen. PCA Algorithm is signified as trails

PCA Algorithm

Input: $x = (x_1, x_2 \ldots . x_M)$

$M -$ Maximum number of features

1. Find mean $\bar{x} = \frac{1}{M}\sum_{i=1}^{M} x_i$
2. Calculate deviation $\leftarrow$ mean $\Phi_i = (x_1 - \bar{x})$
3. Find Covariance matrix (A) // A is (N × M) matrix$\rightarrow \Phi_i \quad (i = 1 \ldots M)$// through $C = \frac{1}{M}\sum_{N=1}^{M} \Phi_N \Phi_N \equiv AA^T$
4. Compute eigenvalue (C) as $C{:}\lambda_1 > \lambda_2 \ldots . \lambda_N$
5. Compute eigenvector (C) as $C{:}\mu_1 > \mu_2 \ldots . \mu_N$ // C will form symmetric basis, (i.e. any vector x or $x_1 - \bar{x}$,originally essential as a proportional combination of the eigenvectors): $x_1 - \bar{x} = b_1\mu_1 + b_2\mu_2 \ldots . + b_N\mu_N = \sum_{i=1}^{N} b_i\mu_i$//
6. Arrange (λ_i, μ_N) //descending order//

$$pc_1 > pc_2 \ldots \ldots > pc_l$$

// l is the transformed number of features//

D. Training and Testing Samples

The input data samples are employed for training and examined with the help of Back Propagation network. At first Selection of weights of neural network for input data are done in an arbitrary way. The neural networks are accepted to learn by using the model data set for training and analysis of data set is performed .Classification technique will be performed and the outcome of classification over the examined data is provided with weights to investigate the error in frequency error or the rate at which the error occurs while performing the procedure of classification process and the error are altered by modifying the weights [17].

E. Neural Networks in Medical Classification

ANNs is a fruitful technique that can be applied for investigating complicated issues in which the associations between input and output data are not clearly defined. Artificial neural network was competently enforced on several divisions of health care like identification of disease image study, deciding the drug for the particular disease. Utilizing artificial neural networks, supervising large number of health characteristics is made possible or might be utilized in the forecasting the response by the patient to the particular remedial action. They are utilized in pattern interpretation because of the capability in obtaining the training and to store the expertise. [18].

An ANN is an extremely interlinked network composed by several simple processing units. Each processing in the framework preserves a portion of active evidence and having capacity of maintaining rare simple processes. An ANN performs the process with the help of broadcasting modifications by stimulating the processors [19]. Utilizing the ANN, it is possible to obtain, preserve and utilize the expertise attained with the help of experiments. The expertise is maintained in stable status net of associations between discrete neurons and might be enhanced involuntarily with the help of utilizing the training technique. [20]

A net comprises several paths that will be triggered, with particular proportion, with the help of vector which will be treated as input. The signals created are propagated and incorporated via several stages of ANN, activating the many neurons, and at last producing the outcome signals.

Depending on the manner of Learning; entire artificial neural networks will be segregated by two training types supervised and unsupervised. Supervised learning, refers the network is learned with the help of the provision for both the patterns of input and output. The neural network having the capacity of fine tuning the connection weights to suit its outcome with the genuine outcome in a repetitive procedure till the necessary outcome is accomplished, while the performing in this phase. An ANN of the unsupervised learning category, like the self-organizing map, the neural network is offered with inputs, it will be deprived of recognized responses. The network contains the capacity of constructing the individual samples for the inducements by the inputs with the help of computing the satisfactory linking weights. That is self-arrangement with the help of grouping the input information in addition with determining the features characteristic related to the application [21].

F. Back Propagation Network

The back-propagation technique established pertaining to the proposed work utilized to learn relies over multilayer neural network through least rate of learning, particularly while utilizing huge learning size of the group. The proposed technique might be enforced with broad means for a network with considerable size which utilizes the back-propagation technique via an best duration (seen time). With the help of BPN technique, accomplishment of the superlative functionality by using the least number of training steps in addition with reduced duration for learning is made possible

i) Standard Back Propagation Algorithm

BPN trains using experience. For a Specific input, making the network functioning in a such a way that by fine tuning the weights on completion of the learning, network will provide the output. Back Propagation networks will be perfectly suited for recognition of patterns and tasks that includes mapping of inputs. It is stated for making the network to learn by providing the samples with respect to the desired output so that network will operate towards achieving the desired results. [22]

ii) Steps of Back Propagation Algorithm

1. Reset Entire connection weights W with least arbitrary values with the help of pseudorandom sequence generator.

2. Reiteration till convergence (While the error E is less than the predefined value or till the gradient vE(t)/vW is lesser in comparison with predefined value).

(1) Calculate improvement factor utilizing

$$\Delta W(t) = -\eta \frac{dE(t)}{dW} \quad (1)$$

(2) Improve the weights with

$$W(t+1) = W(t) + \Delta W(t) \quad (2)$$

(3) Calculate the error E(t+1).

While t is the quantity of iteration, *W* is the weight for linking weight, and *h* is the rate at which the training happens. The Parameter, Mean Square Error (MSE) Function, select the value for the error E that is calculated as the difference between the original output y_j and the anticipated output d_j

$$E = \frac{1}{2}\sum_{i=1}^{n_j}(d_j - y_j)^2 \quad (3)$$

The BP technique explained here contains certain issues. If the framed rate of training is least enough for reducing the complete error, the procedure of training is decelerated. But, rate of training with bigger value might accelerate the procedure of training at the danger of feasible oscillation. Additional issue is that, incomplete negligible opinions or steady phases on error surface will be frequently faced while the procedure of training.

