IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 29 September 2023, accepted 29 October 2023, date of publication 1 November 2023,
date of current version 16 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329386

RESEARCH ARTICLE

Adaptive Elastic Net Based on Modified PSO for Variable Selection in Cox Model With High-Dimensional Data: A Comprehensive Simulation Study

NURIYE SANCAR[1], **EFE PRECIOUS ONAKPOJERUO**[2], **DENIZ INAN**[3], **AND DILBER UZUN OZSAHIN**[4,5]

[1]Department of Mathematics, Near East University, 99138 Nicosia, Turkey
[2]Operational Research Center in Healthcare, Near East University, Cyprus, 99138 Nicosia, Turkey
[3]Department of Statistics, Marmara University, 34722 Istanbul, Turkey
[4]Department of Medical Diagnostic Imaging, College of Health Science, University of Sharjah, Sharjah, United Arab Emirates
[5]Research Institute for Medical and Health Sciences, University of Sharjah, Sharjah, United Arab Emirates

Corresponding author: Dilber Uzun Ozsahin (dozsahin@sharjah.ac.ae)

ABSTRACT In contemporary research, high-dimensional data has become more popular in many scientific fields with the rapid advancement of technology in collecting and storing large datasets. As in any modeling process with high-dimensional data, it is very important to accurately identify a subset of the features and reduce the dimensionality in the Cox modeling process in the case of high-dimensionality. Numerous penalized techniques for the Cox model with high-dimensional data have been developed to handle the multicollinearity problem and decrease variability. Adaptive Elastic-net is one of the penalized methods used for feature selection that both handles the grouping effect and has the oracle property. However, providing these advantageous properties of Adaptive Elastic-net for variable selection in the Cox model depends on the optimal selection of hyperparameters, α, and λ values. For this reason, the appropriate selection of these parameters is quite important. Hyperparameters are generally selected by maximizing k-fold cross-validated log partial likelihood based on grid search over (α, λ) for the model. However, this method does not guarantee optimal α and λ values. In grid search, hyperparameters are typically allowed to take values specified in a limited sequence in a grid. The purpose of this study is to propose a novel method to determine the optimum hyperparameters (α, λ) pair of Adaptive Elastic-net for variable selection in the Cox model with high dimensional data based on modified particle swarm optimization (MPSO). The introduced metaheuristic-based method has been evaluated by extensive simulation studies by comparing it with different traditional penalized methods using various evaluation criteria under different scenarios. According to the comprehensive simulation study, the proposed method outperforms other penalized methods in terms of both variable selection and prediction and estimation accuracy performance for the Cox model in investigating the high-dimensional data.

INDEX TERMS Adaptive elastic net, cox model, high-dimensional data, modified particle swarm optimization, variable selection.

I. INTRODUCTION

The recent surge in technological capabilities for data collection and storage has led to exponential growth in large datasets. This rise in data abundance has particularly impacted critical domains such as biological sciences (e.g., genomics), medicine (e.g., electronic health records), network security (e.g., intrusion detection), and engineering (e.g., IoT devices). In these domains, the advancement in data collection has resulted in a proliferation of independent

The associate editor coordinating the review of this manuscript and approving it for publication was Guido Lombardi.

variables, emphasizing the need for robust high-dimensional data analysis techniques. In statistics, data is called high-dimensional when the number of predictors (p) exceeds the sample size (n) ($p \gg n$) [1]. As a result of the continued rapid digitization of data collection technology, the increasingly high-dimensionality of survival data such as high-throughput genomic data has become a hot topic recently.

In survival analysis, it is quite important to determine the prognostic risk factors that affect survival time. Thus, regression models have an important place in survival analysis. The Cox model is a statistical technique used for analyzing survival-time outcomes about one or more predictors. In survival analysis, the primary interest is often in understanding the time until a particular event of interest occurs, such as death, failure of a system, or the occurrence of a disease.

The Cox model estimates how the hazard, or the probability of the event occurring at a given time, is influenced by various predictor variables. In essence, the Cox model provides valuable insights into how different factors affect the time it takes for a specific event to happen, making it a vital tool in survival analysis and related fields [2]. A critical and challenging problem in the Cox modeling process with high dimensional data is to accurately identify a subset of the important covariates on which the hazard function depends and to decrease dimensionality. The process of variable selection is not only used to eliminate irrelevant, unnecessary, or non-distinctive variables but also to obtain important prognostic factors linking survival time. Because keeping insignificant covariates in the model leads to errors in the interpretation of the outcome regression coefficients and adversely affects the prediction performance of the model. Variable selection in survival models has been presented to filter out the unrelated and unnecessary variables to get the most powerful subset of variables to improve the prediction accuracy and to properly and efficiently choose important factors that significantly influence the survival time. Moreover, having a small part of variables in the model reduces the computation time and complexity in the model. Based on the principle of parsimony, a simpler model with fewer variables is chosen over the more complex model with many variables. Therefore, variable selection in the Cox modeling process with high dimensional data is crucial to minimize variability and make the model more interpretable.

High-dimensional data present several difficulties for the traditional Cox modeling process. When the data is high dimensional, parameters in the Cox model cannot be estimated through the maximization of Cox partial likelihood due to their theoretical structure. Maximizing the partial likelihood for data with $p \gg n$ makes the regression coefficients of the Cox model very large and causes overfitting and unstable estimations. The dimensionality problem causes the Cox model to become ineffective [1]. On the other hand, because of the high dimensionality, the predictors are often strongly correlated which causes multicollinearity problems in the predictor matrix. This problem should be considered as it causes inaccurate inferences due to unstable parameter estimations and also produces high standard errors of the parameters. To handle the multicollinearity problem and degrade the variability, many regularized (penalized) methods have been proposed for the Cox model with high dimensional data. Reference [3] proposed LASSO (Least absolute shrinkage and selection operator) which was originally developed by [3], which is one of the penalized regression methods and solves the multicollinearity problem, makes parameter estimation under the L1 constraint in the Cox model. LASSO is a method that selects variables by size reduction. In this method, both the variable selection and the model parameters are estimated simultaneously. In high dimensional data, LASSO inclines to select a few nonzero coefficients. When multicollinearity exists, regardless of which of the correlated variables is chosen, LASSO randomly selects only one and eliminates the others. On the other hand, when data is highly dimensional, the LASSO method does not have the oracle property (consistency of variable selection) and stability [4], [5]. Hence, [6] proposed an Adaptive LASSO in the Cox model, which was originally proposed by [7]. Adaptive LASSO possesses oracle property. Specifically, the regression coefficients that are truly zero are accurately estimated as zero, and the remaining coefficients are estimated as well as if the correct submodel were known in advance. Additionally, [8] were the first to examine the L2 penalty for the Cox model with high-dimensional survival data. Reference [9] proposed ridge Cox model with L2 penalty by cross-validated partial likelihood (CVL) for high-dimensional survival data. One disadvantage of the ridge estimation of the Cox model is that it takes all the covariates in the modeling process and choosing relevant variables is not provided. Although ridge L2 penalty shrinks all the coefficients towards 0, does not get any of them exactly zero, and thus does not provide a sparse set of variables. Opposed to the LASSO, Ridge, and Adaptive LASSO methods, [10] proposed the Elastic-net method in linear regression first and the Elastic Net method in the Cox model then proposed by [11], which is especially advantageous for high dimensional and multicollinear data. Elastic Net regression combines the strengths of LASSO and Ridge Regressions by grouping and shrinking the parameters associated with regularized variables, leaving them in the equation or removing them all at once. It means that the Elastic-net promotes the grouping effect to reduce the disadvantages of the L1 norm and L2 norm, The Elastic Net method is proposed which tries to provide a balance between the L1 and L2 norm and gives effective results even if the pairwise correlations between the independent variables are high. However, as indicated by [12] and [13], Elastic-net method does not possess the oracle property, a fundamental characteristic of certain regularization methods. For this reason, the Adaptive Elastic-net method was first developed by [14] for the linear model, and [5] proposed Adaptive Elastic-net in the Cox model which both overcomes the grouping effect and has an oracle property. Adaptive Elastic-net model, which takes the elastic net method one step further, provides more accurate and efficient variable selection in high-dimensional data. However, the providing of these advantageous properties of the adaptive elastic net for variable selection in the Cox model depends on the optimal selection of hyperparameters, α, and λ values. For this reason, the appropriate selection of these parameters is quite important.

Hyperparameter optimization is the process of identifying the best combination of hyperparameters for the penalized model to satisfy an optimization function. Hyperparameters are generally selected by maximizing k-fold cross-validated log partial likelihood based on grid search over α, and λ to define the most convenient combination of these parameters. However, this method does not guarantee optimal λ and α values. Because in grid search, α is typically allowed to take values specified in a limited sequence in a grid. All the cross-validation procedures will be implemented for each potential combination of parameters λ and α as specified in the parameter grid to define the optimal combination of hyper-tuning parameter values [15]. There may be an optimal α value outside of the selected range for α and a correspondingly different λ value. We can say that the wider and more fragmented the hyperparameter grid is, the more we can approach the optimal value. As the dimension of the grid increases, it requires a lot of unnecessary processing and does not guarantee optimal α and λ values.

The variable selection problem can be defined as the combinatorial optimization problem, which proposes to reduce the number of variables and remove inappropriate data for the improvement of the interpretability and validity of the model. The selection of optimal subset variables with penalized methods depends on the optimal selection of hyper-tuning parameters [16], [17]. The choice of hyper-tuning parameters has always been a challenge in regularized regression models, particularly for censored data. Therefore, choosing optimal values for Adaptive Elastic Net parameters needs an optimization algorithm.

A. CONTRIBUTION OF THE PAPER

The aim of this study is to develop a modified PSO-based adaptive elastic-net method with an appropriate objective function for variable selection in the Cox model with high-dimensional data. Due to a number of important factors, modified Particle Swarm Optimization (MPSO) was preferred to standard PSO in your model for choosing tuning parameters in an Adaptive elastic net model with high-dimensional data. First off, the updated PSO's time-varying acceleration coefficients and inertia weight give the algorithm more dynamic adaptability [18]. When working with high-dimensional data, this characteristic is especially important since it enables the algorithm to better balance exploitation and exploration during the optimization process. To prevent local optimum conditions and guarantee convergence toward the global optimum, this balance is essential. Second, by offering a more effective method for parameter updates, the modified PSO algorithm enables it to better navigate the challenging solution space associated with high-dimensional data [19]. The acceleration coefficients and inertia weight are time-varying, which enables the algorithm to adaptively change its search approach depending on the effectiveness of the present solution. Therefore, modified PSO is recommended over standard PSO for the proposed Adaptive Elastic Cox model with high dimensional data.

