

Retraction

Retracted: *CoySvM-(GeD)*: Coyote Optimization-Based Support Vector Machine Classifier for Cancer Classification Using Gene Expression Data

Journal of Sensors

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Journal of Sensors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] S. S. S. Reddy, A. Kumar, K. Z. Ghafoor, V. P. Bhardwaj, and S. Manoharan, "*CoySvM-(GeD)*: Coyote Optimization-Based Support Vector Machine Classifier for Cancer Classification Using Gene Expression Data," *Journal of Sensors*, vol. 2022, Article ID 6716937, 9 pages, 2022.

Research Article

CoySvM-(GeD): Coyote Optimization-Based Support Vector Machine Classifier for Cancer Classification Using Gene Expression Data

S. Sai Satyanarayana Reddy ¹, Ashwani Kumar ¹, Kayhan Zrar Ghafoor,^{2,3}
Ved Prakash Bhardwaj ⁴ and S. Manoharan ⁵

¹Department of Computer Science and Engineering, Sreyas Institute of Engineering and Technology, Hyderabad 500068, India

²Department of Software & Informatics Engineering, Salahaddin University-Erbil, Erbil 44001, Iraq

³Department of Computer Science, Knowledge University, University Park, Kirkuk Road, 44001 Erbil, Iraq

⁴School of Computer Science, University of Petroleum and Energy Studies, Dehradun 248007, India

⁵Department of Computer Science, School of Informatics and Electrical Engineering, Hachalu Hundesa Campus, Ambo University, Ambo, Ethiopia

Correspondence should be addressed to S. Manoharan; manoharan.subramanian@ambou.edu.et

Received 22 February 2022; Revised 15 April 2022; Accepted 22 April 2022; Published 9 May 2022

Academic Editor: Paulo Jorge Sequeira Gonçalves

Copyright © 2022 S. Sai Satyanarayana Reddy et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cancer, by any means, is a significant cause of death worldwide. In the analysis of cancer disease, the classification of different tumor types is very important. This test initiates an attitude to the classification of cancer through the data in gene expression by modeling the support vector machine. Genetic material expression data of individual tumor types is designed by the SVM classifier, which tends to increase the potential of genetic data. Feature selection has long been considered a practical standard since its introduction in the field, and numerous feature selection methods have been used in an effort to reduce the input dimension while enhancing the classification performance. The proposed optimization has pertained to the gene expression data that selects the fusion factors for the hybrid kernel function in the SVM classifier and the genes as informative for cancer classification. The analysis of cancer classification is performed using colon cancer and breast cancer, and the performance of CoySVM is tested by taking the measures as precision, recall, and F-measure, and it achieves 87.598%, 95.669%, and 98.088% for colon cancer in addition to 93.647%, 92.984%, and 95% for breast cancer. It shows the best performance due to its highest classification in selected measures than the conventional methods.

1. Introduction

One of the major research areas in the medical field is cancer, and it can be accurately predicted for providing suitable treatment as well as reducing the toxicity of patients [1]. The method using gene expression profiles is more objective, accurate, and reliable than standard tumor diagnostic methods that focused primarily on the physical appearance of the tumor [2, 3]. The number of genes in microarray data is frequently significantly bigger than the number of samples [4]. As a result, standard approaches find such data to be inappropriate or computationally infeasible to analyze. Not

all of the thousands of genes that make up the genes in humans are intolerant and required for the classification. The majority of genetic material is unrelated to the development of cancer and which do not influence the categorization accuracy [5]. Considering such genetic material which increases the trouble's complexity, estimation load to be high also introduces needless crash into the allocation activity. As a result, it is too critical to choose a smaller amount of genetic materials known as instructive genes that will be acceptable for the accurate organization. The splendid collection of genetic material, on the other hand, is frequently unknown [6, 7].

The capabilities of DNA chips enabled the concurrent observation of expression levels to an enormous amount of genes [8], as well as the rise of computer evaluation approaches such as machine learning. These approaches are very much beneficial for cancer prediction [9, 10]. They are also used for prognosis [11] and extracting figures from gene expression data classification models. DNA microarray technology has been widely used in cancer research for illness prediction. It is a fantastic platform that has been utilized to analyze genetic material expression by a variety of experimental investigators [12]. Easily evaluate the gene expression data in hundreds of genes obtained from two different sample cells using microarray technology. Essential investigations, such as illness advancement, precise recognition, precise after therapy, and drug replication to be reached regardless of how the samples are obtained [13]. Many feature selection-based algorithms are developed; in addition, an overview of the feature extracting techniques may be found in [14]. Several previous researchers [15–19] were active in determining the integrity of an attribute subset to determine the best one [20].

