IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 24, 2021, accepted March 5, 2021, date of publication March 10, 2021, date of current version March 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3065341

A Hybrid Feature Selection Optimization Model for High Dimension Data Classification

MOHAMMED QARAAD[1], SOUAD AMJAD[1], IBRAHIM I. M. MANHRAWY[2,3], HANAA FATHI[3], BAYOUMI ALI HASSAN[4], AND PASSENT EL KAFRAWY[3,5]

[1]Department of Computer Science, Faculty of Science, Abdelmalek Essaadi University, Tetouan 93030, Morocco
[2]Department of Basic Sciences, Modern Academy for Engineering and Technology, New Maadi 32511, Egypt
[3]Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Shebin El-Kom 32511, Egypt
[4]Department of Operations Research decision support, Faculty of Computer Science and Information, Cairo University, Giza 12613, Egypt
[5]School of Information Technology and Computer Science, Nile University, Giza 12613, Egypt

Corresponding authors: Mohammed Qaraad (m.qaraad@gmail.com) and Ibrahim I. M. Manhrawy (ibrahimmanhrawy@gmail.com)

ABSTRACT Feature selection is an NP-hard combinatorial problem, in which the number of possible feature subsets increases exponentially with the number of features. In the case of large dimensionality, the goal of feature selection is to determine the smallest possible features considering the most informative subset. In this paper, we proposed a hybrid feature selection optimization model for Cancer Classification called, ENSVM. Our model is based on using the Elastic Net (EN) method that regulates and selects variables for gene selection of genomic microarray data. We applied three different optimization techniques namely Social Ski-Driver (SSD), Randomized SearchCV (RS) and Elastic NetCV (ENCV) for determining Elastic Net with traditional Support Vector Machines for classification. To evaluate the model, we compared the results of applying ENSVM to seven genomic microarray data with the SSD-SVM model and SVM with (RBF) kernel without any feature selection method. The results of the comparison revealed the effect of ENSVM in selecting the optimal feature subset that maximized the classification performance. Accordingly, minimizing the number of features is significant when analyzing high dimensional data for performance nevertheless accuracy. Moreover, the ENSVM model is superior compared with the SSD-SVM model.

INDEX TERMS Cancer classification, feature selection, genomic microarray data, parameter optimization, elastic net (EN), social ski-driver (SSD).

I. INTRODUCTION

Feature selection is called the NP-hard combinatorial problem [1], where subsets of potential features increase exponentially with the number of features [2]. To increase the classification performance, it was necessary to rely on selecting a significate set of features, as it is a major step in most classification methods. A statistical technique is considered successful when relying on independent features. Therefore, the most informative subset of features is the independent features that are the target in feature selection. Elastic Net (EN) is a regularization method, with which high-dimensional data sets can be modelled in addition to variable selection. It is said that EN modelling can use high dimensional data sets for problems with small sample sizes [3]. However, this is not entirely true, as EN modelling is based on the best estimate of model parameters. If the number of variables (or features) p is much higher than the number of observations (or samples) n, that is ($n < p$) a the high-dimensional case. Optimization algorithms have been used to adjust the Elastic Net tuning parameter and choose the best value for transactions. A significant improvement has been shown for the Elastic Net parameter using optimization algorithms [4], [5]. SSD is a scalable optimization algorithm, its behaviour is inspired by various evolutionary optimization algorithms such as sine cosine algorithm (SCA) [6] and Gray wolf optimization (GWO) [7]. The main goal of the SSD is to attain optimum or almost optimum solutions in space. In this paper, three algorithms, social ski-driver (SSD), Randomized SearchCV (RS) [8] and Elastic NetCV (ENCV) [9], [10] are utilized for tuning the parameter of Elastic Net optimization alpha (α). The generated optimization model is used to select the optimum subset of features causing enhancing the performance of cancer classification. To prove the efficiency of our model we compared the results of applying the ENSVM model with three techniques (SSD, RS, ENCV) to seven genomic microarray data with SSD-SVM, and SVM-RBF kernel without any feature selection method. The result shows

The associate editor coordinating the review of this manuscript and approving it for publication was Khin Wee Lai.

the effect of ENSVM compared with the ski-driver (SSD), Randomized Search CV, Elastic Net CV with SVM classifier and SVM-RBF kernel on high-dimensional data analysis. The research paper has been arranged as follows: section 2. refines the related work. In this section, we describe the earlier research in the study. The Elastic Net, SSD-SVM and other used methods are explained in section 3 The proposed model is introduced in section 4; Experimental scenarios and discussions are introduced in section 5. The final observations and future work can be found in section 6.

II. RELATED WORK

Many significant research efforts have been produced in the last years to study the cancer microarray data feature selection and classification. We will provide an overview of some of this work as a context for the research discussed in the remainder of the paper. Esra Pamukcu [11] proposed a new adaptive elastic net (AEN) model using hybrid covariance estimators (ECE) and information complexity (ICOMP). The proposed model examined the implementations of Monte Carlo experimental with simulation data to study the performance of the EN model. The results showed that AEN regression modelling is a good approach that can be used in high-dimensional data with undersized data. Algamal and Lee [12] suggested AAElastic for consistent gene selection while promoting the aggregation outcome on high-dimensional cancer classification. Based on the results based on three real microarray data sets, AAElastic proved to be competitive, effective and gave a positive outcome in terms of a) classification accuracy, sensitivity, and specificity b) consistency of gene selection.

Li *et al.* [13] suggested a simple and efficient resampling technique called RVOS to solve the problem of sample size analysis. Large sample size and category imbalance are the main features of microarray data. First, they developed the RFE and called it the VSSRFE version. VSSRFE proposes to decrease the recursion time by using a larger step size and a keep step size, while the number of features to be eliminated becomes less and less to ensure the quality of the selected significant genes. VSSRFE offers an interesting idea that can speed up the gene selection process. Also, they combined the purely linear support vector classifier LLSVM with the VSSRFE compared to existing methods, LLSVM-VSSRFE has evolved into an efficient and effective trait selector with potential in gene selection.

Wang *et al.* [14] suggested an adaptive elastic net with conditional mutual information. The algorithm shows that the suggested learning algorithm can contribute to the adaptive clustering effect and is robust to the selection of outliers in the microarray dataset. It is also shown that AEN-CMI contributes to the adaptive clustering effect by assessing the significance of gene ranking. Cui *et al.* [15] suggested a technique that can encapsulate the structural correlation between feature methods in the feature selection process, and display features in the form of graphics and vertices. Therefore, the obtained information matrix is utilised to produce an optimization model to determine the object with the greatest relevance and least redundancy to the objective function. This is to eliminate the loss of information due to the use of the graphical illustration of the features. They used an elastic net regression model to formulate the feature selection problem. The experimental ADMM results were used on the real dataset to solve the model, which shows that their method overcomes several well-known feature selection techniques. In [16] introduced a novel algorithm named Ensalg to assess the best regularization path for an elastic net associated with elements caused by sparsity in a target. Compared to existing algorithms, their method can handle the so-called ''many-at-a-time'' situations, when several variables go to zero and/or zero to zero at the same time.

Dhrif *et al.* [17] and others have developed a combinatorial PSO algorithm called COMB-PSO that adapts to multivariate data by selecting the smallest subsets of genes that can reliably classify samples. In particular, COMB-PSO develops coding, convergence rate, divergence control and diversity management of the classic OSP algorithm and balances exploration. Tharwat *et al.* [18] proposed a Dragonfly Algorithm (DA) model, called DA-SVM, which is used to optimize SVM parameters. The proposed model can find the optimal values of the SVM parameters and avoid the optima local problem. In [19], Tharwat and Hassanien used quantum behaved particle swarm optimization (QPSOs to optimize SVM parameters) to reduce classification error, the proposed model can obtain the optimal values of parameters in an SVM model. The outcomes also explained lower classification error rates correlated to the standard GA and PSO algorithms.