II. LITERATURE REVIEW

The PSO method has been exploited for variable selection in many studies While the PSO method was employed by [20] for feature selection and the determination of parameters of support vector machines, with high dimensional datasets, [21] utilized a modified discrete PSO algorithm for feature subset selection in binary classification using logistic regression with high dimensional datasets, and the results show that the proposed modified discrete PSO algorithm has competitive classification accuracy and computational performance. In survival analysis using PSO, [22] used an adaptive LASSO logistic regression model for the diagnosis of Alzheimer's disease. Reference [23] used the same logistic regression with adaptive LASSO with PSO for the diagnosis of leukemia. Reference [4] used an alternative initial weight for adaptive penalized logistic regression to overcome the selection bias issue faced by the adaptive LASSO in a high-dimensional cancer dataset. Finally, [24] used the PSO optimization algorithm for breast cancer diagnosis. Reference [25] proposed a federated evolutionary feature selection algorithm that is based on PSO with low dimensional datasets to demonstrate that the proposed PSO-based algorithm has superior characteristics. Moreover, [26] proposes a PSO algorithm with a logistic regression model for variable selection in large-featured classification problems. The study also presents the Bayesian information criterion (BIC) as a fitness function and evaluates its effectiveness in comparison to other fitness functions. The proposed method was tested on a variety of datasets, and the results showed that it improved classification performance while simultaneously requiring fewer features. According to the findings of the study, the PSO algorithm is comparable to other already existing fitness functions. Similarly, [27] came up with a new way to make linear regression models for data with symbolic interval values. They showed how the PSO algorithm can be used to get around the problems with existing methods like the Centre method, the MinMax method, and the Centre and Range method. The PSO algorithm was used to estimate the parameters of linear regression models, and simulations showed that the proposed method works well. Because of the high dimensionality of the data and the fact that it is censored, selecting variables for a survival analysis, where the outcome is time-to-event data, is a difficult task. In recent studies, there has been a growing interest in making use of PSO for variable selection in the Cox model when working with high-dimensional data. PSO was used in the research carried out by [28] to select the best possible subset of features to use in the Cox model when working with high-dimensional genomic data. The authors demonstrated that the PSO-Cox model they proposed was superior to other approaches in terms of both accuracy and stability. In the same vein, a high-dimensional Cox model that included multiple competing risks was analyzed by [29], and PSO was used to select the best subset of features to include in the model. The authors demonstrated that the proposed PSO-Cox model performed better than other methods in terms of the accuracy of predictions as well as the stability of feature selection.

Variable selection in the Cox modeling process with high dimensional data is crucial to minimize variability and make the model more interpretable. Penalized methods are used for variable selection, especially when dealing with high-dimensional data. Adaptive Elastic net, one of the penalized methods, is more advantageous in variable selection compared to other penalized methods, as it is a method that both overcomes the grouping effect and has an oracle property. The best choice of α and λ hyper-tuning parameters determines the selection of optimal subset variables with penalized approaches. In regularized regression models, choosing the appropriate hyper-tuning parameters has always been difficult, especially for censored data. Hyperparameters are often chosen by maximizing k-fold cross-verified log partial likelihood based on grid search over and to identify the best practical combination of these parameters. However, this approach does not ensure that and are at their best. Due to the fact that in grid searches, is frequently permitted to accept values provided in a grid's restricted sequence. All the cross-validation procedures will be implemented for each potential combination of parameters α and λ as specified in the parameter grid to define the optimal combination of λ hyper-tuning parameter values [15]. There may be an optimal α value outside of the selected range for α and a correspondingly different λ value. We can say that the wider and more fragmented the hyperparameter grid is, the more we can approach the optimal value. As the dimension of the grid increases, it requires a lot of unnecessary processing and does not guarantee optimal α and λ values. Therefore, choosing an optimal combination for Adaptive Elastic Net hyper-tuning parameters needs an optimization algorithm for the selection of an optimal subset of variables. In addition, alterations to the conventional PSO algorithm have been suggested as a means of improving the performance of the algorithm. A variation of the common PSO methodology used to solve optimization issues is the Modified PSO approach. This approach improves the effectiveness and searchability of the PSO algorithm, it combines new tactics and methodologies. These alterations could affect the initialization procedure, the rules for updating the particle locations and velocities, and the techniques for dealing with boundaries and restrictions. These changes are intended to accelerate convergence and increase the accuracy of the PSO algorithm, especially for challenging optimization problems with numerous variables and constraints. The Modified PSO method has been successfully applied in various fields. The time-varying inertia weight and acceleration coefficients have been utilized to improve the search capability of the algorithm. A modified version of the PSO algorithm was suggested to be used for feature selection in the support vector regression (SVR) model in research carried out by [30]. The authors demonstrated that the proposed modified PSO-SVR model performed more accurately and efficiently than other feature selection methods. Similarly, [19] proposed a modified PSO-ENSVM model for high-dimensional cancer microarray datasets for feature selection. Elastic Net and SVM classifier tuning parameters are optimized by the model using the PSO algorithm. The model is assessed and indicates superior performance in terms of lowering the number of features and raising performance rates, according to the results. Reference [18] conducted research in which they proposed a modified version of the PSO algorithm to be used for time series forecasting of autoregressive and moving average structures. The authors demonstrated that the proposed modified PSO method performed better than other traditional PSO methods in terms of the accuracy of its predictions and the efficiency with which it processed those predictions.

In the following parts, Cox PH Model and Penalized Partial Likelihood Methods: From LASSO to Adaptive Elastic Net have been decribed:

A. Cox PH MODEL

Let n be the number of observations in survival data and each of the jth subjects can be defined by (T_j, δ_j, X_j) where T_j is the time that the occurred first: the survival time or the censoring time, δ is the censoring indicator, and $X_j = [X_{j1}, X_{j2}, \ldots, X_{jp}]$ is the predictor vector for the jth subject. Assume a specific hazard function of jth subject $\lambda(t, X_j)$ in the Cox's Proportional Hazard (PH) Model [2], is

$$\lambda(t, X_j) = \lambda_0(t)\, e^{X_j\beta} \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard function, $\beta^T = (\beta_1, \beta_2, \ldots, \beta_p)$ is vector of coefficients. The β of the Cox regression model is estimated by maximizing the logarithmic partial likelihood function ($logpl(\beta)$) expressed by Equation (2) [31].

$$logpl(\beta) = \sum_{j=1}^{n} \delta_j x_j \beta - \sum_{j=1}^{n} \delta_j log\left[\sum_{i\in R(t_j)} e^{X_j\beta}\right] \tag{2}$$

where $R(t_j) = \{m | t_m \geq t_j\}$, is the set of units at risk at time t_j. The Cox regression model is a frequently used approach in the analysis of low-dimensional survival data ($n \gg p$), as it is a semi-parametric method and the regression coefficients can be easily interpreted.

The rapid advancement of technology in the collection and storage of large datasets has facilitated to gathering of an enormous number of predictor variables information of survival data in clinical studies. However, it is most probable that not all predictors in the data are related to survival time. Virtually, a small part of predictor variables is related to the clinical response. Therefore, a critical and challenging problem in the Cox modeling process with high dimensional data is to accurately identify a subset of the important predictors on which the hazard function depends and to decrease dimensionality. Identification of the set of predictor variables for the final accurate model is called the variable selection process. The process of variable selection is not only used to eliminate irrelevant, unnecessary, or non-distinctive variables but also to obtain important prognostic factors linking survival time to increase accuracy and decrease the complexity of the model. In other words, this procedure ensures stability between simplicity and model fitting. Moreover, having a small part of variables in the model reduces the computation time and complexity in the model. Based on the principle of parsimony, a simpler model with fewer variables is chosen over the more complex model with many variables. Therefore, variable selection in the Cox modeling process with high dimensional data is crucial to minimize variability and make the model more interpretable. Statistically, the identification of zero and nonzero coefficients is required [32].

High-dimensional data present several difficulties for the traditional Cox modeling process. When the data is high dimensional, parameters in the Cox model cannot be estimated through the maximization of Cox partial likelihood. Maximizing the partial likelihood for data with p ≫ n makes the regression coefficients of the Cox model very large and causes overfitting and unstable estimations [1], [33], [34]. On the other hand, because of the high dimensionality, the predictors are often strongly correlated which causes multicollinearity problems in the predictor matrix. This problem should be considered as it will cause inaccurate inferences due to unstable parameter estimations and also produce high standard errors of the parameters. To handle the multicollinearity problem and degrade the variability, many regularized (penalized) methods have been proposed for the Cox model with high dimensional data.

B. PENALIZED PARTIAL LIKELIHOOD METHODS: FROM LASSO TO ADAPTIVE ELASTIC NET

In the process of Cox modeling with the high dimensional survival data, it is essential to handle the model complexity [35]. Penalized partial likelihood methods are the most popular approaches for simultaneous variable selection and parameter estimation. The penalized partial likelihood function ($logpl^{*}(\beta)$), is defined as follows:

$$logpl^{*}(\beta) = logpl(\beta) - \sum\nolimits_{i=1}^{p} p_{\lambda}(|\beta_i|) \quad (3)$$

where $p_{\lambda}(|.|)$ is the penalty function with regularization parameter λ.

Minimizing the negative of penalized partial likelihood function with L2 penalty , $p_{\lambda}(|\beta_i|) = \lambda\beta_i^2$, ridge Cox estimator is obtained [9], [31]. Cox model with L2 penalty is not used for variable selection because although ridge L2 penalty shrinks all the coefficients towards 0, does not get any of them exactly zero and thus, does not provide a sparse set of variables.

