International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Lung Cancer Prediction Using Improved SALP Swarm Optimization and LSTM Human Gene Classification.

[1]Kommana Swathi*, [2]Subrahmanyam Kodukula

Submitted: 16/09/2023 **Revised**: 29/10/2023 **Accepted**: 15/11/2023

Abstract: The medical domain requires a gene selection model to manage cancer effectively. The large amount of gene information makes it difficult for the existing model to analyze the relationships between the features. In addition to local optima traps, lower convergence, and overfitting, the existing models have other limitations. The Quantum Ant Lion (QAL) is proposed to select features from gene datasets along with the Improved Salp Swarm Algorithm (ISSA) to initiate the classification. Salp swarm optimization is used in the salp swarm method to increase search efficiency, thereby increasing exploration and overcoming the local optima trap. To increase exploitation in feature selection, the ISSA hyperparameter is applied. The proposed ISSA-LSTM technique increases exploration and exploitation, thus increasing the accuracy rate of the lung cancer classification. LSTM are used to classify the emotional features of the lung cancer. The performance of the ISSA-LSTM method in analyzed in the terms of accuracy, specificity, recall, F-measure and MCC. The ISSA-LSTM method has 99.54 % accuracy for gene classification using micro array gene dataset.

Keywords*: Lung cancer prediction, Human Gene Selection, Quantum Ant Lion, Improved SalpSwarm Algorithm, Hyper parameter initialization.*

1.Introduction:

Now a day,there are many types ofcancer could to be occurred by either genetic or epigenetic changes of our body [1]. Cancer is a compound disease and according to the WHOLung cancer was the terminal cancer among the other cancer types of the world [2]. In 2012, 13% of all new cases of human cancer and 1.590 million cancer-allied deaths were attributable to lung cancer and making it the worst cancer in the world [3]. In spite of recent developments in initialstage of discovery in the therapeutic choices, and anticipation, more than 529 200 patients expired in China from lung cancer in 2011as wellsecretarial for 25% of all fatalities from cancer [4-5] in the world. Additionally, based on histopathological and clinical standardslung cancer can be separated into small cell lung cancer and non-small cell lung cancer in which Lung adenocarcinoma is the most prevalent histopathologicalsubtype of NSCLC of the function [6-7]. In current years, the occurrence of LUAD has increased and the prediction is typically poor because the disease is frequently diagnosed at an advanced stage and its reason for the leading cause of death in the world [8]. Therefore, it is urgently necessary to evaluate and identify early detection biomarkers in order to improve LUAD therapy responses and predict prognosis of the medical system [9-10].

A human gene microarray is one of the most widely accepted tools for identifying and analyzing biological data, and it is nothing more than monitoring hundreds of genes simultaneously in a single transfection experiment. [11-13]. Thereby the DNA microarray gene knowledge was the great support to the investigators to identify the genes caused by the cancer as well recognizes the sub-types of the gene related cancer diseases [14]. Moreover, classification of the cancer is the important for prevent and identify the cancer treatment. In this order they are two phases of classification of cancer first; is the identification of an existing allocates the tumor to the cancer class. However, due the computational and clustering of the function did not well perform the effective classification of the system [15].The classification of the cancer is required to prevent and identify the correct treatment to the patient diagnosis.

The majorcontribution of this research is given as follows:

- The improved SSA algorithm are used to hyper parameter values with optimal weights to enhance as well initialized the emotional classification. In order to improve gene classification efficiency and overcome local optima, a QAL model has been proposed. Quantum bits of a quantum search in the QAL model enhance exploitation.
- To reduce redundant features from hybrid features, ISSA-LSTM-based feature selection is proposed. A redundant feature set affects classification. Moreover, it delivers higher classification accuracy along with the mutation function.

Department of computer science and engineering,Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, A.P,522302, India
Corresponding Author Email: Kommanaswathi@gmail.com, smkodukula@kluniversity.in

2. Literature survey:

A. Sampath Kumar *et al* [16] represented a modified cross over (CSC) algorithm for selecting genes from micro array samples and classify the frequent cancer subtypes of the system. In this work micro array Kent ridge biomedical data repository was presented to analysis and detects the applicable genes along with predictive potency for classification accuracy.

Additionally, using CS algorithm with crossover property for feature selection it was the effective way to lead the population moreover, it was the fraction of the poorest nest and identified the new nest though the cross over function. k-NN classifier was used to detect the effective cancer classification accuracy. Therefore, based the crossover function would help to effective lead the population of the local minima though, using more dimensional datasets for classification was not scalable of the function.