Shekar proposed in [20] a classification of microarray cancer utilising optimized hyper-parameters of random forest tree applying a grid search approach. The optimization of the RF algorithm is given to obtain the best parameters to validate the method. The experimental results of the proposed work showed an improvement compared to modern methods. El-Kafrawy and others in [21] introduced a Multi-Feature selection for De-novo acute myeloid leukaemia. They studied a collection machine learning algorithm: Logistic Regression (LR), Support Vector Machine (SVM), Random Forest Classifier (RF), Naive Bayesian Algorithm (NB), and Decision Tree (DT) for the National Cancer Institute's AML dataset (NCI), in Egypt. Experimental results showed that LR gives tremendous accuracy (92.30%) and giving the lowest possible error rate. In [22], they discussed different techniques for decreasing the size of data on multi-dimensional microarray cancer, which is required for a significant effect when increasing the amount of data. Different techniques for feature selection for these microarrays have been described, as well as their advantages and disadvantages.

A modified mutation strategy-based flower pollination algorithm (MFPA) has been successfully applied to optimize SVM parameters in [23]. The MFPA algorithm is used to determine and optimize the penalty coefficient and kernel of SVM parameters to obtain the best combination of SVM parameters. Their experimental results highlighted the MFPA SVM model had the best learning and generalization capacity

compared to the existing BA model and PSO-SVM model. The combination of Particle Swarm Optimization (PSO) plus Vector Support Machine (SVM) was applied to develop the power grid security, early warning model. In [24] based on the data, the traditional SVM model and the improved PSO-SVM model are used to perform empirical tests in the field of power grid construction safety, improving the predictive accuracy of the developed model.

Our Hybrid feature selection optimization model (ENSVM) for Cancer Classification is different in:

- Hybrid three different techniques to classify cancer microarray data.
- Elastic Net feature selection technique used to reduce the dimensionality and noise of the data to overcome the overfitting problem.
- Social ski-driver (SSD), Randomized SearchCV (RS) and Elastic NetCV (ENCV) optimization techniques are used to optimize hyper tuning parameter alpha of feature selection method (Elastic net) to:
  a) Select a best or nearly optimal subset of the gene which is informative with the greatest relevance and importance.
  b) Increase the performance of cancer classification.

III. PRELIMINARIES

A. SVM WITH RBF KERNEL

One of the Machine learning methods supervised learning, which is considered one of the most important methods of solving regression and classification. Support vector machine (SVM) is one of the pillars for classification and is considered one of the most important factors for solving the classification problem, as in [27] a tumour classifier. SVM is based on obtaining a hyperplane that optimally separates the features into various regions. The SVM-RBF-kernel is one of the functions whose value is based on the distance between the origin and some point C, its parameter means the penalty parameter described by the error term. It dominates the trade-off within a smooth solution boundary and correct classification of the training points, γ gamma is the parameter of the non-linear hyperplane. A higher gamma value will try to fit the training dataset exactly, in other words, the classifier tries to find the best fit. Thus, increasing the gamma value leads to overfitting.

Given the training sample of instance-label pairs $(x_1, y_1), \ldots . (x_i, y_i), i = 1, \ldots .., l, x_i \in R^n, y_i \in \{\pm 1\}$, SVMs demand the solution of the following primal problem [29].

$$\min_{w,\xi,b} \frac{1}{2} W^T W + C. \sum_{i=1}^{l} \xi_i (1) \tag{1}$$

at x_i the learning vector was displayed above the high-dimensional space by the mapping function $(\varphi)\, asz_i = \varphi(x_i).C > 0, z_i$ is the penalty of parameter option for error.

We generally solve equation (1) by determining the following double problem:

$$\min_{w,\xi,b} F(\vartheta) \frac{1}{2} \vartheta^T \Psi \vartheta - e^T \vartheta$$
$$\text{s. t. } 0 \le \vartheta_i \le C, i = 1, \ldots \ldots ., l, y^T \vartheta = 0 \tag{2}$$

where e means the vector of all ones and Ψ is an l by l positive semidefinite matrix. The $(i, j)^{\text{th}}$ element of Ψ is given by $\Psi_{i,j} \equiv y_i y_j K\left(x_i x_j\right), K\left(x_i, x_j\right) \equiv \varphi^T(x_i)\, \varphi\left(x_j\right)$, Ψ i,j is called the kernel function, at $\{\vartheta_i\}_{i=1}^{i}$ are Lagrange multipliers, and $w = \sum_{i=1}^{l} \vartheta_i y_i \varphi(x_i)$, w is the weight vector. The classification decision function is

$$sgn\left(W^T \cdot \phi(x) + b\right) = sgn(\sum_{i=1}^{l} \vartheta_i y_i \cdot K\left(x_i, x_j\right) + b \tag{3}$$

In this research, we utilised the kernel function $K(x_i, x_j)$ has several forms. This research of discussion of the paper is the main function of the RBF, which is widely used among them. The RBF kernel function is as follows:

$$K(x, x_i) = e^{-\frac{\|x - x_i\|^2}{2\sigma^2}} \tag{4}$$

which can transform the parameter sigma σ is defined by the user. The parameters of SVM-RBF indicate the C error penalty option and the RBF parameter σ, i.e. the parameters are namely (C, σ).

B. ELASTIC NET

The elastic network combines the L-1 lasso rate penalty and the L-2 rate penalty of ridge regression [3]. Elastic Net automatically selects variables and allows you to select more than n variables (observable values). The negative binomial is used when selecting and classifying features the probability function is used with a net elastic penalty. Evaluate the model by minimizing the following objective function.

$$\arg\min_{\beta_0,\beta} \left\{ \left[\frac{1}{N} \sum_{i=1}^{N} y_i \left(\beta_0 + x_i^T \beta\right) - \log\left(1 + e^{\beta_0 + x_i^T}\right) \right] + \lambda \left[\frac{(1-\alpha)||\beta||^2}{2} + \alpha||\beta|| \right] \right\} \tag{5}$$

In the preceding equation, t the loss function is the first part that takes the error rate in classification, and the second ingredient is the regularization term. In the equation, (1), λ and α are called tuning parameters. The penalty of the elastic net is controlled by the parameter α, which connects bridge regression ($\alpha = 0$) and lasso ($\alpha = 1$). By increasing alpha α the number of genes (features) of the classifier is reduced.

For example, we apply Elastic Net to a microarray data set called Singh for prostate cancer disease with 12600 features (genes) and 102 samples. We set three values for alpha (α) as 0.01, 0. 001 and 0.001 it's noticed that the number of genes is different according to the value of alpha (α) as seen in figure 1.

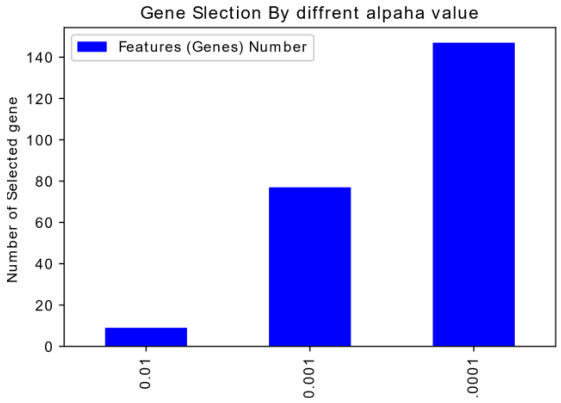

FIGURE 1. Features (Genes) Number.

TABLE 1. Features (genes) number.

Alpha value	Features (Genes) Number
0.01	69
0.001	77
0.0001	147

As we see in table 1 the value of alpha is affecting the gene number and that can eliminate important genes. To ignore that we need to have an optimal value for alpha (α) to get optimal features (genes), which are considered as informative features. Selecting an optimal feature set increases the performance of the classifier. The optimal value of α is evaluated by minimizing the specified objective function. Some small coefficients are reduced to zero, and the corresponding predictor variables are excluded from the final model, which are called ''irrelevant'' features. Other features are considered informative features. The final technique can be applied to predict future results as new data becomes available.

C. ENCV-ENSVM MODEL

EN uses a combination of ($\ell 1$) lasso and ($\ell 2$) ridge regression penalties, this has interesting characteristics of the grouping effect (selective grouping). Hence, one can select a group of related genes and analyse high-dimensional data efficiently [25]. Elastic Net CV (ENCV) is an extension to EN that calculates the k-fold cross-validation mean square prediction error. ENCV encapsulates an inner cross-validation loop to optimize Elastic Net tuning parameter alpha, seen in equation (1), and spare redundant computation. In ENCV-ENSVM we estimate the optimal tuning values for the elastic net alpha α utilising the Elastic Net CV algorithm. The initialization of ENCV parameters presented in Table 3 and the details of the model is presented in section 4.