Reference [3] proposed LASSO (Least absolute shrinkage and selection operator) which was originally developed by [36], which is one of the penalized methods under the L1 penalty, $p_{\lambda}(|\beta_i|) = \lambda|\beta_i|$ in the Cox model. The LASSO and elastic net methodologies often employ penalized techniques for parameter estimation and variable selection. In addition to its extensive utilization, the LASSO has a couple of limitations. One limitation of the model is its lack of robustness in handling large correlations among independent variables, as it randomly selects one variable while disregarding the others. Additionally, in the case of high-dimensional data, the LASSO technique lacks the oracle property, which refers to the consistency of variable selection, as well as stability [4], [5]. Hence, [6] proposed Adaptive LASSO estimator in the Cox model, as seen in Equation (4)

$$\hat{\beta}_{AL} = \mathrm{argmin}_{\beta} - \frac{1}{n} logpl(\beta) + \lambda \sum\nolimits_{i=1}^{p} \frac{|\beta_i|}{|\tilde{\beta}_i|} \quad (4)$$

where $\lambda \sum_{i=1}^{p} \frac{|\beta_i|}{|\tilde{\beta}_i|}$ is the penalty function, and $\tilde{\beta}$ is any consistent estimator of β in high dimensional data. The ridge estimator is generally used as a consistent estimator of β. Adaptive LASSO possesses oracle properties [37]. When the oracle property is provided, the correct model coefficients of zero are estimated as zero, and the remaining coefficients are estimated as if the true sub-model was known beforehand [14]. The LASSO estimator is more favorable than the ridge estimator by pulling certain of the coefficients in the model to zero [38]. In high dimensional data, LASSO inclines to select a few nonzero coefficients. When multicollinearity exists, regardless of which of the correlated variables is chosen, LASSO randomly selects only one and eliminates the others.

Opposed to the LASSO, Ridge, and Adaptive LASSO methods [10] proposed the Elastic-net method in linear regression first and the Elastic Net method in the Cox model is then proposed by [11], which is especially advantageous for high dimensional and multicollinear data. Elastic Net regression combines the strengths of L1 and L2 penalties by grouping and shrinking the parameters associated with regularized variables, leaving them in the equation or removing them all at once. It means that the Elastic Net promotes the grouping effect. Elastic net as a penalized approach for variable selection, to overcome the disadvantages of LASSO. The penalized partial likelihood function with two hyperparameters is given by [33].

$$\hat{\beta}_{EN} = \mathrm{argmin}_{\beta} - \frac{2}{n} logpl(\beta) + \lambda\left[\alpha \sum\nolimits_{i=1}^{p} |\beta_i| + \frac{1}{2}(1-\alpha)\sum\nolimits_{i=1}^{p} \beta_i^2\right] \quad (5)$$

where $\lambda\left[\alpha \sum_{i=1}^{p}|\beta_i| + \frac{1}{2}(1-\alpha)\sum_{i=1}^{p}\beta_i^2\right]$ is the penalty function with $0 \leq \alpha \leq 1$. It is seen that when $\alpha = 1$ in the penalty function of the elastic net estimator, it is reduced to $L1$ norm, that is, to the LASSO penalty. Similarly, when $\alpha = 0$, $L2$ is reduced to the norm, that is, to the Ridge penalty. Therefore, the Elastic network estimator is a method that tries to find a balance between the Ridge and LASSO estimators [33]. Elastic net combines the L2 and L1 penalties, by utilizing the ridge penalty to overcome the multicollinearity problem while utilizing the benefits of LASSO penalty in the process of variable selection in high dimensional data. It provides effective results even if the pairwise correlations between the independent variables are high. On the other hand, [12], [13] showed that the estimator of the elastic Net method does not have the oracle property. Oracle property is attained by the adaptive LASSO, while collinearity is dealt with by the Elastic-net. But [10] also noted that the adaptive LASSO acquired the LASSO 's instability for high-dimensional data, whereas [37] contended that the elastic net fails the oracle features. As a result, the Adaptive Elastic Net method was first developed by [14] for the linear model, and [5], proposed an adaptive elastic net estimator in the Cox model defined in Equation 6 which overcomes the grouping effect and has oracle properties. It has been demonstrated that, in contrast to previous oracle-like techniques, the adaptive elastic net not only benefits from the oracle characteristic but also handles the collinearity issue effectively.

$$\hat{\beta}_{AEN} = \operatorname{argmin}_{\beta} - \frac{2}{n} logpl\,(\beta) + \lambda\left[\alpha \sum_{i=1}^{p} \hat{w}_i\,|\beta_i| + \frac{1}{2}(1-\alpha)\sum_{i=1}^{p}\beta_i^2\right] \quad (6)$$

where $\lambda\left[\alpha \sum_{i=1}^{p}\hat{w}_i\,|\beta_i| + \frac{1}{2}(1-\alpha)\sum_{i=1}^{p}\beta_i^2\right]$ is penalty function, and $\hat{w}_j = \left(\left|\hat{\beta}_{(EN)_i} + 1/n\right|\right)^{-r}$ with $r > 0$. In this study, we set $r = 1$. The experiment's findings demonstrate that this parameter does not significantly affect estimation [4], [39].

This approach, which takes the elastic net method one step further, provides more accurate and efficient variable selection in high-dimensional data. The Adaptive Elastic Net estimator is consistent not only for prediction but also for variable selection [40], [41]. Extensive analyses of high-dimensional datasets have shown that Adaptive Elastic Net generally provides more accurate and stable predictions. However, the providing of these advantageous properties of the Adaptive Elastic Net in the Cox model depends on the optimal selection of hyperparameters, α and λ values. Hyperparameters are generally selected by maximizing k-fold cross-validated log partial likelihood based on grid search over α and λ to define the most convenient combinations [9], [31]. However, this method does not guarantee optimal α and λ values. Because in grid search, α is typically allowed to take values specified in a limited sequence in a grid. All the cross-validation procedures will be implemented for each potential combination of parameters α and λ as specified in the parameter grid to define the optimal combination of hyper-tuning parameter values [15]. There may be an optimal α value outside of the selected range for α and a correspondingly different λ value. We can say that the wider and more fragmented the hyperparameter grid is, the more we can approach the optimal value. As the dimension of the grid increases, it requires a lot of unnecessary processing and does not guarantee optimal α and λ values. Therefore, choosing the optimal combination for Adaptive Elastic Net hyper-tuning parameters needs an optimization algorithm for the selection of an optimal subset of variables.

III. METHODOLOGY

In this study, MPSO technique has been selected as an optimization algorithm to estimate the hyper-tuning parameters in the Adaptive Elastic Net Cox model for variable selection with high dimensional data. Utilizing Modified PSO is justified by its exceptional capacity to strike a balance between search space exploration and exploitation [18]. Finding a reliable and accurate solution is essential in high-dimensional environments where the number of factors might be huge and the relationships between them can be complex. Traditional PSO might not be able to perform this task well due to its set inertia weight and acceleration coefficients. The Modified PSO, on the other hand, is built to adaptively modify its search method over time thanks to its time-varying inertia weight and acceleration coefficients.

A. MODIFIED PARTICLE SWARM OPTIMIZATION METHOD (MPSO)

PSO is a swarm-based metaheuristic algorithm that has been widely used in solving various optimization problems, including variable selection in regression analysis. The algorithm was first introduced by [42]. It has been utilized in many studies due to its simplicity, efficiency, and robustness [42]. According to their locations and velocities in a PSO, particles travel around the search space. The optimal solution, the experiences of neighbors, and each particle's past position all have an impact on its movement.

Additionally, modifications to the standard PSO algorithm have been proposed as a way to enhance the method's performance. The Modified PSO technique is a variant of the typical PSO methodology used to address optimization problems. It integrates fresh strategies and approaches to raise the PSO algorithm's efficacy and searchability. These changes might impact the initialization process, the guidelines for updating particle positions and velocities, and the methods for coping with boundaries and constraints. The PSO method will benefit from these modifications as they will hasten convergence and improve accuracy, especially for difficult optimization problems with multiple variables and constraints.

Another study by [43] applies PSO to ascertain the parameters of simple polarized bodies from their self-potential signals. The parameters include depth to the electric source, electric charge polarization angle, electric dipole moment, and shape factor of a buried body. The method's validity is affirmed through testing on synthetic data, including noise. The results demonstrate stability even at high noise levels and accuracy in estimating parameters, especially depth, showcasing its potential compared to Genetic Algorithm-based techniques. Similarly, [44] introduces the utilization of PSO with function stretching, termed SPSO, to estimate parameters like depth, density, radius, or thickness from residual gravity anomalies in geophysics. The objective function is formulated and transformed using a stretching strategy. Test runs demonstrate the applicability of SPSO in estimating these unknown parameters, even under high noise levels in field data. This marks the first application of SPSO in geophysics and showcases its competitiveness and flexibility compared to genetic algorithms. The method is effectively applied to both synthetic and field data. [18] conducted research in which they proposed a different modified form of the PSO algorithm that incorporates both the time-varying acceleration coefficient and the time-varying inertia weight [7]. The authors demonstrated that the proposed modified PSO method performed better than other traditional PSO methods in terms of the accuracy of its predictions and the efficiency with which it processed those predictions.

MPSO stands as a powerful tool for optimizing various models, showcasing potential improvements in accuracy and efficiency, making it a pertinent addition to the field of optimization algorithms. however, there are concerns about the time complexity associated with the PSO algorithm, especially when dealing with high-dimensional data. To effectively address this concern, we have developed a tailored variant of PSO which is the MPSO algorithm. The MPSO algorithm has been precisely engineered to enhance computational efficiency in the realm of high-dimensional spaces, ensuring a more practical and efficient approach. A key aspect of the MPSO approach involves an optimized initialization strategy for the particles within the swarm. This tailored initialization method ensures a highly efficient commencement of the optimization process, which is particularly suited to the intricacies of high-dimensional data. Further, the MPSO algorithm integrates adaptive mechanisms, dynamically adjusting parameters during optimization. This adaptability strikes a delicate balance between exploration and exploitation, a critical feature, especially in navigating the complexities of high-dimensional spaces. To expedite the convergence process, we have incorporated advanced strategies within the MPSO algorithm. These strategies guide the particles more efficiently toward optimal or near-optimal solutions, thus significantly reducing the overall convergence time. collectively, these modifications effectively tackle the computational challenges posed by high-dimensional data within the PSO framework.

