Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7614264, 14 pages
https://doi.org/10.1155/2022/7614264

Research Article

A New Feature Selection Method Based on Hybrid Approach for Colorectal Cancer Histology Classification

Mohanad A. Deif,[1] **Hani Attar**,[2] **Ayman Amer**,[2] **Haitham Issa**,[3] **Mohammad R. Khosravi**,[4] **and Ahmed A. A. Solyman**[5]

[1]*Department of Bioelectronics, Modern University of Technology and Information (MTI) University, 11571 Cairo, Egypt*
[2]*Department of Energy Engineering, Zarqa University, Jordan*
[3]*Department of Electrical Engineering, Zarqa University, Jordan*
[4]*Department of Computer Engineering, Persian Gulf University, Bushehr, Iran*
[5]*Department of Electrical and Electronics Engineering, Istanbul Gelisim University, 34310 Avcılar, Turkey*

Correspondence should be addressed to Mohammad R. Khosravi; m.khosravi@mehr.pgu.ac.ir

Received 19 January 2022; Revised 3 April 2022; Accepted 13 April 2022; Published 5 May 2022

Academic Editor: Xin Ning

Copyright © 2022 Mohanad A. Deif 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.

Colorectal cancer (CRC) is one of the most common malignant cancers worldwide. To reduce cancer mortality, early diagnosis and treatment are essential in leading to a greater improvement and survival length of patients. In this paper, a hybrid feature selection technique (RF-GWO) based on random forest (RF) algorithm and gray wolf optimization (GWO) was proposed for handling high dimensional and redundant datasets for early diagnosis of colorectal cancer (CRC). Feature selection aims to properly select the minimal most relevant subset of features out of a vast amount of complex noisy data to reach high classification accuracy. Gray wolf optimization (GWO) and random forest (RF) algorithm were utilized to find the most suitable features in the histological images of the human colorectal cancer dataset. Then, based on the best-selected features, the artificial neural networks (ANNs) classifier was applied to classify multiclass texture analysis in colorectal cancer. A comparison between the GWO and another optimizer technique particle swarm optimization (PSO) was also conducted to determine which technique is the most successful in the enhancement of the RF algorithm. Furthermore, it is crucial to select an optimizer technique having the capability of removing redundant features and attaining the optimal feature subset and therefore achieving high CRC classification performance in terms of accuracy, precision, and sensitivity rates. The Heidelberg University Medical Center Pathology archive was used for performance check of the proposed method which was found to outperform benchmark approaches. The results revealed that the proposed feature selection method (GWO-RF) has outperformed the other state of art methods where it achieved overall accuracy, precision, and sensitivity rates of 98.74%, 98.88%, and 98.63%, respectively.

1. Introduction

Among the types of gastrointestinal cancers, colorectal cancer (CRC) is the third most common cancer representing globally 13% of all malignant tumors [1], and it is the fourth most popular reason for cancer-related death with 700,000 deaths per year [2, 3]. To reduce cancer mortality, early diagnosis and therefore early treatment is essential in leading to a greater improvement and survival length of patients. Treatment of CRC is strongly dependent on the stage of the disease at the time of diagnosis. To evaluate patients, many imaging methods as computed tomography (CT) [4, 5], magnetic resonance imaging (MRI) [6], and microscopy [7] have been employed to provide information about the structure and type of tumor. Currently, analytic computation of slide images formed from digital pathology slides

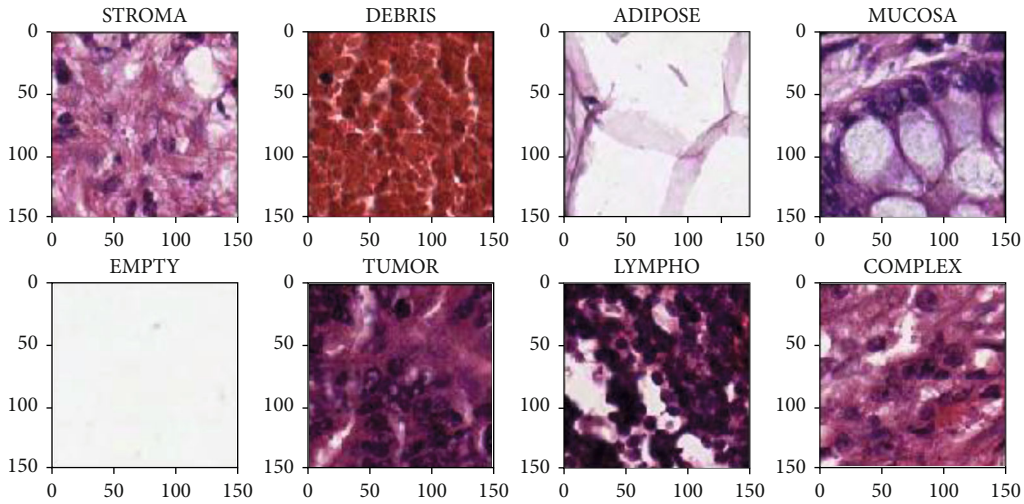

Figure 1: Histopathological images for different categories of tissue.

was reported [8, 9], where extracted texture features are relevant to important applications for computational image analysis for effective classification and diagnosis.

In medical imaging, the classification of different tissue types based on texture analysis was presented by many reports [10, 11]; typically, these methods first extract texture features [12–14] and then introduce the features into a classifier to identify tissue types [15–17]. Histological images may contain more than two tissue types, but most published methods for classifying tissue types on CRC histological images have two categories of the tissue type which are as tumor and Stroma [18–19, 20]. In recent years, machine learning classification has been one of the most successful computational tools in disease investigation and early detection. This includes analyzing a large number of datasets containing a large number of features. Moreover, many noisy, redundant, and unessential features in data sets must be reduced to improve classification accuracy and processing speed. Feature selection optimization was proposed to solve this problem where it is employed to properly select the minimal most relevant subset of features out of a vast amount of complex noisy data. Generally, wrappers and filter methods are the two most common types of feature selection methods [21, 22]. A standard feature selection approach follows a four-step process [21, 23]: (1) candidate subset generation, (2) subset evaluation, (3) picking the criterion for stopping, (4) the subset validation check. As reported from previous studies [21], the quality of the feature selection on the classification algorithm is addressed by wrapper methods and showed superior classification performance than filter methods.

This article proposes a new approach for choosing wrapper features to determine the best subset of features for classifying texture analysis in multiple classes of colon cancer histology. This approach is divided into two phases: feature selection and classification. Gray wolf optimization (GWO) is used in the first place to find the best features of the disease detection dataset by optimizing the hyperparameters of the RF algorithm. In the second stage, an artificial neural network (ANN) is used to classify multiclass texture analysis for colorectal cancer. The proposed method is compared with other competitor's feature selection methods to determine the most efficient method in terms of CRC classification accuracy, sensitivity, and specificity. The comparison aims to evaluate the efficiency of the proposed method in handling high dimensional and redundant data for early diagnosis of CRC and to reach a high classification accuracy with the minimal possible size of a subset of features.

To improve the efficiency and accuracy of performance, CRC has proposed many methods for classifying different tissue types. Wu et al. [24] used color channel histograms, GLCM, and structural features to classify four types of colon tissue (mild, cancerous, adenomatous, and inflammatory) with an accuracy of 75.15 μ. Liu et al. [25] examined an automated colon color recognition system that uses GLCM for texture extraction and a support vector machine (SVM) for classification. The accuracy of this approach in distinguishing between cancerous and noncancerous images was 96.67%. Liu et al. [26] used a 2D discrete wavelet (DW) transform function to accurately identify slide-wide colon cancer images of normally cancerous and adenomatous polyp cases, with 91.11% achieved overall accuracy. To test CRC heterogeneity [27], complete tumor texture features are calculated with Laplace or Gaussian filters (LoG). In addition, as per Rao et al. [28], we examined the texture characteristics of LoG to see if we could tell the difference between CRC patients with liver metastases and CRC patients without metastases. Other studies using local descriptors have been proposed, along with approaches that rely on scale-invariant feature transformations (SIFT) [29], shape contexts [30], and histograms of orientation gradient (HOG) descriptors [31]. increase. Liao et al. [30] used a deep convolutional neural network (CNN) to improve traditional classification methods. The author considered a transition in the learning approach to learn a better predictive model and eventually achieved 95% accuracy.

