Received: April 25, 2022. Revised: June 30, 2022. 11

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Histopathological Lung Cancer Detection Using Enhanced Grasshopper Optimization Algorithm with Random Forest

Manaswini Pradhan[1]* Alauddin Bhuiyan[2] Subhankar Mishra[3] Thanh Thieu[4] Ioana L. Coman[5]

[1]*P.G. Department of Computer Science, Fakir Mohan University, Balasore, Odisha, India*
[2]*Department of Ophthalmology, Icahn School of Medicine at Mount Sinai, NY, USA*
[3] *School of Computer Sciences, National Institute of Science Education and Research, Bhubaneswar, Odisha, India*
[4]*Department of Computer Science, Oklahoma State University, Stillwater, Oklahoma, USA*
[5]*Department of Computer Science, State University of New York, Oswego, New York, USA*
* Corresponding author's Email: mpradhan.fmu@gmail.com

Abstract: In the recent decades, the lung cancer is the most dangerous disease that causes 1.6 million deaths per year. Therefore, the early recognition of lung cancer is an effective way to reduce the mortality rate. Compared to other imaging techniques, the histopathology images are effective in identifying the location and level of cancer. In this manuscript, a novel model is implemented for an automatic classification of histopathological images related to lung tissues. Initially, the color normalization technique is applied for improving the contrast of the histopathological images, which are acquired from the LC25000 lung histopathological image dataset. Additionally, the cancer segmentation is accomplished utilizing saliency driven region edge based top down level set (SDREL). Further, the feature descriptors: Alexnet and gray level co-occurrence matrix (GLCM) features were used for extracting the feature vectors from the segmented histopathology images. Lastly, the enhanced grasshopper optimization algorithm (EGOA) and random forest classifier were used for optimal feature selection and lung tissue classification. The simulation result shows that the EGOA-random forest model obtained 98.50 % of accuracy on the LC25000 lung histopathological image dataset.

Keywords: AlexNet, Color normalization technique, Grasshopper optimization algorithm, Lung cancer detection, Random forest classifier.

1. Introduction

In the recent decades, the lung cancer is the 2nd deadliest disease, so the early detection and treatment of the cancerous cells is important [1]. Compared to the existing imaging techniques, the digital pathology systems produce higher-resolution histopathology images that shows the actual cause of the patient illness [2, 3]. However, it is hard to identify the presence of lung cancer in the imaging techniques such as ultrasound, computed tomography, positron emission tomography, and magnetic resonance imaging (MRI) scan [4]. In the recent periods, the computer aided diagnosis (CAD) system is utilized for an effective diagnosis and detection in the medical fields, because it improves the classification accuracy and speed of histopathological lung cancer diagnosis [5]. The CAD system uses machine-learning techniques for improving the diagnostic accuracy. The image feature extraction is needed for machine-learning techniques to utilize medical images in the CAD system [6]. The tumor heterogeneity is an essential factor for evaluating the tumor aggressiveness. The texture analysis is employed in image feature extraction for assessing the tumor heterogeneity in the CAD system [7, 8]. In this manuscript, a novel model is implemented to improve histological lung cancer detection by addressing two major problems such as higher semantic gap between the extracted feature vectors and non-linear nature of the feature vectors [9]. The

major contributions of this manuscript are listed as follows:

- After collecting the images from LC25000 lung histopathological image dataset, a superior pre-processing technique: color normalization is applied for enhancing the contrast of the images. The color normalization makes the classification of cancerous and normal regions much easier.
- Additionally, the SDREL technique is employed for segmenting the cancerous and normal regions from the denoised histopathological images. Then, the feature extraction is carried-out using GLCM features and AlexNet for extracting discriminative feature vectors.
- A feature selection technique: EGOA is used to reduce the curse of dimensionality concern. In this manuscript, the EGOA is applied for selecting the optimal feature information from the extracted feature vectors.
- Developed random forest classifier for improving the disease classification accuracy. The classifier reduces the false positive rate: usually refer to the probability of falsely rejecting the presence of the patient's disease.

This manuscript is prepared as follows: some articles on the topic histopathological lung cancer detection are reviewed in section 2. The mathematical derivations and the experimental analysis of the proposed EGOA-random forest model is denoted in the sections 3 and 4. The conclusion of this manuscript is represented in section 5.