The following are the basic steps and algorithms for performing the modified PSO:

Stage 1: The $k_{th,}$ position of particles is determined randomly and are kept in a vector X_k

$$X_k = \left(X_{k,1}, X_{k,2}, \ldots, X_{k,m}\right), \quad k = 1, 2, \ldots, np \tag{7}$$

where $X_{i^k}(i = 1, 2, \ldots, m)$ is the *ith* position of $k_{th,}$ particle. And np is the number of positions while m is the number of particles.

Stage 2: Vector velocities are determined randomly and given as,

$$V_k = \left(V_{k,1}, V_{k,2}, \ldots, V_{k,m}\right), k = 1, 2, \ldots, np \tag{8}$$

Stage 3: $pbest$ and $Gbest$ are determined according to the evaluation function as follows

$$pbest_k = \left(P_{k,1}, P_{k,2}, \ldots, P_{k,m}\right), \quad k = 1, 2, \ldots, np \tag{9}$$

$$Gbest_k = \left(P_{g,1}, P_{g,2}, \ldots, P_{g,m}\right), \tag{10}$$

where $Gbest_k$ denotes the best particle, which has the highest value for the evaluation function thus far, and $pbest_k$ denotes a vector containing the positions corresponding to the kth particle's best personal performance.

Stage 4: Let c_1 and c_2 represent social coefficient and cognitive coefficient, while w represent the inertia weight parameter. Let (c_{1f}, c_{1l}), (c_{2f}, c_{2l}) and (w_1, w_2), be intervals of c_1, c_2 and w.

For every iteration, the calculations of each of these parameters is given as

$$\begin{aligned} c_1 &= [(c_{1l} - c_{1f}) \times (t/maxt)] + c_{1i} \\ c_2 &= [(c_{2l} - c_{2f}) \times (t/maxt)] + c_{2i} \\ w &= [(w_2 - w_1) \times ((\max t - t)/maxt)] + w_1 \end{aligned} \tag{11}$$

where *maxtt* are the maximum and current iteration numbers

Stage 5: To update the values of velocity and positions the following formulas are used;

$$\begin{aligned} V_{i,d}^{t+1} &= (w \times V_{i,d}^{t} + c_1 \times r_1 \times (P_{i,d} - X_{i,d}) \\ &\quad + c_2 \times r_2 \times (P_{g,d} - X_{i,d})) \end{aligned} \tag{12}$$

$$X_{i,d}^{t+1} = X_{i,d} + V_{i,d}^{t+1} \tag{13}$$

where r_1 and r_2 represent random numbers the interval [0, 1].

Stage 6: Until a maximum number of iterations ($maxt$) is reached steps 3 to 5 is repeated.

Figure 1 illustrates the flowchart for the basic steps and algorithms for MPSO.

B. PROPOSED METHOD

Due to its superior, adaptability and efficiency, MPSO was chosen over classic PSO for estimating hyper-tuning parameters of Adaptive elastic-net method with an appropriate objective function for variable selection in the Cox model with the high-dimensional data. Its time-varying acceleration coefficients and inertia weight balance exploration and exploitation, allowing convergence to the global optimum while successfully avoiding local optimum conditions in the challenging high-dimensional data solution space.

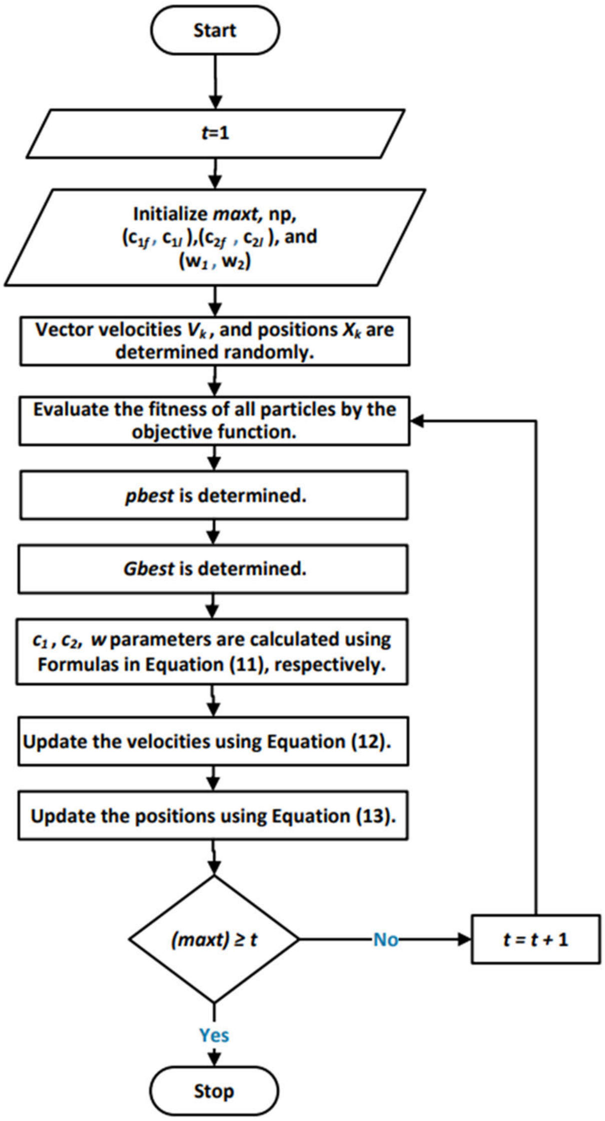


FIGURE 1. Flow chart of MPSO.

For the proposed method, the Extended Bayesian Information Criterion (EBIC) has been used as an objective function. A variation of the Bayesian Information Criterion (BIC) created expressly for high-dimensional data, is the Extended Bayesian Information Criterion (EBIC) [27]. In high-dimensional settings, it is frequently employed for variable selection. In several fields, including genomics and other areas of biomedical research, high-dimensional data, where the number of variables (p) is bigger than the number of observations (n), is becoming more prevalent. Because they were not intended for high-dimensional settings, conventional model selection metrics like BIC or AIC may not perform well in these circumstances. The EBIC improves the BIC's ability to handle high-dimensional data by adding an additional penalty term. A hyperparameter that can be adjusted to achieve a desired trade-off between model fit and model complexity controls this penalty term. In Cox models with high-dimensional data, the EBIC offers a systematic method for variable selection that successfully balances model fit and complexity, controlling for false positives even when the number of variables is significantly greater than the number of observations as defined in Equation (14) [45].

$$EBIC = -2 * (logpl\ (\beta) - \alpha \sum_{i=1}^{p} \hat{w}_i |\beta_i| + \lambda \sum_{i=1}^{p} \beta_i^2) + d * log\ (n) + 2 * \gamma * \log(p) \quad (14)$$

where d is the number of non-zero coefficients in the model (i.e., the number of variables included in the model), n is the number of observations, p is the total number of potential independent variables and γ is a hyperparameter which is between 0 and 1. A better model must have a lower EBIC because it indicates a better balance between model fit and complexity. In this case, γ is a fixed constant that can be set to 0.5, as proposed by [27].

The purpose is to determine the optimum hyperparameters (λ, α) pair of Adaptive Elastic-net for variable selection in for Cox model with high dimensional data which proposes to reduce the number of variables and remove inappropriate data for improvement of interpretability and validity of the model. The objective function defined in Equation (14) is minimized by using modified PSO in the proposed method. The analyses in the study were conducted R Studio version 4.3.1 for Windows software. All experiments in the study are trained offline on a PC equipped with 3.7 GHz i7-8700k core processors, 32G RAM, and NVIDIA 1080 Ti GPU. Trial and error are used to determine the parameters of the MPSO. The following stages have been introduced the proposed method's algorithm:

Stage 1. The selection of the tuning parameters of the MPSOA-ENet algorithm are made as follows: np: 50, maxit: 500, $(c_{1f}, c_{1l}) = (2, 3)$, $(c_{2f}, c_{2l}) = (2, 3)$, $(w_1, w_2) = (0.9, 2)$.

Stage 2. The initial positions of each of the jth (j = 1, 2, ..., 50) particles are designed through random generation using a uniform distribution. The first position of each particle corresponds to the tuning parameter α and is generated from a uniform distribution with parameters (0, 1). Similarly, the second position of each particle represents the tuning parameter λ, which is generated from a uniform distribution with parameters (0, 200).

Stage 3. Velocities are generated using a uniform distribution with a range of (0, 4).

Stage 4. EBIC defined in Equation (14) is taken as objective function. All particles' objective function values are accumulated. Pbest and Gbest particles are identified using the values of the objective function.

Stage 5. Use the formulas in Equation (11) in to update the cognitive coefficient c_1, social coefficient c_2, and inertia parameter w at each iteration.

Stage 6. Formulas in Equation (12) and Equation (13) are used to update the particle velocities and positions, respectively.

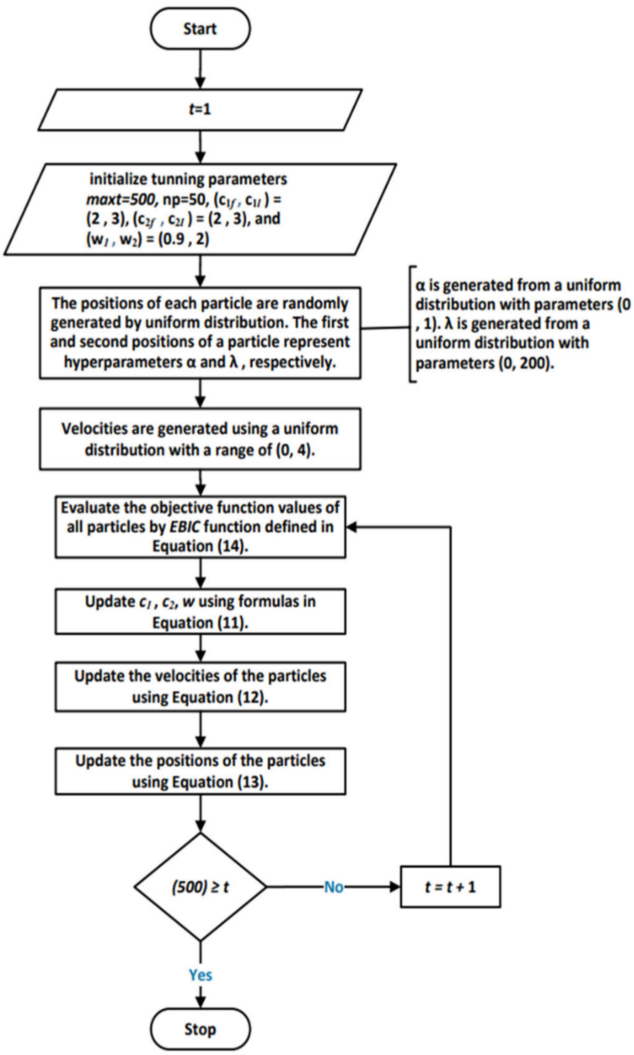


FIGURE 2. **Flow chart of the proposed MPSOA-ENet method.**

Stage 7. Steps 4 through 6 are repeated until the maxit is attained.