All of the previous studies mentioned above applied different image processing methods and different classifier

Table 1: Description of the different categories of datasets used in the experiments.

Categories	Description	Category label
Tumor epithelium	Types of tumors that develop from epithelial cells	0
Simple stroma	The tumor arises in the supporting connective tissue of an organ.	1
Complex stroma	Contains single tumor cells and/or few immune cells	2
Immune cells	Includes immune-cell conglomerates and submucosal lymphoid follicles	3
Debris	Includes necrosis, hemorrhage, and mucus	4
Normal mucosal glands	Normal tissues found in mucosal glands	5
Adipose tissue	Is a loose connective tissue composed mostly of adipocytes	6
Background	No tissue	7

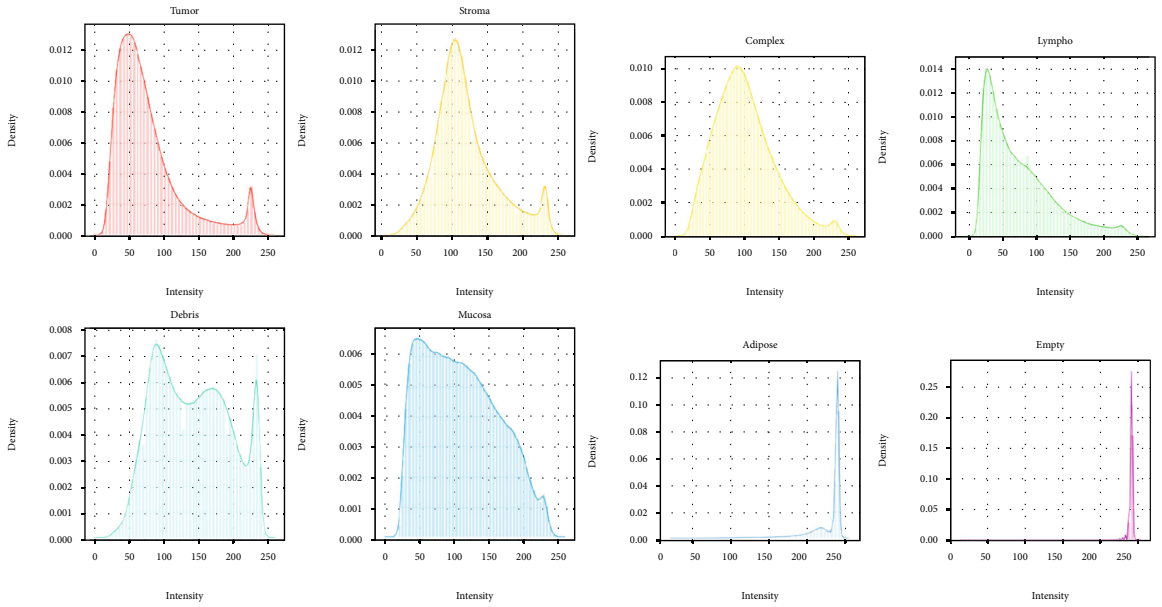

Figure 2: Overall intensity distribution per class.

algorithms to enhance classification accuracy. However, they did not focus on the relation between the number of features and classification accuracy. Hence, a selection method to measure the usefulness of a subset of features by actually training the classifier on it was developed. Therefore, in this study, a new feature selection technique was investigated to improve the accuracy of traditional machine learning approaches.

1.1. Datasets. This work used a dataset of 5,000 human colon cancer histology, including eight types of tissue. This dataset is from the Pathology Archive of the University of Mannheim Medical Center [18]. Different types of tissues were illustrated in Figure 1 and tabulated in Table 1.

2. Methodology

2.1. Preliminaries

(1) Random forests algorithm review

Random forests developed by Breiman, 2001 [32], are a collection of classification and regression decision trees (CART) [32, 33] that are trained on datasets of the same size called bootstraps, created from a random resampling on the training set itself. In Breiman's study, each tree in the collection is formed firstly by selecting k features randomly at each node, and secondly, by calculating the Gini index to determine the best split based on these features in the training set. The decision will be taken by aggregating the predictions of all the decision trees by majority voting. A small value of the Gini index indicates a better classification result. Gini(v), which is node v Gini index, is expressed as follows:

$$\text{Gini}(v) = \sum_{C=1}^{C} \widehat{p}_C^V \left(1 - \widehat{p}_C^V\right), \tag{1}$$

where $\widehat{p}_C^V$ denotes the percentage of class-c observations at node v. Gain (X_i, v), which is the Gini information gain of X_i for splitting node v, is the variance among the impurity at the node v and the sample mean of imperfection at each node of v. Gain (X_i, v) is calculated from:

$$\text{Gain}(X_i, v) = \text{Gini}(X_i, v) - w_L \,\text{Gini}\left(X_i, v^L\right) - w_R \,\text{Gini}\left(X_i, v^R\right), \tag{2}$$

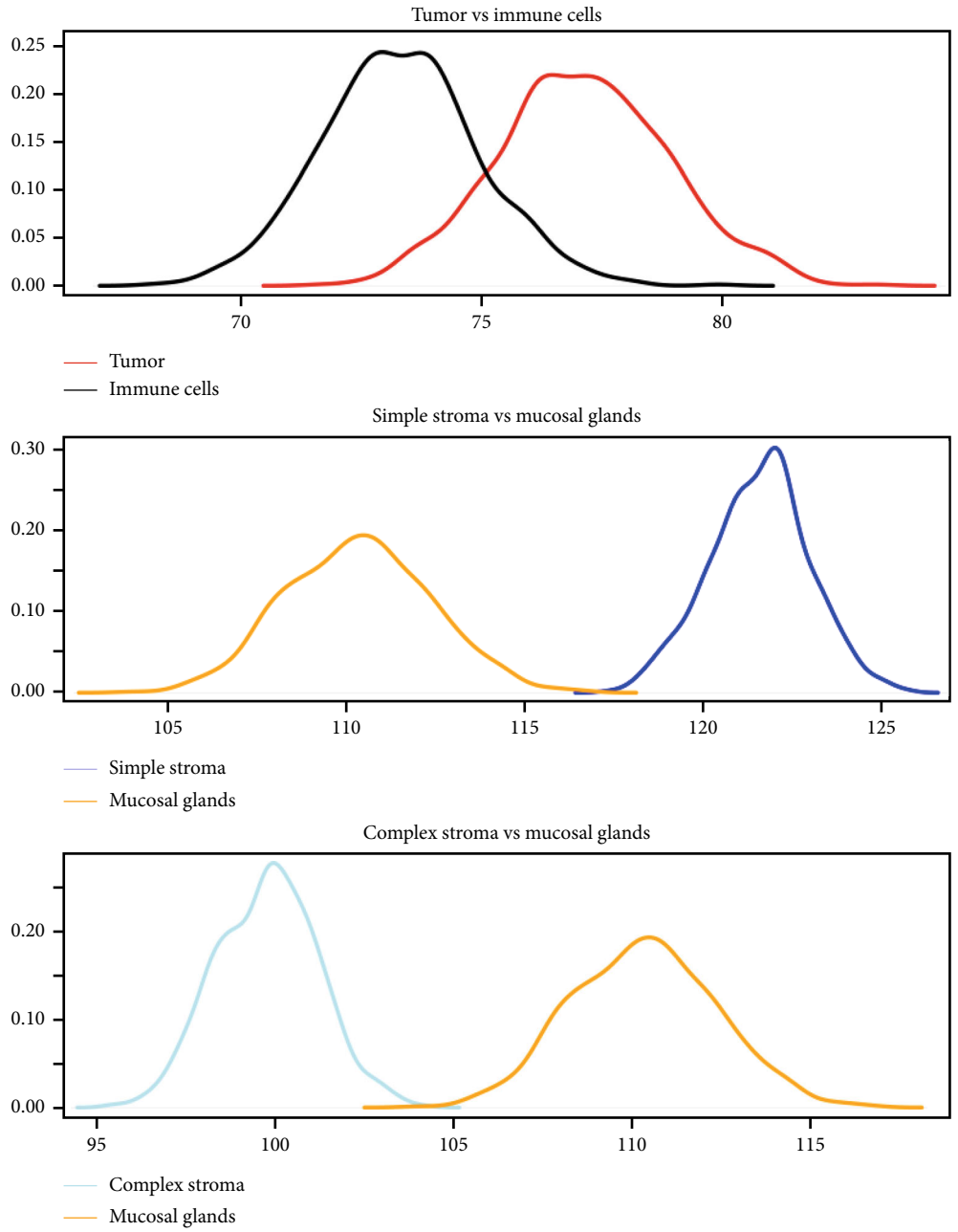

FIGURE 3: Overlaps between different categories of tissue (x and y-axis denote the density and intensity, respectively).