1.1. Motivation. Cancer classification is widely employed for the detection of cancer; the detection of cancer accurately in the early stage is helpful to provide the proper medications, and hence, the risk can be reduced. There are several existing methods for the classification of cancer from the gene expression data, but still there are many challenges such as inaccurate classification, failure to consider the most significant features, computational complexity failure to consider the large databases that may degrade the performance of the system and several other factors. Hence, there is a need for a more accurate automatic classification technique necessary for cancer classification. This test initiates an application to cancer classification through the genetic material expression data by modeling the support vector machine. The potential of each data is maximized by the genetic-related expression data of individual tumor types designed by the SVM classifier. Feature selection has long been considered a practical standard since its introduction in the field, and numerous feature selection methods have been used in an effort to reduce the input dimension while enhancing the classification performance. The proposed coyote-based optimization has pertained in the gene expression data that selects the kernels in the SVM classifier and the genes as informative for cancer classification. The benefit of the proposed method is as follows:

- (1) The precision, recall, and F-measure are greatly improved than the assumed conventional methods
- (2) The modeled algorithm assumed that it is most suited for social-economic conditions

The residue of the paper is structured as follows: Section 2 provides a review related to the existing research. Section 3 gives detail about our proposed method, and Section 4 shows the results and discussion of the proposed work. Eventually, this paper is concluded with its outcome in Section 5.

2. Motivation

In this constituent, there are numerous algorithms for cancer classification and feature selection being proposed by various researchers listed with their benefits and drawbacks. The challenges they faced are also enumerated here.

2.1. Literature Review. Nguyen et al. [7] developed a high-potential gadget for the classification of cancer which combines the characteristics of the hidden Markov model and improved analytic hierarchy process. Compared with other fewer methods, it achieves high accuracy and area under the curve. HMM consumes more time than the existing methods. Mao et al. [1] used a randomization test for the gene selection, and also, to select a few genes from several genes, PLS discriminant analysis is applied. The method is best suited for classification by utilizing the expression data. Ayyad et al. [21] introduced an organization technique termed improved k-nearest neighbor which is applied to the largest modified KNN (LMKNN) and smallest modified KNN (SMKNN). A smaller testing time is generated when compared with both KNN and weighted KNN. Dwivedi [3] employed the artificial neural network for the classification and get compared with the existing five different machine learning techniques. It is treated on independent test data and correctly classifies all the samples with high accuracy.

Salem et al. [20] hybridized both the standard genetic algorithm (SGA) and information gain (IG) in which the feature selection is done through the information gain, genetic algorithm for feature extraction. Finally, for the classification of cancer types, genetic programming (GP) is utilized. Wang et al. [22] developed a cancer classification based on the fuzzy technique, in which the regularization was devised initially based on the fuzzy measure, and then, the gene selection was done for the classification of cancer. The failure in considering the feature selection is the drawback of the system. Aslam et al. [23] devised a cancer classification based on the gene expression data, in which the breath samples were utilized. Then, the feature extraction was done and the classification was performed through the neural networks. Failing to use the optimization strategy which may enhance the classification accuracy is considered a drawback of the system. Kumar et al. have given many solutions for detecting the object from the images using machine learning algorithms [24–27].

2.2. Challenges

- (i) There is less number of samples compared with the huge available features; it is too difficult to apply conventional classifiers for such an unbalanced dimension space [21]
- (ii) It begins with the entire in-born surroundings involved in the dataset related to gene expression; the size of the sample is under 200 vs. several genes involved in every data type [20]
- (iii) In the cancer classification process, accuracy is the significant specification but that is not really the

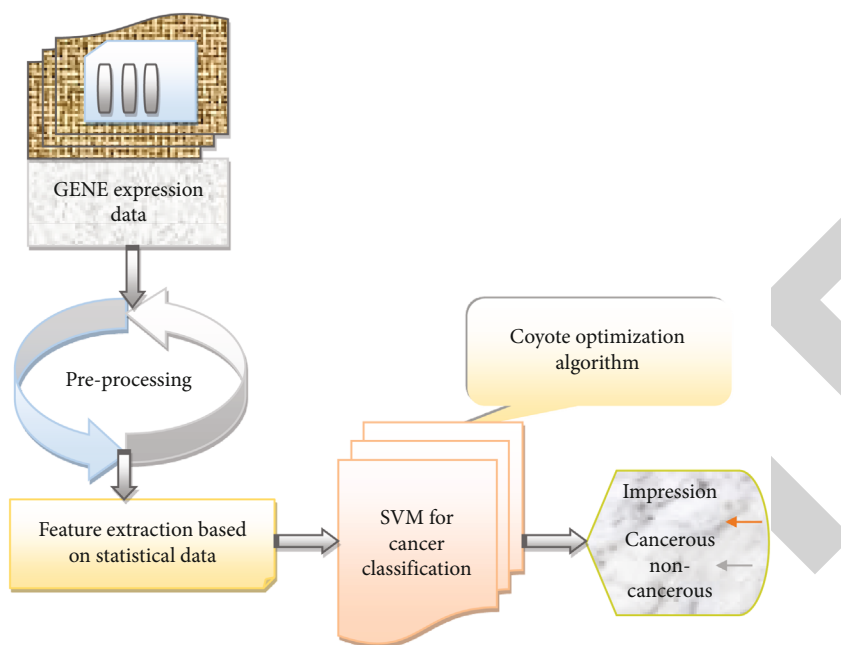


FIGURE 1: Block diagram of the cancer classification using gene expression data.

only goal; there is also a need to achieve classification accuracy and trustworthiness [20]

Thus, from the abovementioned analysis, it is clear that the challenges faced by the conventional cancer classification techniques are inaccurate classification, computational complexity, and failure in choosing the most informative feature, which may enhance the classification accuracy and reduce the computational complexity. In addition, the time complexity and failure to consider the optimization exit. In the proposed cancer classification, the optimal tuning of the SVM is done through the coyote optimization, which minimizes the training rate and enhances the classification accuracy. Besides, the feature selection technique reduces the computational complexity. Thus, by using the proposed method, an efficient cancer classification is performed. Tiwari et al. proposed the hybrid-cascaded framework for image reconstruction [28–31].

