IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 1, February 2025, pp. 213~221
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i1.pp213-221

Microarray gene expression classification: dwarf mongoose optimization with deep learning

Shyamala Gowri Balaraman[1], Anu H. Nair[1], Sanal Kumar[2]
[1]Department of Computer Science and Engineering, Faculty of Engineering and Technology, Annamalai University, Annamalai Nagar, India
[2]Department of Computer Science, Rajeswari Vedachalam Government Arts College, Chengalpattu, India

Article Info

Article history:

Received Jan 29, 2024
Revised Aug 6, 2024
Accepted Aug 30, 2024

Keywords:

Convolutional autoencoder
Deep learning
Deoxyribonucleic acid
Dwarf mongoose optimization
Microarray gene expression

ABSTRACT

The deoxyribonucleic acid (DNA) microarray model holds significant promise for revealing expression data from thousands of genes. It serves as a valuable tool for investigating gene expressions in diverse biological research fields. This study explores advancements in gene selection for cancer detection through artificial intelligence, with a focus on the challenge of extracting pertinent information from vast databases. The application of deep learning architecture in detecting chronic diseases and aiding medical decision-making has proven effective across various domains. Therefore, this study designs an enhanced microarray gene expression classification by utilizing a dwarf mongoose optimization with deep learning (MGEXC-DMODL) approach. The MGEXC-DMODL approach intends to classify the microarray gene expression (MGE). For this, the MGEXC-DMODL technique initially applies the wiener filtering (WF) technique to eradicate the noise. In addition, the MGEXC-DMODL technique employs a deep residual shrinkage network (DRSN) to learn feature vectors. Meanwhile, the convolutional autoencoder (CAE) model was executed for identifying and classifying the MGE data. Furthermore, the dwarf mongoose optimization (DMO)-based hyperparameter tuning is performed to enhance the detection outcomes of the CAE model. The investigational evaluation of the MGEXC-DMODL model is validated using a benchmark database. The comprehensive comparison outcome highlighted the betterment of the MGEXC-DMODL model over recent approaches.

Corresponding Author:

Shyamala Gowri Balaraman
Research Scholar, Department of Computer Science and Engineering, Annamalai University
Annamalai Nagar, Chidambaram 608002, India
Email: shyamalagowribalaraman@gmail.com

1. INTRODUCTION

The microarray gene expression (MGE) data classification problem midpoints around the task of precisely classifying biological samples dependent upon their profiles of gene expression [1]. The microarray model permits scientists to evaluate many gene expression levels together, offering a wealth of data that can be vital for recognizing diseases, finding biomarkers, and increasing targeted treatments [2]. However, this wealth of information also delivers an important analytical and computational task. The problematic report contains developing strong and effectual classification methods that can distinguish among samples like healthy and unhealthy persons, based on their gene expression information [3]. In this situation, the main tasks contain feature selection; dealing with high-dimensional data, and attaining great classification accuracy while securing

biological interpretability. Precise gene expression data classification is very essential for furthering our consideration of difficult diseases, allowing initial diagnosis, and directing personalized treatment plans [4].

Microarray cancer data study is a vital study area across various fields like machine learning (ML), pattern recognition, statistics, computational biology, and other associated areas. It plays a vital part in recognition, analysis, and cancer treatment [5]. The studies are nowadays targeting the improvement of existence rates in cancer patients by developing the process and knowledge of checking and treatment [6]. The foremost trouble with microarray dataset identification arises from numerous issues like shortage of enough samples, imbalanced class, noisy data, and high trouble of feature dimensionality that managed to be difficult to diagnose and have outcomes of wrong classification. Several research works associated with dual-class data classification of microarray cancer have been conducted [7]. Classifying multiclass data of microarray is still an open research area due to an outcome of tasks in class imbalance. Classes with a tiny amount of models have been generally ignored due to the bias of many methods near classes having more amount of elements [8]. ML models are commonly used in resolving various difficult real issues and have been verified to be effective in examining gene expression data. MGE data classification with deep learning (DL) is an innovative technique that connects the power of neural networks to classify biological samples precisely based on their MGE profiles [9]. In this procedure, high-dimensional gene expression data is changed into a plan that will be appropriate for DL, and convolutional neural network (CNN) molecular basis of illnesses and allow applications in precision medicine, biomarker identification, and drug discovery [10].