where v^R and v^L are, correspondingly, the right and left nodes of v and w_R, and wL is the quantities of occurrences allocated to the right and left child nodes, respectively. A random collection of mtry features from P (predictive features) is evaluated at each node, and the function with the highest Gain (X_i, mtry) is used to split the node v. Variable Xi importance score can be calculated as

$$Imp_i = \frac{1}{ntree} \sum_{mtry \in S_{X_i}} \text{Gain}\,(Xi, mtry), \tag{3}$$

where S_{X_i} denotes the set of nodes divided by Xi in the RF using ntree. The RF importance scores are widely used to test the function contributions to class prediction. The number of trees (ntrees) used for each forest and the number of characteristics used at each split are the two most important parameters that determine random forest efficiency (mtry).

Error rates can be evaluated in two ways. One way to do this is to split the dataset into two parts: training and testing. Forests can be created from the training phase, and error rates can be calculated from test parts. OutofBag (OOB) error estimation is another option. Because the random forests algorithm calculates the OOB error during the training phase, we do not need to partition the training data to

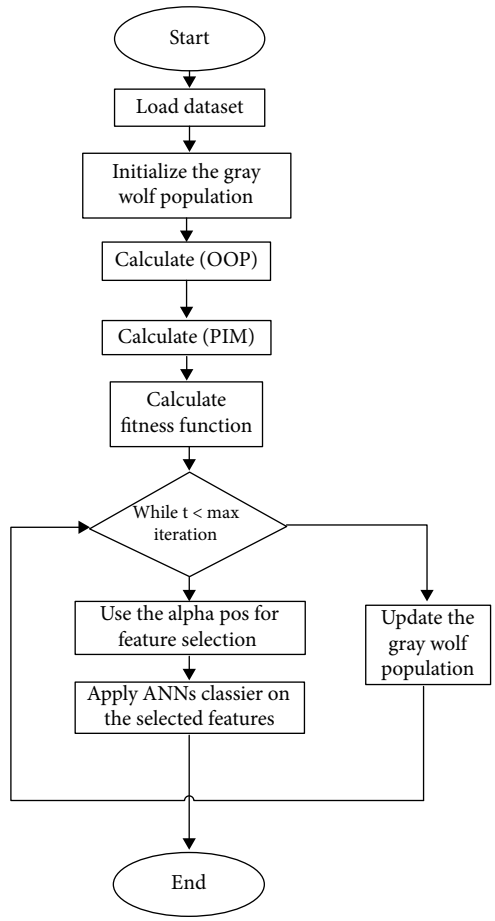

Figure 4: Flowchart of hybrid proposed algorithm (GWO-RF).

retrieve the OOB error. The "OOB error" can be determined as follows:

$$\text{OOB error} = \sum_{i}^{n\,\text{Tree}} \frac{\text{OOB error}_i}{\text{nTree}}. \tag{4}$$

(2) Gray wolf optimization algorithm review

Based on the previous study proposed by Lowe [34], the gray wolf optimization (GWO) algorithm is discussed in the following subsections.

2.1.1. GWO Inspiration. GWO resembles the hunting behavior of gray wolf packs which is established using three basic steps: (1) inclosing, (2) hunting, and (3) attacking. Gray wolves have a social hierarchy that classifies the domination and authority of wolves. The leading and dominating wolf in the gray wolf hierarchy is the alpha wolf which is responsible for decision making, hunting, and other pack activities. Beta wolf follows next in the hierarchy which represents the second and powerful wolf and will assist the alpha in decision making. Omega and delta are less effective, and they have the lowest ranking [33].

Table 2: Parameter setting for (GWO) and (PSO).

Algorithm	Parameter	Numbers
GWO	Number of iterations	100
	Wolves number	5
	Dimensions number	14
	Search domain	[0 1]
	Swam size	50
PSO	Number of iterations	100
	Cognitive parameter	2
	Social parameter	2
	Inertia weight	0.9

2.1.2. Mathematical Model of GWO. The GWO mathematical model is mainly comprised of three basic steps: inclosing, hunting, and attacking which are represented in the following equations [34, 35].

(i) Encircling the prey

In the GWO mathematical model, the inclosing behavior is represented in the following equations:

$$\vec{X}(t+1) = \overrightarrow{X_p}(t) + \vec{A}.\vec{D}, \tag{5}$$

where $\vec{D}$ is defined in equation (5), t is the iteration number, $\vec{A}$, $\vec{C}$ are coefficient vectors, $\overrightarrow{X_p}$ is the prey position, and $\vec{X}$ is the gray wolf position.

The calculation for the vectors $\vec{A}$, $\vec{C}$ is presented in the next equations:

$$\vec{D} = \left|\vec{C}.\overrightarrow{X_p}(t) - \vec{X}(t)\right|, \tag{6}$$

$$\vec{A} = 2\vec{A}.\vec{r_1} - \vec{a}, \tag{7}$$

$$\vec{C} = 2\vec{r_2}, \tag{8}$$

where components of $\vec{a}$ are linearly decreased from 2 to 0 throughout iterations, and $r1$ and $r2$ are random vectors in [0, 1].

(ii) Hunting the prey

In the hunting phase, gray wolves are leaded by alpha. The beta and delta could also help in hunting. To mathematically simulate the hunting behavior of gray wolves, it was assumed that the alpha (α) (fittest candidate solution), beta (β), and delta (δ) had better knowledge of the prey's possible position. The first three best solutions guide the ω wolf to change its location in the decision space based on the position of the best search agents. Representing hunting

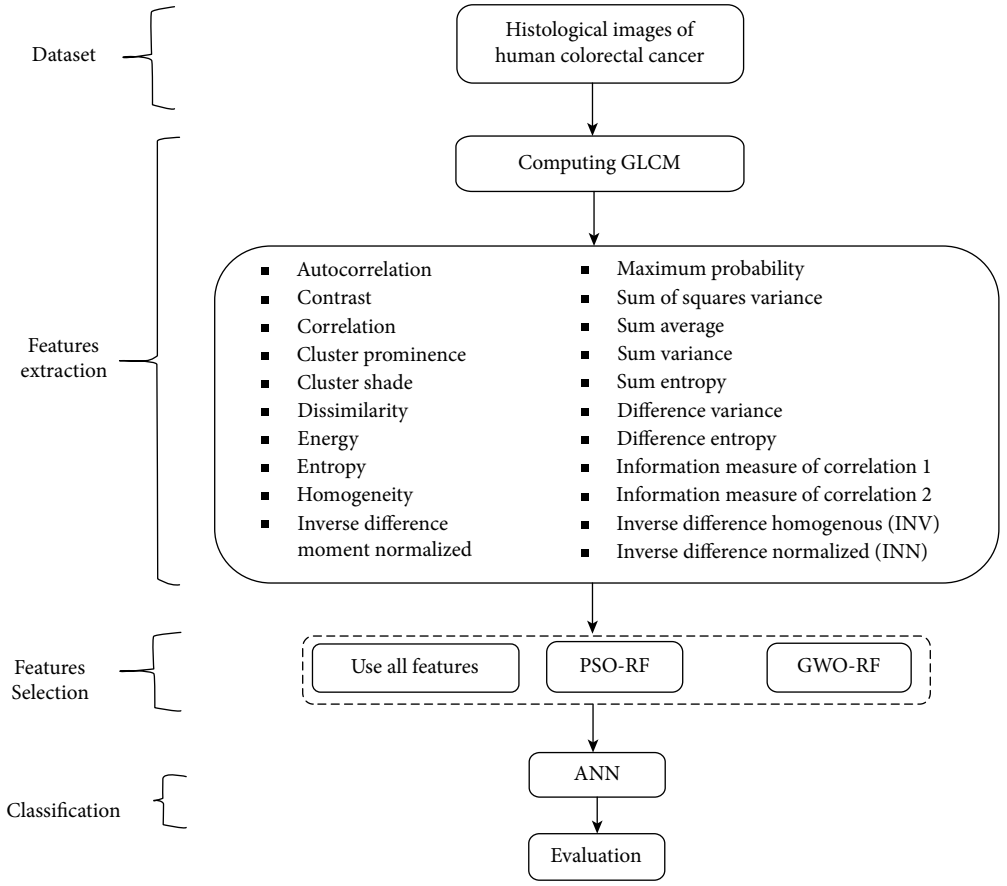

Figure 5: The experimental steps of the proposed approach.