Stage 8. The best (α, λ) pair is obtained as Gbest.

Figure 2 illustrates the flowchart for of the proposed MPSOA-ENet method.

IV. SIMULATION STUDY

The simulation study shows the performance of the proposed modified PSO-based Adaptive Elastic net (MPSOA-ENet) algorithm for variable selection in Cox Model by comparing it with adaptive elastic net Cox model(A-ENet), Elasticnet Cox model (ENet), Adaptive LASSO Cox model (A-LASSO) and LASSO Cox model (LASSO) under different simulation settings in high dimensional data. Here the regularization parameters for the methods are tuned via fivefold cross-validation based on grid search. Survival time t_i is generated from a Weibull distribution as

$$t = -\left(\frac{\log(U)}{\lambda e^{\beta' X}}\right)^{v} \tag{15}$$

where shape parameter v = 0.30, scale parameter $\lambda = 1$ and U is uniformly distributed as Uni∼U(0.10,0.95). The true model coefficients are as follows:

$$\beta = (\underbrace{2}_{15}, \underbrace{0, \ldots, 0}_{p-15}) \tag{16}$$

The model's actual size is 15. For simulated datasets, the percentage of censored observations has been set to 20%, and 40%, using Uniform distribution. We set n = 100 and p = {200,400}. Most of the settings for the generation of independent variables are adapted from [10], [29], and [46].

The following 8 scenarios are considered in the simulation study:

Scenario 1: The independent variables in predictor matrix, X are derived from the multivariate normal distribution, marginally N(0,1) and with pairwise correlation among X_i and X_k as cor(X_i, X_k) = ρ^{i-k} where i ≠ k and with ρ = 0.20 for p = 200.

Scenario 2: The same as Scenario 1, except $\rho = 0.5$.

Scenario 3: The same as Scenario 1, except $\rho = 0.8$.

Scenario 4: The same as Scenario 1, except $p = 400$.

Scenario 5: The same as Scenario 2, except $p = 400$.

Scenario 6: The same as Scenario 3, except $p = 400$.

Scenario 7: The independent variables (p = 200) are generated as follows including grouped variable situations.

$$\begin{aligned} x_i &= w_1 + e_i, \quad w_1 \sim N(0,1), \ i = 1, 2, 3, 4, 5 \\ x_i &= w_2 + e_i, \quad w_2 \sim N(0,1), \ i = 6, 7, 8, 9, 10 \\ x_i &= w_3 + e_i, \quad w_3 \sim N(0,1), \ i = 11, 12, 13, 14, 15 \end{aligned} \tag{17}$$

where x_i independently and identically distributed as $x_i \sim N(0,1), i = 16, 17, , \ldots, p, e_i \sim N(0, 0.01) fori = 1, 2, \ldots, 15$ as independently and identically distributed.

Scenario 8: The same as Case 7, except p = 400.

In the simulation design, the correlation between the independent variables is weak in Case 1 and Case 4; Cases 2 and 5 have medium correlated variables; Cases 3 and 6 have highly correlated variables; while the variables in Cases 7 and 8 are generated including the grouping effects defined in [10] with three identically significant groups, where the correlation between the identical group is as severe as 0.990. In each case of simulation design, censoring rates of 0.20, and 0.40 are considered separately.

Performance evaluation measurements:

The models 'performance in variable selection has been evaluated by F1score, true negative (TN), false negative (FN), sum of square error for the model coefficients (SSE), and concordance index (CI). F1-score, TN, and FN have been evaluated by the confusion matrix.

A confusion matrix is a tabular depiction of the method's performance in the context of variable selection [47]. By contrasting the actual state of variables (whether they are truly zero or not) with the state obtained through the method, it enables the display of the algorithm's performance. In the confusion matrix, True Positive (TP): accurately identifying

Confusion Matrix		Actual	
		Nonzero	Zero
Estimated	Nonzero	TP	FP
	Zero	FN (IC)	TN (C)

FIGURE 3. Confusion Matrix.

non-zero coefficients as non-zero. True Negative (TN) or C: zero coefficients were correctly identified as zero. False Positive (FP): zero coefficients were mistakenly identified as non-zero, False Negative (FN) or IC: non-zero coefficients were incorrectly identified as zero. F1-score is calculated by 2*(P*R)/(P+R), where P = TP/TP+FP is precision and R = TP/TP+FN is recall (or sensitivity). The accuracy of the models can be assessed using the F1-score, which is a harmonic mean of the Precision and Recall measures. Finding the most relevant subset of variables that can contribute to precise prediction models while avoiding overfitting and lowering computing complexity is the aim of variable selection. The F1 score is a good measure for contrasting various variable selection methods and aids researchers in evaluating the feature selection method's overall effectiveness in reaching this aim.

On the other hand, CI is a measure of goodness of fit that calculates the probability that the one with the higher prognostic score from two randomly selected units will live longer than the other unit. CI takes values in the range [0, 1] [48], [49]. A CI value close to 1 indicates that the prediction performance of the model is high. And SSE is the sum of square errors for parameter β where $SSE = \sum_{j=1}^{p}\left(\beta_j - \hat{\beta}_j\right)^2$ where β_j is the vector of actual coefficients, $\hat{\beta}_j$ is the vector of the model coefficients estimated by the models.

The simulations are repeated 100 times randomly. Within each repeated simulation, each simulated data set is divided into a training set and a test set at 70:30 proportion, with tuning parameters chosen from the training set and the estimators have been calculated on the training set. Performance metrics have been computed on test sets after models have been fitted with training data on the testing set, and median value of the performance measurements has been reported.

V. RESULTS

The developed MPSOA-ENet method for variable selection in the Cox model with high-dimensional data has been compared with conventional penalized methods; LASSO, A-LASSO, A-ENet, and ENet using different performance evaluation measures under 8 different simulated data with 20% and 40% censoring rates, separately. The median of each performance evaluation measure for each approach has ultimately been presented after we repeated the simulation 100 times for each scenario in Table 1 and Table 2 for 20% and 40% censoring rates, respectively. The best approach on MSSE, F1-score, TN, FN, and CI have been indicated with a bold font in each simulation case. According to simulation results presented in Table 1 and Table 2, the MPSOA-ENet approach consistently yields the lowest MSSE among all the compared penalized approaches in each case. Furthermore, when the dimensionality remains constant, all methods except LASSO have shown a decrease in MSSE values as the correlation increases. On the other hand, in the case including the grouping effect, our proposed method outperforms other methods, while the LASSO exhibited the lowest performance when dealing with grouped highly correlated variables. The LASSO is not sufficiently steady, and it tends to randomly select certain significant factors while ignoring the other relevant variables when there is a strong association or grouping effect. For this reason, it is not recommended to use the LASSO method for variable selection in high-dimensional data where there is a high correlation or grouping effect. When the performances of the other methods are compared according to the MSSE values, A-ENet and A-LASSO which follow the MPSOAE-net method give good performance, respectively. The ENet method, on the other hand, has been observed to improve performance as the correlation increased and to perform quite poorly at a low correlation (p = 0.3), which is consistent with the literature.

On the other hand, the variable selection capabilities of the methods were evaluated in terms of TN, FN, and F1-score values. The results clearly show that the MPSOA-Enet method is much more successful in different scenarios in variable selection compared to other methods. We observed that the TN values of the proposed method are consistently higher than other existing methods. This encouraging outcome highlights the proposed method's greater capacity to precisely identify true negatives. In cases where there is a low and medium correlation between the variables, A-ENet and A-LASSO give good results after MPSOA-ENet, and in case of high correlation, A-ENet and ENet give good results, respectively. LASSO gives the worst results in medium and high correlation and also in grouping effect. In terms of TN values, an improvement was observed in the performance of all methods as the correlation increased. While the MPSOA-ENet method gives near-perfect results in low and medium correlation, it gives excellent results under the effect of high correlation and grouping. That is, the TN values of our proposed method are higher, and the false negative (FN) values are 0 at the lowest line. Therefore, the F1 score values of the proposed method are higher than the other methods in each scenario. A high F1-score indicates that the variable selection method is successfully identifying the important variables while minimizing the inclusion of unimportant ones.

