Measurement: Sensors 30 (2023) 100902

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Lung cancer classification based on enhanced deep learning using gene expression data

V. Yuvaraj [a,b,*], D. Maheswari [c]

[a] Department of Computer Science, Rathnavel Subramaniam (RVS)College of Arts and Science, Sulur, Coimbatore, India
[b] Department of Computer Applications, Dr.N.G.P. Arts and Science, Coimbatore, 641048, India
[c] Head and Research Coordinator, PG Department of Computer Science, Rathnavel Subramaniam (RVS) College of Arts and Science, Sulur, Coimbatore, India

ARTICLE INFO

Keywords:
Lung cancer
Classification
Gene selection
Microarray technology and normalization

ABSTRACT

Lung cancer is one among the life threatening of cancer globally. The estimation of the World Cancer Research Fund International that in 2022, this disease has recorded the diagnosis of 1.8 million fresh cases of this disease. Medical professionals may safely and effectively treat the patient when they make a proactive diagnosis and classification of their condition. The availability of microarray technology has paved the way for investigations of genes and their relationships to many illnesses, including lung cancers. Prior studies have suggested gene selections using IWOA (Improved Whale Optimization Algorithm). After that, an MLSTM (Modified Long Short-Term Memory) Network is taken into account for the categorization of lung cancer. Nevertheless, because the input data is represented in different scales, LSTMs are prone to overfitting and the salient trait (with a decreased scale) may become useless as a result of other features having values on a higher scale. To get over those problems in this research, introduce an enhanced model for lung cancer classification. In this work, initially, data pre-processing is performed to normalization the data scale using min max normalization. Gene selection is done by using IWOA. Finally, ECNN (Enhanced convolution neural networks) is considered for classifying lung cancer. The experimental results are executed on Matlab 2013b using the Kent Ridge Bio-Medical Dataset, the proposed ICNN framework in this study was contrasted with the pre-existing MIMOGA, SMO, and MLSTM techniques. The results of this work's proposed model demonstrate its effectiveness when evaluated using performance measures such as values for recall, precision, accuracy, and f-measure.

1. Introduction

Recently, there has been a rise in the counts of lung cancer deaths as per the reports from various medical organizations. In comparison with the onset of the century, there has been a huge increase in the death numbers of cancer [1]. Existing therapies like chemotherapies, radiations, and surgeries, in spite of their outcomes have been found to be wanting. Accurate categorization of the various categories of lung cancer is essential for the efficacy of therapy and the reduction of its adverse effects on patients. With the use of microarray technology, millions of genes may be studied to glean important details about how cells work. These data can be used to diagnose and predict cancer [2].

It is imperative to create good quality approaches which select significant gene subsets mainly due to gene data characteristics like smaller sample sizes, higher noise rates, and greater dimensions), useful in better categorising cancer. This method helps doctors categorize a smaller sample of physiologically important gene-based tumours and highlights a few specialised genes for the construction of low-cost trials in addition to being able to reduce computing costs [3,4], along with the added advantage of early diagnostics of cancers and their related therapies. Moreover, studies have proved the usefulness of biological historical information in statistical evaluations in terms of identifying unique biological characteristics while eliminating unclear relationships amongst them [5,6].

Also, the past information can be utilized to check the consistency for the current knowledge, in addition to experimental data verification to substitute the probable gaps or include additional information. In the last few years, several gene expression profiles have been shown to accurately differentiate distinct tumour subtype subcategories. Researchers have discovered that using biological data meaningfully prevents one-sided outcomes from separate tests and does not merely successfully reduce or decrease noise in gene chips. Nevertheless, the

* Corresponding author. Department of Computer Science, Rathnavel Subramaniam (RVS)College of Arts and Science, Sulur, Coimbatore, India.
E-mail addresses: yuvarajvelliangiri@gmail.com (V. Yuvaraj), maheswari@rvsgroup.com (D. Maheswari).

https://doi.org/10.1016/j.measen.2023.100902
Received 10 April 2023; Received in revised form 16 August 2023; Accepted 27 September 2023
Available online 9 October 2023
2665-9174/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

significance of cancer data from earlier classifications has not been thoroughly addressed. Neural network systems can be useful in dealing with above stated concerns [7].

The effectiveness of ANNs' depend on their learning capacities. A popular neural network is the MLP (Multi-Layer Perceptron) which is a conventional feed forward mechanism and uses supervised training. The other well-known supervised techniques in addition to MLPs are gradient-based and stochastic algorithms. Back propagations and their variants are frequently used by academics as standard examples of gradient-based methods, in spite of a few drawbacks of gradient-based algorithms, such as their propensity to only find local minima, their slowness, and their increasing reliance on the starting parameters.