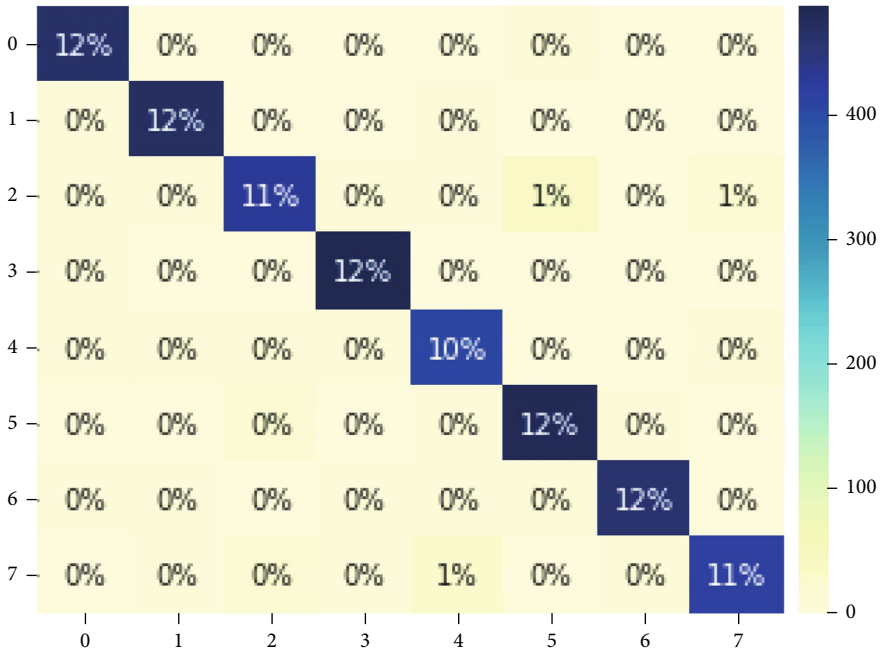

Figure 6: Confusion matrix of ANN model for the test set when using all features (without applying feature selection techniques).

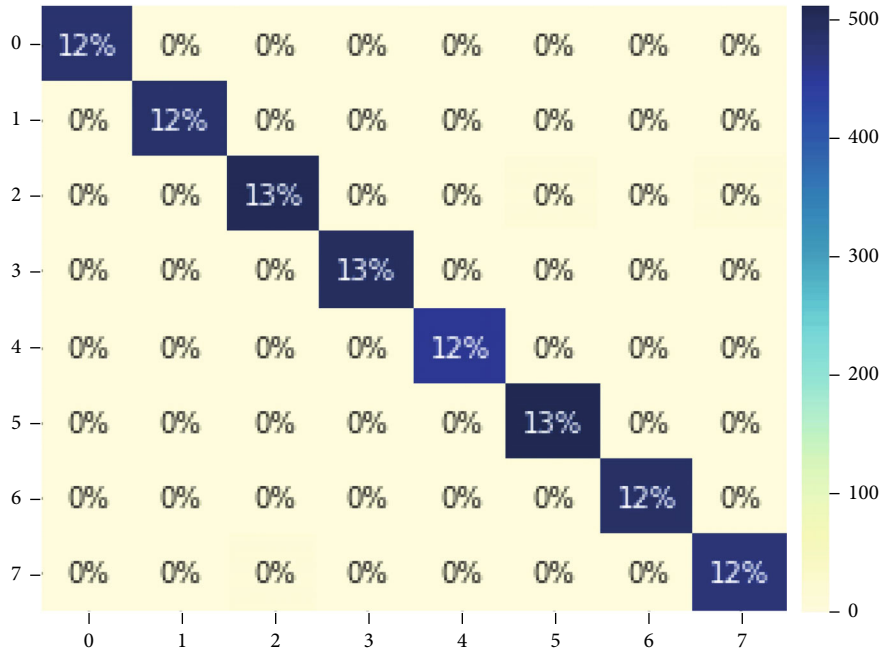

Figure 7: Confusion matrix of ANN model for the test set when using GWO-RF feature selection techniques.

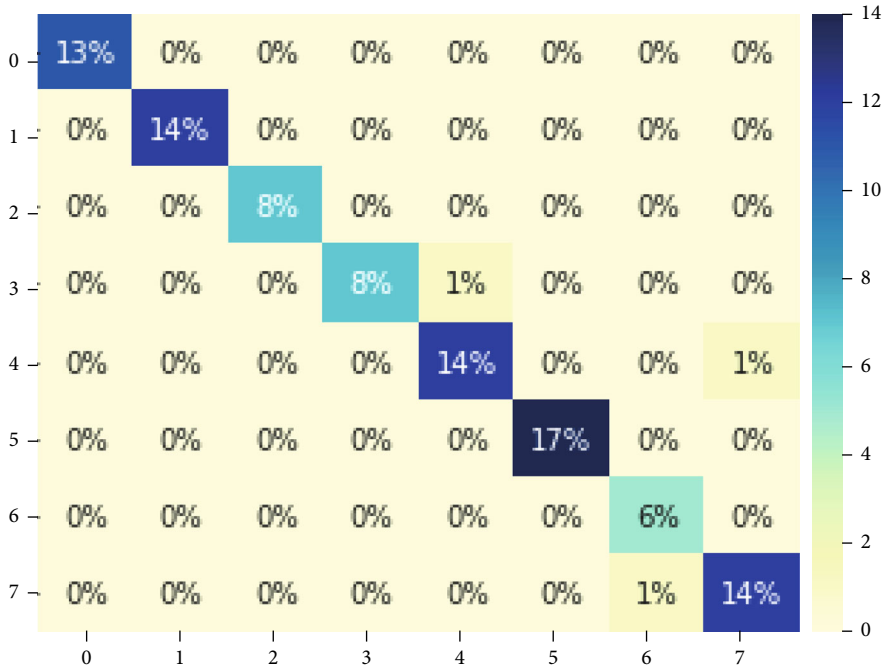

Figure 8: Confusion matrix of ANN model for the test set when using PSO-RF feature selection techniques.

mathematically can take the following equations:

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{C_1}.\overrightarrow{X_\alpha} - \overrightarrow{X}\right|, \overrightarrow{D_\beta} = \left|\overrightarrow{C_2}.\overrightarrow{X_\beta} - \overrightarrow{X}\right|, \overrightarrow{D_\delta} = \left|\overrightarrow{C_3}.\overrightarrow{X_\delta} - \overrightarrow{X}\right|, \tag{9}$$

$$\overrightarrow{X_1} = \left|\overrightarrow{X_\alpha} - \overrightarrow{A_1}.\overrightarrow{D_\alpha}\right|, \overrightarrow{X_2} = \left|\overrightarrow{X_\beta} - \overrightarrow{A_2}.\overrightarrow{D_\beta}\right|, \overrightarrow{X_\delta} = \left|\overrightarrow{X_\delta} - \overrightarrow{A_3}.\overrightarrow{D_\delta}\right|, \tag{10}$$

$$\vec{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3}. \tag{11}$$

(iii) Attacking the prey

The attacking of the prey is dependent on the vector $\vec{a}$ where its value is linearly updated in each iteration to range