In variable selection, the "oracle property" refers to the ideal situation when a variable selection approach can precisely identify the genuine important variables from a broader range of predictors. This is very important property for variable selection techniques. The oracle property is an important and sought-after trait in the process of developing models since it guarantees that the resulting model is reliable, interpretable, efficient in computation, and resilient to noise. The methods with oracle property in the variable selection are

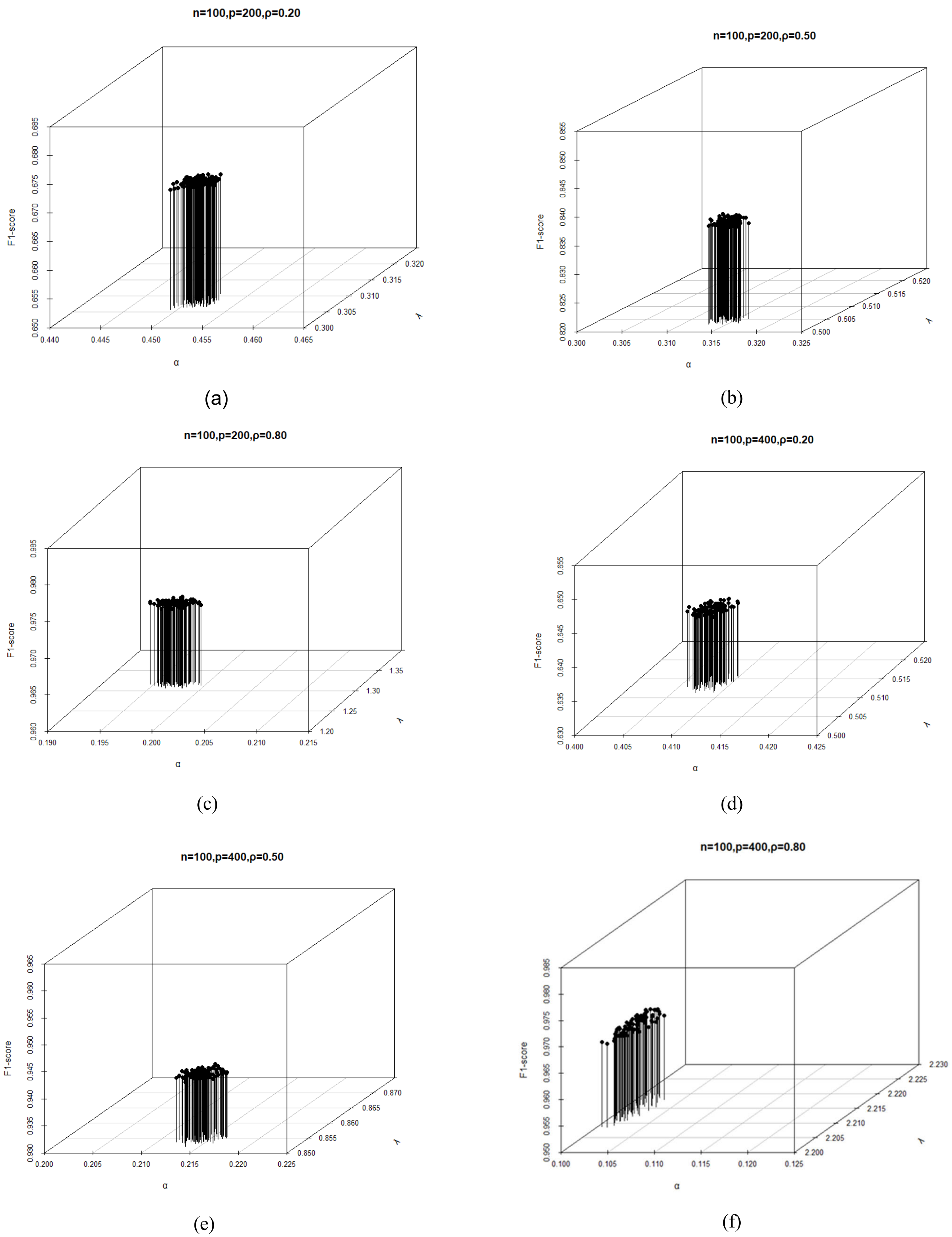
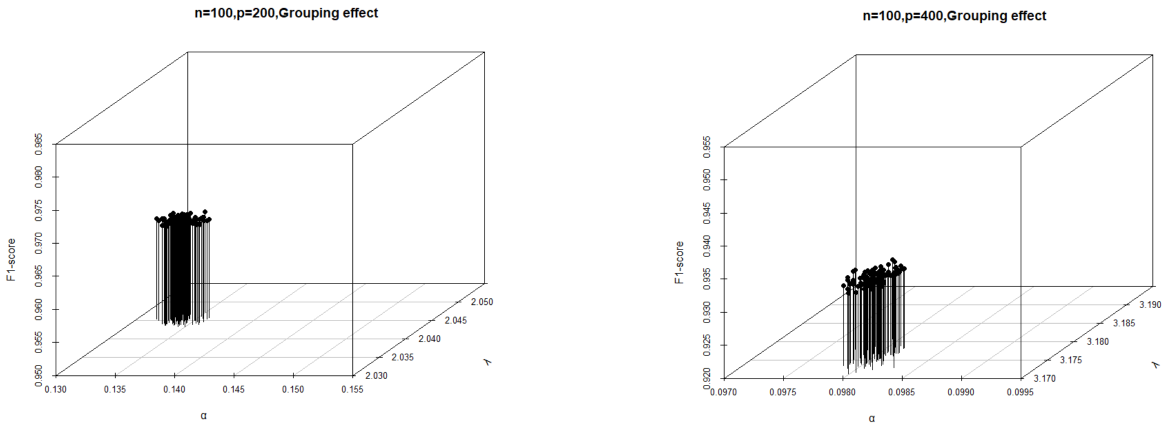


FIGURE 4. **3D graphs of sensitivity analysis results for the hyperparameters (α, λ) for each simulation scenario. (a)-(c) depict results when p = 200 ρ = 0.2, 0.5, and 0.8, respectively. (d)-(f) depict results when p = 400 ρ = 0.2, 0.5, and 0.8, respectively. (g) and (h) depict results regarding grouping effect when p = 200 and 400, respectively.**



(g) (h)

FIGURE 4. ***(Continued.)*** **3D graphs of sensitivity analysis results for the hyperparameters (α, λ) for each simulation scenario. (a)-(c) depict results when p = 200 ρ = 0.2, 0.5, and 0.8, respectively. (d)-(f) depict results when p = 400 ρ = 0.2, 0.5, and 0.8, respectively. (g) and (h) depict results regarding grouping effect when p = 200 and 400, respectively.**

evaluated according to TN (True Negative) values. MPSOA-ENet with larger TN values outperforms the traditional A-ENet and other penalized methods for Cox model in high-dimensional settings. In the simulation study, it was observed that the MPSOA-ENet method greatly improved the oracle feature of the traditional A-ENet method. Furthermore, by the literature, it has been observed once again in different scenarios that ENet and LASSO methods lack this feature.

The variable selection methods have been also compared based on the FN metric. Both MPSOA-ENet and A-ENet methods have demonstrated excellent performance, successfully detecting zero FN values in all scenarios. This outcome indicates their robustness and reliability in minimizing the false negatives, which is crucial for accurate variable selection. E-Net performs better than the LASSO and A-LASSO methods with the increase in correlation. That is, the E-Net method is more successful than LASSO and A-LASSO methods in accurately predicting true positives. According to the FN values, LASSO and A-LASSO methods showed poor performance. This indicates that the LASSO and A-LASSO methods tend to produce more false negatives. In addition, as the correlation increases, the FN values of these methods tend to increase.

In addition, the performance of the methods has been investigated as the dimensionality increases when the correlation between variables remains constant. In general, there is an increase in MSSE values in all methods. However, the increase in MPSOA-ENet and AE-Net methods is quite small. On the other hand, as the dimensionality increases, the F1-score values of the LASSO and A-LASSO methods tend to decrease. However, other methods have improved variable selection performances, especially at medium and high correlations as dimensionality increases. When the CI values of the compared methods are also examined, the proposed MPSOA-ENet method generally gives the highest value for each scenario. Generally, CI values tend to increase as the correlation increases. AE-Net, A-LASSO, and LASSO follow the proposed method in low and medium correlations.

TABLE 1. Simulation results for each simulation scenario for censoring 20%.

Scenario	Methods	MSSE	F1 Score	TN	FN	CI
Scenario 1 p: 200 ρ: 0.2	MPSOA-Enet	**3.042**	**0.714**	**173**	**0**	0.824
	A-ENet	3.818	0.625	167	**0**	**0.825**
	ENet	4.260	0.277	113	1	0.741
	LASSO	3.894	0.329	134	2	0.766
	A-LASSO	3.761	0.462	160	3	0.774
Scenario 2 p: 200 ρ: 0.5	MPSOA-ENet	**2.471**	**0.938**	**183**	**0**	**0.847**
	A-ENet	3.143	0.811	178	**0**	0.795
	ENet	3.921	0.341	132	1	0.749
	LASSO	3.546	0.319	142	4	0.770
	A-LASSO	3.299	0.558	169	3	0.789
Scenario 3 p: 200 ρ: 0.9	MPSOA-ENet	**2.017**	**1.000**	**185**	**0**	**0.863**
	A-ENet	2.721	0.909	182	**0**	0.840
	ENet	3.143	0.638	168	**0**	0.783
	LASSO	3.824	0.449	162	4	0.762
	A-LASSO	2.899	0.514	174	6	0.780
Scenario 4 p: 400 ρ: 0.2	MPSOA-Enet	**3.566**	**0.698**	**372**	**0**	**0.811**
	A-ENet	4.928	0.588	364	**0**	0.754
	ENet	5.997	0.210	289	2	0.719
	LASSO	5.741	0.238	311	3	0.723
	A-LASSO	5.360	0.297	337	4	0.747
Scenario 5 p: 400 ρ: 0.5	MPSOA-Enet	**3.043**	**1.000**	**385**	**0**	**0.826**
	A-ENet	4.842	0.857	380	0	0.767
	ENet	5.254	0.280	314	1	0.725
	LASSO	5.003	0.262	347	7	0.731
	A-LASSO	4.972	0.353	358	6	0.734
Scenario 6 p: 400 ρ: 0.9	MPSOA-Enet	**2.611**	**1.000**	**385**	**0**	**0.858**
	A-ENet	3.398	0.938	383	**0**	0.801
	ENet	4.877	0.811	378	**0**	0.791
	LASSO	5.240	0.410	369	7	0.738
	A-LASSO	3.995	0.500	382	9	0.745
Scenario 7 Grouping effect with p: 200	MPSOA-Enet	**2.587**	**0.968**	**184**	**0**	**0.870**
	A-ENet	3.524	0.732	174	**0**	0.818
	ENet	4.270	0.667	170	**0**	0.788
	LASSO	5.279	0.328	149	5	0.707
	A-LASSO	4.221	0.391	163	6	0.775
Scenario 8 Grouping effect with p: 400	MPSOA-Enet	**3.101**	**1.000**	**385**	**0**	**0.852**
	A-ENet	4.298	0.811	378	**0**	0.793
	ENet	5.447	0.600	365	**0**	0.752
	LASSO	6.561	0.184	321	7	0.695
	A-LASSO	5.850	0.480	371	9	0.753

In the low and medium correlations, the ENet method gives the lowest value among the compared methods. In high correlation, after the proposed method, A-ENet, ENet, and

TABLE 2. Simulation results for each simulation scenario for censoring rate 40%.