3. Proposed Cancer Classification Model Using the Gene Expression Data

Figure 1 revealed the cancer classification using gene expression data. The gene expression data is collected, and the protein molecule gets synthesized in the gene expression by the encoded information in the gene. The data is transferred to the next phase which is the preprocessing in which the quality of the data is modified to an understandable format. The preprocessed data utilizes principal component analysis which reduces the dimension of the enormous amount of data. After preprocessing, the data get transmitted to the feature extraction process. Here, the coyote optimization is performed for feature selection. The SVM classifier collects the extracted features and classifies the labeled data. In the end, the display shows the impression of the classified data.

3.1. Data Preprocessing. The actual data carries noises, unsupported format, and mislaid data which limit the accuracy and efficiency of the model. The gene expression data should be processed in advance to assure the superiority of the data. The stages involved in data preprocessing are cleaning, integrating, reduction, transferring, and discretization of data. The cleaning process involves spotting and eliminating error to enrich the quality. The spellings are miswritten or invalid data reduces the quality of the data. Integration involves gathering and integrating various data from various databases. Data reduction takes place when the data is extremely high and when sometimes it analyzes the fittest data from various amounts of data types. Modification and accumulation take place in the transformation process which depends on the requirement of data. Data discretization involves removing the statistical characteristics from the theoretical one.

3.2. Feature Extraction. Feature extraction involves reducing the dimension of the primary set of original data into a more achievable grouping for the proceeding operation. These wide ranges of data characteristics need a large number of evaluating resources. The performance is enriched by fusing the statistical features which include mean, variance, standard deviation, and entropy with the preset features.

3.2.1. Statistical Features. To evaluate and compute the unpredictability of data and provide extensive perception of the data, statistical features involving mean, variance, standard deviation, and entropy are extricated.

(1) *Mean.* The mean character is represented as σ ; thus, the average of the properties associated with personal data is estimated as

$$\sigma_l = \frac{1}{N} \sum_{j=1}^N n_{lj}, \quad (1)$$

where n_{lj} represents the j^{th} character in the l^{th} data in the gene expression data and N symbolizes the total characters in the data.

(2) *Standard Deviation.* The standard deviation indicates just a minute deviation in the type which affects the classification accuracy. Thus, the standard deviation-based features are estimated as

$$\mu = \sqrt{\frac{\sum (n_{lj} - \sigma_l)^2}{N - 1}}, \quad (2)$$

where the standard deviation of the character is indicated as μ and N_i represents each value in the character, n represents the size of the character, and the mean is given by σ .

(3) *Variance.* The variance is related to the standard deviation which determines the variability of the mean and it is represented by Q_v .

$$Q_v = \mu^2 = \frac{\sum (n_{lj} - \bar{\sigma}_l)^2}{N - 1}, \quad (3)$$

where the value of a single attribute is represented by N_i and the mean value of all observations is represented by $\bar{\sigma}$ and the total number of observations is represented by n .

(4) *Entropy.* Entropy evaluates the unpredictable physical quantity, and it can be determined as

$$X(Y) = X(a_1, \dots, a_N) = - \sum_{j=1}^N a_j \log_2 a_j, \quad (4)$$

where Y is an event with a possible outcomes having probabilities a_1, \dots, a_N .

3.3. Cancer Classification Using Coyote SVM Classifier. In this section, the proposed coyote SVM classifier is proposed for cancer classification, where the SVM is developed with the hybrid kernel function. The hybrid kernel is designed using the fusion factor that is designed based on the coyote optimization, which renders an optimal solution through the advantages, like the effective trade-off between the exploration and exploitation phases.

3.3.1. SVM Working Structure. A collection of gene expression data is given as $(U_1, V_1), (U_2, V_2), \dots, (U_S, V_S), U_h \in E^H$. Assume that $H = 1$. The main objective of the SVM is to evaluate a linear function

$$G(U) = (r \cdot U) + s. \quad (5)$$

The sample data involved in training has λ variations

from the derived targets $V_h (h = 1, 2, \dots, S)$, where r represents the vector of hyperplane coefficients, U is the variable vector, $(r \cdot U)$ is the dot product of r and U , and also, s is the bias in equation (5). The nonnegative slack variables ζ_h and ζ_h^* are introduced, and it is reduced by a chastised objective function.

