ISSN 2616-5775 Vol. 8, Issue 3: 36-42, DOI: 10.25236/AJCIS.2025.080305

Sparse Linear Multi-criteria Optimization Classifier for Disease Prediction

Kai Liu

College of Information Engineering, Nanjing University of Finance and Economics, Nanjing, Jiangsu, China
liukai20210208@163.com

Abstract: *With the changes in environment and lifestyle, the threats of diseases faced by humanity are increasingly growing. In this context, early disease prediction has become a critical issue in the medical field. However, accurate disease diagnosis based on clinical symptoms remains a highly challenging task for healthcare professionals. To address this challenge, data mining technology has demonstrated significant application value in the field of disease prediction. Currently, the amount of data generated annually in the medical field is growing exponentially, and these vast amounts of medical data provide an important foundation for precision medicine. By utilizing advanced data mining techniques, researchers can extract valuable disease characteristic patterns from large medical datasets and establish reliable predictive models. This study innovatively proposes a Sparse Linear Multi-criteria Optimization Classifier (SLMCOC) model for disease prediction. Comparative experiments with classical models such as Decision Trees and KNN demonstrate that SLMCOC achieves higher prediction accuracy. Moreover, its inherent sparsity enables the identification of critical features for classification, thereby enhancing the interpretability of the prediction results.*

Keywords: *Disease prediction; Multi-criteria optimization classifier; Sparsity; Linear*

1. Introduction

In the field of supervised learning methods based on statistical learning theory and optimization techniques, Support Vector Machines (SVMs) have undergone rapid research and development over the past two decades and have been widely applied to problems such as classification, regression, and clustering. Following SVMs, the Multi-criteria Optimization Classifier (MCOC) method has emerged as another novel approach based on optimization theory. This classifier method constructs a multi-objective mathematical programming model by calculating the overlap between data of different categories and the total distance from the data to the decision hyperplane, ultimately generating a corresponding decision function to predict the categories of new data. The MCOC has long been regarded as a useful method in the field of machine learning duing to its simplicity. In recent years, it has been widely applied to solve practical problems across various domains, such as healthcare, finance, and image recognition, demonstrating its versatility and effectiveness.

However, despite its advantages, MCOC exhibits significant limitations in terms of feature reduction, particularly when dealing with high-dimensional datasets that contain redundant or irrelevant information. This limitation not only affects the model's performance but also weakens its interpretability, making it challenging to identify the most influential features in the classification process. Therefore, it leverages sparsity methods to improve prediction accuracy and the interpretability of the results. In summary, while MCOC has proven to be a powerful and efficient classifier, its limitations in feature reduction and interpretability highlight the need for innovative approaches, such as sparsity-based regularization.

To address these challenges, this paper proposes a Sparse Linear Multi-criteria Optimization Classifier (SLMCOC) for disease prediction. The experiment proved that SLMCOC not only improves the accuracy of disease prediction, but also provides a more interpretable model by identifying the most critical features that contribute to classification decisions.

The remainder of this paper is organized as follows: Section 2 reviews related work on multi-criteria optimization. Section 3 introduces the mathematical formulation of the proposed SLMCOC model. Section 4 presents the experimental setup, including datasets and evaluation metrics. Section 5 discusses the results, highlighting the performance improvements and interpretability gains achieved by the

ISSN 2616-5775 Vol. 8, Issue 3: 36-42, DOI: 10.25236/AJCIS.2025.080305

SLMCOC. Finally, Section 6 concludes the paper and outlines potential directions for future research.