D. SSD_SVM MODEL

SSD_SVM [26], proposed by Tharwat, is used to optimize RBF-kernel hyperparameters. The principal aim of the social ski-driver (SSD) is to obtain the best or near-optimal solution in space. The SSD_SVM model runs in two steps: the first step is the process of optimizing the SSD algorithm for the kernel function parameters (C and g) and the penalty function. The second step is the process of predicting the regression of the SVM model for sampled data. The SSD can be represented by $M_i^t = \frac{X_\alpha + X_\gamma + X_\beta}{3}$, $X_i^{t+1} = V_i^t + X_i^t$, at

$$V_i^{t+1} = \begin{cases} c\sin(r_1)\left(P_i^t - X_i^t\right) + \sin(r_1)\left(M_i^t - X_i^t\right), \; if \; r_2 \leq 0.5 \\ c\cos(r_1)\left(P_i^t - X_i^t\right) + \cos(r_1)\left(M_i^t - X_i^t\right), \; if \; r_2 > 0.5 \end{cases} \quad (6)$$

where M_i is the mean global solution as in Gray Wolf Optimizer GWO), X_α, X_γ and X_β the best solutions. V_i is the velocity of X_i, r_i is the uniformly random number is generated in the range of [0.1], P_i the agent is the agent's best solution and c is a parameter that is utilised to balance exploitation and exploration. The calculation is as follows: $c^{t+1} = \alpha c^t$, at t is the current iteration, $0 < \alpha < 1$ is applied to decrease the value of C. Consequently, $C \to 0$ and t is the maximum number of iterations.

E. RS-SVM MODEL

Randomized Search CV [28] is used in the same way as Grid Search CV [29]. However, it replaces the grid search CV for random sampling parameters in the parameter space for continuous variable parameters. Randomized Search CV will scan it as a distribution, which is not possible for a grid search. The search capability depends on the n_iter parameters defined. If n_iter parameters are higher, the prediction accuracy of the algorithm is higher, but the search time is longer. In RS-SVM, Randomized Search CV used to optimize the kernel function parameters C and g as a penalty function. The second step is to predict the SVM regression process of sampled data.

IV. THE PROPOSED MODEL (ENSVM)

The core of the ENSVM model was defined in PSO-ENSVM [44] as shown in figure 2 for the classification of cancer microarray data. In the new ENSVM model, we replaced the optimization technique PSO with SSD. This replacement is implemented to enhance model performance.

In the proposed ENSVM, we use the three optimizers, social ski-driver, Randomized Search CV and Elastic Net CV to estimate the best value for the elastic net tuning parameter α. The principal process of this model is:

1) Initialize the total number of particles, each particle has a random position in one-dimensional space, and each dimension has a random velocity. 2) The dataset microarray is pre-processed. This step is designed to (a) exclude features in a larger range of values, dominant features in a smaller range, (b) avoid numerical difficulties in the calculation process, and (c) scale each feature within the range [0, 1]. 3) The elastic net uses a combination of ($\ell 1$) and ($\ell 2$) penalties to reduce the coefficients of the ''unimportant'' features to 0 or near zero. Consequently, it selects the top n features according to the absolute value of the coefficients according to the penalty α of the elastic net within its range [0, 1]. A value that is superior

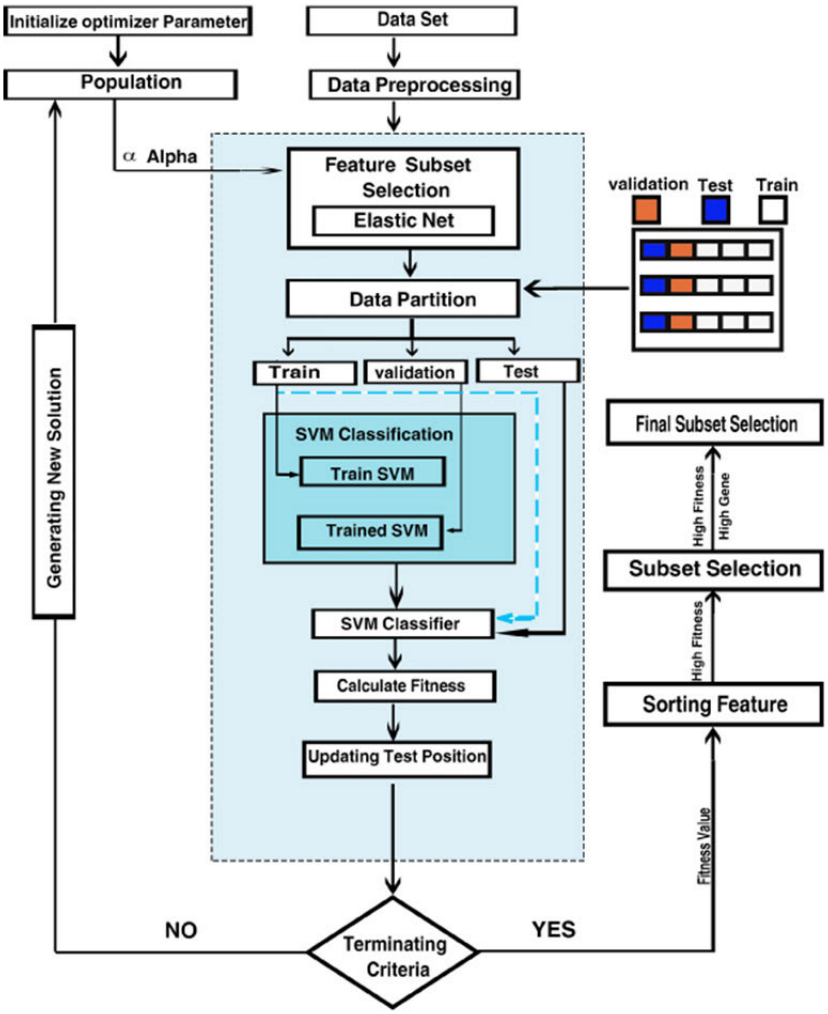

FIGURE 2. The proposed ENSVM Model.

to or close to the optimal control α indicates the best subset of the selected features. 4) Adjusting the SVM hyperparameters, the sample data is split into training data and verification data. 5) Rate the applicability of each particle to SVM. The applicability of each particle in this model is sensitivity. If the fitness is better than the best fitness of the particle, the position vector for the particle is preserved. If the fitness of the particle is better than the globally best fitness, the position vector remains globally best. Finally, the speed and position of the particles are updated until the stop criterion is met. If the closure criterion is met, the features are sorted by fitness function value, then the subset features are selected according to the highest fitness value, and finally selected subset feature has the highest fitness value and the gene value with high correlation. If the completion condition is not fitted, the optimizer generates a new solution and repeats operations 1 through 5 until the completion condition is satisfied.

A. PARAMETERS' INITIALIZATION

At this point, the proposed ENSVM model is initialized for each method, such as the population number and the maximum number of iterations, as shown in Table 3. In the proposed model, ENSVM provides parameter values for the SVM classifier, the training data is used to train the SVM classifier based on α value.

Our model has only one parameter, therefore, the exploration space is one-dimensional, representing every point in the space. The population of particle positions is randomly initialized, and the search range of parameter C is the default setting because we have not optimized the classifier function.

TABLE 2. The detailed settings.

Name	Detailed settings
Hardware	
CPU Core (TM)	Core (TM) i5-7500
Frequency	3.10 GHz
RAM	4 GB
Hard Drive	160
Software	
Operating system	Windows 7
Language	programming langue Python 2.7.

TABLE 3. Parameters' initialization.

Population Size	Max Iteration Number	Dataset Scaling	Validation Method	Penalty Factor (alpha)	Trails
20	50	[0,1]	10-fold	(1e-3, 1e-4)	10
Particular Parameters					
SVM(RBF): $C\ min\ = 0.01, Cmax = 1000, gammaMin = 0.0001, gammaMax = 0.1$					
SSD: $CognitionLearningFactor1 = 2.0; SocialLearningFactor2 = 2.0; a = 2.0 - current_iteration * (2/Max_iterations);$					
SVM(RBF): $C\ min\ = 0.01, Cmax = 1000, gammaMin = 0.0001, gammaMax = 0.1$					
Randomized Grid Search: $estimator = ElasticNet; scoring =' r2'; n_jobs = -1$					
SVM(RBF): $C\ min\ = 0.01, Cmax = 1000, gammaMin = 0.0001, gammaMax = 0.1$					
Elastic NetCV: $tol = 0.0001; selection = cyclic; n_jobs = -1$					
SVM(RBF): $C\ min\ = 0.01, Cmax = 1000, gammaMin = 0.0001, gammaMax = 0.1$					

TABLE 4. Dataset (DS) characterization.