$$\begin{aligned} \min \quad & \frac{1}{2} \|r\|^2 + D \sum_{h=1}^S (\zeta_h + \zeta_h^*) \\ \text{subject to} \quad & \begin{cases} (r \cdot Z(U_h)) + s - V_h \leq \lambda + \zeta_h, \\ V_h - (r \cdot Z(U_h)) - s \leq \lambda + \zeta_h^*, \\ \zeta_h \geq 0, \zeta_h^* \geq 0, \end{cases} \quad (h = 1, 2, \dots, S), \end{aligned} \quad (6)$$

where $Z(U_h)$ is a function of U_h and the chastised constant $D > 0$ is represented as D that identifies the trade-off between minimizing and maximizing the training error and the margin. Initially, the value of D is set as high to obtain the learning process more stable. By the use of optimal technique, the dual-optimization problem is evaluated as follows:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{h,p=1}^S (\beta_h^* - \beta_h) (\beta_p^* - \beta_p) H(U_h \cdot U_p) + \lambda \sum_{h=1}^S (\beta_h^* + \beta_h) \\ & - \sum_{h=1}^S V_h (\beta_h^* - \beta_h) \end{aligned} \quad (7)$$

$$\text{subject to} \quad \begin{cases} \sum_{h=1}^m (\beta_h - \beta_h^*) = 0, \quad (h = 1, 2, \dots, S), \\ 0 \leq \beta_h, \beta_h^* \leq D, \end{cases} \quad (8)$$

where the Lagrange multiplier vectors are β_h and β_h^* and $H(U_h, U_p) = Z(U_h)Z(U_p)$ is the kernel function. The kernel value is the inner product of two vectors U_h and U_p . In this analysis, the Gaussian radial basis function is considered as the kernel function and it is formulated as

$$H(U, U_h) = \alpha * \exp\left(-\frac{\|U - U_h\|^2}{v^2}\right) + \beta * \exp\left(-\frac{\|U - U_h\|^2}{2v^2}\right), \quad (9)$$

where the width parameter is denoted as v . The SVM regression function is analyzed by the obtained optimal problem solution and is as follows:

$$G(U) = \sum_{h=1}^S (\beta_h^* - \beta_h) H(U \cdot U_h) + s. \quad (10)$$

Only a few coefficients $\beta_h^* - \beta_h$ are considered for quadratic programming which depends on the Karush-Kuhn-Tucker stipulations. If H is greater than one, the i^{th} output feature of the SVM model is defined with respect to the

NSV as

$$G_i(U) = \sum_{q=1}^{NSV} (\beta_{qi}^* - \beta_{qi}) H(U \cdot U_q) + s_i. \quad (11)$$

In the kernel function, α and β are the fusion factors, which contribute to the classifier performance. Hence, the fusion factors are optimally decided using coyote optimization.

3.3.2. Coyote Optimization. Each solution obtained from the coyote is feasible for the optimization problems, and its social status is the price of the objective process. The solutions refer to the based-on-the-social conditions in the coyote designed, and the decision variable of the optimization problem is considered as \vec{y} . The social condition I of the e^{th} coyote of the c^{th} pack in the T^{th} instant of time is as follows:

$$I_e^{c,d} = \vec{y} = (y_1, y_2, \dots, y_K). \quad (12)$$

It also refers to the coyote's adaptation to the environment $F_e^{c,d} \in J$.

3.3.3. Initialization. For each coyote, the social conditions are assumed randomly in the search space for e^{th} coyote of the c^{th} pack of the b^{th} dimension given as

$$I_{e,b}^{c,d} = f_{o_b} + g_b \cdot (x_{o_b} \cdot f_{o_b}), \quad (13)$$

where the lower and upper bounds of the decision variable are represented as f_{o_b} and x_{o_b} and K is the search dimension.

3.3.4. Coyote's Conversion. In the specific present social conditions, the coyote's conversion is evaluated as follows:

$$F_e^{c,d} = w(I_e^{c,d}). \quad (14)$$

Based on the number of coyotes present inside the pack, the coyote elimination takes place with the probability O_u given as

$$O_u = 0.005 \cdot L_e^2. \quad (15)$$

Assume that the probability O_u is higher than 1 for $L_e \leq \sqrt{200}$ and coyote is limited to 14 per pack.

$$\text{Beta}^{c,d} = \left\{ I_e^{c,d} \left| \arg_{e=1,2,\dots,L_e} \min w(I_e^{c,d}) \right. \right\}. \quad (16)$$

Coyote optimization collects all the information from

the coyote and evaluates it as

$$C_b^{c,d} = \begin{cases} P_{((L_e+1)/2),b}^{c,d}, & L_e \text{ is odd,} \\ \frac{P_{((L_e)/2),b}^{c,d} + P_{((L_e+1)/2),b}^{c,d}}{2}, & \text{otherwise,} \end{cases} \quad (17)$$

where $P^{c,d}$ denotes the ranked social status of all coyotes of the c^{th} in the T^{th} instant of time at the range $[1, K]$ for every b . It also calculates the age of the coyote which is represented as $\text{Age}_e^{c,d} \in \mathbb{C}$. The combination of two randomly selected parents is used to calculate the age of new birth and death.

$$AB_b^{c,d} = \begin{cases} I_{g_1,b}^{c,d}, R_b < O_z \quad \text{or} \quad b = b_1, \\ I_{g_2,b}^{c,d}, R_b \geq O_z + O_t \quad \text{or} \quad b = b_2, \\ J_b, \quad \text{otherwise,} \end{cases} \quad (18)$$

where g_1 and g_2 are random coyotes from the c^{th} pack, b_1 and b_2 are the two dimensions, scattering and the associated probability are denoted as O_z and O_t , J_b is the random number in b^{th} dimension, and the scattering and the associated probability are evaluated as

$$O_z = \frac{1}{K}, \quad (19)$$

$$O_t = \frac{(1 - O_z)}{2}. \quad (20)$$

If the new social status is quite better than the preexisting one, it can be written as

$$I_e^{c,d+1} = \begin{cases} \text{new } I_e^{c,d}, & \text{new } F_e^{c,d} < F_e^{c,d}, \\ I_e^{c,d}, & \text{otherwise,} \end{cases} \quad (21)$$

4. Results and Discussion

The present part elucidates the results and the preparation of a coyote optimization-based SVM classifier for the classification of cancer disease. The performance evaluation is dependent on the recall, precision, and F-measure. Moreover, the comparative evaluation is performed with the existing conventional methods to justify the achievement of the proposed method.