Scenario	Methods	MSSE	F1-Score	TN	FN	CI
Scenario 1 p: 200 ρ: 0.2	MPSOA-Enet	**6.068**	**0.667**	**170**	**0**	0.763
	A-ENet	7.780	0.536	159	**0**	**0.794**
	ENet	9.021	0.262	107	1	0.727
	LASSO	8.855	0.289	129	3	0.754
	A-LASSO	8.524	0.421	155	3	0.760
Scenario 2 p: 200 ρ: 0.5	MPSOA-Enet	**5.859**	**0.833**	**179**	**0**	**0.831**
	A-ENet	6.842	0.750	175	**0**	0.788
	ENet	7.991	0.277	119	2	0.738
	LASSO	7.456	0.289	135	4	0.765
	A-LASSO	7.097	0.480	162	3	0.772
Scenario 3 p: 200 ρ: 0.9	MPSOA-Enet	**4.595**	**0.968**	**184**	**0**	**0.854**
	A-ENet	5.007	0.833	179	**0**	0.796
	ENet	5.723	0.411	142	**0**	0.790
	LASSO	6.889	0.333	140	3	0.751
	A-LASSO	6.784	0.390	167	7	0.766
Scenario 4 p: 400 ρ: 0.2	MPSOA-Enet	**6.864**	**0.638**	**368**	**0**	**0.782**
	A-ENet	8.002	0.500	355	**0**	0.740
	ENet	9.744	0.169	270	3	0.708
	LASSO	9.201	0.180	299	5	0.733
	A-LASSO	8.920	0.262	327	4	0.724
Scenario 5 p: 400 ρ: 0.5	MPSOA-Enet	**6.342**	**0.938**	**383**	**0**	0.754
	A-ENet	7.118	0.750	375	**0**	**0.796**
	ENet	8.037	0.234	302	2	0.715
	LASSO	8.101	0.184	331	8	0.735
	A-LASSO	7.939	0.188	342	9	0.731
Scenario 6 p: 400 ρ: 0.9	MPSOA-Enet	**5.327**	**0.968**	**384**	**0**	**0.811**
	A-ENet	6.018	0.833	379	**0**	0.787
	ENet	6.568	0.667	370	**0**	0.749
	LASSO	7.657	0.265	341	6	0.737
	A-LASSO	7.262	0.323	374	10	0.729
Scenario 7 Grouping effect with p: 200	MPSOA-Enet	**5.108**	**0.968**	**184**	**0**	**0.863**
	A-ENet	5.983	0.698	172	**0**	0.809
	ENet	6.254	0.566	162	**0**	0.780
	LASSO	7.886	0.141	81	6	0.683
	A-LASSO	7.107	0.291	153	7	0.738
Scenario 8 Grouping effect with p: 400	MPSOA-Enet	**5.843**	**0.938**	**383**	**0**	**0.844**
	A-ENet	6.486	0.750	375	**0**	0.777
	ENet	7.122	0.545	360	**0**	0.723
	LASSO	8.692	0.200	328	7	0.643
	A-LASSO	7.857	0.216	367	11	0.745

A-LASSO follow, respectively. In high correlation, LASSO method gives the lowest CI value. Furthermore, evaluation metric values in Table 1 and Table 2 have been examined for each measure under the same simulation scenarios according to censoring rates. The performances of all approaches become less successful as the censoring percentage rises. In summary, our proposed MPSOA-ENet method have quite good ability in variable selection in Cox model with high dimensional data and can be competitive with other related methods in investigating the high dimensional data with from low to high multicollinearity.

A sensitivity analysis for the hyperparameters (α, λ), has been conducted to assess the validity, stability, and consistency of the proposed MPSOA-Enet method. First of all, the data set was fixed for the sensitivity analysis of α and λ in each simulation scenario. The proposed method for estimating α and λ in the Adaptive Elastic-net Cox model was applied 100 times over the fixed high-dimensional data set. Subsequently, α and λ values estimated by the proposed method, along with the selected performance metric, which is the F1-score, were recorded for each test. Then, 3D graphs were drawn with α, λ and F1 score values for each simulation scenario, with α on the x-axis, λ on the y-axis and FI score on the z-axis. When the graphs as shown in Figure 4 have been examined, for each simulation case, it has been observed that the proposed method gives consistent estimates in pretty narrow ranges for simultaneous α and λ estimates. Thus, it has been observed that these results reveal a well-defined region in the 3-dimensional (3D) space where F1-score values are relatively high. In addition, it can be seen in the 3D graphics that increasing or decreasing (α and λ) values in these narrow ranges cause very slow and steady changes in the F1-score. That is, changes in α and λ did not cause unpredictable fluctuations in the F1-score, and thus, it was observed that the F1-score was stable as long as α and λ changed in very small ranges. These results show that the proposed method is valid, consistent and stable in estimating α and λ.

VI. CONCLUSION

In this study, a modified PSO-based adaptive elastic-net method in the Cox model with high-dimensional data has been proposed as a novel variable selection for Cox model with high-dimensional data. According to the comprehensive simulation study, MPSOA-Enet outperforms other penalized methods in terms of both variable selection and prediction and estimation accuracy performance for the Cox model in investigating the high-dimensional data with low, medium, and high correlated data. The MPSOA-ENet method not only achieves excellent results under high correlation and grouping effects but also performs well in scenarios with low and medium correlation. The simulation results highlight that the MPSOA-Enet approach effectively manages both true negative and false negative rates simultaneously, surpassing the traditional adaptive elastic-net method and all other compared penalized methods.

Moreover, as the dimensionality increases, it improves the variable selection performance, making it a preferred choice for high-dimensional variable selection in the Cox model. In addition, the comprehensive simulation results showed that the MPSOA-Enet method greatly improved the oracle property of the traditional A-ENet method and also achieved more successful results than A-ENet when grouping effect was present. In short, the proposed MPSOA-ENet method, which minimizes the EBIC function and selects α and λ simultaneously, can be preferred to the traditional CV based on the grid search method for the Adaptive Elastic-net Cox model with high-dimensional data. In this study, we extensively compared the proposed method with existing penalized variable selection methods using comprehensive simulated datasets. However, it is important to acknowledge that the study's limitation lies in not applying the proposed method to real-life datasets. In future research, we aim to address this limitation by applying our proposed method to various real-world datasets and conducting comparative analyses with other methods.

REFERENCES

[1] M. Pourahmadi, *High-Dimensional Covariance Estimation: With High-Dimensional Data*. Hoboken, NJ, USA: Wiley, 2013.

[2] D. R. Cox, "Regression models and life-tables,'" *J. Roy. Stat. Soc., B, Methodol.*, vol. 34, no. 2, pp. 187–202, Jan. 1972, doi: 10.1111/j.2517-6161.1972.tb00899.x.

[3] R. Tibshirani, "The LASSO method for variable selection in the Cox model," *Statist. Med.*, vol. 16, no. 4, pp. 385–395, Feb. 1997, doi: 10.1002/(SICI)1097-0258(19970228)16:4.

[4] Z. Y. Algamal and M. H. Lee, "Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification," *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9326–9332, Dec. 2015, doi: 10.1016/j.eswa.2015.08.016.

[5] C. Li, X. Wei, and H. Dai, "Adaptive elastic net method for Cox model," Jul. 2015, *arXiv:1507.06371*. Accessed: May 2, 2023.

[6] H. H. Zhang and W. Lu, "Adaptive LASSO for Cox's proportional hazards model," *Biometrika*, vol. 94, no. 3, pp. 691–703, Aug. 2007, doi: 10.1093/biomet/asm037.

[7] Y. Ma, C. Jiang, Z. Hou, and C. Wang, "The formulation of the optimal strategies for the electricity producers based on the particle swarm optimization algorithm," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1663–1671, Nov. 2006. Accessed: Aug. 9, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1717569/

[8] Y. Luan and H. Li, "Clustering of time-course gene expression data using a mixed-effects model with B-splines," *Bioinformatics*, vol. 19, no. 4, pp. 474–482, Mar. 2003, doi: 10.1093/bioinformatics/btg014.

[9] H. C. van Houwelingen, T. Bruinsma, A. A. M. Hart, L. J. van't Veer, and L. F. A. Wessels, "Cross-validated Cox regression on microarray gene expression data," *Statist. Med.*, vol. 25, no. 18, pp. 3201–3216, Sep. 2006, doi: 10.1002/sim.2353.

[10] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[11] Y. Wu, "Elastic net for Cox's proportional hazards model with a solution path algorithm," *Stat. Sinica*, vol. 22, no. 1, p. 27, Jan. 2012, doi: 10.5705/ss.2010.107.

[12] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Taylor Francis*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001, doi: 10.1198/016214501753382273.

[13] J. Fan and R. Li, "Variable selection for Cox's proportional hazards model and frailty model," *Ann. Statist.*, vol. 30, no. 1, pp. 74–99, Feb. 2002, doi: 10.1214/aos/1015362185.

[14] H. Zou and H. H. Zhang, "On the adaptive elastic-net with a diverging number of parameters," *Ann. Statist.*, vol. 37, no. 4, p. 1733, Aug. 2009, doi: 10.1214/08-aos625.

[15] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: A practical introduction," *BMC Med. Res. Methodol.*, vol. 19, no. 1, Mar. 2019, Art. no. 64, doi: 10.1186/s12874-019-0681-4.

[16] D. U. Ozsahin, E. P. Onakpojeruo, B. Uzun, M. T. Mustapha, and I. Ozsahin, "Mathematical assessment of machine learning models used for brain tumor diagnosis," *Diagnostics*, vol. 13, no. 4, p. 618, Feb. 2023, doi: 10.3390/diagnostics13040618.

[17] D. U. Ozsahin, E. P. Onakpojeruo, B. B. Duwa, A. G. Usman, S. I. Abba, and B. Uzun, "COVID-19 prediction using black-box based Pearson correlation approach," *Diagnostics*, vol. 13, no. 7, p. 1264, Mar. 2023, doi: 10.3390/diagnostics13071264.

