

**A kernel-free L1 norm regularized  $\nu$ -support vector machine model with application****Junyuan Xiao<sup>a</sup>, Guoyi Liu<sup>a</sup>, Min Huang<sup>a\*</sup>, Zhihua Yin<sup>b</sup> and Zheming Gao<sup>a</sup>**<sup>a</sup>College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China<sup>b</sup>Department of Epidemiology, School of Public Health, China Medical University, Shenyang, Liaoning, People's Republic of China**CHRONICLE***Article history:*Received March 10 2023  
Received in Revised Format  
May 10 2023  
Accepted August 3 2023  
Available online  
August 19 2023*Keywords:**Binary classification*  
*Quadratic surface support vector machines*  
*L1 norm regularization*  
 *$\nu$ -SVM***ABSTRACT**

With a view to overcoming a few shortcomings resulting from the kernel-based SVM models, these kernel-free support vector machine (SVM) models are newly promoted and researched. With the aim of deeply enhancing the classification accuracy of present kernel-free quadratic surface support vector machine (QSSVM) models while avoiding computational complexity, an emerging kernel-free  $\nu$ -fuzzy reduced QSSVM with L1 norm regularization model is proposed. The model has well-developed sparsity to avoid computational complexity and overfitting and has been simplified as these standard linear models on condition that the data points are (nearly) linearly separable. Computational tests are implemented on several public benchmark datasets for the purpose of showing the better performance of the presented model compared with a few known binary classification models. Similarly, the numerical consequences support the more elevated training effectiveness of the presented model in comparison with those of other kernel-free SVM models. What's more, the presented model is smoothly employed in lung cancer subtype diagnosis with good performance, by using the gene expression RNAseq-based lung cancer subtype (LUAD/LUSC) dataset in the TCGA database.

© 2023 by the authors; licensee Growing Science, Canada

**1. Introduction**

Recently, machine learning has been extensively applied to various fields of contemporary science, such as computer science, statistics, medicine, economics, engineering, and applied mathematics (Di & Joo, 2007; Mainali et al., 2021), and one of which crucial research orientations is just data categorization is just one of most crucial research orientation; additionally, there are numerous research and applications about it. As a computationally powerful tool, support vector machine (SVM) has shown to be superior to other classification methods in various applications and is gaining popularity (Cholette et al., 2017; Danenas & Garsva, 2015).

SVM was originally proposed by Cortes and Vapnik (1995). It has been well developed in the last two decades and utilized in numerous practical problems. With the aim of handling one binary classification task, a separation hyperplane with a maximized margin is discovered in the SVM model through solving a convex quadratic programming problem. Later, Schölkopf et al. (2000) propounded the so-called  $\nu$  support vector machine ( $\nu$ -SVM) regulating the number of support vectors by one unique parameter  $\nu$ . The parameter  $\nu$  of  $\nu$ -SVM is bounded, so it is more convenient to tune the parameters of  $\nu$ -SVM.

Masses of binary data sets applied to practical problems do not become linearly separable. For the case of linearly inseparable data sets, Cortes and Vapnik (1995) proposed an SVM model with kernel functions. Through adopting those functions of the kernel, the SVM with kernel can yield the nonlinear separation surfaces for some data sets. Our thought illustrates that the related data is mapped into a more elevated dimensional characteristic space; subsequently, the mapped data points are categorized linearly. The  $\nu$ -SVM can be provided with the kernels to manage nonlinear cases as well (Yan & Xu, 2007). It is clear how to select a suitable kernel function; in addition, to deeply tune the involved hyper-parameters depletes a considerable

\* Corresponding author

E-mail: [mhuang@mail.neu.edu.cn](mailto:mhuang@mail.neu.edu.cn) (M. Huang)

ISSN 1923-2934 (Online) - ISSN 1923-2926 (Print)

2023 Growing Science Ltd.

doi: 10.5267/j.ijiec.2023.8.002

quantity of computational efforts (Cristianini & Shawe, 2000; Schölkopf & Smola, 2002). As a result, it is natural to consider a practical approach seeking the nonlinear classifiers within the source space.