Utilizing the term called as momentum will be the easiest technique in circumventing fluctuation issues while the exploration for the least value over the surface of error. The weight improvement in BP algorithm with momentum term can be expressed as trails:

$$\Delta W(t) = -\eta \frac{dE(t)}{dW} + \alpha\, \Delta W(t-1) \quad (4)$$

The flexible rate of training might be implemented for accelerating the convergence technique of the proposed method. For group learning technique, the rate of training might be modified as trails

$$\eta(t) = \begin{cases} \beta\eta(t-1) & if\ E(t) < E(t-1) \\ \theta\eta(t-1) & if\ E(t) > kE(t-1) \\ \eta(t-1) & Otherwise \end{cases}$$

where h(t) will be rate at which the process of training happens during the t^{th} iteration, and β, θ and k will be selected in order to $\beta > 1, 0 < \theta < 1\ and\ k > 1$ Whereas for the improved approach of learning, rate of training might be improved utilizing

$$\eta(t) = \eta_0 + \lambda E(t-1) \quad (5)$$

The training technique with the help of overlooking procedure will be the technique which might forget unemployed associates. With the help of overlooking procedure, the weights which will not strengthened with the help of training that will vanish. The attained network, therefore, contains emaciated framework which imitates the orderliness encompassed in the data, helpful in enhancing the convergence in addition with the network correctness. It is usual, that the improving the associated weights with overlooking procedure term is provided with the help of

$$\Delta W'(t) = \Delta W(t) - \varepsilon sgn(W(t)) \quad (6)$$

While ε is the quantity at which overlooking happens, and sgn(x) is the sign function. The absolute value of associated weight will be framed for reducing with the help of ε because of the second term on the RHS). The analysis of the conjugate gradients technique is employed, since this one least expenditure on processing demonstrates better results (Polak,1971). The associated weights therefore might be stated with the help of]

$$W(t+1) = W(t) + \eta(t)d(t) \quad (7)$$

$$d(t) = -\nabla E[W(t) + \beta(t)d(t-1) \quad (8)$$

$$d(0) = -\nabla E[W(0) \quad (9)$$

While E will be the gradient, d(t) will be conjugate gradient, h(t) will be step wide, b(t) will be calculated by providing with the help of Polak–Ribiere function demonstrated.

G. Ant Lion Optimizer

Ant Lions are used in optimizing the parameter optimization of Back Propagation Network. This method improves the accuracy of prediction and accelerating the convergence mechanism and avoiding the issue of getting caught within the clutches of local minima.

i) Inspiration

Lifespan for antlions encompasses two prominent stages: larvae and adult. A normal entire lifetime might prevail up to 3 years, that happens greatly in larvae. Antlions experience metamorphosis within a sheath for developing into adult. They regularly hunt for the prey in larvae and the adulthood life hood.

The exclusive hunting conduct and the selection of preferred prey by them will be the inspiration for obtaining the names as Antlions. An antlion larva sneers with shape like a cone, hoke inside the earth by navigating in the path that is circle in shape, and chucking out sands using the hugely structured jaw [24]. After excavating the snare, the maggot's fleeces beneath the nethermost of the cone and delays for insects nearing so that the target can be stuck in the hole. The border of the cone is sufficiently strident for insects in making them fall inside the bottommost of the snare effortlessly. After recognition by the antlion that the target is in the snare, it attempts to get hold of it. Nevertheless, creatures typically will not be getting hold directly and attempt to break out from the trap. Antlions perceptively chuck sands in the direction of boundary in the pit for the purpose of making the target again slide in the direction of the antlion where it is waiting for the prey inside the hole created by the antlions. While the target will be getting hold inside the mouth of the animal, it will be dragged below the soil and will be eaten. After eating the target, antlions chuck the remnants external from the hole and modify the hole for the subsequent hunt.

ALO algorithm's fullest concentration lies in the fact of fighting conduct of antlion's maggots in addition with subsequent subdivision the performance of antlions along with the target that are available in nature will be initially explained and undergo modelling utilizing mathematical concepts. A prototype will be developed that relies on optimization technique is suggested in the subsequent section.

ii) Development of Antlion Algorithm using Mathematical concepts in Optimization

While optimizing, the subsequent settings are enforced:

- Ants navigate around the exploration space utilizing diverse arbitrary walks.
- Arbitrary navigations will be enforced for the comprehensive measurement of ants.
- Arbitrary navigations will be impacted with the help of tricks that antlions create.
- The mentioned animal might construct traps comparative to their suitability.
- Antlions having greater holes contain the increase chance for getting hold of ants.
- Every antlion will endeavor in getting hold of antlions in every step, hence try to become the fittest antlion.
- Variety in arbitrary navigation will be reduced flexibly for Modelling the falling ants in the direction of mentioned animals.
- While the ant develops fitness in comparison with antlion, this refers by means of getting hold and dragged below the sand by the antlion.
- Relocating to another place is performed by antlion for capturing the newest prey and constructs another trap in the new location by enhancing the alteration model with the chance of capturing additional prey subsequent to every hunt.

iii) Random walks of ants:

The ALO technique copies communication that exists among the antlions and ants inside the trap. For the purpose of prototyping the communications, ants should be compulsorily navigating over the exploration space, and antlions will be provided permission for performing the process of hunting the ants in addition with accomplishing fitness utilizing traps. Because of the ants navigate in a stochastical manner in nature while exploring for food, an arbitrary walk has to be selected for creating the prototype of ants' navigation as trails

$$X(t) = [0, cumsum(2r(t_1) - 1), cumsum(2r(t_2) - 1 \ldots cumsum2rtn-1 \quad (10)$$

while cumsum computes the cumulative sum, n will be highest quantity of steps, t displays the phase at which the arbitrary navigation belongs, and r(t) will be stochastic function explained that takes up the expression in the following format

$$r(t) = \begin{cases} 1 & if\ rand\ > 0.5 \\ 0 & if\ rand\ \leq 0.5 \end{cases} \quad (11)$$

while t demonstrates the phase at which the arbitrary navigations and rand will be any arbitrary value created alongside uniform distribution that lies in between [0, 1].

Random walks will be entirely reliant over the Equation10. Ants improve their location by arbitrary walk at each stage of optimization. Because of the reason that each exploration space contains the limit, nevertheless, Expression10 might not be straightaway utilized to improve the location of ants. For retaining the arbitrary walks within the exploration space, they will be standardized utilizing the subsequent expression:

$$X_i^t = \frac{(X_i^t - a_i) * (d_i - c_i^t)}{(d_i^t - c_i)} \quad (12)$$

while a_i will be the least of arbitrary navigation of i^{th} variable, b_i will be largest value at which the arbitrary walk belongs, in i^{th} variable, c_i^t will be the least value for the i^{th} variable during t^{th} iteration, and d_i^t specifies the highest value for the ith variable during the t^{th} step. Expression (12) have to be enforced in every step in assuring the existence of arbitrary navigations within the exploration space.

iv) Getting hold of prey in antlion's pits

As per the mentioned fact in the above explanation, arbitrary walks of ants will be impacted by antlions' snares. For creating the mathematical prototype, the subsequent presumption in the form of equation will be expressed

$$c_i^t = Antlion_i^t + c^t \quad (13)$$

$$d_i^t = Antlion_i^t + d^t \quad (14)$$

while c^t is the least of entire variables at t^{th} iteration, d^t designates the vector computing the highest of entire variables during t^{th} step, c_j^t is the least of entire variables for i^{th} ant, d_j^t will be highest of entire variables for i^{th} ant, and $Antlion_j^t$ expresses the location of the chosen j^{th} antlion at t^{th} iteration. Equations (13) and (14) present which ants arbitrarily walk in a hyper sphere explained by the vectors c and d around a chosen antlion.

v) Building trap

For creating the antlions's hunting competence, a roulette wheel will be implemented. Ants will be presumed to be stuck in single chosen antlion. The ALO technique will be needed for utilizing a roulette wheel operator to choose antlions depending on the fitness while performing the optimization task. The strategy provides greater probability of achieving the fitter antlions to capture ants.

vi) Sliding ants towards antlion

With the help of strategies projected, antlions will be having the capacity to construct snares in comparison with the appropriateness and ants will be compulsorily needed to navigate arbitrarily. Nevertheless, antlions chuck sands away towards the center of the hole after recognizing that the ant is within the snare. The conduct by antlion makes the ants to slip down the ensnared ant that will be attempting to break free from the snare. To create the prototype by utilizing mathematics of this performance the radius of ants's arbitrary walks hyper-sphere will be reduced flexible. The subsequent expressions will be projected with respect to the above-mentioned concept.

$$c^t = \frac{c^t}{I} \quad (15)$$

$$d^t = \frac{d^t}{I} \quad (16)$$

While I will be ratio, c^t is the least of entire variables during t^{th} step, and d^t refers the vector comprising , largest value of entire variables during t^{th} step.

vii) Catching prey and re-building the pit

The concluding phase of hunt will be while an ant arrived the lowermost part of the hole and will be captured within the mouth of antlion. Subsequent to the above-mentioned phase, the antlion drags the ant and take it inside the sand and consumes the body of ants. To exploit the proposed procedure, it will be presumed that capturing the target might happen while ants develops fitness than the respective antlion.

An antlion will be needed to function in such a way that for enhancing the place to another place for the purpose of hunting the ant to increase the probability of capturing next target. The subsequent expressions will be projected with respect to the above-mentioned concept.[23]

$$Antlion_j{}^t = Ant_i{}^t if\ f(Ant_i{}^t) > f(Antlion_j{}^t \quad (17)$$

while t depicts the present step, $Antlion^t_j$ displays the place for which the designated j^{th} antlion during the t^{th} step belongs to, and Ant^t_i specifies the place for which the of i^{th} ant belongs during t^{th} step.