[18] E. Egrioglu, U. Yolcu, C. H. Aladag, and C. Kocak, "An ARMA type fuzzy time series forecasting method based on particle swarm optimization," *Math. Problems Eng.*, vol. 2013, pp. 1–12, Jul. 2013, doi: 10.1155/2013/935815.

[19] 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, doi: 10.1109/ISCV49265.2020.9204218.

[20] S.-W. Lin, K.-C. Ying, S.-C. Chen, and Z.-J. Lee, "Particle swarm optimization for parameter determination and feature selection of support vector machines," *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1817–1824, Nov. 2008, doi: 10.1016/j.eswa.2007.08.088.

[21] A. Unler and A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," *Eur. J. Oper. Res.*, vol. 206, no. 3, pp. 528–539, Nov. 2010, doi: 10.1016/j.ejor.2010.02.032.

[22] X. Cui, R. Xiao, X. Liu, H. Qiao, X. Zheng, Y. Zhang, and J. Du, "Adaptive LASSO logistic regression based on particle swarm optimization for Alzheimer's disease early diagnosis," *Chemometric Intell. Lab. Syst.*, vol. 215, Aug. 2021, Art. no. 104316, doi: 10.1016/j.chemolab.2021.104316.

[23] J. Li, K. Liang, and X. Song, "Logistic regression with adaptive sparse group LASSO penalty and its application in acute leukemia diagnosis," *Comput. Biol. Med.*, vol. 141, Feb. 2022, Art. no. 105154, doi: 10.1016/j.compbiomed.2021.105154.

[24] L.-Y. Chuang, G.-Y. Chen, S.-H. Moi, F. Ou-Yang, M.-F. Hou, and C.-H. Yang, "Relationship between clinicopathologic variables in breast cancer overall survival using biogeography-based optimization algorithm," *BioMed Res. Int.*, vol. 2019, pp. 1–12, Apr. 2019, doi: 10.1155/2019/2304128.

[25] Y. Hu, Y. Zhang, X. Gao, D. Gong, X. Song, Y. Guo, and J. Wang, "A federated feature selection algorithm based on particle swarm optimization under privacy protection," *Knowl.-Based Syst.*, vol. 260, Jan. 2023, Art. no. 110122, doi: 10.1016/j.knosys.2022.110122.

[26] O. S. Qasim and Z. Y. Algamal, "Feature selection using particle swarm optimization-based logistic regression model," *Chemometric Intell. Lab. Syst.*, vol. 182, pp. 41–46, Nov. 2018, doi: 10.1016/j.chemolab.2018.08.016.

[27] J. Chen and Z. Chen, "Extended Bayesian information criteria for model selection with large model spaces," *Biometrika*, vol. 95, no. 3, pp. 759–771, Sep. 2008, doi: 10.1093/biomet/asn034.

[28] J. Ren, Y. Du, S. Li, S. Ma, Y. Jiang, and C. Wu, "Robust network-based regularization and variable selection for high-dimensional genomic data in cancer prognosis," *Genetic Epidemiol.*, vol. 43, no. 3, pp. 276–291, Apr. 2019, doi: 10.1002/gepi.22194.

[29] N. Li, H. Yang, and J. Yang, "Nonnegative estimation and variable selection via adaptive elastic-net for high-dimensional data," *Commun. Statist.-Simul. Comput.*, vol. 50, no. 12, pp. 4263–4279, Dec. 2021, doi: 10.1080/03610918.2019.1642484.

[30] Y. Liu, G. Wang, H. Chen, H. Dong, X. Zhu, and S. Wang, "An improved particle swarm optimization for feature selection," *J. Bionic Eng.*, vol. 8, no. 2, pp. 191–200, Jun. 2011, doi: 10.1016/S1672-6529(11)60020-6.

[31] P. J. M. Verweij and H. C. Van Houwelingen, "Penalized likelihood in Cox regression," *Statist. Med.*, vol. 13, nos. 23–24, pp. 2427–2436, Dec. 1994, doi: 10.1002/sim.4780132307.

[32] J. Fan, Y. Feng, and Y. Wu, "High-dimensional variable selection for Cox's proportional hazards model," *Math. Problems Eng.*, vol. 2013, pp. 70–86, Jan. 2010, doi: 10.1214/10-IMSCOLL606.

[33] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for Cox's proportional hazards model via coordinate descent," *J. Stat. Softw.*, vol. 39, no. 5, p. 1, 2011, doi: 10.18637/jss.v039.i05.

[34] Y. Lai, M. Hayashida, and T. Akutsu, "Survival analysis by penalized regression and matrix factorization," *Sci. World J.*, vol. 2013, pp. 1–11, Apr. 2013, doi: 10.1155/2013/632030.

[35] A. Benner, M. Zucknick, T. Hielscher, C. Ittrich, and U. Mansmann, "High-dimensional cox models: The choice of penalty as part of the model building process," *Biometrical J.*, vol. 52, no. 1, pp. 50–69, Feb. 2010, doi: 10.1002/bimj.200900064.

[36] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc., B, Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[37] H. Zou, "The adaptive LASSO and its oracle properties," *Taylor Francis*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006, doi: 10.1198/016214506000000735.

[38] P. Bühlmann and E. Zürich, "High-dimensional graphs and variable selection with the LASSO," *Ann. Statist.*, vol. 34, no. 3, pp. 1436–1462, 2006, doi: 10.1214/009053606000000281.

[39] N. Xiao and Q.-S. Xu, "Multi-step adaptive elastic-net: Reducing false positives in high-dimensional variable selection," *J. Stat. Comput. Simul.*, vol. 85, no. 18, pp. 3755–3765, Dec. 2015, doi: 10.1080/00949655.2015.1016944.

[40] J. Hu, J. Huang, and F. Qiu, "A group adaptive elastic-net approach for variable selection in high-dimensional linear regression," *Sci. China Math.*, vol. 61, no. 1, pp. 173–188, Jan. 2018, doi: 10.1007/s11425-016-0071-x.

[41] K. Jiratchayut, and C. Bumrungsup, "A study of adaptive elastic net estimators with different adaptive weights," *Sci. Technol. Asia*, vol. 20, no. 3, pp. 1–7, Sep. 2015 [Online]. Available: https://ph02.tci-thaijo.org/index.php/SciTechAsia/article/view/39794

[42] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, Perth, WA, Australia, vol. 4, 1995, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.

[43] N. H. Sweilam, M. A. El-Azeem, and K. Metwally, "Self potential signal inversion to simple polarized bodies using the particle swarm optimization method: A visibility study," *J. Appl. Geophys.*, vol. 6, no. 1, pp. 195–208, Mar. 2007.

[44] N. H. Sweilam, M. M. Gobashy, and T. Hashem, "Using particle swarm optimization with function stretching (SPSO) for inverting gravity data: A visibility study," *Proc. Math. Phys. Soc. Egypt*, vol. 86, no. 2, pp. 259–281, 2008.

[45] S. Luo, J. Xu, and Z. Chen, "Extended Bayesian information criterion in the Cox model with a high-dimensional feature space," *Ann. Inst. Stat. Math.*, vol. 67, no. 2, pp. 287–311, Apr. 2015, doi: 10.1007/s10463-014-0448-y.

[46] F. Giordano, S. Milito, and M. Restaino, "A variable selection method for high-dimensional survival data," in *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. Cham, Switzerland: Springer, 2022, pp. 303–308, doi: 10.1007/978-3-030-99638-3_49.

[47] Y. Kim, J. Hao, T. Mallavarapu, J. Park, and M. Kang, "Hi-LASSO: High-dimensional LASSO," *IEEE Access*, vol. 7, pp. 44562–44573, 2019, doi: 10.1109/ACCESS.2019.2909071.

[48] F. E. Harrell, "Evaluating the yield of medical tests," *J. Amer. Med. Assoc.*, vol. 247, no. 18, p. 2543, May 1982, doi: 10.1001/jama.1982.03320430047030.

[49] F. E. Harrell, K. L. Lee, and D. B. Mark, "Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statist. Med.*, vol. 15, no. 4, pp. 361–387, 1996.

NURIYE SANCAR received the B.Sc. degree in applied mathematics and computer science from Eastern Mediterranean University, Cyprus, in 2010, and the M.Sc. and Ph.D. degrees from the Department of Mathematics, Near East University, Cyprus, in 2012 and 2018, respectively.

She is currently an Associate Professor with the Department of Mathematics, Near East University. Her research interests include data analysis, statistical modeling, metaheuristic optimization methods, regression analysis, validity and reliability analysis, and survival analysis.

EFE PRECIOUS ONAKPOJERUO received the B.Sc. degree in medical biochemistry and genetics and the M.Sc. and Ph.D. degrees in biomedical engineering.

He is currently a Passionate Biomedical Engineer, a Machine Learning Engineer, a Decision Analyst, a Data Scientist, a Prolific Researcher, and a Lecturer Assistant with the Biomedical Engineering Department and the Operational Research Center in Healthcare, Near East University, Cyprus, Turkey. He has an extensive knowledge of machine/deep learning models and decision-making analytical evaluation models for disease diagnosis/treatment approaches.

DENIZ INAN received the M.Sc. and Ph.D. degrees in mathematics from Marmara University, Turkey. She holds the position of a Professor with the Department of Statistics, Marmara University, where she is currently the Head of the Department of Statistics. Her research interests include regression analysis, shrinkage estimators, survival analysis, and meta-heuristic algorithms.

DILBER UZUN OZSAHIN received the B.Sc. degree in physics, the M.Sc. degree in natural and applied science, and the Ph.D. degree (Hons.) in physics.

She is currently an Academician and a Researcher with over 915 published articles, an H-index of 118, i10-index of 343, and citations above 51816. She is also an Associate Professor with the Department of Medical Diagnostic Imaging, College of Health Science, University of Sharjah, United Arab Emirates, and the Biomedical Engineering Department, Near East University, Cyprus. She is also the Director of the Operational Research and Integrated Artificial Intelligence in Healthcare Group, University of Sharjah, also the Director of the Operational Research Center in Healthcare, Near East University, and also an Advisor to the Rector of the reputable Near East University.