This study designs an enhanced microarray gene expression classification by utilizing a dwarf mongoose optimization with deep learning (MGEXC-DMODL) approach. The MGEXC-DMODL approach intends to classify the MGE. For this, the MGEXC-DMODL technique initially applies the Wiener filtering (WF) technique to eradicate the noise. In addition, the MGEXC-DMODL technique employs a deep residual shrinkage network (DRSN) to learn feature vectors. Meanwhile, the convolutional autoencoder (CAE) model was executed for identifying and classifying the MGE data. Furthermore, the dwarf mongoose optimization (DMO)-based hyperparameter tuning is performed to enhance the detection outcomes of the CAE model. The investigational evaluation of the MGEXC-DMODL model is validated using a benchmark database.

The remaining sections of the article are arranged as: section 2 illustrates the related works. Section 3 portrays the proposed model. Then, section 4 elaborates on the experimental validation and section 5 completes the work.

2. RELATED WORKS

Saheed [11] intended to develop an ML–based approach to categorize acute myeloid and acute lymphoblastic leukemia dependent upon MGE profiles. The authors utilized linear discriminant analysis (LDA), Ada boost, logistic regression (LR), k-neighbor method, extreme randomized trees algorithm, ridge classifier, gradient boosting, and random forest (RF). The principle component analysis (PCA) was employed for dimensionality reduction. The authors utilize 2 various cross-validation processes due to they make higher-accurate ability evaluations than prior approaches. Vaiyapuri *et al.* [12] designed an innovative red fox optimizer with a deep learning-based microarray gene expression classification (RFODL-MGEC) technique. This model targets increasing classification effectiveness by choosing suitable features. The RFODL-MGEC method employs an innovative request for offer (RFO)-based feature selection (FS) technique for determining optimum feature subsets. Additionally, the RFODL-MGEC method includes a bi-directional cascaded deep neural network (BCDNN) for classifying data. The constraints executed in the BCDNN method could be tuned by employing the chaos game optimizer (CGO) technique.

Rostami *et al.* [13] introduced an innovative social network investigation-based gene selection technique. The developed technique the relevance maximization and redundancy minimization (mRMR) model. Here, at every round, a supreme community was preferred continually. Ke *et al*. [14] considered a swarm-optimizer-assisted filter-wrapper gene selection comprising 2 stages: The primary stage will be the filter step that chooses small top-n percentages of genes and attains decreased information; later, the secondary stage examines for the optimum gene subsets depend upon a wrapper system in the residual genes by employing a swarm optimization related technique. Research by Bacha *et al.* [15], an innovative decreased computer-aided diagnosis (CAD) technique was applied with the MATLAB (version R2016a) platform for categorizing the four cancer subcategories. The outcomes of the experiment have been performed with 4 groups of baseline data under the appearance of cancerous genes.

Pandit *et al.* [16] projected an effective and hybrid DL method for classifying molecular cancer with the help of expression data to resolve these borders. The input data was pre-processed employing a scalable range adaptive bilateral filter (BF). Subsequently, clustering has been accomplished by employing an enriched binomial clustering technique. Followed by the data must be removed through the multifractal Brownian motion (MBM) technique. Later, the significant features should be chosen by utilizing an improved cuckoo

search optimizer (ICSO) method. Lastly, the data classification was executed employing a wavelet-based deep convolutional neural network (DCNN). Hilal *et al*. [17] presented new feature subset selection (FSS) with optimum adaptive neuro fuzzy inference system (OANFIS) for classifying gene expression. The main goal is to identify as well as categorize the gene expression information. To achieve this, the approach develops an enhanced improved grey wolf optimizer-based feature selection (IGWO-FS) technique for achieving optimum feature subsets. Further, the OANFIS technique was exploited in the classification of genes and the hyperparameter tuning of the adaptive neuro-fuzzy inference system (ANFIS) system can be modified by applying a coyote optimization algorithm (COA).

3. THE PROPOSED METHOD

In this study, an enhanced MGEXC-DMODL approach is designed. The MGEXC-DMODL approach intends to classify the presence of the MGE classification. To accomplish this, the MGEXC-DMODL method encompasses pre-processing, feature extractor, classification, and tuning processes. Figure 1 depicts the structure of the MGEXC-DMODL method.