Dataset	Disease	No. Samples	No. Features	IR
Singh (D1) [31]	Prostate Cancer	102	12600	0
Gordon (D2) [32]	Lung Cancer	181	12533	4.39
Chowdary (D3) [34]	Breast Cancer	104	22283	1.48
Alon (D4) [35]	Colon Cancer	62	2000	1.82
Chin (D5) [30]	Breast cancer	118	22215	1.74
West (D6) [36]	Breast Cancer	49	7129	1.0
Golub (D7) [33]	Leukemia	72	7129	1.88

The search range of α is limited to 1e-3 and 1e-4 minutes maximum; the seat position is initialized randomly. As these restrictions increase, the search space will expand. Therefore, more particles are needed to find the best solution, which leads to an increase in calculation time and a decrease in the convergence rate. Table 3 lists the initialization of the variable for each method.

B. FITNESS EVALUATION

In the model, the data is divided into three groups. A training set utilized for training, and a validation set utilized to assess the trained model during iterations. Finally, a test set applied to test the final model. In other words, the training and validation sets are used to build a classification model and determine the appropriate parameters for it, then the test set used to test the fully trained SVM model. Therefore, our model used invisible data for estimation.

After training a machine learning model with high dimensional data, measuring accuracy only is not enough to test the model. Therefore, our goal is to maximize the sensitivity S, because in the minority class the number of samples is minimal, as shown below:

$$\text{Minimize::}\ F = -S = \left(\frac{TP}{TP + FN}\right) \tag{7}$$

where TP denotes the number of true positive values, and TN denotes the number of true negative values. Therefore, for each agent position, the training set is utilized to train the SVM classifier, and the validation set is utilized to assess the sensitivity rate. The best solution is where the α value reaches its maximum sensitivity when the termination condition is satisfied, then the algorithm ends. Otherwise, it will continue to develop the next generation and when the maximum number of iterations is reached, the proposed algorithm ends.

V. RESULTS AND DISCUSSION

In this section, we present the results of the various experiments conducted to evaluate the performance of the proposed ENSVM model. The platform used to test the ENSVM model is a PC with the detailed settings as shown in Table 2.:

Table 4 uses seven benchmark sample microarray reference datasets and binary classes. These datasets involve research on human cancer, including breast cancer, colon cancer, Leukemia, lung cancer and prostate cancer. The input specification of the different datasets is given in table 4, such as the number of features, number of samples, imbalance rate (IR). IR defines the number of instances of the majority class for each instance in the class.

In our experiments, ten-fold cross-validation is used to evaluate the results of all methods, in which the data is randomly divided into (approximately) 10 equal-sized subsets for k = 10 runs. Hence it is performed multiple times over time, with one subset serving as a test group, one as a validation group, and the rest as training groups as shown in Figure 2. The results obtained are expressed as mean ± standard deviation. Besides, in all experiments, 50 iterations are conducted. In our experiment, we used three

TABLE 5. SVM (RBF) Kernel, RS-SVM optimize RBF kernel C and gamma, and RS- RS-ENSVM in terms of (specificity, sensitivity, and AUC).

Dataset	SVM(RBF)Kernel (default parameter)			RS-SVM optimize RBF kernel C and gamma			RS-ENSVM		
	AUC	Sen	Spec	AUC	Sen	Spec	AUC	Sen	Spec
D1	0.87 ± 0.09	0.90 ± 0.13	0.85 ± 0.11	0.86 ± 0.08	0.92 ± 0.10	0.81 ± 0.18	0.95 ± 0.05	0.98 ± 0.06	0.92 ± 0.10
D2	0.98 ± 0.04	0.99 ± 0.03	0.97 ± 0.07	0.96 ± 0.07	0.99 ± 0.02	0.93 ± 0.13	0.98 ± 0.05	1.00 ± 0.00	0.97 ± 0.10
D3	0.91 ± 0.15	0.97 ± 0.06	0.85 ± 0.30	0.96 ± 0.05	0.98 ± 0.05	0.93 ± 0.11	0.93 ± 0.15	0.99 ± 0.04	0.88 ± 0.30
D4	0.78 ± 0.18	0.62 ± 0.30	0.95 ± 0.15	0.73 ± 0.21	0.57 ± 0.45	0.90 ± 0.17	0.89 ± 0.16	0.83 ± 0.31	0.95 ± 0.10
D5	0.82 ± 0.09	0.70 ± 0.16	0.94 ± 0.09	0.80 ± 0.13	0.62 ± 0.29	0.97 ± 0.05	0.86 ± 0.10	0.76 ± 0.20	0.97 ± 0.05
D6	0.28 ± 0.20	0.35 ± 0.29	0.22 ± 0.35	0.57 ± 0.13	0.80 ± 0.21	0.35 ± 0.39	0.79 ± 0.18	0.83 ± 0.22	0.75 ± 0.31
D7	0.92 ± 0.09	0.96 ± 0.09	0.88 ± 0.18	0.85 ± 0.15	1.00 ± 0.00	0.70 ± 0.31	0.98 ± 0.05	1.00 ± 0.00	0.97 ± 0.10
P-value							**AUC**	**Sen**	**Spec**
Friedman testing				**SVM & RSSVM & RSENSVM**			0.010723	0.01312	0.03912

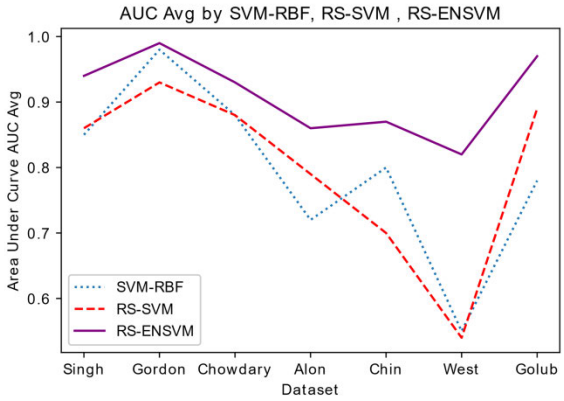

FIGURE 3. AUC of SVM (RBF) Kernel, RS-SVM, and RS-ENSVM in terms of AUC.

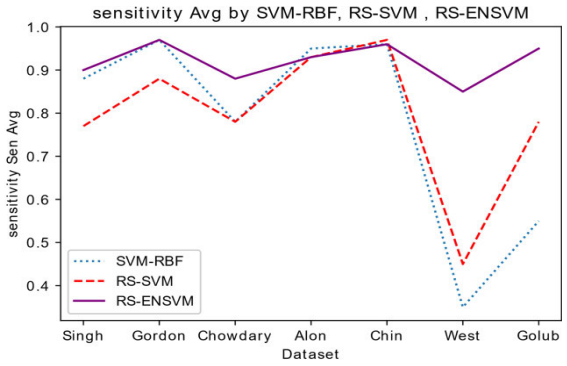

FIGURE 4. The sensitivity of the SVM (RBF) Kernel, RS-SVM, and RS-ENSVM.

assessment methods, namely specificity (Spec.). Specificity is the negative sampling rate for correct classification. It is defined as TNR $= \frac{\text{TN}}{\text{TN+FP}}$, where TN represents the number of true negative results and FP is the number of false positives. In our model, specificity is the ability to correctly identify a person who is not as sick as unhealthy using (TEST). AUC indicates the area under the receiver operating curve (ROC). Measured

$$\text{AUC} = \frac{1 + \text{TPR} - \text{FPR}}{2}.$$

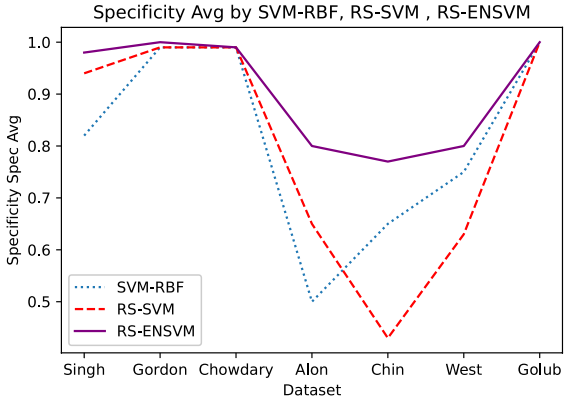

FIGURE 5. The specificity of the SVM (RBF) Kernel, RS-SVM, and RS-ENSVM.