To avoid this problem in existing work introduced an effective gene selection model that depends on IWOA. Next, a MLSTM Network is considered for lung cancer classification. However, input data are presented in various scales and it may make a pivotal attribute (with a reduced scale) to be less effective due to other attributes holding values on a higher scale and LSTMs are prone to overfitting. To get over those problems in this research introduce an enhanced model for lung cancer classification. In this work, initially, data pre-processing is performed to normalization the data scale using min max normalization. Gene selection is done by using IWOA. Finally, ECNN (Enhanced convolution neural network) is applied for lung cancer classification.

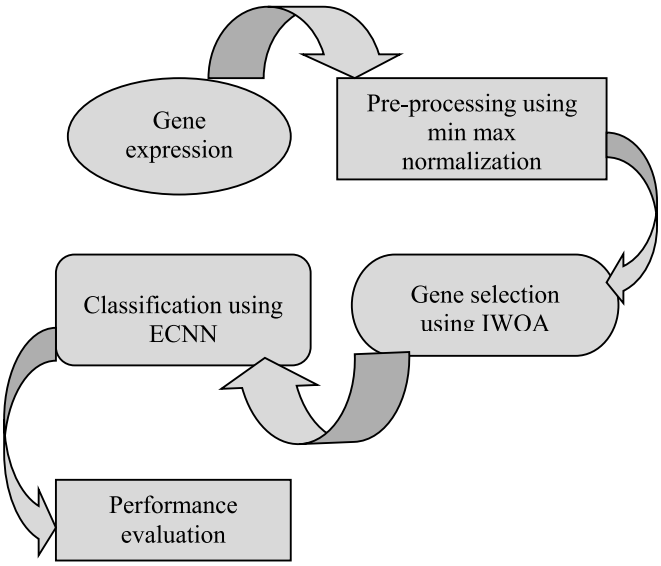

Fig. 1. Typical block diagram of the proposed model.

2. Literature review

Arunkumar and Ramakrishnan [2018] [8] proposed a similarity metric that chooses the characteristics using a fuzzy rough fast reduction approach. The suggested fuzzy rough fast reduction approach specifies a tailored similarity measure to identify lowest counts of beneficial genes in initial stages. Entropies based on IG (Information Gain) assist in reducing dimensionalities in the next stage and subsequently RF (random forest) classifier is utilized to examine gene expression datasets for identifying lung/ovarian cancers and leukaemia. The suggested technique successfully achieves classification accuracies of 99.45%/99.6%, and 97.22%, respectively. R, an open source software package is used to demonstrate the technique where performance metric values including recall, precision, accuracy, f-measure and regions of characteristic show better outcomes in comparison with the existing techniques in classifications.

Rafii and Hassani [2015] [9] proposed usage of MLPs for classifying lung cancers from microarray data. The study used PCA to first reduce data's dimensions and subsequently MLP networks (feed-forward neural networks) classified instances. Their assessments of classification accuracies on lung cancer gene expression datasets, demonstrated the effectiveness of their proposed technique. The study also portrayed comparative values of other techniques using the same dataset.

Azzawi et al. [2018] [10] proposed enhancements to MLP neural network predictions by using IMPSO (improved Particle Swarm Optimization). The study calculated weights/biases of MLPs using their IMPSO. The study enhanced accuracies of classifications in predicting lung cancers from gene expression data using their discriminant schema namely MLP-IMPSO. Using actual microarray lung cancer datasets, the recommended model's prediction accuracy was thoroughly examined and compared with other machine learning techniques including neural networks based on radial basis functions, C4.5, SVM (support vector machine), MLPs, and NB (Naive Bayes). The reliability of their evaluations were improved using data cross-validations. The study found enhanced performances of their suggested approach when prior information was considered. Success has been demonstrated, demonstrating how the method enhances the accuracy of lung cancer diagnosis using historical biological data. The results of the study make it clear that the suggested strategy is successful in identifying lung cancer.

Salem et al. [2015] [11] introduced a unique method for categorising human cancer illnesses according on gene expression characteristics. The suggested method combines DGA (Deep Genetic Algorithm) with IG where initial selections were based on IG followed by the usage of GA (genetic Algorithm) for reducing feature counts. In the final stages, GP (genetic Programming) categorizes cancer types. The study evaluated their schema along with existing techniques on seven different cancer datasets for satisfactory results.