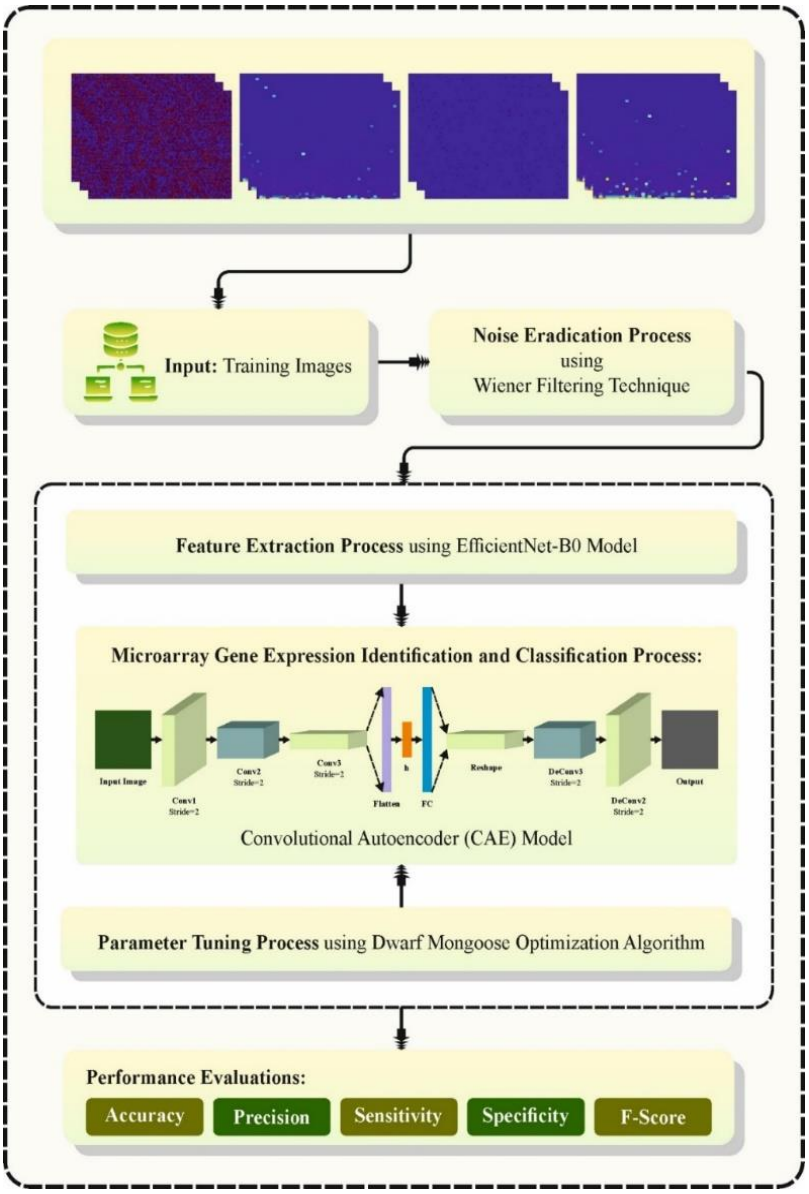

Figure 1. Workflow of MGEXC-DMODL technique

3.1. Preprocessing

Initially, the MGEXC-DMODL technique applies the WF technique to eradicate the noise that exists in it. WF is an efficient way to enhance the accuracy and quality of microarray images [18]. The microarray technique includes the simultaneous analysis of thousands of biological samples, generating large images with imperfections and inherent noise. In this context, WF is used for enhancing the signal-to-noise ratio, enhancing the clarity and reducing artifacts of gene expressions. The WF efficiently sharpens the image by statistically modeling the characteristics of noise and desired signal (microarray spots representing gene expressions), which facilitates quantification and more accurate detection of gene expressions. This technology achieved remarkable success in genomics and bioinformatics research, assisting in the extraction of biological data from microarray images and contributing to advancement in understanding complicated cellular processes.

3.2. Feature extraction

The MGEXC-DMODL technique employs DRSN to learn feature vectors. The concept behind the integration of the RSN with the deep residual network has resulted in the formulation of the DRSN model [19]. DRSN is a complex cascaded deep neural network (DNN) that employs a soft threshold function and attention mechanism to filter out noisy data. The DRSN uses self-attention modules to systematically select valuable feature data while removing noise and ineffectual features, hence boosting the DNN capacity to extract valuable feature data from the noise.

The fundamental unit of DRSN is the residual shrinkage building unit (RSBU). The DRSN consists of one identity mapping, two-batch normalization (BN), one soft threshold learning subnetwork, 2 activation functions (Mish), and 2 convolutional layers ($Conv$). There is a subnetwork in each segment and its role is to independently learn a group of thresholds. Therefore, the threshold is not too large and guaranteed to be positive. The feature map learns various thresholds; hence the above subnetwork is utilized as an attention module, and the soft threshold is used to convert the observed invalid feature into zero, and relevant features are retained. The soft thresholding removes the feature closer to 0 and retains the negative and positive features.

$$y = \begin{cases} x - thr\ x > thr \\ 0\ -thr \le x \le thr \\ x + thr\ x < -thr \end{cases} \tag{1}$$

In (1), x and y are the input and output features and thr shows the thresholding function.