4.1. Experimental Setup. The proposed coyote optimization-based SVM classifier for cancer classification is implemented using PYTHON, and the system configuration of the implementation includes PYTHON 3.7. software running in a Windows 10 operating system with 8 GB RAM.

4.2. Comparative Methods. The methods utilized for the comparison include the artificial neural network (ANN) [7], DT [20], SVM [32], CNN [33], FMR [22], DSSAENN [23], and modified KNN [21], which are compared with the developed CoySVM method.

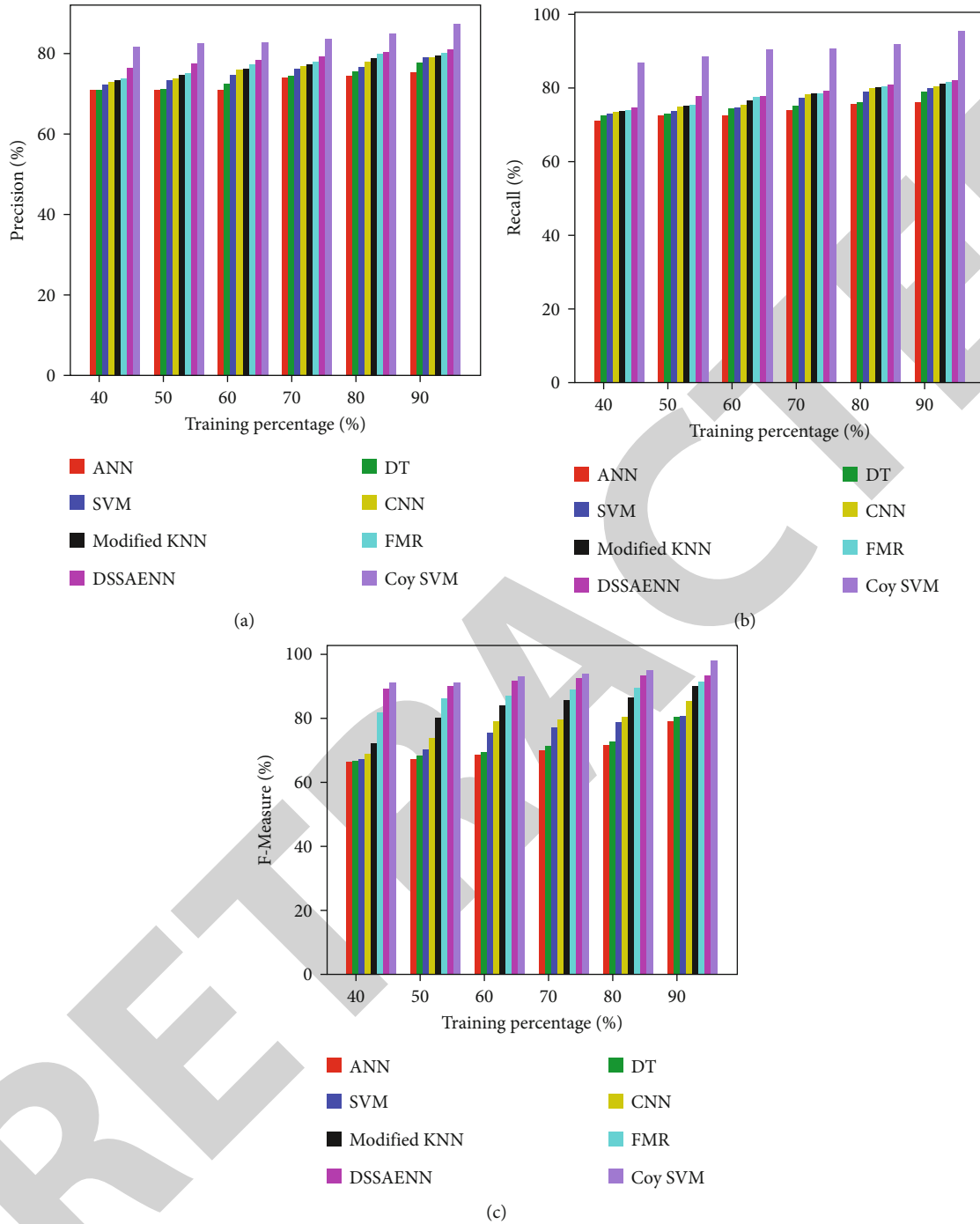


FIGURE 2: Comparative analysis of colon cancer classification based on the performance metrics (a) precision, (b) recall, and (c) F-measure.