The flowchart of the antlion optimizer is presented in Figure 1

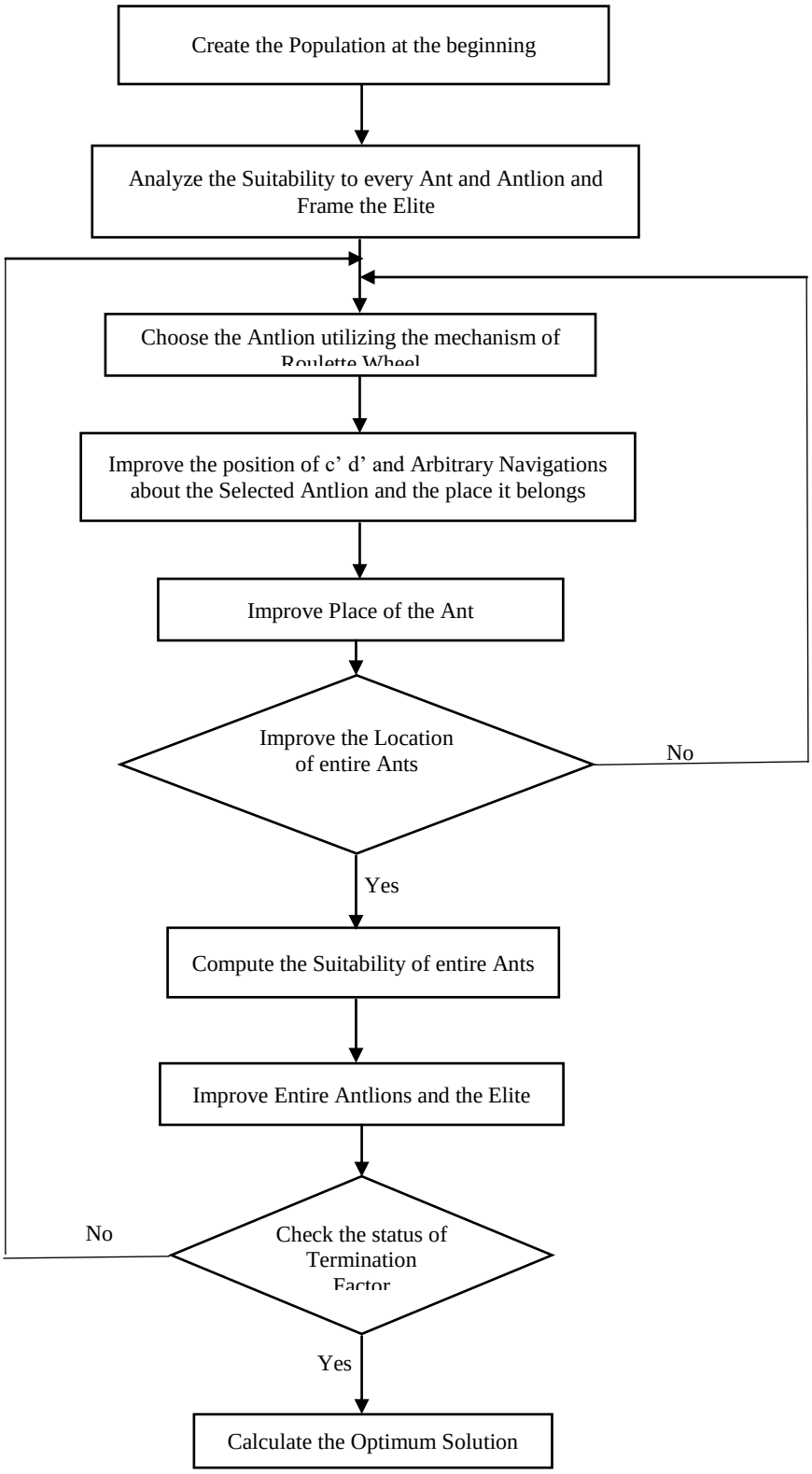

Fig 1 Flowchart of the Proposed Antlion Algorithm

H Parameter Identification and Optimization using ALO

In general, an unidentified parameter recognizing issue with recognized framework and chosen algorithm might altered to an optimization issue. Subsequent to triggering the actual framework, an explained objective function should be specified alongside the actual framework outcomes will be recognized as system outcomes. Subsequently unidentified parameters might consider the vector to be particle and the objective function have to be minimized in the optimization procedure. While feedbacks from actual framework and recognized structure will cooperate with each other in the healthier manner, the predicted parameters will be nearer to actual values.

i) Objective Function

The objective function of the projected technique is to accomplish the maximum rate of classification with training and testing samples. A latest objective function alongside weights which epitomize the significance of every parameter is projected and the weights will be computed with the resulting nonconformity. This latest motivation is the weighted least squares error amongst original and predicted outcome vectors. Expression of the function in the following way:

$$C_{IOF}(\theta) = \sum_{k=1}^{L}\sum_{j=1}^{n}(w_j(z_j(k) - \widehat{z_j(k)})^2 \quad (17)$$

while z = [x y m_t] will be the output vector of actual struture, $\hat{z} = [\hat{x}, \hat{y}, \widehat{mt}]$ will be output vector of predicted prototype, L will be quantity of samples, n will be size of output vector for the structure with the value n = 3. The output vector will be computed with an amount of diverse periods while computing the weights in the motivation. The weight vector w = [w1 w2 w3] will be computed as per the subsequent steps:

(1) Frame each value of every vector parameter $\theta_i, i = 1,2 \ldots m$i, (m will be the size of parameter vector θ,m = 5), and obtain system output vector z = [z_1(k), . . . , z_n(k)], k = 1, . . . , L.

(2) Loop: i = 1:m

Alter the i[th] parameter's value, $\theta_{new} = \theta_i * (1 + \Delta\%)$and attain the system output vectorˆzi = [ˆzi1(k) , ..., ˆzin(k)].

End Loop.