$$\frac{\partial y}{\partial x} = \begin{cases} 1\ x > thr \\ 0\ -thr \le x \le thr \\ 1\ x < -thr \end{cases} \tag{2}$$

In (2) is derived from (1), and after derivation, the soft thresholding becomes 0 or 1.

$$soft = (x, a) = sign(x) * \max\{|x| - thr, 0\} \tag{3}$$

The soft thresholding is converted into (3), where $sign(x)$ denotes the symbolic function. The soft threshold is to independently attain the threshold range. The DNN has better outcomes in self-learning, hence the incorporation of soft threshold and DNN can effectively differentiate features from the irrelevant features. The principle of the attention module includes facilitating the NN model to learn input factors independently and allocate weights to them. This enables to assignment of computational resources to obtain essential features, which results in better performance. The attention module includes a key-value mapping process via query operation of matrix-vector. This involves deriving the corresponding weight values and calculating the similarity between dimension vectors, which undergo normalization through the Softmax function. Consequently, the weight values are multiplied with a matrix dimensional vector and their summation is attained for formulating the last attention matrix. If $K = V = Q$, then it is represented as a self-attention module and is given as (4).

$$Attention(Q, K, V) = soft\max\left\{\frac{QK^T}{\sqrt{d_k}}\right\} \tag{4}$$

Where $Q = (q_1, q_2, q_l) \in R^{nxd}, K = (k_1, k_2, k_l) \in R^{nxd}, V = (v_1, v_2, v_n) \in R^{nxd}, d$ refers to the dimensional of single vector, n indicates the amount of input vectors that are attained by linear conversion of input matrix X. d_k indicates the matrix with dimension k to adjust the inner products. K^T denotes the transposition of K, and the formula of the input sequence linear mapping process is given as (5):

$$\begin{cases} Q = W_q X \\ K = W_k X \\ V = W_v X \end{cases} \quad (5)$$

where the linear mapping parameter matrices w_q, w_k, and w_v are self-learned in the training model.

3.3. Classification using convolutional autoencoder model

At this phase, the CAE model can be executed for identifying and classifying the MGE data. CAE combines the benefits of convolution filtering in CNN with unsupervised pretraining of autoencoders [20]. Rather than the fully connected (FC) layer, the encoder has a convolution layer and the decoder has a deconvolution layer in contrast to the topology for autoencoders. The deconvolution filter is an inverse version of the convolution filter. Additionally, the deconvolution layer should be followed by the unpooling layer. The unpooling process can be done by keeping the location of maximum value during pooling, which preserves the value of that location while unpooling and zeroing the rest.

Spatial locality can be retained by incorporating the convolution function at all the neurons. Thus, for the input matrix P, the encoder computes.

$$e_i = \sigma(P * F^n + b) \quad (6)$$

In (6), σ indicates the activation function, b is encoder bias,$*$ signifies $2D$ convolution, and F^n represents n^{th} 2D convolutional filter. Zero padding applies input matrix P for retaining spatial resolution. Next, the reconstruction is attained by (7).

$$z_i = \sigma\left(e_i * \tilde{\mathrm{F}}^{\mathrm{n}} + \tilde{b}\right) \quad (7)$$

In (7), $\tilde{\mathrm{F}}^{\mathrm{n}}$ shows n^{th} $2D$ convolution filters in the decoder, z_i designates the reconstruction of i^{th} input and b indicates bias of the decoder. Unsupervised pre-training is used in the network that minimizes the (8).

$$E(\theta) = \sum_{i=1}^{m}(x_i - z_i)^2 \quad (8)$$

The FC layer and softmax classifier are added and the decoder part is removed at the end of the network after unsupervised pretraining of the unpooling and deconvolution layers.

3.4. Dwarf mongoose optimization-based hyperparameter tuning

Eventually, the DMO-based parameter tuning method is executed for enhancing the recognition outputs of the CAE method. Chen *et al.* [21] developed a DMO algorithm which is a population-based metaheuristic model. This method splits the mongoose populace into 3 dissimilar groups such as babysitter, scout, and alpha. Below the control of a female leader, the whole populace jointly feeds as an adhesive unit. If the group of alpha flops to find food, an interchange happens among followers of the babysitter and alpha groups. So, associates of the alpha group at the same time are involved in hunting actions while penetrating for a sleeping mound. DMO needs only physically organized limits to decrease the difficulty of the system use. When the associates of the alpha group have inadequate aptitudes, they will interchange followers of babysitters, and alpha groups offer DMO the capability to uphold populace variety. The sleep mound device can stop the algorithm from arriving at local goals.

– Initialize

Set the DMO's mathematical method, as presented in (9).