Fei He*et al* [17] introduced the analysis of microarray profiling of differentially expressed IncRNA and mRNA in lung adenocarcinomas and bioinformatics (LUAD) gene expression microarray of the function.In this micro array was used benjamini-hochberg technique, and the network was constructed based on Pearson correlation co-effectual and visualization function. Additionally, validated with considered the collected IncRNA and mRNA gene expression analysis information to the TGCA data based on the weighted gene co-expression network exploration as well as were assessed by using reactome database of the system.Consequently, the data from the analysis would help to enhance the identification of potential LUAD expression information but their function remains to be elucidated of the gene expression function.

Bulent Haznedar*et al* [18] demonstrated classification of the microarray gene expression cancer data using optimizing adaptive neuron fuzzy inference system and the simulated annealing algorithm. In additionally, feature selection was to be used for reduce the size of the problem by taking the aspects and deliver the efficient classification accuracy of the micro array dataset. The ANFIS was the taken input and output from the *IF-THEN* rule structures and it obtained five different layers with consequent parameters. So, in this method to increase the successful classification with accuracy rate however, the performance of the algorithm remains limited due to the profusion of genes numbers.

Abhilasha Chaudhuri *et al* [19] represented the effective filter wrapper approach to microarray cancer data classification with an integrated binary jaya algorithm for feature selection. This approach helps to select the optimum subset aspects from microarray data to enhance the accuracy of the classification. It was also possible to make the aggregate aspects using a technique for order preference by similarity to the ideal solution function and feature extraction. In addition, a transfer function that changes over time was used for subset selection to find the best subset aspects for population-based optimization of the system. Therefore, the algorithm helps trade-off exploration and exploitation with average execution time, though it takes much more computational cost than the selection method.

B. H. Shekar *et al* [20] proposed the L1 regulated aspects selection and classification approach to micro array cancer detection using deep learning algorithm. In this selection based on LSVM which was included the shrinking strategy of penalty terms to expectation error in order to decrease the weight of the irrelevant aspects of the function. Additionally, to the classification purpose deeplearning neural networks was initialized along with sigmoid activation function to the input function and the accommodate multiclass classification, soft-max activation to the output function of the system. Therefore, in this selection function to effective reduces the irrelevant weights of the microarray classification but this classification dataset was highly imbalanced and very small.

3. Proposed Methodology:

In this research, the lung cancer classification and detection by using hyper parameter optimization function. To increase the accuracy of classification, we apply the normalization method to the gene dataset. LSTM classifiers use hyperparameter optimization to use the selected features after QAL has selected the features. Figure 1 illustrates the flow of ISSA-LSTM in gene classification.

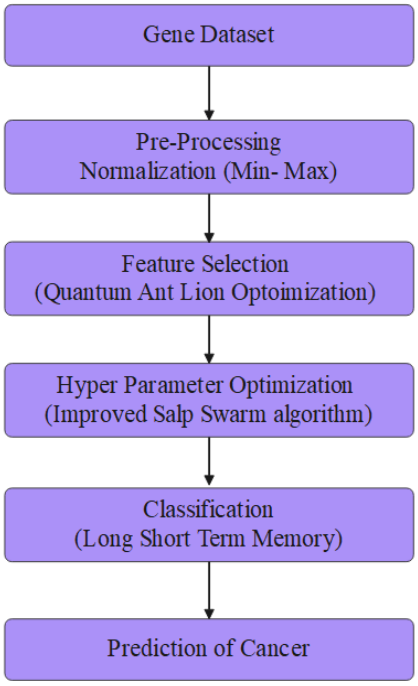

Fig 1. The block diagram of ISSA-LSTMmethod

3.1 Data Acquisition:

In thisISSA-LSTMmethod, utilized the lung cancer classification is done used human gene expression microarray cancer dataset [16]. The micro array gene dataset were perused and managed using Agilent aspects extraction software and the superiority control and normalization was performed using Gene Spring GX software of the function. This dataset obtained the expression level of 1626 genes are taken form the lung samples and these data genes are consisting of two types of lungs cancer information and 1.50 of adenocarcinoma of the lung and 31.0 of malignant pleural mesothelioma of the function [17-18]. Moreover, the input of the array values is given through the MATLAB R 2022for classifying the lung cancer.