Li et al. [2018] [12] proposed classifications of lung cancer gene expression data using sparse overlapped group lasso penalty and adaptive multinomial regressions. The study offered a biologically comprehensible overlapping grouping approach that emphasized the importance of gene groups from underrepresented classes. The significance of gene within groups were evaluated by conditional mutual information, and data-oriented weights. A regularised adaptive multinomial regression is introduced in accordance with the grouping mechanism and the built weights, and the solving algorithm is designed. This method can be used to carry out multi-class classification by selecting the significant gene groups for each class, and it can also assist in adaptive selection of significant genes within each group. It was proven from the results of experiments that the performance achieved by the scheme was considerably enhanced when compared to other 6 techniques in terms of performance metrics in classifications including accuracies, and selected genes are genes responsible for lung cancer incidences.

Anand and Bansal [2021] [13] suggested techniques for choosing best gene subsets that were extremely important for more accurate predictions. The Kent Ridge Biomedical Dataset Repository has provided the data on lung cancer gene expression. The goal of the survey is to increase AUC ROC and F-measure values so that the model can be appropriately evaluated while working with an unbalanced dataset.

Singh and Susan [2022] [14] advocated usage of FMM (Fuzzy Min-Max) classifiers to categorize lung cancer gene expression microarray data which are generally ignored due to their higher dimensions and processing costs. LASSO (Least Absolute Shrinkage and Selection Operator), improved FMM classifier's accuracy and speed in selecting the best gene subsets for classifying lung cancers. The study also compared the performances of their FMM-LASSO with SVM, RF, KNN (K-nearest Neighbour), NB, and LR (Logistic Regression) with and without LASSO. Their outcomes showed that their FMM-LASSO achieved better outcomes than other techniques.

3. Proposed methodology

This section studies about the proposed model extensively. Proposed model consist of three phases first one is pre-processing using min max

normalization, second phase is gene selection dependent on IWOA and third one is ECNN based lung cancer classification. General deign of the proposed schema is depicted as Fig. 1.

3.1. Pre-processing using min-max normalization

Min-max normalization is a simple technique that rescales the data values to a range between 0 and 1, using the minimum and maximum values of the original data. This technique preserves the relative order and distance of the data points, but it also reduces the variance and magnifies the effect of outliers.

Input features of gene data might have present with different scale that can result in incorrect results and to prevent these problems, it is necessary the scale of data be normalized. In this research, Min -max Normalization model is used and the process of normalization involves the translation of numerical values into another range applying a mathematical function. In this proposed study, min-max normalization is helpful in lung dataset normalization [15,16]. Min-max normalization is one among the predominant means of normalizing data. The normalization of the values present in the dataset are performed well within the particular range of the minimum and maximum value from dataset and all the values are replaced in accordance with the expression below.

$$v' = \frac{v - \min_A}{\max_A - \min_A}(new\_max_A - new\_min_A) + new\_min_A \quad (1)$$

Where,

A - Attribute data,

Min(A), Max(A) signifies minimum and maximum absolute value of A correspondingly.

v' indicates the new value of every entry in data

v signifies the Older value of every entry in data.

New_max(A), new_min(A) stand for the max and min value of the range (i.e boundary value of range needed) correspondingly.

3.2. Gene selection based on IWOA

Following normalization, it is necessary to choose the best genes as only a tiny part of gene expression data inputs exhibited differential patterns for different groups or samples. IWOA used in this work assisted in proper gene selections. WOA methods are based on population dynamics and influenced by nature. They use a technique known as "bubble-net hunting" to find preys and imitate humpback whale behaviours. WOAs use stochastic optimizations through their search agents which assist in finding solutions to optimization issues. The three basic steps of WOA include surrounding preys, attacks with bubble nets, and keeping an eye on best preys.

The core mathematical expression of the WOA is given in Eqs. (2) and (3).

$$X(t+1) = X^*(t) - A.|C.X^*(t) - X(t)| \text{ if } p < 0.5 \quad (2)$$

$$X(t+1)|C.X^*(t) - X(t)|.\, e^{bl}\cos(2\pi t) + X^*(t) \text{ if } p \geq 0.5 \quad (3)$$