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d-1} & X_{1,d} \\ X_{2,1} & X_{2,1} & \cdots & X_{2,d-1} & X_{2,d} \\ \vdots & \vdots & X_{i,j} & \vdots & \vdots \\ X_{N,1} & X_{N,2} & \cdots & X_{N,d-1} & X_{N,d} \end{bmatrix} \quad (9)$$

Whereas $X_{i,j}$ denotes the location of the ith mongoose in the jth dimension; N signifies the populace number; X signifies the solution of the candidate and d is the size of the problem. The mathematical method is displayed in (10).

$$X_{i,j} = unifrnd(lb, ub, d) \quad (10)$$

Here $unifrnd$ is employed to make evenly spread random numbers; lb and ub denote the upper and lower limits, correspondingly; and d signifies the dimension.

– Alpha group

The foraging direction of the dwarf mongoose is defined by the female leader, who is formed in the group of alpha. The possibility of every female individual in the alpha group fetching a leader has been defined by (11).

$$\alpha = \frac{fit(i)}{\sum_{i-1}^{n} fit(i)} \tag{11}$$

where $fit(i)$ denotes the fitness output of the ith individual; $n = N - bs$; n signifies the number of individuals in the group of alpha; and bs represents the individuals count in the babysitter group.

The alpha females have preferred foraging ways, and their formulation is as (12):

$$X_{i+1} = X_i + p \times peep \times (X_i - X_k) \tag{12}$$

Whereas X_i signifies the position of the ith individual; X_{i+1} denotes the novel food source place; p signifies the random amount among $[-1,1]$; X_k is an arbitrary individual in the alpha group and $peep$ is set to 2. The sleeping mound (SM) is the relaxing location of dwarf mongooses. Its expression is as (13):

$$smi = \frac{fit(i+1) - fit(i)}{\max\{|fit(i+1), fit(i)|\}} \tag{13}$$

The mathematical method of the mean SM is as (14):

$$\varphi = \frac{\sum_{i=1}^{n} sm_i}{n} \tag{14}$$

– Scout group

The separate followers of the group of scouts will not arrive at their preceding SM. This promises the algorithm's exploration capability. The SM mathematical formula is as (15):

$$X_{i+1} = \begin{cases} X_i - C \times p \times r \times \left|X_i - \vec{M}\right| & if\ \varphi_i + 1 > \varphi_i \\ X_i + C \times p \times r \times \left|X_i - \vec{M}\right| & else \end{cases} \tag{15}$$

Here X_{i+1} denotes the location of the subsequent SM; C signifies the parameter that monitors the flexibility of the mongoose populace.

– Babysitter group

The babysitter group dimension naturally consists of sub-ordinate individuals concerned for their offspring where it is defined as dependent upon the dimension of population. This affects the system by consistently declining the alpha group foraging possible over the period. Parameter L changes the data regarding foraging places for other followers. The babysitter fitness weight is fixed to 0, which safeguards the average weight of the alpha group in the following iteration has been decreased therefore means the group effort is delayed.

The DMO model grows a fitness function (FF) for attaining superior classifier results. It expresses a positive numeral to imply the best output for the candidate's efficiency. Here, the lessening of the classifier errors is measured as FF, as denoted in (18).

$$\begin{aligned} fitness(x_i) &= ClassifierErrorRate(x_i) \\ &= \frac{No.of\ misclassified\ samples}{Overall\ samples} \times 100 \end{aligned} \tag{16}$$

4. EXPERIMENTAL VALIDATION

The outcome evaluation of the MGEXC-DMODL method can be examined utilizing three benchmark datasets [22], such as breast, colon, and ovarian cancer. A comprehensive comparison result of the MGEXC-DMODL method on the breast cancer dataset can be highlighted in Figure 2 [23]–[25]. These outcomes pointed out that the CGRMD-MR-ANFIS, grid-based, Fuzzy c means, and CNN model has shown the least performance. Meanwhile, the RF model gains slightly boosted outcomes. However, the MGEXC-DMODL

technique demonstrates maximum performance with $accu_y$ of 94.59%, $prec_n$ of 94.12%, $reca_l$ of 94.59%, and F_{score} of 94.02%.