2. Related works

Q. Wang [10] integrated random forest classifier and self-paced learning bootstrap method for improving lung cancer classification. In this literature, the developed model effectively selects the important genes that enhances classification performance related to the traditional techniques on five publicly available lung cancer datasets. Where, the developed model superiorly assists the clinicians in lung cancer prognosis and gene selections. The presented model decreases the dataset noise and improves the classification performance, but it includes the issue of dimension disasters. K. Adu [11] implemented a new dual horizontal squash capsule networks (DHS-CapsNets) for classifying the colon and lung cancers on histopathology images. The presented DHS-CapsNets model comprises of two major functions in image classification: (i) horizontal squash (H-squash) function and encoder feature fusion. In this study, the encoder feature fusion integrates the feature vectors, which were extracted from the convolutional layers. The encoder feature fusion generates rich feature information for effective colon and lung cancer classification. Similarly, the H-Squash function generates sparsity for the discriminative capsules to extract active feature values from the histopathology images with different background. The experimental examinations stated that the developed model attained effective performance in colon and lung cancer classification on LC25000 dataset in light of recall and precision. The increasing depth of the DHS-CapsNets model causes the vanishing gradient problem.

Masud [12] implemented the convolutional neural network (CNN) model, which finds five dissimilar types of tissues (two non-cancerous and three cancerous) in colon and lung cancers using histopathology images. In the resulting section, the CNN model obtained better classification performance on the LC25000 dataset in terms of f-measure score and classification accuracy. The CNN model was computationally expensive, because it needs enormous amount of histopathology images for CNN model training. M. Ali and R. Ali, [13] implemented a multi-input CapsNet model for colon and lung cancer classification, where it comprises of both separable convolutional layers block and conventional convolutional layers block. The unprocessed histopathology images were considered as the input of convolutional layers block and the pre-processed histopathology images were considered as the input of separable convolutional layers block. In this study, the multi-scale fusion, color balancing, image sharpening and gamma correction techniques were used as the pre-processing techniques. The empirical results showed that the multi-input CapsNet model obtained better classification performance on the LC25000 dataset, but the computational complexity of the developed model was comparably higher related to the traditional machine learning techniques.

Nishio [14] integrated homology based method and the texture analysis (gray level size zone matrix, GLCM, gray level dependence matrix, neighboring gray tone difference matrix, and gray level run length matrix) were implemented in order to extract features from the histology images. Further, the machine learning techniques: Support vector machine (SVM), decision tree, logistic regression, K-Nearest neighbor (KNN), gradient tree boosting and random forest classifier were utilized for classifying five lung tissue classes such as invasive adenocarcinoma, emphysema, normal, lepidic pattern of

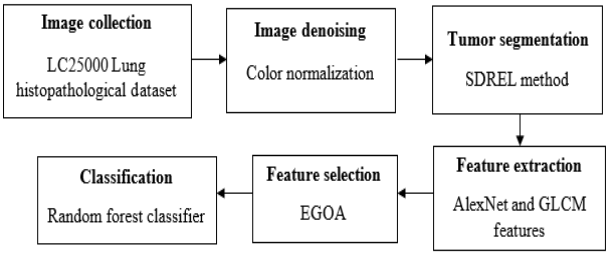


Figure. 1 Flowchart of the proposed EGOA-random forest model

adenocarcinoma and atypical adenomatous hyperplasia. M. Yildirim and A. Cinar, [15] implemented a 45 layer model: MA-Colon NET model for diagnosing colon cancer. However, the machine learning classifiers and MA-Colon NET model manually controls an enormous amount of histology images and the interpretation consumes more time. B. K. Hatuwal and H. C. Thapa [16] implemented CNN for histopathological lung cancer detection. As stated earlier, the CNN model was computationally expensive, where it requires enormous amount of data for model training and testing. To overcome the above-stated problems, a novel automated model: EGOA-random forest model is implemented in this manuscript.

3. Methodology

In the histopathological lung cancer detection, the proposed system comprises of six phases such as image collection: LC25000 lung histopathological image dataset, image denoising: color normalization, segmentation: SDREL method, feature extraction: GLCM features and AlexNet, feature selection: EGOA, and classification: random forest classifier. The flowchart of the proposed EGOA-random forest model is depicted in Fig. 1.