(3) Compute

$$g_j = \sum_{i=1}^{m}\sum_{k=1}^{L}\left(z_j(k) - \hat{z}_{ij}(k)\right)^2$$

(4) Compute the j[th] weight

$$w_j = \frac{g_j}{\sum_{j=1}^{n} g_j}$$

ii) Parameter Identification Strategy

Figure 2 depict the picture that encompasses the procedure of ALO dependent BPN parameter realization. Initially an appropriate input signal will be selected for triggering actual system and the approximated system. The attained processed outcome and modelled outcome as inputs in the suitability evaluator will be utilized in calculating the suitability. Then, in ALO-reliant realizerthe unidentified parameter vector $\hat{\theta}$ will be recognized by reducing the suitability function$C_{IOF}(\hat{\theta})$. Therefore, the process repeats, the recognized parameters estimated in the direction of or nearer to actual values slowly [25]. The parameter identification strategy is displayed in figure 3.

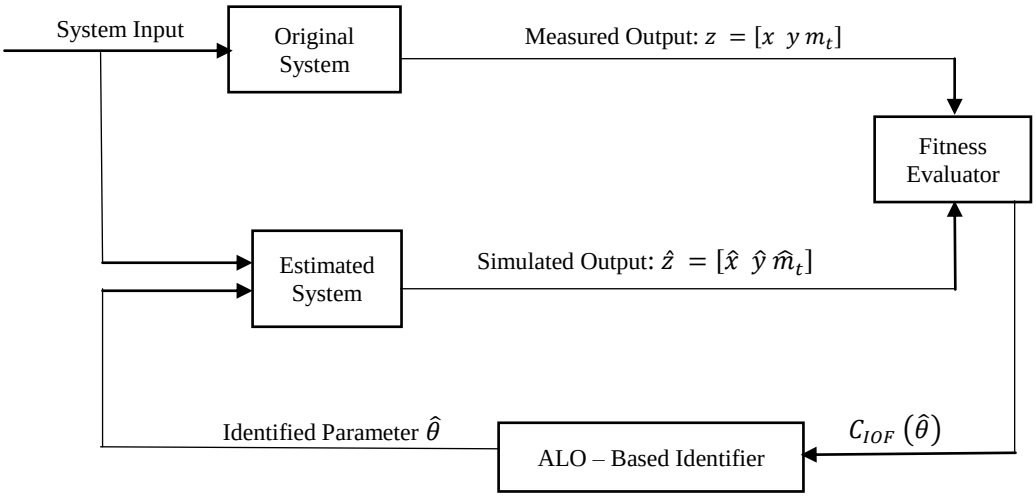

Fig 2 Block Diagram of Parameter Identification using ALO

Parameter Recognition exactness is computed with the help of the following expression Parameter Error (PE):

$$PE = \left|\frac{\theta_i - \hat{\theta}_i}{\theta_i}\right| * 100\% \quad (18)$$

and average parameter error (APE):

$$APE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\theta_i - \hat{\theta}_i}{\theta_i}\right| * 100\%$$

where $\theta_i$is the parameter in actual system, $\hat{\theta}_i$is the parameter in approximated system, m is the size of θ.

V. EXPERIMENTAL ANALYSIS

Lung Cancer Data set downloaded UCI Machine learning repository [16] which is open source and benchmarked for training and testing for so many machine learning techniques utilized to validate the performance of the projected algorithm.

The lung data set comprises huge number of features. The certain features will be irrelevant and redundant. These features will consume much time while processing them and producing the results of classification. Optimum quantity of features will be chosen using Principal Components Analysis and given to BPN for training.

Out of the entire lung cancer data set obtained for investigations 70 % of data set is utilized in making the BPN Classifiers to learn and the remaining 30% of dataset will be employed to verify the functionality of the classification process of the designed technique. The functionality of the projected ALO – BPN Algorithm is investigated by analyzing the outcomes with BPN – GA [27] and BPN – WO[26] with reference to metrics attained from Confusion matrix in Table 1.

In the research paper group of 7 features are chosen utilizing the Principal Component Analysis BPN framework will be provided as 7-14-1 that is BPN structure containing input layer with 7 neurons and 14 neurons in the layer that will be in hiding mode and 1 neuron in the output layer the framework is chosen

TABLE 1
CONFUSION MATRIX

		Predicted Class	
	Total Population	Prediction Positive	Prediction Negative
Actual Class	Condition Positive	True Positive (TP)	False Negative (FN)
	Condition Negative	False Positive (FP)	True Negative (TN)

A **confusion matrix** is a table that will be useful in describing the functionality of a selected classifier over cluster of test data for which the actual values are already recognized.

Performance Metrics

- True Positive (TP) - The Mined dataset comprising cancer nodule that will be decided as cancerous.
- False Positive (FP) - The Mined dataset devoid of cancer nodule that will be identified as cancerous.
- True Negative (TN) - The Mined dataset deprived of cancer nodule will be recognized as non-cancerous.
- False Negative (FN) - The Mined dataset comprising cancer nodule will be forecasted as non-cancer

Accuracy

This characterizes as numerous times the diverse illustrations will be examined by the same process and the machine or system accomplishes outcomes which will be exact. The accuracy is the percentage of factual outcomes in whole data.

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

Sensitivity

Sensitivity signifies by what means precisely a cancer investigation recognizes persons as existence of lung cancer.

$$\text{Recall(R) or Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

Specificity

Specificity symbolized that by what means precisely a cancer investigation identifies person who do not suffer from lung cancer.

$$\text{Precision(P) or } (1 - \text{Specificity}) = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

It was decided to test all these parameters under three stages of cancer. Since Detection of Lung Cancer at primitive state will be having profound impact in deciding the treatment strategy for increasing the chance of survival.