Two statistical testing methods are also used to evaluate the performance of our model. The Wilcoxon test [42] is a nonparametric statistical test that compares two paired groups, the goal of the test is to determine if two or more sets of pairs are different from one another in a statistically significant manner. Friedman test [43] is used for data with three or more correlated or repeated outcomes whose distribution is not normal. The null hypothesis is that the distribution is the same across repeated measures.

The source code and relevant data can be downloaded from https://github.com/MohammedQaraad/optimizeENSVM.

A. EXPERIMENTAL RESULTS

This section will present the results of five experiments to assess the proposed ENSVM model. The purpose of the first experiment, in section 1, is to evaluate the ENSVM model using Randomized SearchCV (RS), called RS-ENSVM with RS-SVM and SVM (RBF) Kernel. In the second experiment in section 2, the goal is to use the social ski driver (SSD), called SSD-ENSVM with the SSD-SVM and SVM (RBF) Kernel. The third experiment in section 3, Elastic NetCV (ENCV), called ENCV-ENSVM, evaluated with SVM (RBF) Kernel. In all three experiments, we performed a statistical p-value test to determine the significance of the results.

TABLE 6. The SVM (RBF) Kernel, SSD -SVM optimize RBF kernel C and gamma, and SSD - ENSVM in terms of (specificity, sensitivity, and AUC).

	SVM(RBF)Kernel (default parameter)			SSD-SVM optimize RBF kernel C and gamma			SSD-ENSVM		
Dataset	**AUC**	**Sen**	**Spec**	**AUC**	**Sen**	**Spec**	**AUC**	**Sen**	**Spec**
D1	0.85 ± 0.11	0.88 ± 0.10	0.82 ± 0.17	0.89 ± 0.09	0.87 ± 0.12	0.92 ± 0.13	0.93 ± 0.09	0.90 ± 0.10	0.96 ± 0.12
D2	0.98 ± 0.04	0.97 ± 0.07	0.99 ± 0.03	0.76 ± 0.24	0.53 ± 0.48	1.00 ± 0.00	0.99 ± 0.04	0.97 ± 0.07	1.00 ± 0.00
D3	0.88 ± 0.15	0.78 ± 0.31	0.99 ± 0.04	0.87 ± 0.19	0.75 ± 0.39	0.99 ± 0.04	0.93 ± 0.15	0.88 ± 0.30	0.99 ± 0.04
D4	0.72 ± 0.18	0.95 ± 0.15	0.50 ± 0.32	0.80 ± 0.19	0.93 ± 0.16	0.67 ± 0.32	0.85 ± 0.19	0.90 ± 0.17	0.80 ± 0.24
D5	0.80 ± 0.11	0.96 ± 0.06	0.65 ± 0.20	0.62 ± 0.18	0.97 ± 0.05	0.26 ± 0.40	0.85 ± 0.09	0.95 ± 0.07	0.74 ± 0.20
D6	0.55 ± 0.28	0.35 ± 0.45	0.75 ± 0.31	0.52 ± 0.15	0.25 ± 0.34	0.78 ± 0.32	0.73 ± 0.24	0.73 ±0.28	0.72 ± 0.33
D7	0.78 ± 0.18	0.55 + 0.36	1.00 ± 0.00	0.80 ± 0.24	0.60 ± 0.49	1.00 + 0.00	0.97 ± 0.07	0.95 + 0.15	1.00 ± 0.00
p-value							**AUC**	**Sen**	**Spec**
Friedman testing					**SVM & SSD-SVM & SSD-ENSVM**		0.0430	0.0120	0.1025

TABLE 7. AUC, specificity, and Sensitivity of SVM (RBF) Kernel, Elastic Net CV (ENCV).

Dataset	SVM(RBF)Kernel (default parameter)			ENCV-ENSVM		
	AUC	Sen	Spec	AUC	Sen	Spec
D1	0.85 ± 0.11	0.88 ± 0.10	0.82 ± 0.17	0.94 ± 0.07	0.90 ± 0.10	0.98 ± 0.06
D2	0.98 ± 0.04	0.97 ± 0.07	0.99 ± 0.03	0.99 ± 0.04	0.97 ± 0.07	1.00 ± 0.00
D3	0.88 ± 0.15	0.78 ± 0.31	0.99 ± 0.04	0.93 ± 0.15	0.88 ± 0.30	0.99 ± 0.04
D4	0.72 ± 0.18	0.95 ± 0.15	0.50 ± 0.32	0.85 ± 0.19	0.90 ± 0.17	0.80 ± 0.24
D5	0.80 ± 0.11	0.96 ± 0.06	0.65 ± 0.20	0.87 ± 0.08	0.96 ± 0.06	0.77 ± 0.18
D6	0.55 ± 0.28	0.35 ± 0.45	0.75 ± 0.31	0.82 ± 0.26	0.85 ± 0.24	0.80 ± 0.33
D7	0.78 ± 0.18	0.55 + 0.36	1.00 ± 0.00	0.97 ± 0.07	0.95 ± 0.15	1.00 ± 0.00
P-value				**AUC**	**Sen**	**Spec**
Wilcoxon testing				0.02799	0.01796	0.04639

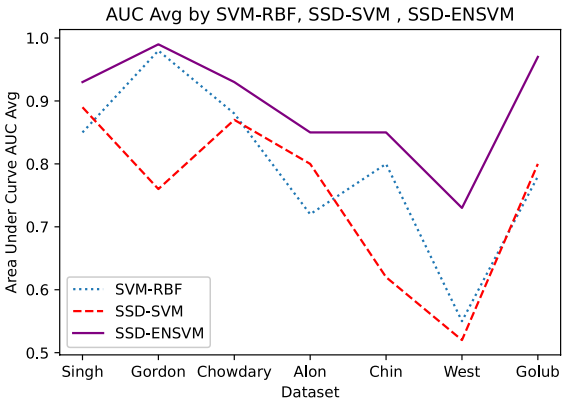

FIGURE 6. AUC of SVM (RBF) Kernel, SSD-SVM, and SSD-ENSVM in terms.

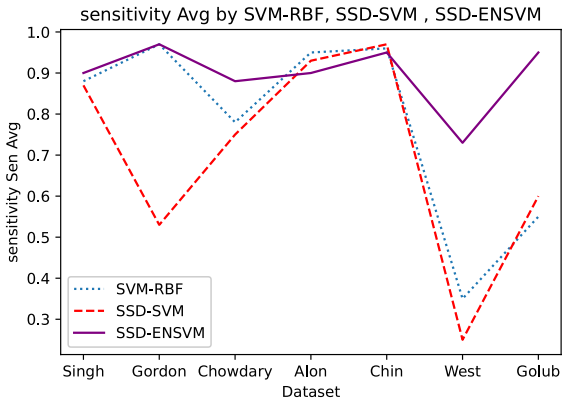

FIGURE 7. The Sensitivity of SVM (RBF) Kernel-SSD-SVM, and SSD-ENSVM.

Section 4 explained the result that evaluates the three techniques (RS-ENSVM), (SSD-ENSVM) and (ENCV-ENSVM) for selecting the average value of optimal alpha and the number of selected genes. The experiment in section 5 used different types of data sets to evaluate the performance of the proposed model.

1) RS-ENSVM VS RS-SVM VS SVM(RBF) KERNEL

This experiment aims to compare the proposed model using Randomized Search CV (RS) (RS-ENSVM) with RS-SVM and SVM(RBF) Kernel. The computation time is also compared for each model. Table 5 shows the results of this experiment for specificity, sensitivity, and AUC it can be seen from this table that the sensitivity of the proposed model is much higher than the results obtained using the RS-SVM and SVM (RBF) kernel models. For the comparison, the non-parametric Friedman test was applied for every dataset.