2. Related Work

All the text must use the font, Times New Roman. On Macintosh, please choose font, Times. Except in special circumstances, such as program code. All the text must use the font, Times New Roman. On Macintosh, please choose font, Times. Except in special circumstances, such as program code. Researchers have extensively explored Multi-criteria Optimization (MCO) from diverse perspectives [1,2,3]. The MCO method has been applied to network intrusion detection and credit card customer behavior analysis. A multi-criteria mathematical programming model for multi-class classification has been proposed, achieving high classification accuracy and low false positive rates in the classification of multi-class network intrusions[4]. The multi-stage multi-criteria fuzzy MCO method, grounded in fuzzy set theory and methodology, has been investigated and implemented, demonstrating a notable enhancement in the classifier's separability[5]. Nonetheless, the approach is not without its limitations; it necessitates the resolution of several distinct linear programming problems and is characterized by the instability of the solutions procured. In response to the issue of solution instability in the MCO classification model, a Regularized Linear Multi-criteria Optimization classification model (RLMCO) has been proposed[6]. The stability of the model's solution is enhanced by incorporating a regularization term into the model. Comparative analyses conducted on actual datasets between this method and MCO, RLMCO, as well as SVCs, have demonstrated an improvement in performance. Building upon the foundation of MCO classifiers, a Multi-criteria Quadratic Optimization (MCQO) classification model has been proposed to address the issues of speed and scalability inherent in multi-criteria linear programming algorithms. This method has been applied to credit scoring[7]. In order to identify subsets of features that significantly contribute to classification, a classification model and algorithm based on rough set preprocessing and multi-objective optimization have been developed. This method has been applied to practical fields such as medical diagnosis and prediction, as well as the prediction of hot spot residues in protein interactions [8,9]. The comprehensive exploration and application of MCO and its variants highlight the versatility and effectiveness of MCO techniques in addressing complex classification problems across various domains. The continuous refinement and adaptation of these methods underscore their potential for further advancements and applications in both theoretical and practical contexts.

This paper employs a Linear MCO classifier (LMCOC) for disease prediction. To identify important attributes, an approximate function of ℓ_0 -norm regularization is introduced into the linear MCO model to select feature weights, aiming to enhance the interpretability of the model.

3. Multi-criteria Optimization Classifier (MCOC)

In recent years, the Multi-criteria Optimization Classifier (MCOC) method based on optimization theory has attracted considerable research and practical application. This method obtains a compromise solution for classification problems by minimizing the overall overlap between two classes and maximizing the total distance of data points from both classes to the decision boundary. By balancing these two goals, the method aims to achieve a robust and interpretable classification model that generalizes well to unseen data. In classification problem, given a training dataset consisting of a set of observation points $X=\{x_1,x_2,...,x_n\}$ and a corresponding set of attributes $F=\{f_1,f_2,...,f_d\}$, each observation point $x_i(x_i\in R^d)$ is associated with a label $y_i(y_i\in\{-1,1\})$ indicating its belonging to one of the two classes, where d represents the number of attributes and n denotes the size of the observation set. The goal is to find a decision function to determine the classes of new observation points.

Let α_i ($\alpha_i\geq 0$) represent the deviation distance between an observation point x_i and the separating hyperplane. The distance vector $\alpha=(\alpha_1,\alpha_2,...,\alpha_n)^T$ represents the deviation distances for all onservation points in the dataset $D=\{X,Y\}$. The sum of α_i is calculated by the function $f(\alpha)=\|\alpha\|_p^p(p\geq 1)$ which should be minimized with respect to α. Similarly, let $\beta_i(\beta_i\geq 0)$ denote the distance between an observation point x_i and the separating hyperplane. The distance vector $\beta=(\beta_1,\beta_2,\ldots,\beta_n)^T$ represents the distances for all points in the dataset D. The sum of β_i is

ISSN 2616-5775 Vol. 8, Issue 3: 36-42, DOI: 10.25236/AJCIS.2025.080305

computed by the function $f(\beta)=\|\beta\|_q^q (q\geq 1)$ which should be maximized with respect to β. Based on this, a classifier model based on multi-criteria optimization can be defined as follows:

$$\begin{aligned}&\min_{\alpha}\|\alpha\|_p^p \text{ and } \max_{\beta}\|\beta\|_q^q (p,q\geq 1))\\&s.t. y_i(\omega^T x+b)=\beta_i-\alpha_i,\alpha_i,\beta_i\geq 0, i=1,\ldots,n.\end{aligned} \tag{1}$$

Where ω $(\omega\in\mathbb{R}^d)$ is the weight vector and b is the intercept.