Table 2,3, and 4 describes the analysis of functionality for the projected technique neural network and ant lion optimization with respect to functionality metrics like accuracy, Specificity and Sensitivity for prediction of all stage's caner.

TABLE 2
PERFORMANCE ANALYSIS PREDICTION OF LUNG CANCER AT THIRD STAGE USING BPN- ALO TECHNIQUE

Classifiers	Accuracy (%)	Specificity (%)	Sensitivity (%)
BPN-GENETIC ALGORITHM	95.4	91.7	93.5
BPN-WO	98.1	95.2	90.3
PROPOSED BPN – ALO	99.1	96.8	88.6

TABLE 3
PERFORMANCE ANALYSIS PREDICTION OF LUNG CANCER AT SECOND STAGE USING BPN- ALO TECHNIQUE

Classifiers	Accuracy (%)	Specificity (%)	Specificity (%)
BPN-GENETIC ALGORITHM	84.3	82.6	90.5
BPN-WO	87.6	86.4	86.7
PROPOSED BPN – ALO	93.2	89.6	84.3

TABLE 4
PERFORMANCE ANALYSIS PREDICTION OF LUNG CANCER AT FIRST STAGE USING BPN- ALO TECHNIQUE

Classifiers	Accuracy (%)	Specificity (%)	Sensitivity (%)
BPN-GENETIC ALGORITHM	78.7	75.2	83.7
BPN-WO	85.2	80.7	78.6
PROPOSED BPN – ALO	87.1	85.3	74.7

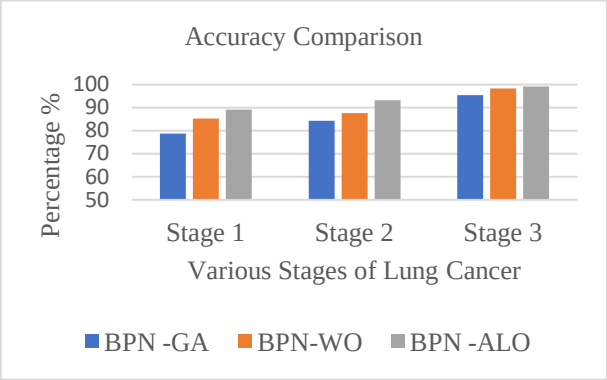

Fig 3 Accuracy Comparison of Prediction of Lung Cancer at Various Stages

Figure 3 demonstrates the accuracy, y Comparison of Lung Cancer at various stages of Lung Cancer. Out of the Techniques Available for the Prediction of Lung Cancer the proposed BPN – ALO performs better by having a considerable accuracy of 99.1 for detecting the lung cancer at stage 3 and at stage 2 it is having the prediction accuracy of 93.2 and even at stage 1 it is having the prediction accuracy of 87.1. From the results it is inferred that the proposed BPN – ALO technique performs better than other techniques.

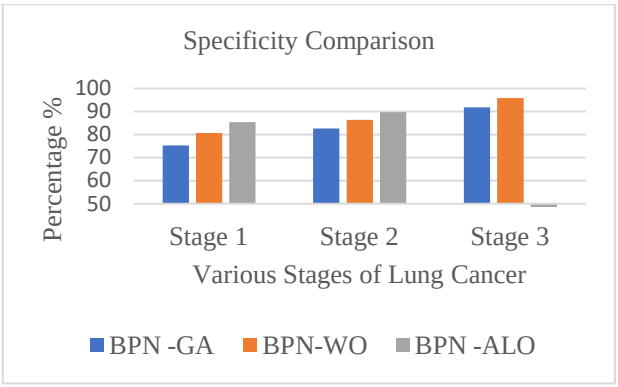

Fig 4 Specificity Comparison of Prediction of Lung Cancer at Various Stages

Figure 4 establishes the Specificity Analysis of Lung Cancer at three phases of Lung Cancer. Among the methods Available for the Prediction of Lung Cancer the proposed BPN – ALO performs better by having a substantial precision of 96.8for detecting the lung cancer at stage 3 and at stage 2 it is having the prediction precision of 89.6 and even at stage 1 it is having the prediction accuracy of 85.2. From the attained outcomes it is observed that the proposed BPN – ALO technique performance is the best on comparison with other techniques.

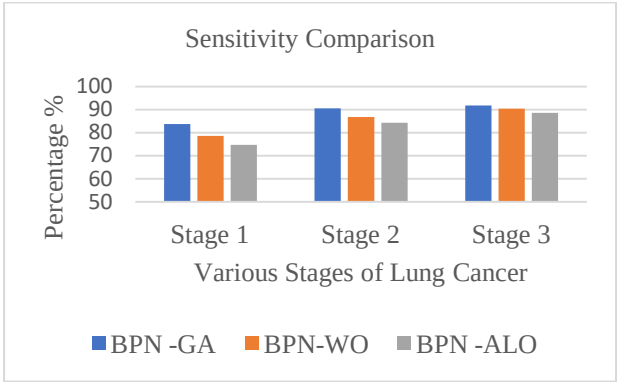

Fig 5 Sensitivity Comparison of Prediction of Lung Cancer at Various Stages

Figure 5 show the Sensitivity examination of Lung Cancer at all stages of Lung Cancer. Between the available strategies Prediction of Lung Cancer, the proposed BPN – ALO performs better by obtaining the sensitivity values of 88.6 for detecting the lung cancer at stage 3 and at stage 2 it is having the prediction precision of 84.6 and even at stage 1 it is having the prediction accuracy of 74.7. From the accomplished results it will be noticed that the performance of proposed BPN – ALO procedure is having the improved values by comparing with other techniques.