Fixed values specifying shapes of logarithmic spirals, b, are initialised to 1 for linear reductions of coefficient vectors a, such that A = 2a. (r-a); C = 2.r, from 2 to 0 throughout iterations which is the optimal solution, the output by X*. X* refers to vectors that represents positions of whales. Values between 0 and 1 are allocated to the random vector, r. As the locations of the whales are updated, any random number between −1 and 1 is utilized to switch between equations (2) and (3); in equations (2) and (3), probabilities are 50% and 50%, correspondingly, meaning the whales choose either path at random and at the same rate during the optimization stage. During the bubble-net phase, vector A can have any random value between [-1, 1], but during the search phase, it can also have a value above or below 1. Eq (4) displays the search technique in action

$$X(t+1) = X_{rand} - A.|C.X_{rand} - X(t)| \quad (4)$$

The WOA algorithm carries out a global search by using this random search approach. At first, random solutions are produced when the WOA search begins. Next, the technique is utilized for revising these solutions, iteratively. Till a predefined highest count of iterations is attained, the search will go on.

3.2.1. Improved Whale Optimization Algorithm

It is necessary for an optimization algorithm to have a right trade-off between exploration and exploitation to yield effective results. Hence, there is a reduction in the step size with the increment in iterations in WOA. Parameter A controls this step size. But, it has been found that during the iterations of WOA, poor divergence makes the traps to get confined to the local optima.

To address these issues, this work makes use of IWOA. The levy fly function can be used to change the value of A. As a result, WOA's exploration and exploitation talents are simultaneously strengthened.

The mathematical expression of the Levy distribution is as given.

$$L(s,\gamma,\mu) = \begin{cases} \sqrt{\frac{\gamma}{2\pi}}\exp\left[-\frac{\gamma}{2(s-\mu)}\right]\frac{1}{(s-\mu)^{3/2}} & \text{if } 0 < \mu < \infty \\ 0 \quad \text{if } s \leq 0 \end{cases} \quad (5)$$

Here μ, γ, and s represent the position parameter, the scale parameter that manages the scale of distribution and the set of samples, correspondingly.

3.2.2. Algorithm for improved whale optimization

START

1. Import data
2. Initialize the positions of the whale population X
3. compute the fitness of each whale
4. Initialize a and r, Compute A and C
5. Determine X* to be the location of the best hunter whale
6. Set $t = 1$
7. while $t \leq$ max iterations do
8. for each hunting whale, do
9. if $p < 0.5$
10. if $|A| < 1$
11. update the location of the current hunting whale applying (2)
12. else if $|A| \geq 1$
13. randomly select multiple search agents
14. update the position of the current hunting whale applying (4)
15. end if
16. else if $p \geq 0.5$
17. upgrade the location of the present hunting whale applying (3)
18. end if
19. end for
20. upgrade X* if there the best solution exists
21. $t = t + 1$
22. end while
23. output X*

END

3.3. Lung cancer classification using ECNN

Once feature selection is done, it is important to classify the selected features. In this study, Improved CNN (Convolution Neural Networks) used for lung cancer prediction [17,18]. A CNN and a common artificial neural network are structurally diverse. In contrast to traditional ANN that spreads out the input onto a vector, the layers in a CNN are selected such that they have a spatial match with the input data. CNNs are made up of one or more independent or several completely linked layers, one

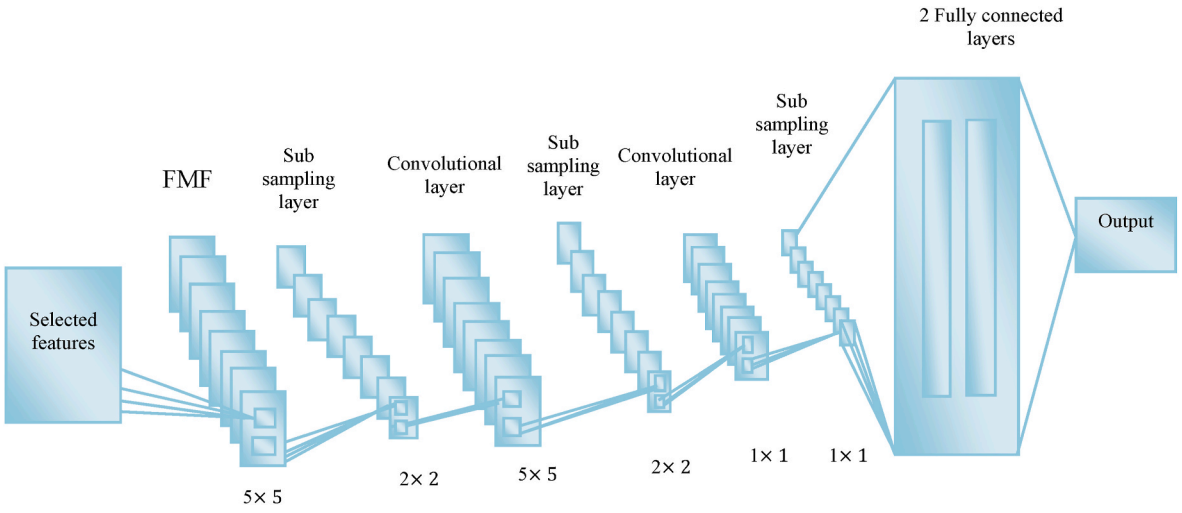

Fig. 2. CNN architecture.

or more counts of convolutions, sub-sampling layer sequences, and output layers [19,20].