In terms of sensitivity, the p-values for all datasets are less than the 0.05 significance level. In terms of AUC and specificity results, our proposed RS-ENSVM model outperforms

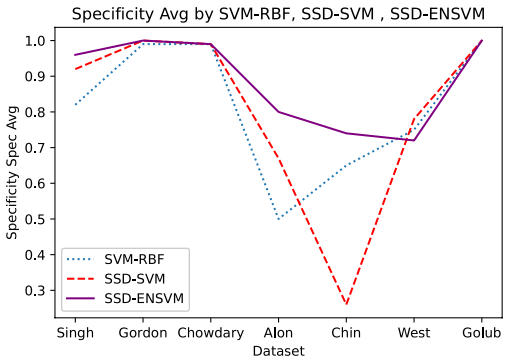

FIGURE 8. The Specificity of SVM (RBF) Kernel, SSD-SVM, and SSD-ENSVM.

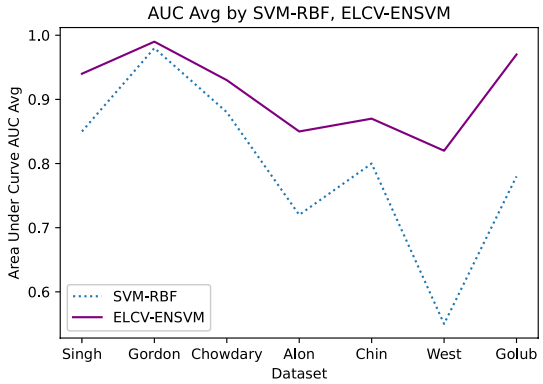

FIGURE 9. AUC of (RBF) Kernel and ENCV-ENSVM.

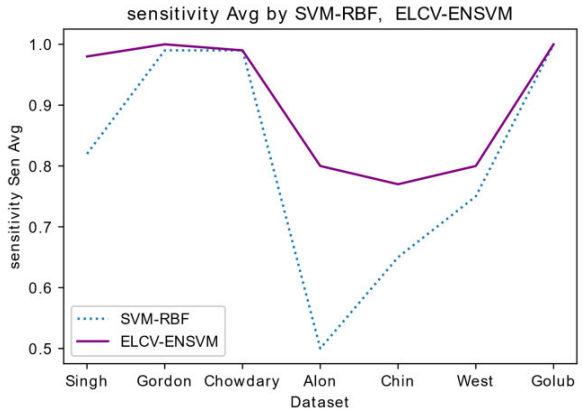

FIGURE 10. The sensitivity of the SVM (RBF) Kernel and ENCV-ENSVM.

the RS-SVM and SVM kernel model (RBF); the p-values for all datasets are less than the 0.05 significance level.

Compared to the RS-SVM and SVM (RBF), the proposed RS-ENSVM provides high specificity, sensitivity, and AUC results. It is also worth noting that the required computation time in seconds of the RS-ENSVM model with the highest dimension dataset (Gordon) is 3808.34 ms where the time for all data sets is 17263.63 ms. The computation time for all data sets of the SVM (RBF) model is 48.141581296920776 and the computation time for all data sets RS-ENSVM model is 48.141581296920776.

Figures 3, 4 and 5 presents a comparison between the SVM(RBF) kernel, RS-SVM optimizes RBF kernel

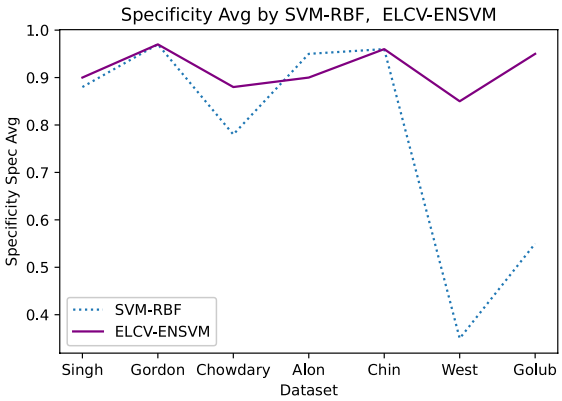

FIGURE 11. The specificity of the SVM (RBF) Kernel and ENCV-ENSVM.

TABLE 8. The number of selected genes.

dataset	The number of selected, genes				
	SVM	RSSVM	RS-ENSVM	SSDENSVM	ELEN-SVM
D1	12600	12600	144	140	142
D2	12533	12533	175	176	234
D3	22283	22283	124	127	160
D4	2000	2000	102	102	113
D5	22215	22215	180	190	235
D6	7129	7129	79	78	77
D7	7129	7129	92	91	148

TABLE 9. The values of the optimal.

Dataset	SSD- ENSVM	RS-ENSVM	ECV-ENSVM
D1	0.00063	0.00051	0.00052
	(0.00021)	(0.00027)	(0.00041)
D2	0.00057	0.00058	0.00015
	(0.00014)	(0.00022)	(0.00004)
D3	0.00039	0.00059	0.00016
	(0.00017)	(0.00029)	(0.00007)
D4	0.00051	0.00051	0.00025
	(0.00012)	(0.00029)	(0.00012)
D5	0.00047	0.00064	0.00016
	(0.00021)	(0.00029)	(0.00012)
D6	0.00051	0.00050	0.00096
	(0.00022)	(0.00036)	(0.00005)
D7	0.00056	0.00051	0.00013
	(0.00018)	(0.00029)	(0.00002)

(C and gamma), and RS-ENSVM of the average in terms of sensitivity, specificity, and AUC.

As shown in figures 3, 4 and 5 the average of specificity, sensitivity and area under the curve are higher in our proposed model RS-ENSVM than SVM (RBF) kernel and RS-SVM model.

2) SSD-ENSVM VS SSD -SVM VS SVM(RBF)KERNEL

This experiment aims to compare the proposed model using social ski-driver (SSD) (SSD -ENSVM) with SSD -SVM and SVM (RBF) Kernel model and view computation time for each model. Tables 6 show the results of this experiment.

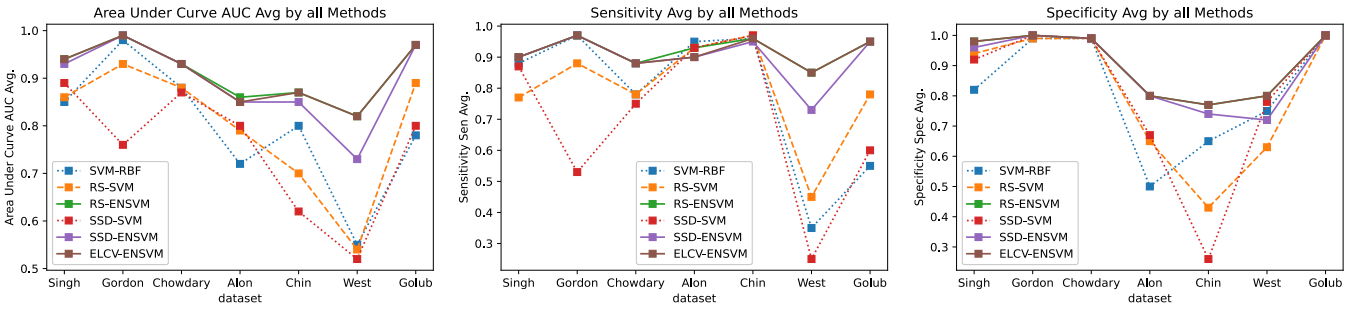

FIGURE 12. The ENSVM model techniques.

TABLE 10. Experimental results of the proposed model with different datasets.

Dataset	multiclass	SSDENSVM			RSENSVM			ELENSVM		
		Test_acc AVG	No. features	Time	Test_acc AVG	No. features	time	Test_acc AVG	No. features	Time
Iris	Yes	96.0± 6.11	4	15.46	96.0± 6.11	4	10.47	96.0± 6.11	4	54.02
Poker 9_vs_7	No	97.55± 2.00	10	12.26	97.55± 2.00	10	21.40	97.55± 2.00	10	54.44
winequality-white-9_vs_4	No	97.05± 2.94	11,9,10	14.90	97.05± 2.94	11,9,10	10.83	97.05± 2.94	11,9,10	54.10
Glass6	No	97.2± 3.04	7,8	17.65	97.22± 3.04	7,8	11.76	97.22± 3.04	7,8	53.93
Glass	Yes	86.88± 5.81	9,8,10	30.06	86.88± 6.19	9,8,10	12.36	85.97± 6.52	9,8,10	54.93

As shown in the figure 12, the SSD-ENSVM model typically provides results with higher sensitivity than the SSD-SVM and SVM (RBF) kernel models. All datasets use the non-parametric Friedman test. When using the SSD-ENSVM model, the SSD-ENSVM model performs better than the SSD-SVM and SVM(RBF) Kernel model, in terms of AUC and sensitivity results, and the p-value is below the predicted statistical significance level of the entire dataset 0.05. In terms of specificity, the SSD-ENSVM model performs better than the SVM-RBF, but the p-value is greater than the significance level of 0.05. we used a non-parametric Wilcoxon signed-rank test to interpret this result by comparing our model SSD- ENSVM with SSD–SVM and SVM (RBF) SVM (RBF) Kernel model. As we read in table 7.