3.2 Data Normalization:

As part of the preprocessing, data is normalized to eliminate the effects of significant aspects and outliers. In generally, these data are composed from the various source of data base and thus values are varies with wide range. Thereby, the values are required to scale with the stated range by using min – max normalization function within the range of 0 to 1 as in equation (1).

$$Normalized\_value = \frac{Original\_value - \min\ (original\ value\ in\ the\ series)}{\max(original\ value\ in\ the\ series) - \min\ (original\ value\ in\ the\ series)} \quad (1)$$

In this work 10 fold cross validation method was used to training and testing the processed values as in equation (2) and (3) are:

$$x_{std} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

$$x_{scaled} = x_{std} \times (max - min) + min \quad (3)$$

Moreover, in this method, the min-max transformation method reduces the feature range differences and is further applied to the QAL for effective classification and feature selection.

3.3 QAL Feature Selection:

Using the Ant Lion Optimization method, parametric values and feature weights are determined. According to this method, ant lions hunt by lining cone-shaped pits in the sand and catching their prey (ants). Ant lions' hunger levels determine the size of their traps. When the need for food is high, the pit will be larger, and vice versa. The ant-lion hunting process is mathematically modeled, followed by the search for feature subsets, the definition of the classifier parameter, and the algorithm used for model learning.

The P ant lions method and ants in dimensional problem space are denoted as $ALO = \{AL_i, A_i | i \in P$, where as ant lions and ant are denoted as P. In a given search space, the initial positions of the ant lions are determined randomly. Based on the stochastic and random walk movement of ants in a given search space, equation (4) shows the random variable for finding food in the given search space.

$$s = \begin{cases} 1 & if(r > 0.5) \\ 0 & else \end{cases} \quad (4)$$

Randomly, r is generated in the range [0, 1]. Furthermore, where the current iteration denotes and the

maximum iteration denotes. An ant's position is maintained within a predefined search space by using min-max normalization. Equation (5) and boundary conditions are related, so the newer calculated position exceeds the boundary conditions. An ant's newer position is calculated using equation (5).

$$X^t = \frac{(x^t - a_i) \times (a_i^t - c_i^t)}{b_i - a_i} + c_i^t \quad (5)$$

To the further, ratio are represents as$I$, as given in equation (6).

$$I = L\frac{t}{T} \quad (6)$$

A constant parameter controls the degree of exploitation. This attribute requires feature weights to be modified by several digits. To present the consumption of lion ants, a fitness value is used. When ant exploitation is better for the newer ant lion position than the current position, the correct position is updated as follows: exploit is a constant parameter. With a few decimal places of precision, the parameter meets feature weight modification requirements. Equation (7) describes the consumption of a lion ant.

$$AL_m^{t+1} = A_h^t \; if \; fit(AL_m^t) < fit(A_h^t) \quad (7)$$

The classification accuracy fitness function is given in equation (8).

$$fit = \frac{correctly\ classified\ instances}{Total\ instances} \quad (8)$$

The best ant lion is measured from all ant lions in each iteration that have the highest fitness. In each iteration movements of ants are affected by elite ant lion and ants tend to move towards elite ant lion.This might lead to ant lions trapping into local optima and elitism is applied to maintain the best solution at any iteration in optimization of the function. Therefore, based the QAL with fitness function effectively reduced the range of dimensionality of the features of the function.

3.4 Hyper parameter Initialization to LSTM classification using ISSA:

After reducing the dimensionality of the array features, the selected hyper parameter aspects are initialized to achieve the higher the performance, in terms of emotion classification, of the function. Themajor impartial is to enhance the hyper parameter based on an LSTM classifier ISSA optimization algorithm to originate efficient classification of the microarray features. The parameters in this scope are batch size and concealed neurons. The ISSA are starts from the earlyexplanations that are initialized the hyper parameter, in that which is arbitrarily created and then tries to enhance the accurateness of the emotionorganization model. Moreover, the LSTM system's fitness is responsible for executing the evaluation and returning the classification function's accuracy. In general, SSA aims at obtaining a swarm of salp features, and each salp fluctuates continuously, known as the salp chain. Each salp in a swarm is a representation of an agent, which conducts search operations on a particular improvement issue. Additionally, salping swarms are based on major categories of the process, namely, those that lead and those that follow. During search creation, follower salps follow the leading salps to allocate the best solution. The following swarm obtaining of sapls in signified as follows equation (9):

$$S = \begin{bmatrix} x_{11} & x_{12} & \cdots x_{1D} \\ a_{21} & a_{22} & \cdots x_{2D} \\ \vdots & \vdots & \vdots \\ a_{N1} & a_{N2} & \cdots x_{ND} \end{bmatrix} \quad (9)$$