3.1 Image collection and denoising

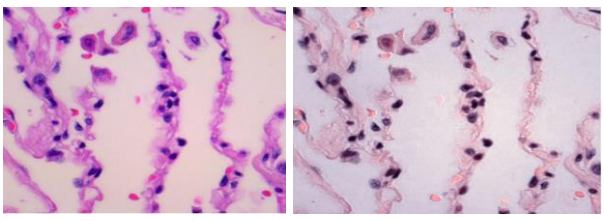
In this manuscript, the proposed EGOA-random forest model's performance is validated on the LC25000 lung histopathological dataset. The LC25000 dataset comprises of 25,000 histology images with 5 classes like 5000 benign colon tissue, 5000 benign lung tissue, 5000 lung squamous cell carcinomas, 5000-colon adenocarcinomas, and 5000 lung adenocarcinomas. The generated histopathological images are 768×768 pixel size and in JPEG file format. The histopathological images belong to benign lung tissue, lung squamous cell carcinomas, and lung adenocarcinomas are used for experimental investigation.

Dataset link: https://academictorrents.com/details/7a638ed187a6180fd6e464b3666a6ea0499af4af

After histopathological image collection, the color normalization technique is applied for finding the deformation and variation in the images that improves the visual ability of the collected images. The general formula of the color normalization technique is specified in Eq. (1).

$$IN_r = (I - Min) + \frac{newMax - newMin}{Max - Min} + newMin \quad (1)$$

Where, I indicates collected lung histopathological images, IN_r denotes normalized images, $Min = 0$ and $Max = 255$ represents minimum and maximum pixel range of the images, and $newMax - newMin$ denotes new pixel range of the normalized images. The sample collected and normalized histopathology images are represented in Fig. 2.



(a) (b)

Figure. 2: (a) collected histopathology image and (b) normalized histopathology image

3.2 Segmentation

After normalizing the histopathology images, the SDREL method is implemented to segment the nuclei and non-nuclei cells from the normalized histopathology images. In the SDREL method, the gradient information and region based information are initially taken into account [17]. Further, the saliency knowledge of the histopathological images is embedded into the proposed model that is very sensitive to the human visual systems. In the image segmentation, the SDREL method has two major

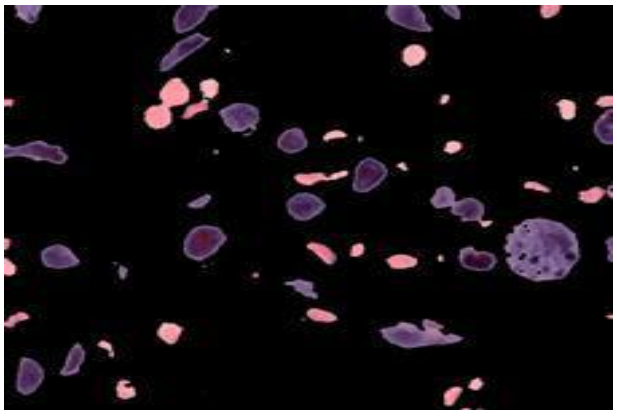


Figure. 3 Output of SDREL method

phases: (i) external energy that is constituted by edge based information (gradient maps) and region based information (color intensity and saliency maps), and (ii) regularization term that is used for penalizing discrepancy and contour length between signed distance function and level set function. As mentioned earlier, the evolution is dominated by external energy in the 1st phase, and the evolution is completely depending on the internal energy in the 2nd phase. The major benefits of using the SDREL method for image segmentation are listed as follows:

- The SDREL method includes a two-stage evolution protocol that reduces the sensitiveness to the initialization condition, which is the main problem in several level set based methods. The top-down evolution method makes the proposed model insensitive to the initialization condition that results in fast and robust level set evolution.
- A new energy term is introduced that incorporates the level-set energy function with the saliency maps, which efficiently improves the level set based image segmentations. Hence, the segmented image is graphically shown in Fig. 3.

3.3 Feature extraction and selection

After the nuclei and non-nuclei cell segmentation, the GLCM features and AlexNet model are utilized to extract active feature vectors. In this manuscript, around 21 GLCM features are implemented to extract active feature vectors from the segmented histopathology images. The 21 GLCM features are: sum average, homogeneity, inverse difference moment normalized, variance, sum of squares, information measure of correlation, dissimilarity, difference variance, maximum probability, inverse difference, cluster shades, autocorrelation, sum variance, difference entropy, correlation, cluster prominence, energy, sum entropy, contrast, inverse difference normalized, and entropy [18]. Correspondingly, the AlexNet model consists of eight layers such as 5 convolutional layers and 3 fully connected layers. Each convolutional layer is followed by max pooling (MP) operation and fully connected layer is followed by leaky rectifier linear unit (ReLU) activation function. The parameter settings of the AlexNet model are listed as follows: L2 regularization is 0.01, training algorithm is stochastic gradient descent algorithm, validation frequency is 30, learning rate is 0.015, momentum is 0.6 and maximum epoch is 10 [19]. The pre-trained design of the AlexNet model is represented in Table 1.