3.3.1. Drawbacks of traditional CNN

In the traditional model of CNN, convolution kernel is utilized for feature mapping which at times leads to information loss. To get over this problem, in this study, fuzzy membership function for input layer is brought into use.

3.3.2. Improved CNN

CNNs encompass Convolution, sub-sampling, and fully linked layers. The general block diagram of a CNN is shown in Fig. 2. The sections that follow offer a brief summary of each type of layer.

3.3.3. Convolution layer

In this phase of convolutions, input features and kernels (filters) are convolved together. The outcomes of convolutions of input features and kernels are utilized to produce output feature maps with sizes i*i where kernels belonging to convolution matrices are typically called filters.

In this layer Fuzzy Membership Function (FMF) is used for feature mapping which can be defined as ($w_1 = 0.3, w_2 = 0.4, w_3 = 0.5, w_4 = 0.7$) and will be calculated as follows

$$o^2 = u_i^{(j)}\left(a_i^{(2)}\right) \tag{6}$$

Where $\boldsymbol{u}_i^{(j)}(.)$ is a membership function $\boldsymbol{u}_i^{(j)}(.) : \boldsymbol{R} \longrightarrow [\boldsymbol{0}, \boldsymbol{1}]$, i = 1,2, …,M, j = 1,2, ….,N. Using Gaussian membership function.

The output obtained from the *l*-th convolution layer, represented as $C_i^{(l)}$, includes feature maps. It is established as

$$C_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{a_i^{(l-1)}} K_{i,j}^{(l-1)} * C_j^{(l)} \tag{7}$$

Where, $B_i^{(l)}$ indicates the bias matrix and $K_{ij}^{(l-1)}$ stands for convolution filter or kernel having size *a*a* which provides the connection between the *j*-th feature map in layer $(l-1)$ and the *i*-th feature map present in the same layer.

The output $C_i^{(l)}$ layer includes feature maps. In (10), the first convolution layer $C_i^{(l-1)}$ forms the input space, that is, $C_i^{(0)} = X_i$. The feature map is produced by the kernel. After convolutions, activation functions are used in layers for non-linear conversions of convolution layer outputs.

$$Y_i^{(l)} = Y\left(C_i^{(l)}\right) \tag{8}$$

Where, $Y_i^{(l)}$ refers to the output that the activation function yields and $C_i^{(l)}$ indicates the input received by it.

3.3.4. Sub sampling or pooling layer

The preceding convolution layer helps in extracting this layer's principal goal, which is to geographically minimise the dimensionality of the features map. The feature maps are used as the mask, and the sub sampling function is used between them. Among the novel sub sampling techniques that were introduced were average pooling, sum pooling, and maximum pooling. The maximal pooling, which generates the matching output feature from the greatest value in each block, is the dominant pooling. Because it enables the convolution layer endure rotation and conversion among the inputs, a sub sampling layer must be taken into consideration.

3.3.5. Fully connected layer

The CNN's last layer is a conventional feed forward network that has single or many hidden layers. In the output layer, SoftMax activation function is used:

$$Y_i^{(l)} = f\left(z_i^{(l)}\right), \tag{9}$$

$$\text{Where } z_i^{(l)} = \sum_{i=1}^{m_i^{(l-1)}} w_{i,j}^{(l)} y_i^{(l-1)} \tag{10}$$

Where, $w_{ij}^{(l)}$ indicate the weights which the entire fully connected layer has to tune so that every class can be represented and *f* indicates the transfer function specifying it being nonlinear. It must be noted that the fully **linked** layer's neurons include the nonlinear property, not single layers like convolutions and pooling layers.