Additionally, the leading salps are updating their states to the floating with the help of Equation (10)

$$L_j = \begin{matrix} \{F_j + r_1 \times (1b + r_2 \times (ub - 1b)) r_3 \geq 0.5 \} \\ \{F_j - r_1 \times (1b + r_2 \times (ub - 1b)) r_3 < 0.5 \} \end{matrix} \quad (10)$$

In there, $L_j$and $F_j$are the j^{th} correlate for the state of the leading sapls and the food sources as well, ub and $1b$ are the higher and lesser boundary bounds for the elucidation range of the function. $r_2$and r_3 are the random figures among 0 and 1 also r_1 is a adjustable become decreases while the iteration are increase. The iteration formula are shown in equation (11)

$$r_1 = 2 \times \exp\left[-(\tfrac{4t}{T})^2\right] \quad (11)$$

Where t and T are the current iteration and maximum iteration of the function. in the follower salps update their states are shown in equation (12)

$$x_{i,j} = \tfrac{1}{2} a t^2 + v_0 t \quad (12)$$

In this $a = v_f - v_0/\delta t$and$v_0 = (x - x^0/t$. In this function time was referred as the optimization process therefore, the discrepancy among the iteration in 1 and the equation (12) are becomes the equation (13).

$$X_{i,j} = \tfrac{1}{2}(X_{i,j} + X_{i-1,j}) \quad (13)$$

In there the cost of i was more than 1 as well $X_{i,j}$and $X_{i-1,j}$ are significant the j^{th} correlate of the follower salps$i$ and $i-1$ are correspondingly. Moreover, this SSA was the first initialize salps in a swarm resolution range as well as followers and leaders of salps are update their states to re-position at better positions and states of

the function. In some cases, the current SSA may trap at suboptimal resolutions fairly easily due to features such as fast convergence speed and ease of application. Therefore, the iteration among the leading and followers salps are required to enhance the characterize the performance of SSA. To explore the resolution range are more effectual and effectively prevent the local solution into the sub optimal solution they required the strategy known as mutation into SSA of the function. Furthermore, in the SSA optimization function, three different mutation schemas were performed: the Cauchi SSA mutation, the Gaussian SSA mutation, and the Levitation SSA mutation, which are entrenched in the SSA optimization of the system. In this function before smearing the mutation schema the greedy search method is assumed among the state of shown as equation (14):

$$Y^{t+1} = \begin{cases} X^t \; if \; f(X^t) < f(X^{t+1}) \\ X^{t+1} \; if \; f(X^{t+1}) < f(X^t) \end{cases} \quad (14)$$

The state of X^t of t^{th} iteration and X^{t+1} of $(t+1)$th iteration of the function. Moreover, once the completed the greedy search to the corresponding salps the mutation schema is applied along with mutation rate (m_r). In this section the levy-flight SSA is used to enhance the salps variety in the SSA optimization function. Levy-flight mutations can handle the global search solution more efficiently by mutating the salps as well. Levy-SSA mutates equation (15) as:

$$\hat{x_i} = x_i \times (1 + Levy\,(\delta)) \quad (15)$$

Where $x_i$represent the i^{th} salps and $Levy\,(\delta)$ was the random amountproduced by using the levy distributed function. A general simplified form of the levy distributed as represented as equation (16)

$$Levy\,(\beta) \sim y = t^{-\beta-1}, \quad 0 < \beta \leq 1 \quad (16)$$

In thereβ is the constancy of the index. Additionally, the levy- distributed are obtaining the random number by using the equation (17) as

$$Levy\,(\beta) \sim \frac{\Psi \times u}{|v^{1/\beta}|} \quad (17)$$

Where $u \; and \; v$ are the normaldissemination function and the value of the Ψ as defined as the equation (18) as

$$\Psi = \left[\frac{\Gamma(1+\beta)\sin(\pi\frac{\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{(\beta-1)/2}}\right] 1/\beta \quad (18)$$

In this case the value of the β was 1.5 fixed usually, the levy-flight mutation was creates various offspring salps as it long tailed delivery. Therefore, this feature is helpful to initialize to select and extract the effective aspects to the microarray classification function.