Table 1. Pre-trained design of the AlexNet model

Hidden layers		Design
Convolutional	1	96 filters (11 × 11) with MP (3 × 3)
	2	256 filters (5 × 5) with MP (3 × 3)
	3	384 filters (3 × 3) with MP (3 × 3)
	4	384 filters (3 × 3)
	5	256 filters (3 × 3) with MP (3 × 3)
Fully connected	1	4096 nodes with Leaky ReLU
	2	4096 nodes with Leaky ReLU
	3	1000 nodes with Leaky ReLU

The extracted 2089 feature vectors are given as the input to the EGOA for selecting the optimal feature vectors for image classification. The conventional GOA mimics the swarming behavior of the grasshoppers [20]. The position of the i^{th} grasshopper is indicated as Xi that is mathematically stated in Eq. (2).

$$Xi = Si + Gi + Ai \tag{2}$$

Where, Si indicates social interaction, Gi denotes gravitational force, and Ai states wind advection experienced by the i^{th} grasshopper. These three elements replicate the motion of the grasshoppers in the GOA, which are mathematically specified in the Eqs. (3), (4), and (5).

$$Si = \sum_{j=1, j\neq i}^{N} s(d_{ij})\grave{d}_{ij} \tag{3}$$

$$Gi = -geg \tag{4}$$

$$Ai = uew \tag{5}$$

Where, $\grave{d}_{ij} = \frac{xj-xi}{d_{ij}}$ and $d_{ij} = |xj - xi|$ indicates unit vector, $s(r) = \frac{fe-r}{l-e-r}$ denotes strength of the social forces, f denotes attraction intensity, l denotes attractive length scale, g states gravitational constant, u denotes constant drift, ew states unity vector that utilizes similar direction of the wind, and eg represents unity vector that is directed towards the globe center. The nymph grasshopper's motion is determined using the elements Gi, Ai and Si, which is specified in Eq. (6).

$$Xi = j = 1, jiN\, s\big(d_{ij}\big)(d_{ij}) - geg + uew \tag{6}$$

Where, N represents number of grasshoppers. In

order to solve the optimization problem, the GOA accomplished exploration and exploitation to identify the global optimum approximation. Hence, the Eq. (6) is updated as shown in Eq. (7).

$$Xid = cj = 1, jiNcubd - Ibd2s(|xjd - xid|)xj - xid_{ij} + Td \quad (7)$$

Where, Ibd represents lower bound value, Td states target value, ubd indicates upper bound value, and c represents decreasing coefficient that is directly proportional to the number of iterations, as represented in Eq. (8).

$$C = cmax - itrcmax - cmin \times Maxitr \quad (8)$$

Where, itr states iteration number, $Maxitr = 100$ denotes maximum number of iteration, $cmax$ represents maximum c value and $cmin$ indicates minimum c value. In the conventional GOA, the target values are selected utilizing the best solutions achieved so far. However, in the EGOA, the target is selected from the Pareto optimal solutions, which effectively enhances the distribution rate and premature convergence rate. In the Pareto optimal solutions, the neighboring solutions are considered and counted as the quantitative measures to identify the crowdedness regions. Whereas, the probability of target selection Pr_i is mathematically specified in Eq. (10).

$$Pr_i = \frac{1}{N_i} \quad (10)$$

Where, N_i denotes number of solutions in the vicinity. Additionally, a roulette wheel technique is employed along with Pr_i to choose target from the archive. The parameter settings of the EGOA are listed as follows: maximum value of c is 1, lower bound value is 0, minimum value of c is 0.2, population size is 22, and upper bound value is 11. Finally, the selected 1402 feature vectors are fed to the random forest classifier to perform classification.

3.4 Classification

After selecting the discriminative feature vectors, the tumor classification is performed with the random forest classifier. When using high dimension LC25000 lung histopathological dataset, the random forest classifier is a significant technique that utilizes limited feature vectors to classify benign lung tissue, lung squamous cell carcinomas, and lung adenocarcinomas. The random forest is an effective non-parametric pattern classifier that efficiently decreases the concern of probability density complexity. In the random forest, every decision tree is considered as a base classifier that performs decision-making. Additionally, the growth rules of every decision tree is investigated for generating a robust random forest classifier [21].