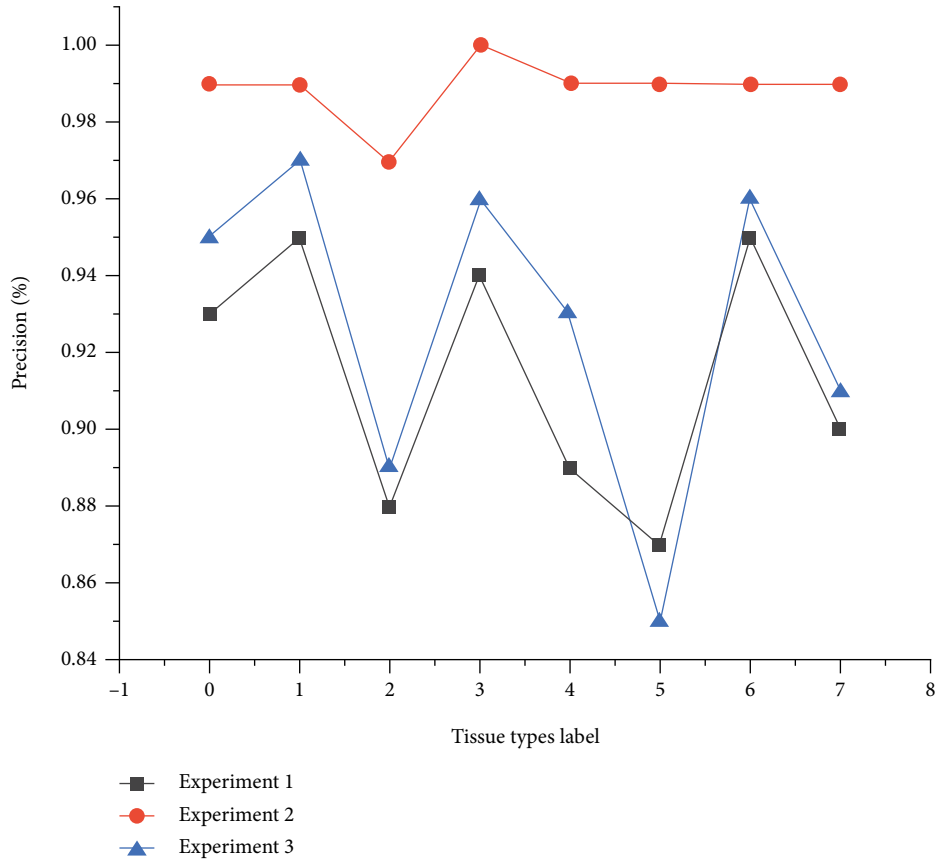

Figure 9: Classification precision of ANN model in all experiments.

2 to 0 according to the following equation

$$\overrightarrow{a} = 2 - t.\frac{2}{Max_iter}, \tag{12}$$

where t is the iteration number, and Max_iter is the total number of iterations allowed for the optimization.

(3) Particle swarm optimization (PSO)

PSO was initially developed by Kennedy and Eberhart in 1995 [36]. Due to its optimization accuracy, many researches were conducted on PSO to solve a variety of engineering optimization problems. PSO is a heuristic global optimization method as well as a swarm intelligence-based optimization algorithm. The concept of PSO emerges from the behavior of swarm particles and the social interaction between particles. When looking for food, the birds either disperse or move together to find it. The birds move from one location to another in search of food, and the bird closest to the food can smell it. PSO's basic algorithm is made up of n swarm particles, and the position of each particle represents a potential solution. According to the three principles, the swarm particle changes its position.

(1) Keep its inertia

(2) Update the condition relating to its optimal position

2.2. Update the Condition to the Swarm's Most Optimal Position

2.2.1. Feature Extraction. Before computing the texture features, all red/green/blue (RGB) images were converted to grayscale and depicted as the overall intensity distribution per class, as shown in Figure 2, to evaluate the difference between samples in each group.

It was clear from Figure 2 that the intensity values of all histological images of the applied dataset are very various. The distributions of categories (tumor, stroma, complex, and lymph) are convergent and contradictory to categories (debris, mucosa, adipose, and empty). Moreover, images of adipose and empty are of very high intensity, and the intensity distributions are very similar which makes it difficult to distinguish. We should find out their meanings as adipose may be very tough to distinguish from empty. And does empty mean, no image at all. To get a clearer understanding of the extent of overlaps, densities were plotted against intensities for each attribute in Figure 3, the categories that demonstrated significant overlaps with one another are (tumor and immune cells), (mucosal glands and simple stroma), and (complex stroma and muscol glands).

Comparing tumors to immune cells in Figure 3, it can be observed that overlap is seen around the value of 75 (both high in density), and the total range of overlap extends from around 70 to 80. Therefore, if pixel values alone were used to generate classifiers, these two classes would have the highest

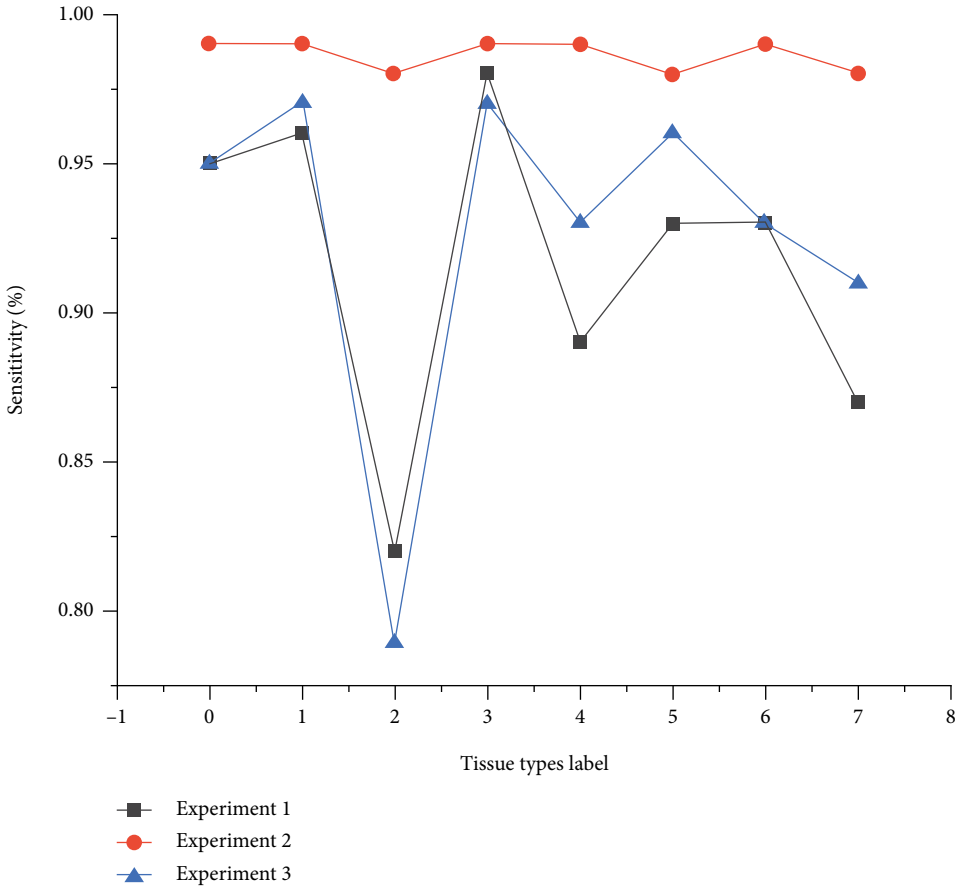

Figure 10: Classification of ANN model in all experiments.