4.3. *Performance Measures.* The metrics used for the discussion with the comparative methods with the proposed model are precision, recall, and F-measure.

4.4. *Analysis of the Comparative Methods.* The comparative analysis of the existing conventional method in predicting the plant disease is based on recall, precision, and F-measure to reveal the importance of the developed model.

4.4.1. *Analysis Based on Colon Cancer.* The comparative analysis of colon cancer classification is illustrated in

Figure 2. The analysis is performed in methods such as ANN, DT, SVM, CNN, modified KNN, FMR, DSSAENN, and CoySVM for the cancer classification. The results obtained from the ANN, DT, SVM, CNN, modified KNN, FMR, DSSAENN, and proposed CoySVM are 75.55%, 77.98%, 79.21%, 79.27%, 79.67%, 80.33%, 81.12%, and 87.60%, respectively; these show the precision rates of the method with the training percentage of 90. The recall rates of the conventional methods ANN, DT, SVM, CNN, modified KNN, FMR, DSSAENN, and proposed CoySVM are 76.20%, 79.14%, 80.01%, 80.51%, 81.30%, 81.71%, 82.09%,

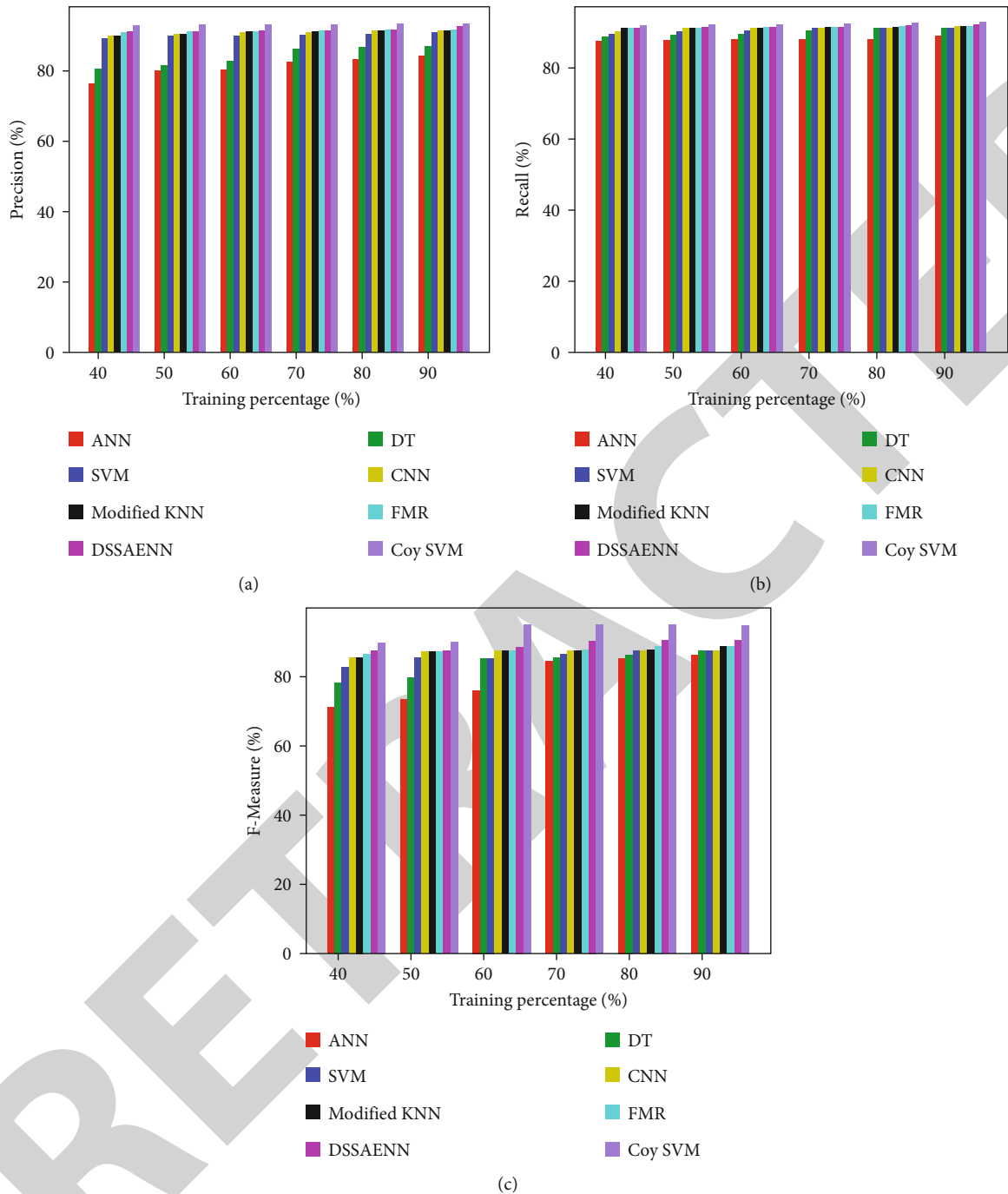


FIGURE 3: Comparative analysis of breast cancer classification based on the performance metrics (a) precision, (b) recall, and (c) F-measure.

and 95.67%, respectively, with the training percentage of 90. The analyses of F-measure in the conventional methods ANN, DT, SVM, CNN, modified KNN, FMR, DSSAENN, and proposed CoySVM with training percentage of 80 are 71.81%, 72.97%, 78.81%, 80.49%, 86.68%, 89.64%, 93.39%, and 95.22%, respectively.