4. Results and discussion

Here, the experimental results of the suggested model tests executed on Matlab 2013b. Using the Kent Ridge Bio-Medical Dataset, the proposed ICNN framework in this study was contrasted with the pre-existing MIMOGA, SMO, and MLSTM models according to metrics such as recall, precision, accuracy, and f-measure. The Kent Ridge Bio-Medical Dataset Repository offers data on lung cancer gene expression. This database contains the gene expression information gathered

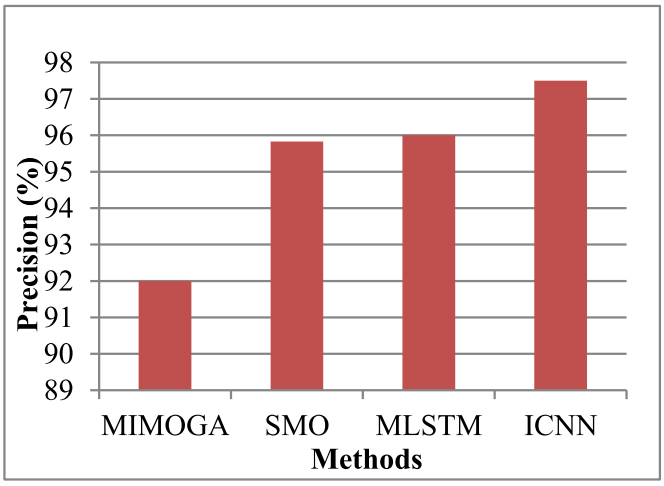

Fig. 3. Precision results comparison.

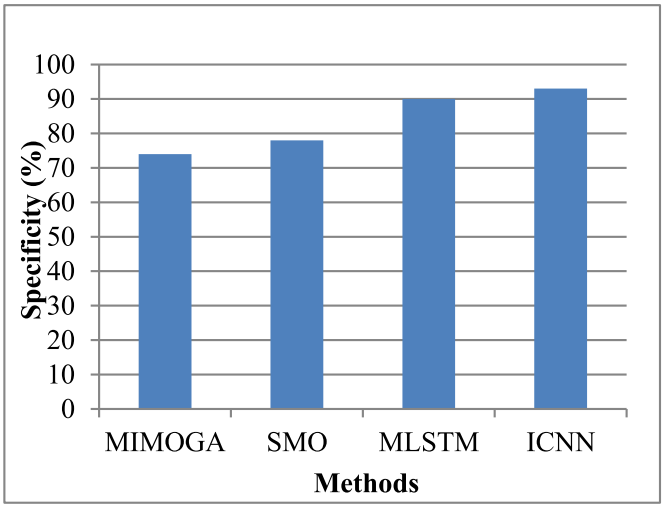

Fig. 4. Specificity results comparison.

from 86 primary lung adenocarcinoma samples and 10 non-neoplastic lung sample. 7129 genes are present in each sample. The full dataset is used to generate 30% of the test set and 70% of the training set. The results are validated using both the training and test datasets.

4.1. Precision

Precision is a measure of accuracy or quality. Algorithms having a better precision typically have increased relevance compared those having reduced precision. For instance, precision may be computed by the ratio of its counts of true positives and its total counts of false positives. It is computed as,

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \tag{11}$$

The above Fig. 3, portrays the comparison evaluation between the proposed approach and recent techniques in terms of precision metric. Methodologies are listed in the x-axis and the precision value is taken on the y-axis. In the above figure proposed ICNN method is compared with the existing MIMAGA, SMO and MLSTM approaches. It is concluded from the outcomes that this proposed ICNN framework produces higher precision results of 97.5%while the other existing MIMAGA, SMO and MLSTM methods produces 92%, 95.83%, 96% accordingly.

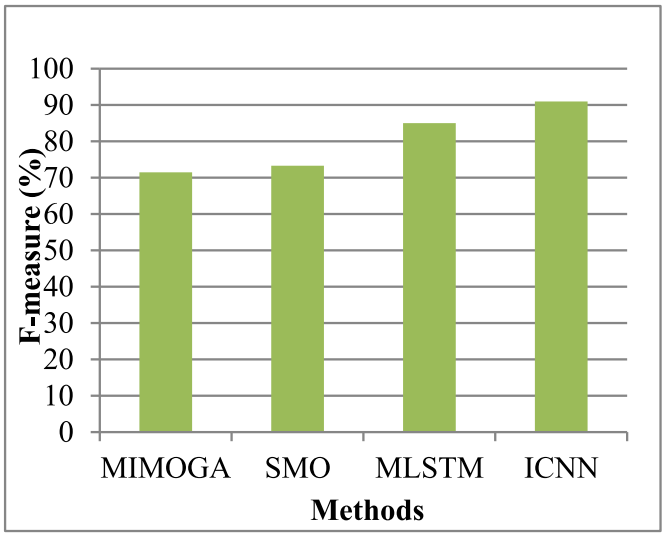

Fig. 5. F measure results comparison.