A new MCOC model can be formulated by introducing a penalty factor $C_1(C_1>0)$, and it can be rewritten as follows:

$$\begin{aligned}&\min_{\alpha,\beta} C_1\|\alpha\|_p^p-\|\beta\|_q^q (p,q\geq 1))\\&s.t. y_i(\omega^T x+b)=\beta_i-\alpha_i,\alpha_i,\beta_i\geq 0, i=1,\ldots,n.\end{aligned} \tag{2}$$

If p=1 and q=1 in model (2), we obtain a Linear Multi-Criteria Optimization Classifier (LMCOC) with linear constraints, which can be described as:

$$\begin{aligned}&\min_{\alpha,\beta} C_1\sum_{i=1}^{n}\alpha_i-\sum_{i=1}^{n}\beta_i\\&s.t. y_i(\omega^T x+b)=\beta_i-\alpha_i,\alpha_i,\beta_i\geq 0, i=1,\ldots,n.\end{aligned} \tag{3}$$

By solving model (3), once the weight vector and intercept are obtained, the separating hyperplane can be expressed as: $\omega^T x+b=0$. Meanwhile, the label of a new observation point can be determined based on the following decision function: $y=sign(\omega^T x+b)$. The observation point x belongs to the positive class when $y=1$, this is, when $\omega^T x+b>0$.

4. Sparse Linear Multi-criteria Optimization Classifier (SLMCOC)

It is essential to incorporate sparsity method into the MCOC models. By introducing sparsity, the model can automatically select the most relevant features while discarding redundant ones, thereby enhancing its robustness, accuracy, and computational efficiency. Furthermore, sparsity improves the interpretability of the model, as it provides clearer insights into the contribution of each feature to the classification outcome.

In classification problems, regularization techniques are widely used to control model complexity, prevent overfitting, and enhance the model's generalization ability. The core idea of regularization is to introduce a penalty term into the objective function, which restricts the range of model parameters, thereby avoiding excessive fitting to the training data. For linear classifiers, Each element ω_m in the weight vector represents the contribution of the corresponding attribute to the classification outcome. The larger the absolute value of ω_m, the more significant the influence of that attribute on the classification.

More precisely, the weight vector ω $(\omega=(\omega_1,\ldots\omega_d)^T)$ is used to determine the importance of each attribute for classification. Specifically, the value of ω_m indicates whether the m-th attribute should be kept, based on whether ω_m is nonzero or zero. The ℓ_0-norm of the augmented weight vector $\bar{\omega}^{(t)}$ $(\bar{\omega}^{(t)}=(\omega_1^{(t)},\ldots,\omega_d^{(t)},\omega_{d+1}^{(t)})^T)$, where $\omega_{d+1}^{(t)}$ is the intercept of the separating hyperplane at iteration t, is defined as an approximate convex function with the initial value $\bar{\omega}^{(0)}$ as follow:

$$\|\bar{\omega}^{(t)}\|_0\approx(\bar{\omega}^{(t-1)})^T diag(\theta^{(t)})\bar{\omega}^{(t-1)}, \tag{4}$$

and the m-th element $\theta_m^{(t)}$ of the $(d+1)\times(d+1)$ martix $diag(\theta^{(t)})$ is defined by

ISSN 2616-5775 Vol. 8, Issue 3: 36-42, DOI: 10.25236/AJCIS.2025.080305

$$\theta_m^{(t)} = \begin{cases} \dfrac{1}{(\bar{\omega}_m^{(t-1)})^2}, if \left|\bar{\omega}_m^{(t-1)}\right| > \rho \\ \dfrac{1}{\rho^2}, otherwise. \end{cases} \tag{5}$$

According to equations (4) and (5), we can get the ℓ_0-norm of the weight vector $\bar{\omega}^{(t)}$ at iteration t when $t \to +\infty$ with the optimal weight vector ω^*. The detailed derivation process is as follows.