By collecting the gene profile dataset of the persons who are detected with First Stage and Second Stage of Lung cancer, the data set is presented for predicted for recurrence of lung cancer with the same work.

Table 5 and 6 establishes the Chance of Prediction of Recurrence affected by Lung Cancer during second at First Stage.

TABLE 5
PERFORMANCE ANALYSIS PREDICTION OF RECURRENT AT SECOND STAGE USING BPN- ALO TECHNIQUE

Classifiers	Accuracy (%)	Specificity (%)	Sensitivity (%)
BPN-GENETIC ALGORITHM	66.7	64.1	71.6
BPN-WO	70.9	68.7	67.2
PROPOSED BPN – ALO	72.3	70.4	65.7

TABLE 6
PERFORMANCE ANALYSIS PREDICTION OF RECURRENT AT FIRST STAGE USING BPN- ALO TECHNIQUE

Classifiers	Accuracy (%)	Specificity (%)	Sensitivity (%)
BPN-GENETIC ALGORITHM	61.2	60.4	63.4
BPN-WO	63.2	62.6	61.7
PROPOSED BPN – ALO	67.8	64.5	59.3

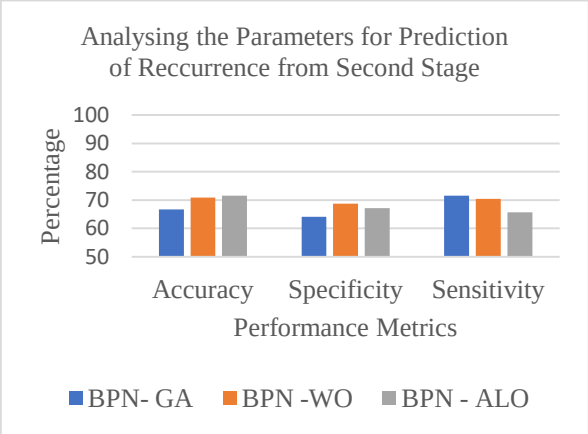

Fig 6 Comparison of Various Parameters for Prediction of Recurrence Lung Cancer at Second Stage

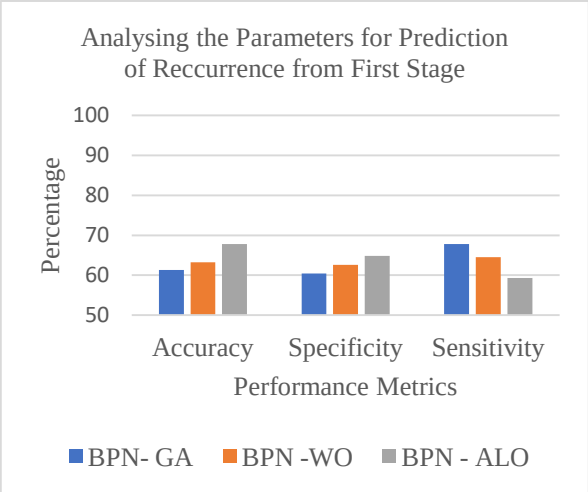

Fig 7 Comparison of Various Parameters for Prediction of Recurrence Lung Cancer at Various Stages

From Figures 6 and 7 it is explained that comparison of various parameters for predicting the chance of recurrence for Stage 2 and Stage 1 respectively. From the figures the proposed BPN – ALO technique has the better prediction accuracy of for the chance of recurrence for stage 2 patients of 72.3 and for stage 1 patients prediction accuracy of 67.8. Though the prediction of chance of recurrence the it is having low value of accuracy it is better than some of the other techniques available.

VI. CONCLUSION

The research paper prominently concentrated to predict the presence of lung cancer possibility utilizing BPN by selecting the parameter optimization of BPN using Nature Inspired Meta Heuristics techniques so as for improving the accuracy and precision of the lung cancer prediction at primitive stage itself. In addition, the proposed BPN – ALO technique is utilized in predicting the chance of recurrence after undergoing the treatment. These both results will be very much helpful in deciding the nature and depth of treatment provided to lung cancers. From the results obtained from the Investigations performed the proposed BPN - ALO has the better performance in predicting the chance of recurrence and achieved highly accurate results in predicting the presence of lung cancer at primitive stage.