The random forest classifier works based on principle of bagging, which utilizes decision tree as a base learner. First, the selected optimal feature vectors are randomly sampled for training sets Z. Then, select the sub feature vectors from the selected optimal feature vectors, if $m(m < M)$, where M indicates selected optimal feature vectors. Next, select the m feature vectors from the M feature vectors and further partition the nodes utilizing best spilt on m dimensional feature vectors. In this scenario, the random forest's error rate depends on two factors such as tree strength needs to be high and correlation between the trees needs to be low for decreasing the error rate [22].

Pseudocode of the random forest classifier

Input:

$Number\ of\ decision\ trees\ in\ the\ random\ forest\ classifier$ "z"
$Number\ of\ selected\ optimal\ feature\ vectors$ "M"
$Training\ samples$ "Z"
$Proportion\ of\ feature\ vectors\ considered\ to\ create\ decision\ tree$ "m"

Output:

$Ensemble$ "m"

1. $E \leftarrow 0$
2. For $i = 1$ to z do
3. $Z^i \rightarrow Bootstrap\ sample\ (Z)$
4. $Q^i \rightarrow Random\ select\ (m)$
5. $O^i \rightarrow Build\ random\ forest\ (Z^i, Q^i)$
6. $EU\{O^i\} \rightarrow E$
7. End for
8. Return E
9. End algorithm

4. Experimental results

In this manuscript, the proposed EGOA-random forest model's effectiveness is validated using Matlab 2020 software environment on a system configuration with windows 10 operating system, 4 TB hard-disk, 16 GB random access memory, and Intel core i9 processor. The proposed EGOA-random forest model's performance is validated using five

performance metrics such as Matthew's correlation coefficient (MCC), sensitivity, false discovery rate (FDR), accuracy, and specificity on the LC25000 lung histopathological image dataset. The MCC is an effective single value metric that generally ranges between 0 to 1, where 1 represents the best agreement between the actual and predicted values and 0 indicates no agreement. Additionally, the FDR is defined as the ratio of the number of false positives to the number of total positive results. The mathematical representations of MCC and FDR are represented in the Eqs. (10) and (11). Where, FP, FN, TP, and TN are represented as false positive, false negative, true positive, and true negative.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \times 10 \quad (10)$$

$$FDR = \frac{FP}{FP+TP} \times 100 \quad (11)$$

The accuracy is the intuitive performance metric in the medical image classification and it is defined as the ratio of predicted observations to the total observations. Specificity and sensitivity are stated as the proportion of negative and positive cases. The mathematical formula of accuracy, specificity, and sensitivity is indicated in the Eqs. (12), (13), and (14).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (12)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (13)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (14)$$

4.1 Quantitative evaluation

In this research article, the proposed EGOA-random forest model's effectiveness is validated on the LC25000 lung histopathological image dataset, where it consists of 25000 histopathological images in that 80:20 % of the medical images is applied for model training and testing. In addition, the 10 fold cross validation is carried out in the research manuscript that helps in better use of the data, which provides more detailed information about the proposed EGOA-random forest model's performance. By validating Table 2, the experimental examination is performed utilizing dissimilar classification techniques such as multi-SVM (M-SVM), deep neural network (DNN), KNN, and random forest with and without using the EGOA feature selection algorithm. By viewing Table 2, the random forest with the EGOA feature selection algorithm obtained maximum performance in the medical image classification with the FDR of 97.64 %, sensitivity of 97.98 %, MCC of 98.57 %, accuracy of 98.50 %, and specificity of 96.34 %. Compared to existing machine learning classifiers, the random forest classifier generates many trees on the subset of the data and merges all the tree's output. This process decreases the overfitting and the variance issues in decision tree that improves the overall classification accuracy. Graphical presentation of the proposed model with and without feature selection is specified in the Figs. 4 and 5.

Table 2. Simulation result of EGOA-random forest model by varying the classification techniques

Without feature selection					
Classifiers	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**MCC (%)**	**FDR (%)**
MSVM	74.53	73.41	78.20	73.89	73.98
KNN	80.16	79.87	79.62	78.06	80.51
DNN	78.28	79.22	79.70	79.79	79.98
Random forest	94.44	95.83	93.50	96.91	93.97
With feature selection					
Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)	MCC (%)	FDR (%)
MSVM	77.34	75.05	79.10	76.67	78.67
KNN	81.59	82.24	81.18	82.27	83.36
DNN	82.75	83.58	84.79	81.88	85
Random forest	98.50	97.98	96.34	98.57	97.64