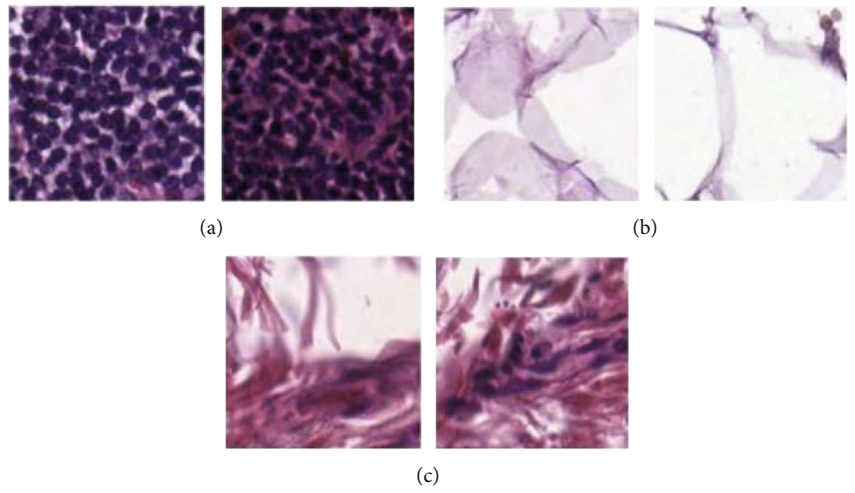

Figure 11: Example images of misclassification: (a) simple stroma, (b) adipose tissue, and (c) debris and mucus.

chance to get mixed up with one another. Beyond that, they have similar shapes of density, similar mean, and standard deviation, which makes it difficult to differentiate by the first moments. To differentiate them, a look at their potential differences at higher moments might be taken, or RGB data might be pursued and color information might be used. As for mucosal glands compared to other stromas, MG's extremely large values overlap a little with simple stroma's extremely small values, and the opposite happens for MG vs. complex.

The variation of image intensity distributions between classes inspired us to perform feature extraction based on the gray-level cooccurrence matrix (GLCM). The GLCM is a highly effective tool for extracting image features that

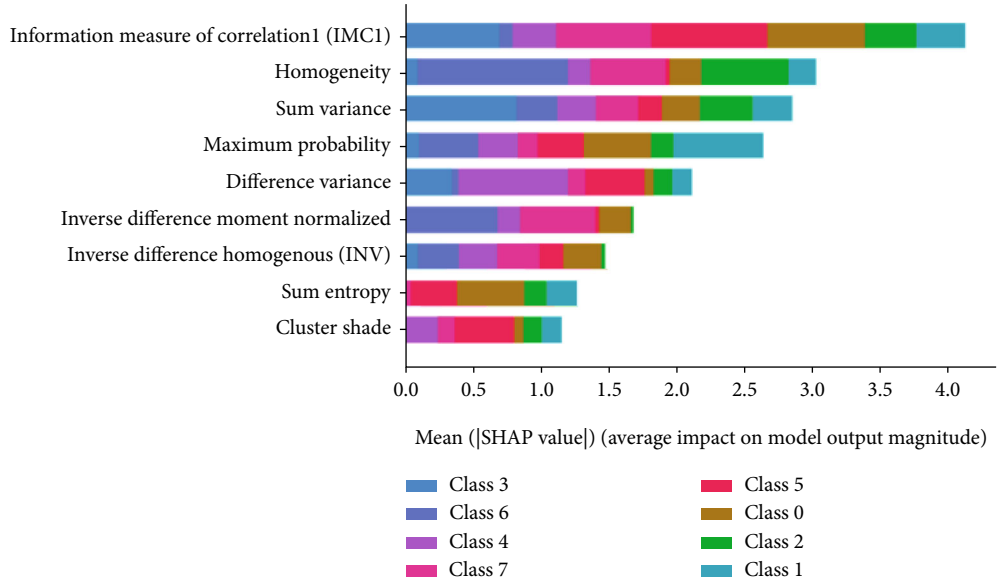

Figure 12: SHAP bar plot to show the global importance for each feature.

Table 3: Comparison between ANN model and state of art methods.

Performance metrics	ANN	KNN	RF
Accuracy	98.74%	94.2%	90.3%
Sensitivity	99%	94.5%	73%
Specificity	99%	93.2%	91.5%

map the probabilities of gray level cooccurrence based on pixel spatial relationships in various angular directions. We used four directions (0°, 45°, 90°, and 135°) and five displacement vectors (from 1px to 5px). To make this texture descriptor invariant concerning rotation, we averaged the GLCMs obtained from all four directions for each displacement vector.

GLCM usage can lead to extract features such as information measure of correlation1, autocorrelation, entropy, dissimilarity, correlation, cluster prominence, contrast, cluster shade, energy, homogeneity, maximum probability, sum of squares variance, sum variance, sum average, sum entropy, difference variance, difference entropy, information measure of correlation 2, inverse difference normalized (INN), inverse difference homogenous (INV), and inverse difference moment normalized.

2.3. The Proposed Algorithm. The hybrid algorithm was proposed based on GWO optimizer and RF (GWO -RF) to ensure recognition of multiclass texture analysis in colorectal cancer with better accuracy. The proposed algorithm takes the input of a set of features and returns the reduced subset of features and therefore upgrading the performance of the classifier model. The proposed algorithm process was summarized in Figure 4 and described in the following steps:

(1) Load the dataset

(2) Initialize the position of (α, β, and δ) with zero

(3) Initialize the variable score (α, β, and δ) with zero

(4) Randomly initialize the population of the wolves' store

(5) Calculate the fitness function (f) using equation (13)

(6) Find the sigmoid of X using equation (6)

(7) Update the positions of the agents

(8) Use the alpha pos for feature selection

(9) Apply the artificial neural networks (ANNs) classifier on the selected features

The fitness function f is calculated using the following equation:

$$f = \mathrm{PIM} + W_f \times \left(1 - \left(\frac{N_S}{N_t}\right)\right). \tag{13}$$

The permutation importance measure (PIM) for the permuted feature X_j was computed using the following equation:

$$\mathrm{PIM}_k\left(X^j\right) = \frac{1}{n_{\mathrm{tree}}} \sum_t \left(\mathrm{errOOB}_t^j - \mathrm{errOOB}_t\right), \tag{14}$$

where X_j is the permutated feature. errOOB_t^j is the perturbed sample. errOOB_t is the error of a single tree t on this OOB sample. W_f is the weight factor, creates a balance between the feature importance, PIM, and the number of selected features. The value of W_f is 0.7. N_S is the number of selected features. N_t is the total number of features present in the dataset.

Table 4: Comparison of works conducted on classify multiclass texture analysis in colorectal cancer.

Author(s)	Number of class classification	Method used	Classifier	Accuracy (%)
Wu et al. [24]	Four types	Color channel histograms, GLCM,	Logistic regression classifier	75.15%
Liu et al. [25]	Two types	GLCM	Support vector machines (SVM)	96.67%.
Liu et al. [26]	Three types	2D discrete wavelet (DW)	Neural networks	91.11%
J Malik et al.	Four types	Deep learning	CNN	95%
The proposed method	Eight types	GWO-RF	ANN	98%

3. Experimental and Parameter Setting

The human colorectal cancer dataset containing 5,000 histological images was employed for training and testing. We divide the dataset into two parts, 90% for training and validation (80% training, 100 10% validation) and 10% for testing, in 10-fold cross-validation. The training and testing procedures were conducted in Google Colab. The GWO algorithm and PSO parameters settings are outlined in Table 2.

For random forest, we tuned m_{tree} for selected 9 features by the approach described in Section (2.3), and we attained the best performance at m try = 8 and the value of n_{tree} = 100 was obtained.

To evaluate the proposed approach, three experiments were accomplished. Each experiment was divided into two phases. Feature selection methods were used in the first step to remove redundant and irrelevant features by looking for the best features in the colorectal cancer dataset. The highly successful classifier (ANN) was used in the second process, which depends on the optimum function subset obtained in the first phase. Figure 5 shows the general structure of the proposed method. In the first experiment, all features were used without applying feature selection techniques. Then, two experiments were performed by employing the proposed feature selection method (GWO-RF) and a counterpart (PSO-RF) method for comparison.

In addition, to evaluate our prediction ANN model classifier, other states of art classifiers KNN (k-nearest neighbor) and RF (random forest) were implemented and compared them with the proposed model.