3.5 LSTM classification:

After initialized the classification features Long Short Term Memory was used to classify the emotional classification of microarray cancer database. In this work, we applied the LSTM with ISSA optimization algorithm.In whichwas developed for solving difficulties of classification and regression inquiry training techniques based on supervised learning various categories of figures from various sample data. The model uses a multi-class classification problem with non-lined and lined classifications of data. An LSTM An LSTM develops multiple hyperplanes in multidimensional space, and the optimal hyperplane divides various classes optimally. Kernel functions in non-linear classification maximize. From the increasing interest in LSTMs, researchers have developed many highly promising submissions. In pattern recognition and image processing, LSTM models have been widely used. The LSTM architecture of the proposed model is also applied to the process of classifying genes, and the RBF is used to apply the LSTM model to the proposed model.

4. Result and Discussion:

The outcomes of the proposed ISSA-LSTMmethod using the dataset was generated above in this section. The implementation and simulation of the ISSA-LSTM method are done using MATLAB R2022 software. This research used an i5 processor with 8GB of RAM.

All analysis and assessmenttrials are performed on public field databases, widely used in the related works, namely Human microarray Lung cancer database and output limits are utilized to construct the results of the function. The dataset used to analyze the ISSA-LSTMmethod and it performs best with LSTM classifiers where 80% of testing and 20% of testing the function. In that, the data is randomly taken for training and testing according to the iterations.The performance of the ISSA-LSTMmethod is analyzed by means is accuracy, Precision, Recall,F-measure and MCC which are expressed in equations (19) - (23).

Accuracy

It is easiest to understand the accuracy performance metric as the ratio of correctly predicted observations to all observations.

Equation shows the equation for accuracy (19)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (19)$$

Precision

In terms of positive observations, precision measures how often explanations are correctly predicted.

Equation shows the precision equation (20)

$$Precision = \frac{TP}{TP+FP} \times 100$$

(20)

Recall

By classifying it as positive (true positive), recall determines how many actual positives are captured by our system. The Recall equation is (21)

$$Recall = \frac{TP}{TP+FN} \times 100$$

(21)

F- Measure

The F- measure is calculated by averaging Precision and Recall. Consequently, it takes into account both false positives and false negatives when calculating this score. Equation (22) shows the equation for F1-score.

$$F - Score = \frac{2TP}{2TP+FP+FN} \times 100 \quad (22)$$

MCC

The MCC is calculating the best measures to evaluate as the correlation coefficient among the predicted value and true value.

The equation for MCC is shown in the equation (23)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

(23)

4.1 Performance analysis of ISSA-LSTM method:

This section shows the performance analysis of gene lung classification using the Human gene micro array dataset. In micro array gene dataset, the ISSA optimization with LSTM based feature dimension reduction achieves the recognition accuracy of 99.50% whereas the individual features such as Cuckoo Search Crossover, Differential Evolution, Artificial Bee Colony and skeleton achieves 89.44%, 91.38%, 90.76% . According to the results, combining the features improves classification accuracy. Additionally, the performance of ISSA-LSTMmethod is improved by using feature dimension reduction. The LSTM is used by the ISSA-LSTMmethod for classifying the microarray gene lung cancer.

In this section, the performance of the ISSA-LSTMmethod is analyzed with different classifiers and with different feature dimension reduction approaches. The different classifiers used to evaluate the ISSA-LSTMmethod are Recurrent Neural Network (RNN), Neural Network (NN), and Generative Adversarial Network (GAN). Moreover, the performance evaluation of proposed LSTM with different classifiers are analyzed for with and without feature selection which is shown in Table 1. Figure 2shows the graphical comparison of the LSTM withoutfeature selection with different classifiers where the Table 2 and Figure 2shows is for classifiers with SSAE. The analysis concludes that the LSTM is effective with feature selection achieves higher classification accuracy 99.73% than the LSTM without feature selection. Moreover, the LSTM with feature selection provides better performance than the other classifiers such as RNN, NN and GAN with and without feature selection. For example, the LSTM with feature selection achieves 99.73%, whereas the feature selection with RNN obtains 89.05%, NN obtains 93.76% and GAN obtains 95.22%. The LSTM with the capacity of handling high dimension spaces and optimal feature selection using ISSA-LSTMis used to improve the classification of micro array lung cancer classification.

Table 1. Performance analysis of different classifiers in ISSA-LSTM method without feature selection.