By investigating Table 3, the random forest classifier's performance is validated with different optimization techniques such as particle swarm optimization (PSO), artificial bee colony (ABC), whale optimization algorithm (WOA), GOA, and EGOA in terms of FDR, specificity, accuracy, MCC, and sensitivity. As denoted in Table 3, the combination: random forest classifier with EGOA obtained maximum classification performance related to the comparative techniques such as PSO, GOA, ABC and WOA. The simulation results of the proposed model by varying the feature selection

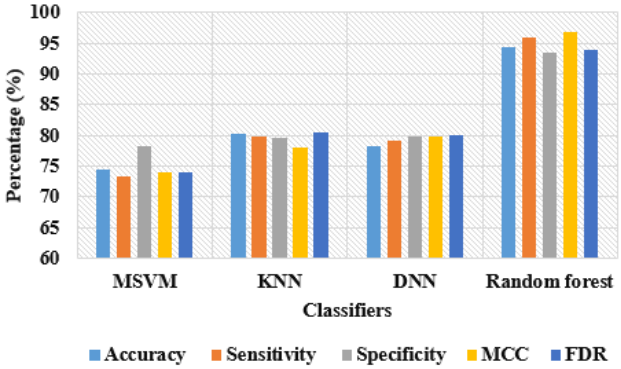


Figure. 4 Experimental results of the proposed model without feature selection

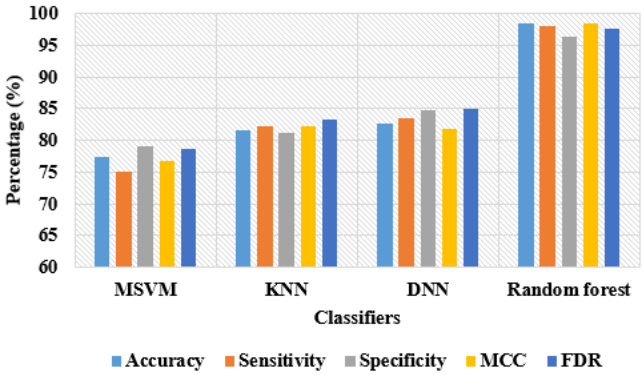


Figure. 5 Experimental results of the proposed model with feature selection

Table 3. Simulation result of EGOA-random forest model by varying the feature selection techniques

Random forest classifier					
Optimizers	Accuracy (%)	Sensitivity (%)	Specificity (%)	MCC (%)	FDR (%)
PSO	88.28	88	85.57	87.04	87.63
GOA	83.59	84.94	83.89	82.64	84.46
ABC	94.06	92.02	92.19	94.44	93.09
WOA	95.44	95.59	94.49	96.65	96.92
EGOA	98.50	97.98	96.34	98.57	97.64

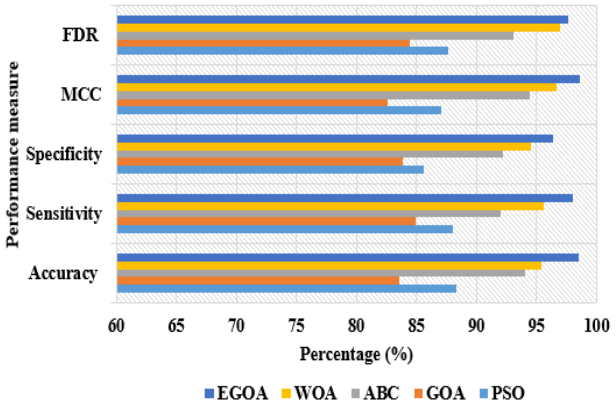


Figure. 6 Experimental results of the proposed model by varying the feature selection techniques

techniques which is graphically indicated in Fig. 6. The proposed EGOA selects discriminative feature vectors from the selected feature vectors, which efficiently diminishes the system complexity and running time of the classifier. In this manuscript, the EGOA-random forest model consumed 48.20 seconds to validate the LC25000 lung histopathological image dataset, which is limited compared to other classification techniques.

4.2 Comparative evaluation

The comparative analysis between the proposed EGOA-random forest model and the prior models is stated in Tables 4 and 5. M. Masud [12], B. K. Hatuwal, and H.C. Thapa [16] developed CNN model for histopathology lung cancer detection. In the M. Masud [12], the CNN model attained 96.33 % of accuracy, 96.38 % of f-measure, 96.39 % of precision and 96.37 % of recall. Correspondingly, in the B. K. Hatuwal and H. C. Thapa [16], CNN model achieved 97.20 % of accuracy, 97.33 % of f-measure, precision, and recall value on the LC25000 dataset. Related to the comparative models, the EGOA-random forest model achieved a maximum classification accuracy of 98.50 %, f-measure of 99.08 %, precision of 98.92 % and recall of 97.98 % on LC25000 dataset.