The ANN network model consisted of three layers, namely, input, hidden, and output layer. The input layer in experiment 1 had several nodes equal to the number of all features that are extracted from the dataset, while in experiments 2 and 3, the number of input nodes proportional to the size of features was extracted from the GWO- RF model and PSO-RF. Three hidden layers are used. Experiments are also performed with one and two hidden layers, but the best results are obtained by using three hidden layers. The output layer consists of eight neurons that represent all categories of a dataset. Activation functions were RLU (rectified linear units) and sigmoid for the hidden and output layers. The network was trained in 1000 epochs using the stochastic gradient descent (SGD) optimization algorithm and a learning rate of 0.0001.

To ensure algorithm robustness, each experiment was run five times on each data set, and a confusion matrix was developed, and then the accuracy, precision, and sensitivity of the classifier were assessed using true positive (TP), true negative (TN), false-negative (FN), and false-positive (FP) values. These measures are calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\%, \tag{15}$$

$$\text{A Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\%, \tag{16}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \tag{17}$$

4. Results and Discussion

To evaluate the effectiveness of the implemented feature selection technique (GWO-RF) to enhance the ability of the ANN classifier to discriminate between eight distinct tissue types of human colorectal cancer histological images, a comparative analysis of the ANN classifier in three experiments was conducted using the confusion matrix and shown in Figures 6–8.

The output class in the y-axis of the confusion matrix for the test set refers to the classifier's prediction, while the target class in the x-axis refers to the true reference class. The confusion matrix's diagonal shows the percentages of correct predictions for each category. The percentage of predicted errors is shown by the figures in the off-diagonal.

By comparing the three matrices, the highest percentage of overall accuracy (98.74%) of the ANN classifier was achieved when using the GWO-RF feature selection technique in the experiment (2) which is shown in Figure 7. GWO-RF model recommended 9 features to be fed into the ANN classifier which were information measure of correlation1 (IMC1), cluster shade, homogeneity, maximum probability, sum variance, sum entropy, difference variance, inverse difference homogenous (INV), and inverse difference moment normalized.

The confusion matrix for the test set in Figure 8 indicates that the percentage of overall accuracy was reduced from 98.74% to 94.11% when using the PSO-RF feature selection technique in the experiment (3). The PSO-RF model selected 14 features which were contrast, autocorrelation, energy, information measure of correlation1 (IMC1), cluster shade, homogeneity, maximum probability, sum variance, sum entropy, difference variance, inverse difference homogenous (INV), and inverse difference moment normalized. On the

other hand, the results observed in Figure 6 revealed that the overall accuracy of the ANN significantly dropped in the experiment (1) to 91.45% when using all features (21 features) without applying feature selection techniques.

Based on the results of three experiments, the advantages of GWO over other PSO algorithms are simple implementation with a simple structure, low memory and computational requirements, fast convergence with continuous reduction of search space, and few coefficients of determination [37], two to avoid local minima and tune algorithm performance. A function that has only control parameters. This improves stability and robustness.

To examine the study more extensively, the precision and sensitivity for the ANN classifier were computed in each experiment, and the results were shown in Figures 9 and 10. In general, a significant gap between the percentage of precision and sensitivity for the ANN classifier was apparent when using the GWO-RF feature selection technique as compared to the PSO-RF feature selection method and when employing all features. The ANN classifier achieved overall precisions and sensitivities of 99%, 99% when applying GWO-RF, 94%, 94% when applying PSO-RF, and 92%, 92% when not using any feature selection method, respectively.

Furthermore, Figures 9 and 10 indicate that the performance of the ANN classifier for all eight tissue types in colorectal cancer was significant when using the proposed GWO-RF feature selection method which reveals the efficient performance of the proposed model. Example images of ANN classifier misclassification were shown in Figure 11. While the precision and sensitivity for the ANN classifier for all eight classes when applying the PSO-RF feature selection technique and all features were very close to each other. Also, we can observe that when employing all features and the PSO-RF feature selection method, the values of precisions were minimal when the ANN classifier discriminated classes 2 and 5 tissue types, and the lowest value of sensitivity was attained when discriminating class 2 tissue type. This outcome indicates the inefficient performance of the ANN classifier when employing the PSO-RF model and without applying any feature selection technique.

To interpret the ANN model and to show the relative importance of each feature and its effect on the model prediction performance, an aggregate SHAP bar graph was performed and shown in Figure 12. The SHAP bar graph plots the mean absolute SHAP value for each feature [38, 39]. It was noticed from the SHAP bar graph that the most important features that had a significant effect on the ANN model's prediction were ranked from the most important to the least important. The order of the features is the information measure of correlation1 (IMC1), homogeneity, sum variance, maximum probability, difference variance, inverse difference moment normalize, inverse difference homogenous (INV), sum entropy, and cluster shade.

It is also noticeable in Table 3 that the ANN classifier attained a significantly higher accuracy than the corresponding counterpart classifiers.

Table4 shows the comparison of similar works conducted to classify multiclass texture analysis in colorectal cancer. The highest accuracy results were achieved using our proposed GWO-RF feature selection process (98%). In addition, the ANN classifier could classify the largest number of classes (eight distinct tissue types) as compared to the other studies which signify the impressive outcomes for the proposed GWO-RF feature selection method.

5. Conclusion

A novel wrapper feature selection approach is proposed in this paper to decide the best feature subset for multiclass texture analysis classification in colorectal cancer histology. Gray wolf optimization (GWO) and random forest (RF) algorithm were utilized to find the best features in the histological images of the human colorectal cancer dataset. Then, based on the best-selected features, the artificial neural networks (ANNs) were applied to classify multi-class texture analysis in colorectal cancer. The results revealed that the presented work (GWO-RF) feature selection method has outperformed other state of art methods where it achieved overall accuracy, precision, and sensitivity rates of 98.74%, 98.88%, and 98.63%, respectively. These results are very promising and show the superior and efficient performance of the presented method in handling high dimensional and redundant data for early CRC diagnosis with high classification accuracy. The proposed model can be further applied in the future for the classification of another various cancer histology, reducing the time, cost, and effort of large searching space. The proposed model can be further applied in the future for the classification of another various cancer histology, reducing the time, cost, and effort of large searching space [40–46]. Moreover, the presented technology by the authors in [47–50] can be implemented to obtain better results.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. J. Granados-Romero, A. I. Valderrama-Treviño, E. H. Contreras-Flores et al., "Colorectal cancer: a review," *International Journal of Research in Medical Sciences*, vol. 5, no. 11, pp. 4667–4676, 2017.

[2] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global patterns and trends in colorectal cancer incidence and mortality," *Gut*, vol. 66, no. 4, pp. 683–691, 2017.

[3] W. Bernard, "Stewart and Christopher P. Wild.: world cancer report, ISBN 978-92-832-0429-9," Technical Report, 2014.

[4] M. A. Deif and R. E. Hammam, "Skin lesions classification based on deep learning approach," *Journal of Clinical Engineering*, vol. 45, no. 3, pp. 155–161, 2020.

[5] J. R. Petrella, L. M. Shah, K. M. Harris et al., "Preoperative functional MR imaging localization of language and motor areas: effect on therapeutic decision making in patients with potentially resectable brain tumors," *Radiology*, vol. 240, no. 3, pp. 793–802, 2006.

[6] R. N. Low, "MRI of colorectal cancer," *Abdominal Radiology*, vol. 27, no. 4, pp. 418–424, 2002.

[7] A. N. Esgiar, R. N. G. Naguib, B. S. Sharif, M. K. Bennett, and A. Murray, "Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa," *IEEE Transactions on Information Technology in Biomedicine*, vol. 2, no. 3, pp. 197–203, 1998.

[8] L. A. D. Cooper, J. Kong, D. A. Gutman, W. D. Dunn, M. Nalisnik, and D. J. Brat, "Novel genotype-phenotype associations in human cancers enabled by advanced molecular platforms and computational analysis of whole slide images," *Laboratory Investigation*, vol. 95, no. 4, pp. 366–376, 2015.

[9] M. B. Loughrey, P. J. Kelly, O. P. Houghton et al., "Digital slide viewing for primary reporting in gastrointestinal pathology: a validation study," *Virchows Archiv*, vol. 467, no. 2, pp. 137–144, 2015.