The required computational time in second of the SSD-ENSVM model with the highest dimension dataset (Gordon) is 14897.562027215958 where the time for all data sets is 17263.624940395355. The computation time for all data sets of the SSDSVM model is 8899.498739242554.

Figures 6,7 and 8 shows a Comparison between three models in terms of sensitivity, specificity, and AUC.

3) ENCV-ENSVM VS SVM(RBF) KERNEL

This experiment goal to compare the models proposed using the ENCV-ENSVM with SVM(RBF) Kernel and view computation time for each model. Table 7 shows the results of this experiment for specificity, sensitivity, and AUC. The table shows that the sensitivity of the proposed model is much higher than the results obtained using the SVM kernel model (RBF). For additional comparison, the non-parametric Wilcoxon signed-rank test was utilized for all data sets. In terms of sensitivity, the P-values of all datasets are below the significance level of 0.05. In terms of AUC and specificity results, our suggested model ENCV-ENSVM performs better than the SVM-RBF model. The p-values of all data sets are below the significance level of 0.05. In general, the suggested (ENCV-ENSVM) provides high specificity, sensitivity, and AUC results compared to the SVM-RBF model., the proposed (ENCV-ENSVM) achieves high sensitivity, specificity, and AUC results. The required computational time in second of the ENCV ENSVM model is 6803.924229621887.

Figures 9, 10 and 11 show a comparison between two models in terms of sensitivity, specificity, and AUC.

4) OPTIMAL ALPHA

This experiment aims to view the value of optimal alpha using each one of the three techniques (ski-driver (SSD), Randomized Search CV (RS), Elastic Net CV (ENCV)) using in our proposed model (ENSVM). The experiment used 50 iterations with 10Fold cross-validation. table 8 shows the number of selected genes using each one of the techniques used in our model. The average value of the optimal alpha shown in table 9.

The below figure 12 show a comparison summary of all ENSVM model techniques

5) ENSVM MODEL EVALUATION

This experiment aims to investigate the performance of our suggested ENSVM model.

a: MULTICLASS DATA

In our experiment, every dataset has only two classes. As a further test, using the proposed model to a multiclass dataset, two established multiclass datasets, for example, Iris [39] has four classes in the UCI data repository, and Glass [40] has six.

There are 10 sample classes in a category. the results in [26] the experiment (shown in table 10) gave competitive results.

b: IMBALANCED DATA

In this experiment, three established unbalanced datasets (poker-9_vs_7 and winequality-white-9_vs_4 from Keele's data repository [41]) and (from UCI data repository [Glass 6]) The performance of our proposed model concerning imbalanced data. The results of the competition are presented in Table 10. These results prove that our model also applies to imbalanced data.

VI. CONCLUSION AND FUTURE WORK

In this article, we suggest a hybrid feature selection optimization model for high dimensional microarray data classification. The proposed model used Elastic Net as a gene selection method, considering three different optimizer techniques to tune parameter alpha of the EN to enhancement the performance of the model; with SVM as a classifier. The proposed model was effective in selecting the optimal or near-optimal informative and important subset of genes. The model obtained high classification results. Various experiments were performed to compare the suggested EN-SVM model against three optimization techniques: social ski-driver (SSD), Randomized Search CV (RS) and Elastic NetCV (ENCV) with SVM(RBF) Kernel, RS-SVM and SSD-SVM respectively. The results of this study showed that the proposed ENSVM model outperformed the SVM(RBF) Kernel RS-SVM as well as SSD-SVM models in terms of, specificity, sensitivity, and AUC. Also, our model had an effective result with imbalanced and multiclass data. In future studies, we will analyze the biologically significant interpretation of the list of most importance subset gene resulted in our model to help molecular biologists in cancer treatment.

REFERENCES

[1] V. Froese, R. Van Bevern, R. Niedermeier, and M. Sorge, "A parameterized complexity analysis of combinatorial feature selection problems," in *Proc. Int. Symp. Math. Found. Comput. Sci.*, Berlin, Germany: Springer, Aug. 2013, pp. 445–456.

[2] J. Too, A. R. Abdullah, and N. Mohd Saad, "Binary competitive swarm optimizer approaches for feature selection," *Computation*, vol. 7, no. 2, p. 31, Jun. 2019.

[3] H. Zhou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc., Ser. B, Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[4] M. Elhoseny, A. Tharwat, A. Farouk, and A. E. Hassanien, "K-coverage model based on genetic algorithm to extend WSN lifetime," *IEEE Sensors Lett.*, vol. 1, no. 4, pp. 1–4, Aug. 2017.

[5] W. Yamany, A. Tharwat, M. F. Hassanin, T. Gaber, A. E. Hassanien, and T.-H. Kim, "A new multi-layer perceptrons trainer based on ant lion optimization algorithm," in *Proc. 4th Int. Conf. Inf. Sci. Ind. Appl. (ISI)*, Sep. 2015, pp. 40–45.

[6] S. Mirjalili, "SCA: A sine cosine algorithm for solving optimization problems," *Knowl.-Based Syst.*, vol. 96, pp. 120–133, Mar. 2016.

[7] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[8] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.

[9] W. Richert and L. P. Coelho, *Building Machine Learning Systems With Python*. Packt Publishing Ltd., 2013. [Online]. Available: https://books.google.co.in/books?id=C-yglCEcK0sC

[10] I. I. M. Manhrawy, M. Qaraad, and P. El-Kafrawy, "Hybrid feature selection model based on relief-based algorithms and regulizer algorithms for cancer classification," *Concurrency Comput. Pract. Exper.*, 2021, Art. no. e6200, doi: 10.1002/cpe.6200.

[11] E. Pamukcu, "Choosing the optimal hybrid covariance estimators in adaptive elastic net regression models using information complexity," *J. Stat. Comput. Simul.*, vol. 89, no. 16, pp. 2983–2996, 2019.

[12] Z. Y. Algamal and M. H. Lee, "Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification," *Comput. Biol. Med.*, vol. 67, pp. 136–145, Dec. 2015.

[13] Z. Li, W. Xie, and T. Liu, "Efficient feature selection and classification for microarray data," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0202167.

[14] Y. Wang, X.-G. Yang, and Y. Lu, "Informative gene selection for microarray classification via adaptive elastic net with conditional mutual information," *Appl. Math. Model.*, vol. 71, pp. 286–297, Jul. 2019.

[15] L. Cui, L. Bai, Z. Zhang, Y. Wang, and E. R. Hancock, "Identifying the most informative features using a structurally interacting elastic net," *Neurocomputing*, vol. 336, pp. 13–26, Apr. 2019.

[16] H. Bonnel and C. Schneider, "Post-Pareto analysis and a new algorithm for the optimal parameter tuning of the elastic net," *J. Optim. Theory Appl.*, vol. 183, no. 3, pp. 993–1027, Dec. 2019.

[17] H. Dhrif, L. G. S. Giraldo, M. Kubat, and S. Wuchty, "A stable combinatorial particle swarm optimization for scalable feature selection in gene expression data," 2019, *arXiv:1901.08619*. [Online]. Available: http://arxiv.org/abs/1901.08619

[18] A. Tharwat, T. Gabel, and A. E. Hassanien, "Parameter optimization of support vector machine using dragonfly algorithm," in *Proc. Int. Conf. Adv. Intell. Syst. Inform.*, Cham, Switzerland: Springer, Sep. 2017, pp. 309–319.

[19] A. Tharwat and A. E. Hassanien, "Quantum-behaved particle swarm optimization for parameter optimization of support vector machine," *J. Classification*, vol. 36, no. 3, pp. 576–598, Oct. 2019.