4.2. Specificity

Specificity is given by the ratio of the real negatives, whose prediction is done as the negative (or true negative). It is computed as below:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True negatives} + \text{False positive}} \tag{12}$$

Fig. 4 illustrates the outcomes pertaining to the performance comparison analysis for the proposed ICNN and recent MIMAGA, SMO and MLSTM methods in terms of Specificity. As per the above graph, various classification methodologies are listed along the x-axis and y-axis represents the Specificity values. Proposed work uses whale optimization model for gene selection in which step size parameter will be calculated based on levy flight function and it increases the true positive rate by which it automatically increases the Specificity. It can be inferred from the above figure that the proposed algorithm shows improved Specificity values which is 93% whereas the existing MIMAGA, SMO and MLSTM methods algorithms provide lower Specificity results which are 74%,78% and 90% accordingly (see Fig. 5).

4.3. F-measure

An F-score computes the harmonic mean of the precision and recall values of a system. *F*1-Score is computed as:

$$\text{F1} - \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{13}$$

Performance comparison results of proposed ICNN and existing MIMAGA, SMO and MLSTM methods interms of f-measure is shown in the above figure. In the graph above, multiple classification methodologies are listed along the x-axis and y-axis represents the f-measure values. It is demonstrated from the above figure that the proposed ICNN algorithm yields better f-measure values which is 91% whereas the existing MIMAGA, SMO and MLSTM methods provides 71.5%, 73.3% and 85% accordingly.

4.4. Accuracy

Accuracy measures the degree to which the technique is correct on an overall and is computed by the ratio between total of the original classification parameters (True positive +True negative) and the sum of the classification parameters (true positive + true negative + false postive + false negative.).

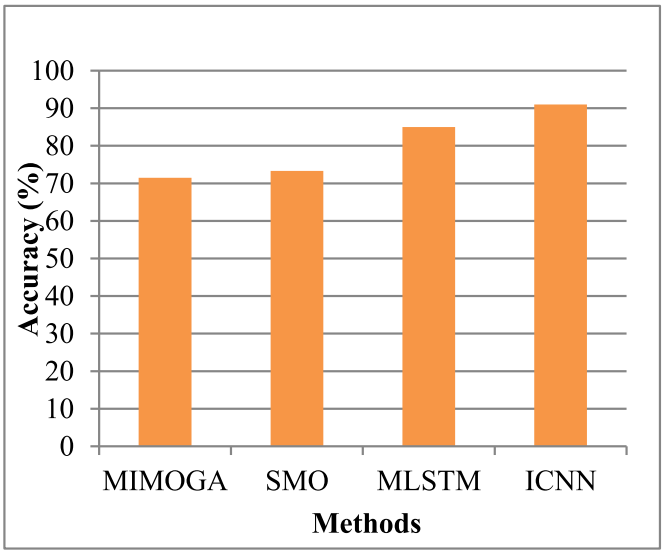

Fig. 6. Accuracy results comparison.

$$\text{Accuracy} = \frac{T_p + T_n}{(T_p + T_n + F_p + F_n)} \tag{14}$$

Accuracy metric comparison for the classifiers such as MIMAGA, SMO and MLSTM is shown in Fig. 6. The x-axis lists the methodologies and the accuracy values are taken along y-axis. It can be confirmed from the above figure that the proposed ICNN algorithm yields improved results of accuracy which is 93% whereas the existing MIMAGA, SMO and MLSTM methods provides lower accuracy results which are 77.77%, 85.33%and 90.23% accordingly.

5. Conclusion and future work

Lung cancer is one among the prevalent cancer kinds all over the globe and becomes the cause behind over one million deaths every year. Identifying differentially represented genes and various gene expression patterns can elucidate the interpretation of the normal and abnormal subjects at the transcriptomic level. In this study, pre-processing is computed using min-max to normalization the data scale. Gene selection is done by using IWOA. ECNN is applied for lung cancer classification. Experimental results reveal that this proposed model produces the accuracy results of 93% which is improved when compared with the other current models. However proposed model does not implement for other diseases and this could be focused in the upcoming future.