REFERENCES

1. https://www.medicalnewstoday.com/articles/323701.php
2. Greco, F. A., & Oldham, R. K. (1979). Small-cell lung cancer. New England Journal of Medicine, 301(7), 355-358.
3. https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/stages
4. Xie, Y., & Minna, J. D. (2008). Predicting the future for people with lung cancer. Nature medicine, 14(8), 812.
5. Uramoto, H., & Tanaka, F. (2014). Recurrence after surgery in patients with NSCLC. Translational lung cancer research, 3(4), 242.
6. Dandıl, E. (2018). A Computer-Aided Pipeline for Automatic Lung Cancer Classification on Computed Tomography Scans. Journal of Healthcare Engineering, 2018.
7. Naresh, P., & Shettar, D. R. (2014). Early Detection of Lung Cancer Using Neural Network Techniques. Prashant Naresh Int. Journal of Engineering Researchand Applications, 4.
8. Wu, Y. C., Wei, N. C., Hung, J. J., Yeh, Y. C., Su, L. J., Hsu, W. H., & Chou, T. Y. (2017). Generating a robust prediction model for stage I lung adenocarcinoma recurrence after surgical resection. Oncotarget, 8(45), 79712.
9. Win, S. L., Htike, Z. Z., Yusof, F., & Noorbatcha, I. A. (2014). Cancer recurrence prediction using machine learning. International Journal of Computational Science and Information Technology (IJCSIT), 6(1).
10. Selvanambi, R., Natarajan, J., Karuppiah, M., Islam, S. H., Hassan, M. M., & Fortino, G. (2018). Lung cancer prediction using higher-order recurrent neural network based on glowworm swarm optimization. Neural Computing and Applications, 1-14.
11. Kose, U. (2018). An Ant-Lion optimizer-trained artificial neural network system for chaotic electroencephalogram (EEG) Prediction. Applied Sciences, 8(9), 1613.
12. Yamany, W., Tharwat, A., Hassanin, M. F., Gaber, T., Hassanien, A. E., & Kim, T. H. (2015, September). A new multi-layer perceptrons trainer based on ant lion optimization algorithm. In 2015 Fourth international conference on information science and industrial applications (ISI) (pp. 40-45). IEEE.
13. Tharwat, A., & Hassanien, A. E. (2018). Chaotic antlion algorithm for parameter optimization of support vector machine. Applied Intelligence, 48(3), 670-686.
14. Teramoto, A., Tsukamoto, T., Kiriyama, Y., & Fujita, H. (2017). Automated classification of lung cancer types from cytological images using deep convolutional neural networks. BioMed research international, 2017.
15. https://machinelearningmastery.com/practice-machine-learning-with-small-in-memory-datasets-from-the-uci-machine-learning-repository/
16. https://archive.ics.uci.edu/ml/datasets.html
17. Kuruvilla, J. and Gunavathi, K., 2014. Lung cancer classification using neural networks for CT images. Computer methods and programs in biomedicine, 113(1), pp.202-209.
18. Brause, R. W. (2001, October). Medical analysis and diagnosis by neural networks. In International Symposium on Medical Data Analysis (pp. 1-13). Springer, Berlin, Heidelberg.
19. Tkacz, E. J., & Kostka, P. (2000). An application of wavelet neural network for classification of patients with coronary artery disease based on HRV analysis. In Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143) (Vol. 2, pp. 1391-1393). IEEE.
20. Al-Shayea, Q. K. (2011). Artificial neural networks in medical diagnosis. International Journal of Computer Science Issues, 8(2), 150-154.
21. Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis.
22. Gupta, A., & Shrivastava, M. (2012). Performance of ANN using Back Propagation Algorithm for Medical Diagnosis System. International Journal of Advanced Computer Research (IJACR), 2.
23. Mirjalili, S. (2015). The ant lion optimizer. Advances in Engineering Software, 83, 80-98.
24. Griffiths, D. (1986). Pit construction by ant-lion larvae: a cost-benefit analysis. The Journal of Animal Ecology, 39-57.
25. Tian, T., Liu, C., Guo, Q., Yuan, Y., Li, W., & Yan, Q. (2018). An improved ant lion optimization algorithm and its application in hydraulic turbine governing system parameter identification. Energies, 11(1), 95.

26. Ibrahim Aljarah · Hossam Faris, · Seyedali Mirjalili, optimizing connection weights in neural networks using the whale optimization algorithm (Nov 2016) International Journal of Soft Computing
27. Bhuvaneswari, P. and Therese, A.B., 2015. Detection of cancer in lung with k-nn classification using genetic algorithm. Procedia Materials Science, 10, pp.433-440.

AUTHORS PROFILE

Dr. S. Senthil is Professor and Director, School of Computer Science and Applications, REVA University, India. Previously, he worked as Associate Professor and Head, Department of Computer Science, Vidyasagar College of Arts and Science, India. He obtained his B.Sc. (Applied Science – Computer Technology) from PSG College of Technology, India, in 1995, Master of Computer Applications from Bharathidasan University, India, in 1999, M.Phil. in Computer Science from Manonmaniam Sundaranar University, India in 2002, and Ph.D. in Computer Science from Bharathiar University , India, in 2014. Another achievement is clearing SET in 2012. He has published 40 research papers in various reputed National and International Journals. He has presented a paper entitled "Lossless Preprocessing Algorithms for better Compression" in an IEEE International Conference at Zhangjiajie, China. His interest is in Database Systems, Data Mining, Data Compression, and Big Data Analytics.

Ayshwarya B, currently pursuing Ph.D. from REVA University, Bangalore . Her research focuses on datamining .She completed her M.Sc. degree from saurastra university, Gujarat. She has more than 11 years of teaching experience from Kristu Jayanti College, Bangalore with subject expertise in C, C++, Java, Python, PHP,database management System,Datamining,Deep Learning and other web development technologies. She published 10 Papers in National and International Journals and Authored three books.

Dr. Shubha.A, Director and Professor-School of Commerce and Management Studies, holds post graduate degree in Commerce and Management studies. She has Master of Philosophy in Management studies. She also holds Doctorate degree in Commerce. She has 15 years of teaching Experience. Her research focuses on media's impact on youth. She is been an active researcher adding to her credit, several national and international articles in journals of high repute. She is also the recipient of best teacher award that was instituted by world CSR group in the year 2015. She actively participates in attending and organizing various seminars and conferences. Her keen interest lies in the areas of accounting, finance and marketing.