$$\begin{aligned} \lim_{t\to+\infty} \left\|\bar{\omega}^{(t)}\right\|_0 &= \lim_{t\to+\infty} (\bar{\omega}^{(t-1)})^T diag(\theta^{(t)}) \bar{\omega}^{(t-1)} \\ &= \lim_{t\to+\infty} \sum_{\bar{\omega}_m^{(t-1)} \neq 0} \frac{1}{(\bar{\omega}_m^{(t-1)})^2} \times (\bar{\omega}_m^{(t-1)})^2 + \sum_{\bar{\omega}_m^{(t-1)} = 0} \frac{1}{\rho^2} \times 0 \\ &= \lim_{t\to+\infty} \sum_{\bar{\omega}_m^{(t-1)} \neq 0} 1 \\ &= \left|\{m \mid \omega_m^* \neq 0, m = 1, ..., d\}\right| \end{aligned} \tag{6}$$

To identify significant features from the feature set, we introduce $\|\omega\|_0$ into model (3). The Sparse Linear Multi-criteria Optimization Classifier (SLMCOC) model can be formulated as:

$$\begin{aligned} &\min_{\omega,\alpha,\beta} C_2 \|\omega\|_0 + C_1 \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \beta_i \\ &s.t. y_i(\omega^T x + b) = \beta_i - \alpha_i, \alpha_i, \beta_i \geq 0, i = 1, \ldots, n. \end{aligned} \tag{7}$$

The parameter C_2 （$C_2 > 0$) controls the sparsity of the model.

5. Experiment

In this section, we evaluate the proposed model using several classic datasets. The experimental design, along with a detailed analysis of prediction accuracy and feature selection performance, will be thoroughly discussed in following subsections.

5.1. Datasets

The datasets used in this experiment include heart diease dataset, wisconsin diagnostic breast cancer dataset (WDBC) and diabetes dataset, are sourced from the UCI Machine Learning Repository. These three datasets have 13, 30 and 16 features, respectively. The HD contains 303 records from the Cleveland Clinic, used to predict whether a patient has heart disease. Although the dataset is relatively small, it is rich in features, making it suitable for feature selection and model interpretability research. It has 165 positive samples and 138 negative samples.The WDBC contains 569 samples and is used to predict whether a breast tumor is benign or malignant. With a moderate amount of data and high feature dimensionality, this dataset is well-suited for feature selection and performance evaluation of classification algorithms. It has 212 positive samples and 357 negative samples. The diabetes dataset contains medical data from 768 patients and is used to predict whether a patient has diabetes. The dataset has a moderate amount of data and a low feature dimensionality. It has 268positive samples and 500 negative samples.

The datasets are splited into a training set and a testing set in an 8:2 ratio. And the sampling of positive and negative samples should generally adhere to the ratio of the positive and negative classes. In this experiment, we randomly select 250, 600, and 450 samples for training to obtain the optimal parameters to get the best model, while 50, 150, and 100 observation points were used to generate prediction results, respectively.

5.2. Experiment Design

Before the experiment begins, all attribute columns of the dataset are standardized using min-max normalization. We use 5-fold cross-validation on training set to get optimal model with the parameter sets ($C_1 = [10^{-5}, ..., 10^5]$ and $C_2 = [10^{-5}, ..., 10^5]$). And after introducing the regualtion function in section

ISSN 2616-5775 Vol. 8, Issue 3: 36-42, DOI: 10.25236/AJCIS.2025.080305

3.1, we set a sparse threshold $\tau(\tau = e^{-4})$ to select features. It indicates that the m-th attribute is not important for calssification if $|\omega_m^*| < \tau$. Otherwise, this attribute is considered an important attribute.

We select several measures to evaluate the performance of models. Total accuracy is the most fundamental performance measures of a model, reflecting its overall performance.. And F_1 score also important in imbalanced class datasets. In the evaluation of feature selection performance, we set a feature weight threshold and count the number of features with non-zero weights to compare the feature reduction capabilities of different algorithms.

5.3. Analysis of Prediction Performance

We conducted comparative experiments on SVC, Decision Tree, K-Nearest Neighbors (KNN), Naïve Bayes, Random Forest, Logistic Regression and SLMCOC models with three datasets.

Table 1: Predictive performance of models in heart diease dataset.

Models	Accuracy (%)	F_1 score
SVC	72.58	0.8325
Decision Tree	81.84	0.8333
KNN	81.48	0.8500
Naïve Bayes	83.33	0.9130
Random Forest	82.09	0.8461
Logistic Regression	79.26	0.8174
SLMCOC	**86.67**	**0.9166**

The bold value indicates that the classifier achieves the best performance compared to other models.