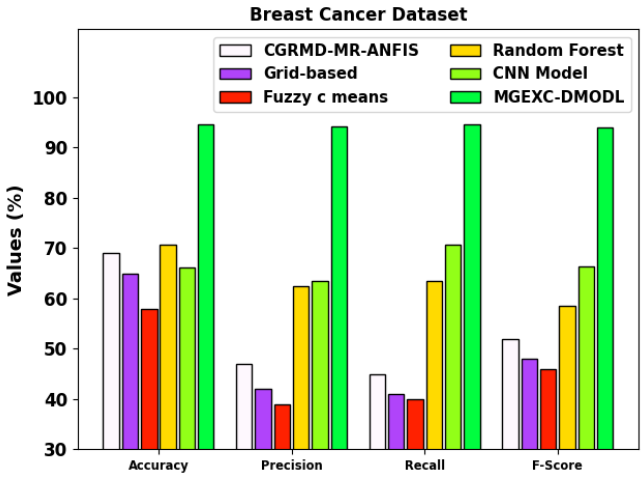

Figure 2. Comparative result of the MGEXC-DMODL system under breast cancer dataset

A wide comparative analysis of the MGEXC-DMODL method with the colon cancer dataset can be emphasized in Figure 3. These obtained findings indicate that the genetic algorithm (GA)-support vector machine (SVM), GA- K-nearest neighbors (KNN), random+SVM, PCA-voting, logistic bootstrap (LogitBoot), and RF methods have shown poorer performance. Meanwhile, the two-way clustering technique achieves moderated increased outcomes. Nevertheless, the MGEXC-DMODL model reveals supreme performance with an $accu_y$ of 96.15%, $prec_n$ of 92.86%, $reca_l$ of 96.15%, and F_{score} of 94.15%.

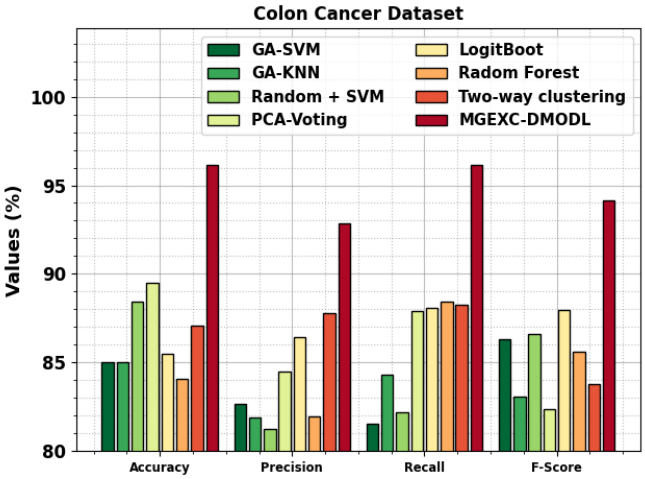

Figure 3. Comparative outcome of the MGEXC-DMODL method under colon cancer dataset

An extensive comparative result of the MGEXC-DMODL method at the ovarian cancer dataset can be underscored in Figure 4. These accomplished findings denote that the linear SVM, RF, ensemble SVM, common feature optimization (CFO)-LDA, laplace approximation (LAPO)-KNN, and gradient boosted classifier (GBCO)-LR techniques get poorer performance. Similarly, the adaptive ant colony optimization (AAO)-multi-layer perceptron (MLP) technique obtains moderated boosted outcomes. However, the MGEXC-DMODL technique shows excellent performance with an $accu_y$ of 95.31%, $prec_n$ of 96.81%, $reca_l$ of 95.31%, and F_{score} of 95.89%. These outcomes confirmed the boosted performance of the MGEXC-DMODL method under gene expression classification.

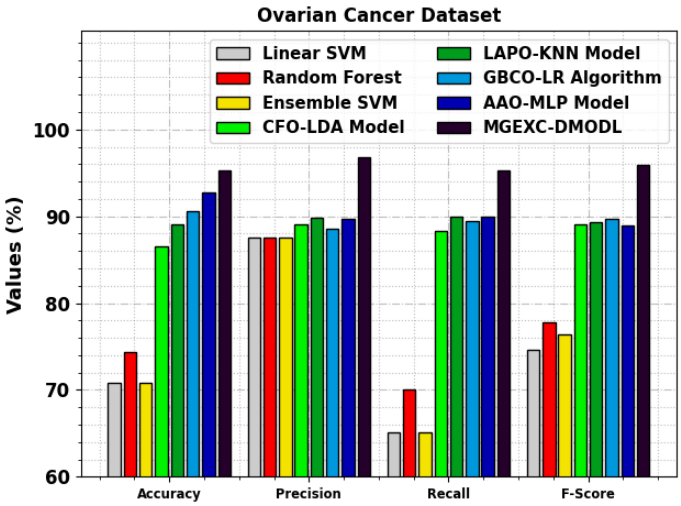

Figure 4. Comparative outcome of the MGEXC-DMODL model with ovarian cancer dataset