[20] B. H. Shekar and G. Dagnew, "Grid search-based hyperparameter tuning and classification of microarray cancer data," in *Proc. 2nd Int. Conf. Adv. Comput. Commun. Paradigms (ICACCP)*, Feb. 2019, pp. 1–8.

[21] P. El-Kafrawy, I. I. M. Manhrawy, H. Fathi, M. Qaraad, and A. K. Kelany, "Using multi-feature selection with machine learning for de novo acute myeloid leukemia in egypt," in *Proc. Int. Conf. Intell. Syst. Adv. Comput. Sci. (ISACS)*, Dec. 2019, pp. 1–8.

[22] M. Qaraad, S. Amjad, H. Fathi, and I. I. M. Manhrawy, "Feature selection techniques for cancer classification applied to microarray data: A survey," in *Proc. Int. Conf. Intell. Syst. Adv. Comput. Sci. (ISACS)*, Dec. 2019, pp. 1–8.

[23] Y.-X. Wang, X.-Z. Li, and Z.-Y. Wang, "Parameters optimization of SVM based on the swarm intelligence," *J. Phys., Conf. Ser.*, vol. 1437, Jan. 2020, Art. no. 012005.

[24] Y. Sun, J. Zhu, X. Yu, C. Ye, and G. Fan, "Research on early warning of power grid construction safety based on PSO-SVM model," *J. Phys., Conf. Ser.*, vol. 1449, Jan. 2020, Art. no. 012037.

[25] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," *Pattern Recognit.*, vol. 43, no. 1, pp. 5–13, Jan. 2010.

[26] A. Tharwat and T. Gabel, "Parameters optimization of support vector machines for imbalanced data using social ski driver algorithm," *Neural Comput. Appl.*, Apr. 2019.

[27] W. Li, X. Xing, F. Liu, and Y. Zhang, "Application of improved grid search algorithm on SVM for classification of tumor gene," *Int. J. Multimedia Ubiquitous Eng.*, vol. 9, no. 11, pp. 181–188, Nov. 2014.

[28] X. Wang and Q. Zhou, "LTE network quality analysis method based on MR data and XGBoost algorithm," in *Proc. 5th IEEE Int. Conf. Big Data Analytics (ICBDA)*, May 2020, pp. 85–89.

[29] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, Jul. 2003.

[30] K. Chin, S. DeVries, J. Fridlyand, P. T. Spellman, R. Roydasgupta, W. L. Kuo, A. Lapuk, R. M. Neve, Z. Qian, T. Ryder, and F. Chen, "Genomic and transcriptional aberrations linked to breast cancer pathophysiologies," *Cancer cell*, vol. 10, no. 6, pp. 529–541, 2006.

[31] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub, and W. R. Sellers, "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203–209, Mar. 2002.

[32] G. J. Gordon, R. V. Jensen, L.-L. Hsiao, S. R. Gullans, J. E. Blumenstock, S. Ramaswamy, W. G. Richards, D. J. Sugarbaker, and R. Bueno, "Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma," *Cancer Res.*, vol. 62, pp. 4963–4967, Sep. 2002.
[33] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, Oct. 1999.
[34] D. Chowdary, J. Lathrop, J. Skelton, K. Curtin, T. Briggs, Y. Zhang, J. Yu, Y. Wang, and A. Mazumder, "Prognostic gene expression signatures can be measured intissues collected in RNAlater preservative," *J. Mol. Diagnostics*, vol. 8, no. 1, pp. 31–39, 2006.
[35] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. Nat. Acad. Sci. USA*, vol. 96, no. 12, pp. 6745–6750, Jun. 1999.
[36] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. A. Olson, J. R. Marks, and J. R. Nevins, "Predicting the clinical status of human breast cancer by using gene expression profiles," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 20, pp. 11462–11467, Sep. 2001.
[37] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.
[38] M. Goldman. (2008). *Spring 2008-Stat C141/Bioeng C141-Statistics for Bioinformatics*. Accessed: Jun. 7, 2020. [Online]. Available: https://www.stat.berkeley.edu/~mgoldman.Section0402.pdf
[39] [Online]. Available: http://archive.ics.uci.edu/ml/index.php
[40] C. Blake. (1998). *UCI Repository of Machine Learning Databases*. [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html
[41] [Online]. Available: https://sci2s.ugr.es/keel/index.php
[42] S. Siegel, "Nonparametric Statistics," *Amer. Stat.*, vol. 11, no. 3, pp. 13–19, Jun. 1957.
[43] D. W. Zimmerman and B. D. Zumbo, "Relative power of the wilcoxon test, the friedman test, and repeated-measures ANOVA on ranks," *J. Experim. Educ.*, vol. 62, no. 1, pp. 75–86, Jul. 1993.
[44] M. Qaraad, S. Amjad, P. El-Kafrawy, H. Fathi, and I. I. M. Manhrawy, "Parameters optimization of elastic NET for high dimensional data using PSO algorithm," in *Proc. Int. Conf. Intell. Syst. Comput. Vis. (ISCV)*, Jun. 2020, pp. 1–7.

MOHAMMED QARAAD received the B.Sc. degree from the Department of Computer Science, Mutah University, Jordan, in 2006, and the M.Sc. degree from the Department of Science, Abdelmalek Essaadi University, in 2018, where he is currently pursuing the Ph.D. degree. His research interests include optimizing features selection techniques for classification high dimension data. He attended the 2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS), in December 2019, Taza, Morocco, and the fourth edition of the International Conference on Intelligent Systems and Computer Vision (ISCV 2020).

SOUAD AMJAD received the B.Sc. degree from the Faculty of Science, Cadi Ayyad University, Semlalia, Morocco, the M.Sc. degree from the Department of Science, University Cadi Ayyad, and the Ph.D. degree from the Faculty of Science, University Cadi Ayyad. He is currently an Assistant professor with the College of Computer Science, University Abdelmalek Essaadi, Tétouan.

IBRAHIM I. M. MANHRAWY received the B.Sc. degree (Hons.) in mathematics and computer science and the M.Sc. degree in computer science from Menoufia University, Egypt, in 2006 and 2015, respectively, where he is currently pursuing the Ph.D. degree in machine learning with the Faculty of Science, School of Mathematics and Computer Science. He has five publications in the fields of machine learning. His research interests include machine learning and optimization.

HANAA FATHI received the B.S. and M.S. degrees in computer science from Menoufia University, in 2006 and 2015, respectively. She has four publications in the fields of software engineering and machine learning. She attended the 2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS), in December 2019, Taza, Morocco, and the fourth edition of the International Conference on Intelligent Systems and Computer Vision (ISCV 2020).

BAYOUMI ALI HASSAN received the bachelor's degree from the Pure Mathematics and Computer Science Department, Faculty of Science, Ain Shams University, Cairo, Egypt, in 1985, the master's degree from the Faculty of Science, Ain Shams University, in 1992, and the Ph.D. degree in mathematical modeling in area of operations research area Bangor University, North Wales, U.K., in 1998. He is currently an Associate Professor with the Operations Research Decision Support Department, Faculty of Computer Science and Information, Cairo University. His research interests include the theory of artificial intelligence, large-scale models, and multi-objective optimization and its applications.

PASSENT EL KAFRAWY received the bachelor's degree from the Computer Science and Engineering Department, American University, Cairo, the master's degree from the Faculty of Science, Menoufia University, and the Ph.D. degree in computer science and engineering in the field of computational geometry and artificial intelligence from the University of Connecticut, USA, in 2006. She taught at Eastern the State University of Connecticut for a period of one year. She has worked as an Assistant Professor with the Mathematics and Computer Science Department, Faculty of Science, Menoufia University, in 2007. She joined the Computer Science and Engineering Department, American University, as an Adjunct Professor, in 2011. She has appointed as an Associate Professor, in 2013. Since 2018, she has been a Professor. She joined the Information Technology and Computer Science School, Nile University, in 2019. She has over 45 publications and editor in three books. She has been supervising several research studies between Ph.D. and M.Sc. in the field of natural language processing, semantic knowledge, bioinformatics, big data analytics, and knowledge mining and acquisition. She is a member of the Egyptian Society of Language Engineering and the Editor-in-Chief of the *Journal of Egyptian Language Engineering*. She has joined the IBRO School for neurodegenerative physician training organized by ENND and personalized medicine workshop organized by AUC. Her research interests include software engineering, bioinformatics, big data analytics, machine learning, and cloud computing.