From the results in Table 1, we observe that SLMCOC achieved the highest accuracy of 86.67% and the best F_1 score of 0.9166, indicating it is the most effective model for predicting heart disease in this dataset. This performance is notably better than that of other classifiers, marking SLMCOC as the best performing model in our experiment. The Naïve Bayes model also showed competitive performance, with an accuracy of 83.33% and an F_1 score of 0.9130, which makes it the second-best model in terms of F_1 score. It outperforms many of the other traditional models, such as Random Forest (accuracy: 82.09%, F1 score: 0.8461) and Logistic Regression (accuracy: 79.26%, F_1 score: 0.8174).

Table 2: Predictive performance of models in wisconsin diagnostic breast cancer dataset.

Models	Accuracy (%)	F_1 score
SVC	97.82	0.9329
Decision Tree	95.58	0.9634
KNN	93.81	0.9417
Naïve Bayes	94.69	0.9832
Random Forest	98.57	0.9755
Logistic Regression	96.33	0.9805
SLMCOC	**99.04**	**1.0000**

The bold value indicates that the classifier achieves the best performance compared to other models.

As demonstrated in Table 2, we observe that the SLMCOC model achieved the highest accuracy of 99.04% and an F_1 score of 1, making it the best-performing classifier in this experiment. The perfect F_1 score indicates that SLMCOC not only correctly classifies the majority of instances but also handles the balance between precision and recall exceptionally well. Following SLMCOC, Random Forest showed impressive results with an accuracy of 98.57% and an F_1 score of 0.9755, placing it as the second-best classifier.

Table 3: Predictive performance of models in diabetes dataset.

Models	Accuracy (%)	F_1 score
SVC	90.15	0.8974
Decision Tree	93.27	0.9808
KNN	91.35	0.9836
Naïve Bayes	88.64	0.9455
Random Forest	95.19	0.9846
Logistic Regression	86.28	0.9167
SLMCOC	**96.80**	**1.0000**

ISSN 2616-5775 Vol. 8, Issue 3: 36-42, DOI: 10.25236/AJCIS.2025.080305

The bold value indicates that the classifier achieves the best performance compared to other models.

As the results shown in Table 3, the SLMCOC model achieved the highest performance with an accuracy of 96.80% and an F_1 score of 1, marking it as the best performing model for this dataset. Random Forest and KNN followed closely, showing the power of ensemble and distance-based methods. Decision Tree also performed well, while traditional models like SVC, Naïve Bayes, and Logistic Regression provided good results but were outperformed by more advanced models.

5.4. Feature Importance Analysis

After a finite number of iterations, the SLMCOC model obtains the optimal weight values. Feature weights with absolute values smaller than the feature threshold are set to zero, and the remaining features are retained as important ones. The number of feature weights that are not zero can simply be used as an indicator of the model's sparsity. To evaluate whether the model effectively enhances interpretability, the proportion of each feature's weight value relative to the total weight of all remaining features can be calculated. The proportion of important feature weights of the the three datasets is shown in Figure 1, 2.

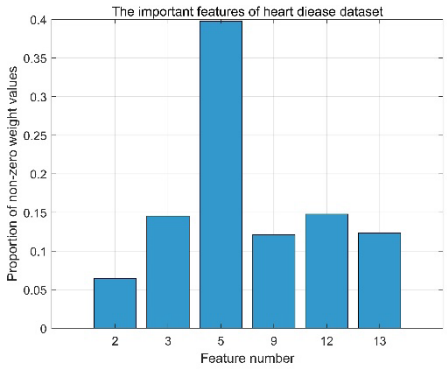

Figure 1: The important features of heart diease dataset.