Different classifier	Accuracy	Specificity	Recall	F1- Score	MCC
RNN	83.05	82.216	83.445	79.627	80.345
NN	73.76	81.1625	84.545	79.2587	81.26535
GAN	85.22	81.5489	81.65445	82.27	82.64596
LSTM	86.43	87.25	87.445	87.67	86.345

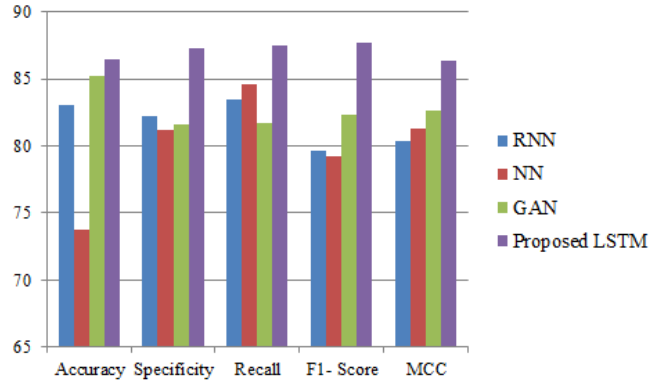

Fig 2. Graphical comparison of the different classifiers without feature selection

Table 2 shows how well ISSA-LSTM works when features are chosen and different ways of reducing feature dimensions are used. Figure 3 compares LSTM with different classifiers. According to the analysis, ISSA-QAL provides a higher classification accuracy of 99.73% than RNN, NN, and GAN. Since the ISSA-QAL selects its features optimally, it provides higher accuracy.

Table 2. Performance analysis of different classifiers in ISSA-LSTM method with feature selection.

Different classifiers	Accuracy	Specificity	Recall	F1- Score	MCC
RNN	89.05	92.216	93.445	89.627	90.345
NN	93.76	91.1625	94.545	89.2587	91.26535
GAN	95.22	91.5489	91.65445	96.27	92.64596
LSTM	99.73	97.25	98.445	98.67	98.345

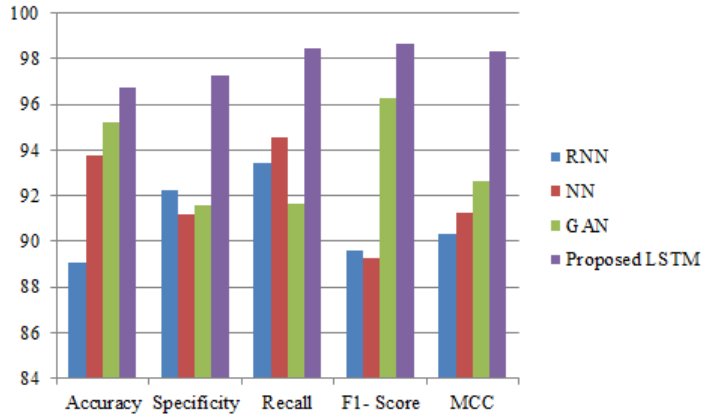

Fig 3. Graphical comparison of the different classifiers with feature selection

Table 3 shows the performance analysis of ISSA-LSTM with feature selection and different optimization algorithms. Figure 4 illustrates a graphical comparison between ISSA and various optimizations. Based on the results of the analysis, we conclude that the proposed ISSA-LSTM provides a higher classification accuracy of 99.54 % than the GA and ABC approaches. Due to the optimal feature selection, the ISSA-LSTM provides higher emotional classification.

Table 3. Performance analysis of different feature selection in ISSA-LSTM method with different optimization algorithm.

Different feature optimization	Accuracy	Specificity	Recall	F1- Score	MCC
GA	93.05	92.216	93.445	93.627	91.345
DE	93.76	91.1625	94.545	.91.2587	91.26535
ABC	91.22	91.5489	91.65445	94.27	92.64596
ISSA	99.5456	97.25	97.445	98.67	96.345

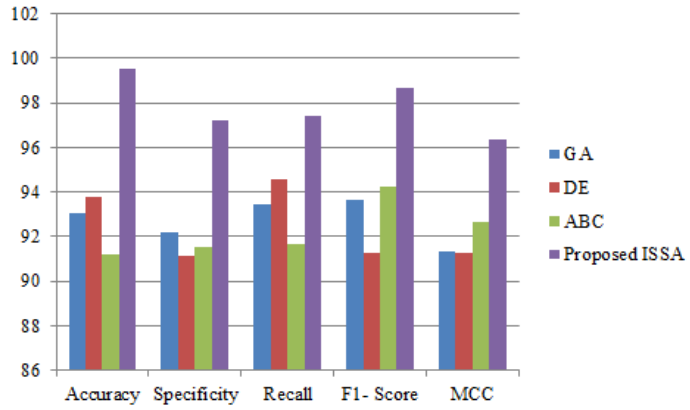

Fig 4. Graphical comparison of the ISSA-LSTM with different optimization algorithm.