[10] L. R. Schad, W. Härle, I. Zuna, and W. J. Lorenz, "Magnetic resonance imaging of intracranial tumors: tissue characterization by means of texture analysis," *Zeitschrift für Medizinische Physik*, vol. 2, no. 1, pp. 12–17, 1992.

[11] R. A. Lerski and L. R. Schad, "The use of reticulated foam in texture test objects for magnetic resonance imaging," *Magnetic Resonance Imaging*, vol. 16, no. 9, pp. 1139–1144, 1998.

[12] Y. Liu, Z. Song, X. Xu et al., "Bidirectional GRU networks-based next POI category prediction for healthcare," *International Journal of Intelligence Systems*, 2021.

[13] F. Bianconi, A. Álvarez-Larrán, and A. Fernández, "Discrimination between tumour epithelium and stroma via perception-based features," *Neurocomputing*, vol. 154, pp. 119–126, 2015.

[14] N. Linder, J. Konsti, R. Turkki et al., "Identification of tumor epithelium and stroma in tissue microarrays using texture analysis," *Diagnostic Pathology*, vol. 7, no. 1, p. 22, 2012.

[15] M. Fang, Y. Chen, R. Xue et al., "A hybrid machine learning approach for hypertension risk prediction," *Neural Computing and Applications*, pp. 1–11, 2021.

[16] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, no. 6, pp. 610–621, 1973.

[17] F. G. Zöllner, K. E. Emblem, and L. R. Schad, "Support vector machines in DSC-based glioma imaging: suggestions for optimal characterization," *Magnetic Resonance in Medicine*, vol. 64, no. 4, pp. 1230–1236, 2010.

[18] L. Wang, C. Wu, L. Tang et al., "Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method," *Acta Geotechnica*, vol. 15, no. 11, pp. 3135–3150, 2020.

[19] J. N. Kather, C.-A. Weis, F. Bianconi et al., "Multi-class texture analysis in colorectal cancer histology," *Scientific Reports*, vol. 6, no. 1, p. 27988, 2016.

[20] A. Chaddad, C. Desrosiers, A. Bouridane, M. Toews, L. Hassan, and C. Tanougast, "Multi texture analysis of colorectal cancer continuum using multispectral imagery," *PLoS One*, vol. 11, no. 2, article e0149893, 2016.

[21] H. Shahabi, A. Shirzadi, S. Ronoud et al., "Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm," *Geoscience Frontiers*, vol. 12, no. 3, article 101100, 2021.

[22] M. Dash and H. Liu, "Feature selection for classification," *Intelligen Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.

[23] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[24] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: greening big data," *IEEE Systems Journal*, vol. 10, no. 3, pp. 873–887, 2016.

[25] Y. Liu, D. Li, S. Wan et al., "A long short-term memory-based model for greenhouse climate prediction," *International Journal of Intelligence Systems*, vol. 37, no. 1, pp. 135–151, 2022.

[26] Y. Liu, A. Pei, F. Wang et al., "An attention-based category-aware GRU model for the next POI recommendation," *International Journal of Intelligence Systems*, vol. 36, no. 7, pp. 3174–3189, 2021.

[27] M. Abbasi and M. Rafiee, "A calibrated asymptotic framework for analyzing packet classification algorithms on GPUs," *The Journal of Supercomputing*, vol. 75, no. 10, pp. 6574–6611, 2019.

[28] M. Abbasi, A. Najafi, M. Rafiee, M. R. Khosravi, V. G. Menon, and G. Muhammad, "Efficient flow processing in 5G-envisioned SDN-based internet of vehicles using GPUs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 5283–5292, 2020.

[29] H. Kalkan, M. Nap, R. P. W. Duin, and M. Loog, "Automated classification of local patches in colon histopathology," in *In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 61–64, Tsukuba, Japan, 2012.

[30] L. Jiao, Q. Chen, S. Li, and Y. Xu, "Colon cancer detection using whole slide histopathological images," *World Congress on Medical Physics and Biomedical Engineering*, vol. 2013, pp. 1283–1286, 2012.

[31] S. D. F. Hilado, L. A. G. Lim, R. N. G. Naguib et al., "Implementation of wavelets and artificial neural networks in colonic histopathological classification," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 18, no. 5, pp. 792–797, 2014.

[32] F. Ng, B. Ganeshan, R. Kozarski, K. A. Miles, and V. Goh, "Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival," *Radiology*, vol. 266, no. 1, pp. 177–184, 2013.

[33] S.-X. Rao, D. M. J. Lambregts, R. S. Schnerr et al., "Whole-liver CT texture analysis in colorectal cancer: does the presence of liver metastases affect the texture of the remaining liver?," *United European Gastroenterology Journal*, vol. 2, no. 6, pp. 530–538, 2014.

[34] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[35] S. Belongie, C. Carson, H. Greenspan, and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026–1038, 2002.

[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection, in 2005," in *IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1, pp. 886–893, San Diego, CA, USA.

[37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[38] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, CRC press, 1984.

[39] I. Ahmed, Y. Zhang, G. Jeon, W. Lin, M. R. Khosravi, and L. Qi, "A blockchain-and artificial intelligence-enabled smart IoT framework for sustainable city," *International Journal of Intelligence Systems*, 2022.

[40] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[41] Q. Al-Tashi, H. M. Rais, S. J. Abdulkadir, S. Mirjalili, and H. Alhussian, "A review of grey wolf optimizer-based feature selection methods for classification," *Evolutionary Machine Learning Techniques*, pp. 273–286, 2020.

[42] M. A. Deif, A. A. A. Solyman, M. H. Alsharif, S. Jung, and E. Hwang, "A hybrid multi-objective optimizer-based SVM model for enhancing numerical weather prediction: a study for the Seoul Metropolitan Area," *Sustainability*, vol. 14, no. 1, p. 296, 2022.

[43] M. A. Deif, R. E. Hammam, and A. Solyman, "Gradient boosting machine based on PSO for prediction of leukemia after a breast cancer diagnosis," *International Journal Advances Science Engeneering. Information. Technology*, vol. 11, no. 2, pp. 508–515, 2021.

[44] M. A. Deif, A. A. A. Solyman, and R. E. Hammam, "ARIMA model estimation based on genetic algorithm for COVID-19 mortality rates," *International Journal of Information Technology & Decision Making*, vol. 20, pp. 1775–1798, 2021.

[45] M. Deif, R. Hammam, and A. Solyman, "Adaptive neuro-fuzzy inference system (ANFIS) for rapid diagnosis of COVID-19 cases based on routine blood tests," *International Journal Intelligent. Engeneering and System*, vol. 14, no. 2, 2021.

[46] M. A. Deif, A. A. A. Solyman, M. H. Alsharif, and P. Uthansakul, "Automated triage system for intensive care admissions during the COVID-19 pandemic using hybrid XGBoost-AHP approach," *Sensors*, vol. 21, no. 19, p. 6379, 2021.

[47] H. Attar, A. T. Abu-Jassar, V. Yevsieiev, V. Lyashenko, I. Nevliudov, and A. K. Luhach, "Zoomorphic mobile robot development for vertical movement based on the geometrical family caterplillar," *Computationa Intelligence and Neuroscience*, vol. 2022, article 3046116, p. 19, 2022.

[48] H. H. Attar, A. A. A. Solyman, A.-E. F. Mohamed et al., "Efficient equalisers for OFDM and DFrFT-OCDM multicarrier systems in mobile E-health video broadcasting with machine learning perspectives," *Physical Communication*, vol. 42, p. 101173, 2020.

[49] A. A. Solyman, H. Attar, M. R. Khosravi, and B. Koyuncu, "MIMO-OFDM/OCDM low-complexity equalization under a doubly dispersive channel in wireless sensor networks," *International Journal of Distributed Sensor Networks.*, vol. 16, no. 6, p. 155014772091295, 2020.

[50] A. A. A. Solyman, H. Attar, M. R. Khosravi et al., "A low-complexity equalizer for video broadcasting in cyber-physical social systems through handheld mobile devices," *IEEE Access*, vol. 8, pp. 67591–67602, 2020.