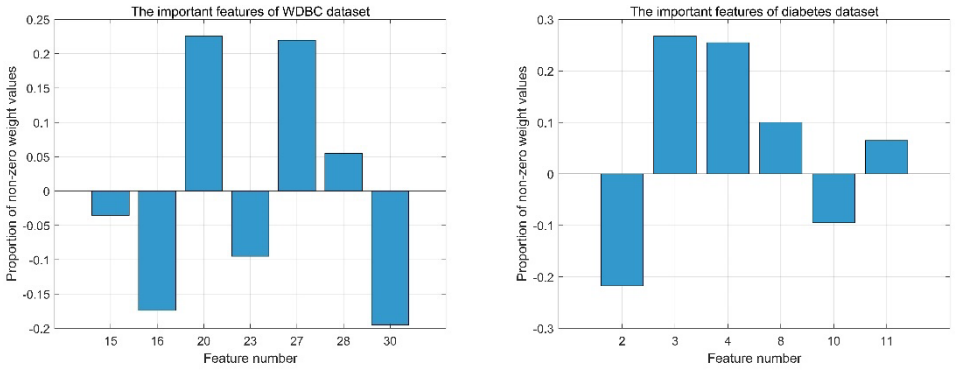

Figure 2: The important features of WDBC and diabetes datasets.

As shown in Figure 1, the original heart disease dataset has 13 features. After feature sparsification, 6 features remain while achieving a sparsity rate of 53.84% (the proportion of redundant features to the total number of features) while maintaining the model's prediction accuracy. Additionally, based on the proportion of the feature weight to the total weight, it was found that Feature 5 is crucial to the model, contributing to the interpretability of the model's predictions. Similarly, based on the analysis from Figure 2, the WDBC dataset initially had 30 features. After eliminating redundant features, 7 features remained, achieving a sparsity rate of 76.67%. Among these, features 16, 20, 27, and 30 are considered highly important for the model. For the diabetes dataset, there were initially 16 features. After sparsification, 6 features remained, resulting in a sparsity of 62.50%. Additionally, the weight values of features 2, 3, and 4 represent a significant proportion of the total weight.

6. Conclusions

In this paper, we discussed the historical research on multi-criteria optimization classifiers and

ISSN 2616-5775 Vol. 8, Issue 3: 36-42, DOI: 10.25236/AJCIS.2025.080305

considered their application in the field of disease prediction. After developing a sparse linear multi-criteria optimization classifier model and comparing it experimentally with other classical classifiers, we found a significant improvement in prediction accuracy. Additionally, the model exhibits strong capabilities in feature sparsification, which is expected to aid in disease prevention. In future research, we will extend the study of multi-criteria optimization to non-linear models and investigate better sparsification functions.

References

[1] Shi, Y., Peng, Y., Xu, W., et al. Data mining via multiple criteria linear programming: Applications in credit card portfolio management[J].International Journal of Information Technology & Decision Making, 2002,1(1):131-151.
[2] Shi Y. Multiple Criteria Optimization based Data Mining Methods and Applications: A Systematic Survey[J]. Knowledge and Information Systems, 24(3), 369–391.
[3] Shi, Y., Tian, Y., Kou, G., Peng, Y., Li, J. Optimization based data mining: theory and applications[M]. Springer, London:DOI:10.1007/978-0-85729-504-0.
[4] Kou G, Peng Y, Chen Z, et al. Multiple criteria mathematical programming for multi-class classification and application in network intrusion detection[J].Information Sciences,2008,179(4):371-381.
[5] He J. et al. Classifications of Credit Cardholder Behavior by using Fuzzy Linear Programming[J]. International Journal of Information Technology and Decision Making,2004,3(4):633-650.
[6] Shi Y., Tian Y., Chen X., Zhang P. Regularized multiple criteria linear programs for classification[J]. Science in China Series F: Information Sciences, 2009, 52(10):1812-1820.
[7] Peng Y., Kou G., Shi Y. and Chen Z. A Multi-Criteria Convex Quadratic Programming Model for Credit Data Analysis[J].Decision Support Systems,2007,44(4):1016-1030.
[8] Zhang Z., Shi Y. and Gao G. (2009b). A Rough Set-based Multiple Criteria Linear Programming Approach for the Medical Diagnosis and Prognosis. Expert Systems with Applications,2008,36(5), 8932–8937.
[9] Chen R., Zhang Z., et al. Prediction of protein interaction hot spots using rough set-based multiple criteria linear programming[J].Journal of Theoretical Biology,2011,269(1):174-180.