4.2 Comparative analysis:

This section shows the comparative analysis of ISSA-LSTM for the human micro array gene cancer classification. The dataset used for the Lung cancer classification is micro array lung dataset in where the performance of ISSA- LSTMis compared to the CSC – Gene [16], ANFIS- SA [18] and L1- LSTM [20]. The comparative analysis between the IAAS-LSTM with others are shows in Table 4and Figure 4. From the analysis, it is known that the ISSA-LSTMprovides better accuracy performance than the CSC – Gene [16], ANFIS- SA [18] and L1- LSTM [20]. For example, the accuracy of the ISSA – QAL is 99.54% whereas the accuracy of CSC – Gene is 96.98%, ANFIS- SA [18] is 96.28% and L1- LSTM [20] is 93.57%. The ISSA – QAL obtains less accuracy, because it processes all the hybrid features extracted during the classification. The elimination of redundant features is used to improve the cancer classification of the human micro array database.

Table 4. Comparative analysis of ISSA-LSTM with other optimization methods.

Different methods	Accuracy	Specificity	Recall	F1- Score	MCC
CSC- Gene [16]	96.98	NA	NA	NA	NA
ANFIS-SA [18]	96.28	97.2	96.35	96.15	91.25
L1-LSTM [20]	93.57	94.312	94.215	94.365	92.64
Proposed ISSA-LSTM	99.5456	97.25	97.445	98.67	96.345

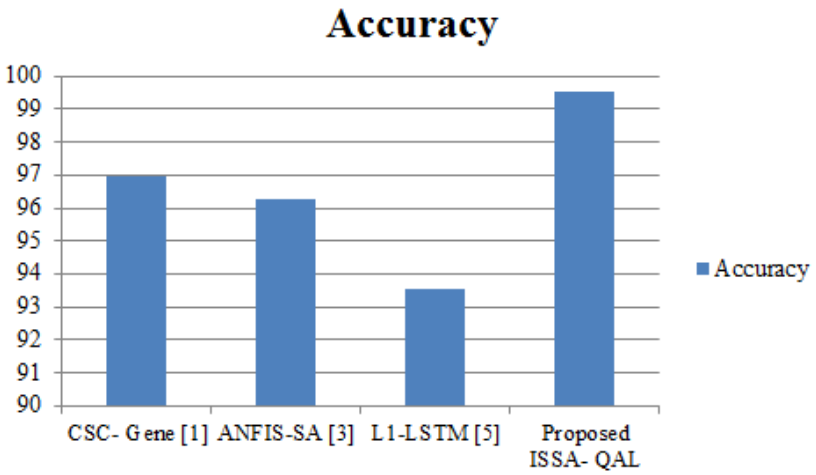

Fig 5. Graphical comparison of the ISSA-LSTM accuracy with different optimization algorithm.

5. Conclusion

A large number of gene information makes lung cancer prediction a difficult task. Local optima traps, lower convergence, and overfitting limit existing gene selection methods. The QAL method improves exploration and exploitation processes by getting around local optima traps and slower convergence rates. Quantum bits facilitate exploration and help escape local optima traps in quantum search. In order to maximize its exploitation and convergence rate, a QAL model is selected based on its fitness function. Additionally, the ISSA-LSTM method improves the exploration and exploitation of relevant gene classification features. In high-dimensional data sets, LSTM is an efficient classifier. In gene classification, the LSTM method has 99.72% accuracy and the ISSA-LSTM method has 99.53 % accuracy. This research will apply the hybrid method to improve gene classification efficiency in the future.

Reference:

[1] Ekanayake, A., Madegedara, D., Chandrasekharan, V. and Magana-Arachchi, D., 2020. Respiratory bacterial microbiota and individual bacterial variability in lung cancer and bronchiectasis patients. *Indian journal of microbiology*, *60*(2), pp.196-205.

[2] Jang, H.J., Lee, H.S., Ramos, D., Park, I.K., Kang, C.H., Burt, B.M. and Kim, Y.T., 2020. Transcriptome-based molecular subtyping of non–small cell lung cancer may predict response to immune checkpoint inhibitors. *The Journal of Thoracic and Cardiovascular Surgery*, *159*(4), pp.1598-1610.