By incorporating the EGOA with random forest classifier in this manuscript, the computational complexity of the proposed EGOA-random forest model is linear that is the main issue addressed in the literature review section.

Table 4. Comparative investigation between the proposed EGOA-random forest model and the existing models

Models	Accuracy (%)	F-measure (%)	Precision (%)	Recall (%)
CNN [12]	96.33	96.38	96.39	96.37
CNN [16]	97.20	97.33	97.33	97.33
EGOA-random forest	98.50	99.08	98.92	97.98

Table 5. Experimental results of EGOA-random forest model with different cancer type

Models	Cancer type	F-measure (%)	Precision (%)	Recall (%)
CNN [16]	Benign tissue	100	100	100
	Adenocarcinoma	96	95	97
	Squamous cell Carcinoma	96	97	95
EGOA-random forest	Benign tissue	99.42	99	97.75
	Adenocarcinoma	99.02	98.92	97.74
	Squamous cell Carcinoma	98.80	98.84	98.45

5. Conclusion

In this article, a new EGOA-random forest model is proposed for an effective histopathological lung cancer detection. The proposed EGOA-random forest model comprises of three major phases: extraction of features, selection and tissue classification. After pre-processing the histopathological images, the feature extraction is performed by GLCM feature and AlexNet, where the combination of textural and deep feature vectors significantly decreases the semantic gap between the extracted feature vectors. In addition, the EGOA is proposed to select the discriminative feature vectors and then the selected feature vectors are given as the input to the random forest classifier for classifying lung adenocarcinoma, lung squamous cell carcinoma, and lung benign tissue. The simulation result represented that the EGOA-random forest model attained 98.50 % of classification accuracy on the LC25000 lung histopathological image dataset, where the achieved experimental result is superior compared to the existing optimizers and classifiers. The computational complexity of the proposed model is linear by selecting discriminative feature vectors by EGOA. As the future extension, the hybrid deep learning system can be implemented for further enhancing the histopathological lung cancer detection.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st and 3rd author. The supervision and project administration, have been done by 2nd, 4th and 5th author.

References

[1] G. Bansal, V. Chamola, P. Narang, S. Kumar, and S. Raman, "Deep3DSCan: Deep residual network and morphological descriptor based framework for lung cancer classification and 3D segmentation", *IET Image Processing*, Vol. 14, No. 7, pp. 1240-1247, 2020.

[2] S. K. Lakshmanaprabu, S. N. Mohanty, K. Shankar, N. Arunkumar, and G. Ramirez, "Optimal deep learning model for classification of lung cancer on CT images", *Future Generation Computer Systems*, Vol. 92, pp. 374-382, 2019.

[3] M. Sun, K. Liu, Q. Wu, Q. Hong, B. Wang, and H. Zhang, "A novel ECOC algorithm for multiclass microarray data classification based on data complexity analysis", *Pattern Recognition*, Vol. 90, pp. 346-362, 2019.

[4] A. H. Chehade, N. Abdallah, J. M. Marion, M. Oueidat, and P. Chauvet, "Lung and Colon Cancer Classification Using Medical Imaging: A Feature Engineering Approach", *Physical and Engineering Sciences in Medicine*, 2022.

[5] A. Saif, Y. R. H. Qasim, H. A. M. A. Sameai, O. A. F. Ali, and A. A. M. Hassan, "Multi Paths Technique on Convolutional Neural Network for Lung Cancer Detection Based on Histopathological Images", *International Journal of Advanced Networking and Applications*, Vol. 12, No. 2, pp. 4549-4554, 2020.

[6] M. Toğaçar, "Disease type detection in lung and colon cancer images using the complement approach of inefficient sets", *Computers in Biology and Medicine*, Vol. 137, pp. 104827, 2021.

[7] K. Suzuki, "Pixel-based machine learning in computer-aided diagnosis of lung and colon cancer", *In Machine Learning in Healthcare Informatics*, Springer, Berlin, Heidelberg, pp. 81-112, 2014.