5. CONCLUSION

In this study, enhanced MGEXC-DMODL approach is designed. The MGEXC-DMODL technique intends to classify the presence of the MGE classification. To accomplish this, the MGEXC-DMODL technique initially applies the WF technique to eradicate the noise that exists in it. In addition, the MGEXC-DMODL technique employs DRSN to learn feature vectors. Meanwhile, the CAE technique can be executed for the identification and classification of MGE data. Furthermore, the DMO-based hyperparameter tuning is performed to improve the recognition outcomes of the CAE algorithm. The comparative analysis of the MGEXC-DMODL methodology highlighted the superior outcome of 94.59%, 96.15%, and 95.31% over recent state of art approaches under benchmark datasets. The MGEXC-DMODL methodology encounters restrictions in handling large datasets and warrants exploration in real-world clinical settings, prompting future enhancement in disease detection via MGE classification.

REFERENCES

[1] S. Osama, H. Shaban, and A. A. Ali, "Gene reduction and machine learning algorithms for cancer classification based on microarray gene expression data: A comprehensive review," *Expert Systems with Applications*, vol. 213, 2023, doi: 10.1016/j.eswa.2022.118946.

[2] M. Abd-Elnaby, M. Alfonse, and M. Roushdy, "Classification of breast cancer using microarray gene expression data: A survey," *Journal of Biomedical Informatics*, vol. 117, 2021, doi: 10.1016/j.jbi.2021.103764.

[3] R. Tabares-Soto, S. Orozco-Arias, V. Romero-Cano, V. S. Bucheli, J. L. Rodríguez-Sotelo, and C. F. Jiménez-Varón, "A comparative study of machine learning and deep learning algorithms to classify cancer types based on microarray gene expression data," *PeerJ Computer Science*, vol. 2020, no. 4, 2020, doi: 10.7717/peerj-cs.270.

[4] E. Alhenawi, R. Al-Sayyed, A. Hudaib, and S. Mirjalili, "Feature selection methods on gene expression microarray data for cancer classification: A systematic review," *Computers in Biology and Medicine*, vol. 140, 2022, doi: 10.1016/j.compbiomed.2021.105051.

[5] J. O. Agushaka, A. E. Ezugwu, O. N. Olaide, O. Akinola, R. A. Zitar, and L. Abualigah, "Improved dwarf mongoose optimization for constrained engineering design problems," *Journal of Bionic Engineering*, vol. 20, no. 3, pp. 1263–1295, 2023, doi: 10.1007/s42235-022-00316-8.

[6] S. Aruna and L. V. Nandakishore, "Empirical analysis of the effect of resampling on supervised learning algorithms in predicting the types of lung cancer on multiclass imbalanced microarray gene expression data," *EAI/Springer Innovations in Communication and Computing*, pp. 15–27, 2022, doi: 10.1007/978-3-030-86165-0_2.

[7] N. Fernandez-Pozo *et al.*, "PEATmoss (Physcomitrella Expression Atlas Tool): a unified gene expression atlas for the model plant Physcomitrella patens," *Plant Journal*, vol. 102, no. 1, pp. 165–177, 2020, doi: 10.1111/tpj.14607.

[8] L. Chen, L. Klebanov, A. Almudevar, C. Proschel, and A. Yakovlev, "A study of the correlation structure of microarray gene expression data based on mechanistic modelling of cell population kinetics," *Statistical Modeling for Biological Systems: In Memory of Andrei Yakovlev*, pp. 47–61, 2020, doi: 10.1007/978-3-030-34675-1_3.

[9] M. Loey, M. W. Jasim, H. M. EL-Bakry, M. H. N. Taha, and N. E. M. Khalifa, "Breast and colon cancer classification from gene expression profiles using data mining techniques," *Symmetry*, vol. 12, no. 3, 2020, doi: 10.3390/sym12030408.

[10] Y. Wang, H. Wei, L. Song, L. Xu, J. Bao, and J. Liu, "Gene expression microarray data meta-analysis identifies candidate genes and molecular mechanism associated with clear cell renal cell carcinoma," *Cell Journal*, vol. 22, no. 3, pp. 386–393, 2020, doi: 10.22074/cellj.2020.6561.

[11] Y. K. Saheed, "Effective dimensionality reduction model with machine learning classification for microarray gene expression data," *Data Science for Genomics*, pp. 153–164, 2022, doi: 10.1016/B978-0-323-98352-5.00006-9.

[12] T. Vaiyapuri, Liyakathunisa, H. Alaskar, E. Aljohani, S. Shridevi, and A. Hussain, "Red fox optimizer with data-science-enabled microarray gene expression classification model," *Applied Sciences*, vol. 12, no. 9, 2022, doi: 10.3390/app12094172.