[3] Thakur, T., Batra, I., Luthra, M., Vimal, S., Dhiman, G., Malik, A. and Shabaz, M., 2021. Gene expression-assisted cancer prediction techniques. *Journal of Healthcare Engineering*, *2021*.

[4] Kochan, N., Tütüncü, G.Y. and Giner, G., 2021. A new local covariance matrix estimation for the classification of gene expression profiles in high dimensional RNA-Seq data. *Expert Systems with Applications*, *167*, p.114200.

[5] Dash, R., 2021. An adaptive harmony search approach for gene selection and classification of high dimensional medical data. *Journal of King Saud University-Computer and Information Sciences*, *33*(2), pp.195-207.

[6] Mandal, M., Singh, P.K., Ijaz, M.F., Shafi, J. and Sarkar, R., 2021. A tri-stage wrapper-filter feature selection framework for disease classification. *Sensors*, *21*(16), p.5571.

[7] Dwivedi, A.K., 2018. Artificial neural network model for effective cancer classification using microarray gene expression data. *Neural Comp*

[8] Sayed, S., Nassef, M., Badr, A. and Farag, I., 2019. A nested genetic algorithm for feature selection in high-dimensional cancer microarray datasets. *Expert Systems with Applications*, *121*, pp.233-243.

[9] Nautiyal, B., Prakash, R., Vimal, V., Liang, G. and Chen, H., 2021. Improved Salp Swarm Algorithm with mutation schemes for solving global optimization and engineering problems. *Engineering with Computers*, pp.1-23.

[10] Ibrahim, R.A., Ewees, A.A., Oliva, D., Abd Elaziz, M. and Lu, S., 2019. Improved salp swarm algorithm based on particle swarm optimization for feature selection. *Journal of Ambient Intelligence and Humanized Computing*, *10*(8), pp.3155-3169.

[11] Algamal, Z.Y. and Lee, M.H., 2019. A two-stage sparse logistic regression for optimal gene selection in high-dimensional microarray data classification. *Advances in data analysis and classification*, *13*(3), pp.753-771

[12] Dalwinder, S., Birmohan, S. and Manpreet, K., 2020. Simultaneous feature weighting and

parameter determination of neural networks using ant lion optimization for the classification of breast cancer. *Biocybernetics and Biomedical Engineering*, *40*(1), pp.337-351.

[13] Nakisa, B., Rastgoo, M.N., Rakotonirainy, A., Maire, F. and Chandran, V., 2018. Long short term memory hyperparameter optimization for a neural network based emotion recognition framework. *IEEE Access*, *6*, pp.49325-49338.

[14] Le, D.N., Parvathy, V.S., Gupta, D., Khanna, A., Rodrigues, J.J. and Shankar, K., 2021. IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification. *International journal of machine learning and cybernetics*, *12*(11), pp.3235-3248.

[15] Sharifrazi, D., Alizadehsani, R., Roshanzamir, M., Joloudari, J.H., Shoeibi, A., Jafari, M., Hussain, S., Sani, Z.A., Hasanzadeh, F., Khozeimeh, F. and Khosravi, A., 2021. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomedical Signal Processing and Control*, *68*, p.102622.

[16] Sampathkumar, A., Rastogi, R., Arukonda, S., Shankar, A., Kautish, S. and Sivaram, M., 2020. An efficient hybrid methodology for detection of cancer-causing gene using CSC for micro array data. *Journal of Ambient Intelligence and Humanized Computing*, *11*(11), pp.4743-4751.

[17] He, F., Huang, L., Xu, Q., Xiong, W., Liu, S., Yang, H., Lu, W., Xiao, R., Hu, Z. and Cai, L., 2020. Microarray profiling of differentially expressed lncRNAs and mRNAs in lung adenocarcinomas and bioinformatics analysis. *Cancer Medicine*, *9*(20), pp.7717-7728.

[18] Haznedar, B., Arslan, M.T. and Kalinli, A., 2021. Optimizing ANFIS using simulated annealing algorithm for classification of microarray gene expression cancer data. *Medical & Biological Engineering & Computing*, *59*(3), pp.497-509.

[19] Chaudhuri, A. and Sahu, T.P., 2021. A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification. *Computers & Electrical Engineering*, *90*, p.106963.

[20] Shekar, B.H. and Dagnew, G., 2020. L1-regulated feature selection and classification of microarray cancer data using deep learning. In *Proceedings of 3rd international conference on computer vision and image processing* (pp. 227-242). Springer, Singapore.