Funding statement

The author(s) of this study did not get any special funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

https://www.kaggle.com/datasets/yusufdede/lung-cancer-dataset

References

[1] S.A. Ludwig, S. Picek, D. Jakobovic, Classification of cancer data: analyzing gene expression data using a fuzzy decision tree algorithm, in: Operations Research Applications in Health Care Management, Springer, Cham, 2018, pp. 327–347.
[2] P. Chaudhari, H. Agrawal, K. Kotecha, Data augmentation using MG-GAN for improved cancer classification on gene expression data, Soft Comput. 24 (15) (2020) 11381–11391.
[3] L. Sun, Z. Xiao-Yu, Q. Yu-Hua, J.C. Xu, S.G. Zhang, Y. Tian, Joint neighborhood entropy-based gene selection method with Fisher score for tumor classification, Appl. Intell. 49 (4) (2019) 1245–1259.
[4] R. Alanni, J. Hou, H. Azzawi, Y. Xiang, A novel gene selection algorithm for cancer classification using microarray datasets, BMC Med. Genom. 12 (1) (2019) 1–12.
[5] M.M. Babu, Introduction to microarray data analysis, in: Computational Genomics: Theory and Application, 2004, pp. 225–249.
[6] J.A. Alzubi, B. Bharathikannan, S. Tanwar, R. Manikandan, A. Khanna, C. Thaventhiran, Boosted neural network ensemble classification for lung cancer disease diagnosis, Appl. Soft Comput. 80 (2019) 579–591.
[7] H. Lu, H. Gao, M. Ye, X. Wang, A hybrid ensemble algorithm combining AdaBoost and genetic algorithm for cancer classification with gene expression data, IEEE ACM Trans. Comput. Biol. Bioinf 18 (3) (2019) 863–870.
[8] C. Arunkumar, S. Ramakrishnan, Attribute selection using fuzzy roughset based customized similarity measure for lung cancer microarray gene expression data, Future Computing and Informatics Journal 3 (1) (2018) 131–142.
[9] F. Rafii, B.D.R. Hassani, MLP network for lung cancer presence prediction based on microarray data, in: 2015 Third World Conference on Complex Systems (WCCS), IEEE, 2015, November, pp. 1–6.
[10] H. Azzawi, J. Hou, R. Alanni, Y. Xiang, A hybrid neural network approach for lung cancer classification with gene expression dataset and prior biological knowledge, in: International Conference on Machine Learning for Networking, Springer, Cham, 2018, November, pp. 279–293.
[11] H. Salem, G. Attiya, N. El-Fishawy, Gene expression profiles based human cancer diseases classification, in: 2015 11th International Computer Engineering Conference (ICENCO), IEEE, 2015, December, pp. 181–187.
[12] J. Li, Y. Wang, X. Song, H. Xiao, Adaptive multinomial regression with overlapping groups for multi-class classification of lung cancer, Comput. Biol. Med. 100 (2018) 1–9.
[13] M. Anand, N. Bansal, Lung cancer prediction using machine learning techniques with statistical approach for gene selection, in: Soft Computing and Signal Processing, Springer, Singapore, 2021, pp. 273–284.
[14] Y. Singh, S. Susan, Optimal gene selection and classification of microarray data using fuzzy min-max neural network with LASSO, in: International Conference on Intelligent and Fuzzy Systems, Springer, Cham, 2022, pp. 777–784.
[15] V.G. Raju, K.P. Lakshmi, V.M. Jain, A. Kalidindi, V. Padma, Study the influence of normalization/transformation process on the accuracy of supervised classification, in: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE, 2020, August, pp. 729–735.
[16] J. Wang, G. Wen, S. Yang, Y. Liu, Remaining useful life estimation in prognostics using deep bidirectional LSTM neural network, in: 2018 Prognostics and System Health Management Conference (PHM-Chongqing), IEEE, 2018, October, pp. 1037–1042.
[17] M.A. Albahar, Skin lesion classification using convolution neural network with novel regularizer, IEEE Access 7 (2019) 38306–38313.
[18] A. Jafari, A. Ganesan, C.S.K. Thalisetty, V. Sivasubramanian, T. Oates, T. Mohsenin, Sensornet: a scalable and low-power deep convolution neural network for multimodal data classification, IEEE Transactions on Circuits and Systems I: Regular Papers 66 (1) (2018) 274–287.
[19] W. Zhang, K. Ma, J. Yan, D. Deng, Z. Wang, Blind image quality assessment using a deep bilinear convolution neural network, IEEE Trans. Circ. Syst. Video Technol. 30 (1) (2018) 36–47.
[20] Y. Liu, Y. Zhong, Q. Qin, Scene classification based on multiscale convolution neural network, IEEE Trans. Geosci. Rem. Sens. 56 (12) (2018) 7109–7121.