[13] M. Rostami, S. Forouzandeh, K. Berahmand, M. Soltani, M. Shahsavari, and M. Oussalah, "Gene selection for microarray data

classification via multi-objective graph theoretic-based method," *Artificial Intelligence in Medicine*, vol. 123, 2022, doi: 10.1016/j.artmed.2021.102228.
[14] L. Ke, M. Li, L. Wang, S. Deng, J. Ye, and X. Yu, "Improved swarm-optimization-based filter-wrapper gene selection from microarray data for gene expression tumor classification," *Pattern Analysis and Applications*, vol. 26, no. 2, pp. 455–472, 2023, doi: 10.1007/s10044-022-01117-9.
[15] S. Bacha, O. Taouali, and N. Liouane, "Reduced CAD system for classifications of cancer types based on microarray gene expression data," *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, SETIT 2022*, pp. 133–137, 2022, doi: 10.1109/SETIT54465.2022.9875863.
[16] D. Pandit, J. Dhodiya, and Y. Patel, "Molecular cancer classification on microarrays gene expression data using wavelet-based deep convolutional neural network," *International Journal of Imaging Systems and Technology*, vol. 32, no. 6, pp. 2262–2280, 2022, doi: 10.1002/ima.22780.
[17] A. M. Hilal *et al.*, "Feature subset selection with optimal adaptive neuro-fuzzy systems for bioinformatics gene expression classification," *Computational Intelligence and Neuroscience*, vol. 2022, 2022, doi: 10.1155/2022/1698137.
[18] R. Liu, Y. Li, H. Wang, and J. Liu, "A noisy multi-objective optimization algorithm based on mean and Wiener filters," *Knowledge-Based Systems*, vol. 228, 2021, doi: 10.1016/j.knosys.2021.107215.
[19] T. Han, Z. Zhang, M. Ren, C. Dong, X. Jiang, and Q. Zhuang, "Speech emotion recognition based on deep residual shrinkage network," *Electronics*, vol. 12, no. 11, 2023, doi: 10.3390/electronics12112512.
[20] C. Campbell and F. Ahmad, "Semi-supervised attention-augmented convolutional autoencoder for radar-based human activity recognition," *SPIE Defense + Commercial Sensing*, 2022, doi: 10.1117/12.2622366.
[21] S. Chen, Y. Zhou, and Q. Luo, "Hybrid adaptive dwarf mongoose optimization with whale optimization algorithm for extracting photovoltaic parameters," *AIMS Energy*, vol. 12, no. 1, pp. 84–118, 2024, doi: 10.3934/energy.2024005.
[22] Zhu *et al.* "Microarray datasets," *Weka ARFF format*. [Online]. Available: http://csse.szu.edu.cn/staff/zhuzx/Datasets.html
[23] P. Mishra and N. Bhoi, "Cancer gene recognition from microarray data with manta ray based enhanced ANFIS technique," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 3, pp. 916–932, 2021, doi: 10.1016/j.bbe.2021.06.004.
[24] A. El-Nabawy, N. El-Bendary, and N. A. Belal, "Epithelial ovarian cancer stage subtype classification using clinical and gene expression integrative approach," *Procedia Computer Science*, vol. 131, pp. 23–30, 2018, doi: 10.1016/j.procs.2018.04.181.
[25] S. K. Prabhakar and S. W. Lee, "An integrated approach for ovarian cancer classification with the application of stochastic optimization," *IEEE Access*, vol. 8, pp. 127866–127882, 2020, doi: 10.1109/ACCESS.2020.3006154.

BIOGAPHIES OF AUTHOR

Shyamala Gowri Balaraman is a Research Scholar in the Department of Computer Science and Engineering, Annamalai University from 2021 January. She has published 2 research papers in international journals and conferences. Currently, her research projects include image processing, machine learning, deep learning, and data science. She can be contacted at email: shyamalagowribalaraman@gmail.com.

Anu H. Nair is an Assistant Professor in the Department of Computer Science and Engineering, Annamalai University from 2005. Her main research areas include image processing in machine learning techniques. She has authored 70 research papers in international journals and conferences, and two book chapters. Her current research projects include big data analytics, biometric person identification, and medical image processing. She can be contacted at email: anu_jul@yahoo.co.in.

Sanal Kumar is currently working as Assistant Professor in the P.G. Department of Computer Science at R.V. Government Arts College, Chengalpattu, Tamilnadu, India (on Deputation). He has authored 65 papers in international journals and conferences and holds life membership in CSI, ISTE, and IAENG. His research interests span image processing, pattern recognition, and wearable computing. He can be contacted at email: sanalprabha@yahoo.co.in.