4.4.2. *Analysis in Brest Cancer.* The comparative analysis of breast cancer classification is illustrated in Figure 3. The analysis is performed in methods such as ANN, DT, SVM, CNN, modified KNN, FMR, DSSAENN, and proposed

CoySVM for the cancer classification. The results obtained from the methods ANN, DT, SVM, CNN, modified KNN, FMR, DSSAENN, and proposed CoySVM are 80.53%, 83.06%, 90.03%, 91.12%, 91.21%, 91.43%, 91.54%, and 93.32%, respectively; these shows the precision rates of the method with the training percentage of 60. The recall rates of the conventional methods ANN, DT, SVM, CNN, modified KNN, FMR, DSSAENN, and proposed CoySVM are 88.16%, 90.71%, 91.23%, 91.44%, 91.54%, 91.67%, 91.77%, and 92.59%, respectively, with the training percentage of 70. The analysis of F-measure in the conventional methods

TABLE 1: Comparative performance of conventional methods with the developed CoySVM.

Methods	Colon cancer			Breast cancer		
	Precision	Recall	F-measure	Precision	Recall	F-measure
ANN	75.55	76.20	79.09	84.46	89.12	86.26
DT	77.98	79.14	80.42	87.18	91.38	87.55
SVM	79.21	80.01	80.78	91.23	91.43	87.58
CNN	79.27	80.51	85.60	91.67	91.78	87.69
Modified KNN	79.67	81.30	90.18	91.68	91.82	88.80
FMR	80.33	81.71	91.44	91.82	91.88	88.92
DSSAENN	81.12	82.09	93.57	92.75	92.42	90.58
Proposed CoySVM	87.60	95.67	98.09	93.65	92.98	95.00

ANN, DT, SVM, CNN, modified KNN, FMR, DSSAENN, and proposed CoySVM with a training percentage of 80 are 85.34%, 86.38%, 87.44%, 87.63%, 87.74%, 88.86%, 90.48%, and 95.00%, respectively.

4.5. Comparative Discussion. This section presents the discussion of the methods employed for the classification of cancer by utilizing the data related to expression of gene. The discussion depicts the performance metric evaluation of recall, F-measure, and precision at various levels. Table 1 reveals the comparative performance of the methods with a training percentage of 90.

Thus, from the analysis, it is concluded that the proposed method obtained better performance compared to the other state-of-the-art techniques. The feature extraction of the method minimizes the computational complexity, and the coyote optimization can tune the weights of the SVM through the global best solution by maintaining the exploration and exploitation phases. Besides, the SVM classifier is memory efficient and performs a better classification based on the accuracy for high-dimensional data by making separation among the classes. Hence, the abovementioned advantages help to obtain enhanced performance over the other methods.

5. Conclusion

This test introduced a newly developed technique for cancer disease classification. CoySVM proves to be the uttermost vigorous technique among the other four scrutinized classifiers based on performance metrics such as recall, F-measure, and precision. Cancer classification which depends on the data in gene expression is a favorable exploration field in processing the data. In this paper, the classification of cancer is based on its types such as colon and breast cancer by the CoySVM classifier. The CoySVM is utilized to select the kernel value existing in the SVM classifier. The achieved recall, F-measure, and precision are quite better than the conventional methods.

Data Availability

The dataset will be provided upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Z. Mao, W. Cai, and X. Shao, "Selecting significant genes by randomization test for cancer classification using gene expression data," *Journal of Biomedical Informatics*, vol. 46, no. 4, pp. 594–601, 2013.
- [2] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," *Science*, vol. 270, no. 5235, pp. 467–470, 1995.
- [3] A. K. Dwivedi, "Artificial neural network model for effective cancer classification using microarray gene expression data," *Neural Computing and Applications*, vol. 29, no. 12, pp. 1545–1554, 2018.
- [4] V. Bolon-Canedo, N. Sanchez-Maroo, A. Alonso-Betanzos, J. M. Bentez, and F. Herrera, "A review of microarray datasets and applied feature selection methods," *Information Sciences*, vol. 282, pp. 111–135, 2014.
- [5] V. Rupapara, F. Rustam, W. Aljedaani, H. F. Shahzad, E. Lee, and I. Ashraf, "Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model," *Scientific Reports*, vol. 12, no. 1, p. 1000, 2022.
- [6] K. Y. Yeung, R. E. Bumgarner, and A. E. Raftery, "Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data," *Bioinformatics*, vol. 21, no. 10, pp. 2394–2402, 2005.
- [7] T. Nguyen, A. Khosravi, D. Creighton, and S. Nahavandi, "Hidden Markov models for cancer classification using gene expression profiles," *Information Sciences*, vol. 316, pp. 293–307, 2015.
- [8] H. M. Alshamlan, G. H. Badr, and Y. A. Alohal, "Genetic bee colony (GBC) algorithm: a new gene selection method for microarray cancer classification," *Computational Biology and Chemistry*, vol. 56, pp. 49–60, 2015.
- [9] E. Bard and W. Hu, "Identification of a 12-gene signature for lung cancer prognosis through machine learning," *Journal of Cancer Therapy*, vol. 2, no. 2, pp. 148–156, 2011.
- [10] A. Ishaq, S. Sadiq, M. Umer et al., "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE access*, vol. 9, pp. 39707–39716, 2021.