[8] S. Wang, D. M. Yang, R. Rong, X. Zhan, J. Fujimoto, H. Liu, J. Minna, I. I. Wistuba, Y. Xie, and G. Xiao, "Artificial intelligence in lung cancer pathology image analysis", *Cancers*, Vol. 11, No. 11, pp. 1673, 2019.

[9] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, E.

Tsougenis, Q. Huang, M. Cai, and P. A. Heng, "Weakly supervised deep learning for whole slide lung cancer image analysis", *IEEE transactions on cybernetics*, Vol. 50, No. 9, pp.3950-3962, 2019.

[10] Q. Wang, Y. Zhou, W. Ding, Z. Zhang, K. Muhammad, and Z. Cao, "Random forest with self-paced bootstrap learning in lung cancer prognosis", *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, Vol. 16, No. 1s, pp. 1-12, 2020.

[11] K. Adu, Y. Yu, J. Cai, K. Owusu-Agyemang, B.A. Twumasi, and X. Wang, "DHS-CapsNet: Dual horizontal squash capsule networks for lung and colon cancer classification from whole slide histopathological images", *International Journal of Imaging Systems and Technology*, Vol. 31, No. 4, pp. 2075-2092, 2021.

[12] M. Masud, N. Sikder, A. A. Nahid, A. K. Bairagi, and M. A. A. Zain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework", *Sensors*, Vol. 21, No. 3, pp. 748, 2021.

[13] M. Ali and R. Ali, "Multi-Input Dual-Stream Capsule Network for Improved Lung and Colon Cancer Classification", *Diagnostics*, Vol. 11, No. 8, pp. 1485, 2021.

[14] M. Nishio, M. Nishio, N. Jimbo, and K. Nakane, "Homology-based image processing for automatic classification of histopathological images of lung tissue", *Cancers*, Vol. 13, No. 6, pp. 1192, 2021.

[15] M. Yildirim and A. Cinar, "Classification with respect to colon adenocarcinoma and colon benign tissue of colon histopathological images with a new CNN model: MA_ColonNET", *International Journal of Imaging Systems and Technology*, Vol. 32, No. 1, pp. 155-162, 2022.

[16] B. K. Hatuwal and H. C. Thapa, "Lung cancer detection using convolutional neural network on histopathological images", *International Journal of Computer Trends and Technology*, Vol. 68, pp. 21-24, 2020.

[17] X. H. Zhi and H. B. Shen, "Saliency driven region-edge-based top down level set evolution reveals the asynchronous focus in image segmentation", *Pattern Recognition*, Vol. 80, pp. 241-255, 2018.

[18] N. Zulpe and V. Pawar, "GLCM textural features for brain tumor classification", *International Journal of Computer Science Issues (IJCSI)*, Vol. 9, No. 3, pp. 354, 2012.

[19] S. Samir, E. Emary, K. E. Sayed, and H. Onsi, "Optimization of a pre-trained AlexNet model for detecting and localizing image forgeries", *Information*, Vol. 11, No. 5, pp. 275, 2020.

[20] S. Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, and I. Aljarah, "Grasshopper optimization algorithm for multi-objective optimization problems", *Applied Intelligence*, Vol. 48, No. 4, pp. 805-820, 2018.

[21] Z. Masetic and A. Subasi, "Congestive heart failure detection using random forest classifier", *Computer methods and programs in biomedicine*, Vol. 130, pp. 54-64, 2016.

[22] V. E. Christo, H. K. Nehemiah, J. Brighty, and A. Kannan, "Feature selection and instance selection from clinical datasets using co-operative co-evolution and classification using random forest", *IETE Journal of Research*, pp. 1-14, 2020.

Notation	Parameter
I	Collected lung histopathological images
IN_r	Normalized images
$newMax - newMin$	New pixel range of the normalized images
$Max - Min$	Maximum and minimum pixel range of the images
Si	Social interaction
Gi	Gravitational force
Ai	Advection experienced by the i^{th} grasshopper
d_{ij}	Unit vector
$s(r)$	Strength of the social forces
f	Attraction intensity
l	Attractive length scale
g	Gravitational constant
u	Constant drift
ew	Unity vector that utilizes similar direction of the wind
eg	Unity vector that is directed towards the globe center
Ibd	Lower bound value
Td	Target value
ubd	Upper bound value
c	Decreasing coefficient
N	Number of grasshoppers
itr	Iteration number
$Maxitr$	Maximum iteration
$cmax$	Maximum c value
$cmin$	Minimum c value
Pr_i	Probability of target selection
FP, FN, TP, and TN	False positive, false negative, true positive and true negative