- [11] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *US National Library of Medicine National Institutes of Health, Cancer Informatics*, vol. 2, p. 117693510600200, 2006.
- [12] R. M. Luque-Baena, D. Urda, J. L. Subirats, L. Franco, and J. M. Jerez, *Analysis of Cancer Microarray Data using Constructive Neural Networks and Genetic Algorithms*, I. Rojas and F. M. Ortuno Guzman, Eds., 'IWBIO', Copicentro Editorial, 2013.
- [13] G. Chakraborty and B. Chakraborty, "Multi-objective optimization using Pareto GA for gene-selection from microarray data for disease classification," in *IEEE International Conference Systems, Man, and Cybernetics (SMC)*, pp. 2629–2634, Manchester, UK, 2013.
- [14] J. Jeyachidra and M. Punithavalli, "A study on statistical based feature selection methods for classification of gene microarray dataset," *Journal of Theoretical and Applied Information Technology*, vol. 53, no. 1, pp. 107–114, 2013.
- [15] A. Ghaheeri, S. Shoar, M. Naderan, and S. S. Hoseini, "The applications of genetic algorithms in medicine," *Oman Medical Journal*, vol. 30, no. 6, pp. 406–416, 2015.
- [16] C. D. A. Vanitha, D. Devaraj, and M. Venkatesulu, "Gene expression data classification using support vector machine and mutual information-based gene selection," *Procedia Computer Science*, vol. 47, pp. 13–21, 2015.
- [17] D. L. Tong and A. C. Schierz, "Hybrid genetic algorithm-neural network: feature extraction for unpreprocessed microarray data," *Artificial Intelligence in Medicine*, vol. 53, no. 1, pp. 47–56, 2011.
- [18] S. Hengpraprom, "GA-based classifier with SNR weighted features for cancer microarray data classification," *International Journal of Signal Processing Systems*, vol. 1, no. 1, pp. 29–33, 2013.
- [19] S. K. Patel, "Attack detection and mitigation scheme through novel authentication model enabled optimized neural network in smart healthcare," *Computer Methods in Biomechanics and Biomedical Engineering*, pp. 1–27, 2022.
- [20] H. Salem, G. Attiya, and N. El-Fishawy, "Classification of human cancer diseases by gene expression profiles," *Applied Soft Computing*, vol. 50, pp. 124–134, 2017.
- [21] S. M. Ayyad, A. I. Saleh, and L. M. Labib, "Gene expression cancer classification using modified K-nearest neighbors technique," *Biosystems*, vol. 176, pp. 41–51, 2019.
- [22] J. F. Wang, Z. Y. He, S. H. Huang, H. Chen, W. Z. Wang, and F. Pourpanah, "Fuzzy measure with regularization for gene selection and cancer prediction," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 8, pp. 2389–2405, 2021.
- [23] M. A. Aslam, C. Xue, Y. Chen et al., "Breath analysis based early gastric cancer classification from deep stacked sparse autoencoder neural network," *Scientific Reports*, vol. 11, no. 1, p. 4014, 2021.
- [24] A. Kumar, "A cloud-based buyer-seller watermarking protocol (CB-BSWP) using semi-trusted third party for copy deterrence and privacy preserving," *Multimedia Tools and Applications*, pp. 1–32, 2022.
- [25] A. Kumar, "Design of secure image fusion technique using cloud for privacy-preserving and copyright protection," *International Journal of Cloud Applications and Computing (IJCAC)*, vol. 9, no. 3, pp. 22–36, 2019.
- [26] A. Kumar, "A review on implementation of digital image watermarking techniques using LSB and DWT," in *The Third International Conference on Information and Communication Technology for Sustainable Development (ICT4SD 2018)*, Hotel Vivanta by Taj, GOA, INDIA, 2018.
- [27] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, p. 18, 2020.
- [28] M. Devi, S. Singh, S. Tiwari, S. Chandra Patel, and M. T. Ayana, "A survey of soft computing approaches in biomedical imaging," *Engineering*, vol. 2021, pp. 1–15, 2021.
- [29] S. Tiwari, "A variational framework for low-dose sinogram restoration," *International Journal of Biomedical Engineering and Technology*, vol. 24, no. 4, pp. 356–367, 2017.
- [30] K. V. Arya, S. Tiwari, and S. Behwalc, "Real-time vehicle detection and tracking," in *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 1–6, Chiang Mai, Thailand, 2016.
- [31] S. Tiwari and R. Srivastava, "An OSEM-based hybrid-cascaded framework for PET/SPECT image reconstruction," *International Journal of Biomedical Engineering and Technology*, vol. 18, no. 4, pp. 310–332, 2015.
- [32] P. Kaur, H. S. Pannu, and A. K. Malhi, "Plant disease recognition using fractional-order Zernike moments and SVM classifier," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8749–8768, 2019.
- [33] W. Yang, Y. Chen, Y. Liu et al., "Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain," *Medical Image Analysis*, vol. 35, pp. 421–433, 2017.