To avoid a challenge of using nonlinear kernel tricks, Dagher et al. (2008) proposed kernel-free SVM. Kernel-free SVM directly uses nonlinear surfaces to separate the two types of data without mapping them to a high-dimensional feature space. It does not require pre-selection of kernels or tuning kernel parameters, which saves plenty of computational efforts and time during the training. Later, Bai et al. (2015) presented one kernel-free least squares QSSVM with the intention of diagnosing diseases. Luo et al. (2016) introduced fuzzy methods and presented one fuzzy kernel-free quadratic surface SVM model. Gao et al. (2019) presented the least squares twin QSSVM through acquiring figures using two quadratic surfaces. Nevertheless, kernel-free QSSVM models might also own drawbacks. Firstly, the quantity of variables inside SQSSVM will exponentially ascend after the equivalent reconstruction (Luo et al., 2016; Gao et al., 2022; Mousavi et al., 2019; Yan et al., 2018), which degrades the training efficiency of SQSSVM. Moreover, SQSSVM cannot be simplified to a hyperplane as the given data set becomes linearly separable (Mousavi et al., 2019). One potential remedy to resolve both issues is adding an L1 norm regularization term to the objective function. The introduction of the L1 norm regularization term can bring sparsity to the model (Mousavi & Shen; 2019; Mousavi et al., 2019; Shen & Mousavi; 2018). As the penalty parameter of the L1 norm regularization term increases, the decision variables of the quadratic surface become more sparse as a way to prevent computational complexity and overfitting. At the same time, with the large penalty parameter of the regularization term, the quadratic surface will be nearly a linear plane (Mousavi et al., 2019). In similar circumstances, the model is capable of capturing the sparsity mode of the Matrix coefficient of the authentic quadratic surface.

In this paper, we put forward the most advanced kernel-free  $v$ -fuzzy is raised, which reduced quadratic surface support vector machine with L1 norm regularization model for nonlinear binary categorization, marked as (L1- $v$ -FRQSSVM). The proposed model was rigorously derived and studied. Similarly, computational tests are implemented with the objective to detect the accuracy of classification, the training effectiveness, along with the parameter sensibility. These models of SVM have been smoothly employed in numerous authentic issues, one of which is to diagnose the disease. Lung carcinoma, acting as a malignant tumor, is one of the most common carcinomas all over the world. Early diagnosis is important for the treatment (Xie et al., 2021). Lung carcinoma has several subtypes, among which lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC) become the most pervasive. LUAD and LUSC differ in clinical presentation, diagnosis, treatment and prognosis. Therefore, it is very important to make accurate pathological diagnosis for the subtypes of patients with subtypes of lung carcinoma. Presently, several QSSVM models with no kernel (Luo et al., 2016) have been adopted in the diagnosis of diseases. However, on account of the vast quantity of features inside gene expression RNAseq-based lung cancer datasets, the kernel-free QSSVM model presented in literary works may not be appropriate to diagnose the lung cancer subtype. Some advanced machine learning models (Malik et al., 2022; Tang et al., 2022) have similarly been put forward and adopted for lung cancer subtype diagnosis. In this paper, by using the gene expression RNAseq dataset of LUAD/LUSC obtained from the TCGA database, The proposed model was demonstrated to cope with the subtype diagnosis of lung cancer successfully. The dominating contributions of our research are summed up as listed:

1. The presented (L1- $v$ -FRQSSVM) model introduces the L1 norm regularization term. It promotes the sparsity of the model, so that computational complexity and overfitting can be avoided. At the same time, it enables the model to infinitely approximate the linear plane when facing the linearly divisible case of the split interface.
2. The proposed (L1- $v$ -FRQSSVM) model combines the  $v$ -SVM idea. As a bounded parameter, the parameter  $v$  is selected in a more fixed range, which makes the model more convenient for parameter tuning. Meanwhile the model is able to control the quantity of support vectors by the parameter  $v$ .
3. The separation quadratic surface originates from the presented (L1- $v$ -FRQSSVM) model with no cross terms in form of the quadratic form such that it's far more effective compared to the other experimented QSSVM models with no kernel on the part of time for computing. What's more, we equip the presented model with a fuzzy membership, which is employed in every one of training data points and assists in lessening the comparative benefit of noise or outliers, producing the optimal separation surface.
4. We apply this proposed (L1- $v$ -FRQSSVM) model to the diagnosis of lung cancer subtypes (LUAD/LUSC). There are nearly two million features in the data of the lung cancer subtype diagnosis issue, limiting the usability of the known kernel-free QSSVM model, however, doesn't affect the usability of the presented model. In accordance with the computational consequences, the great performance for lung cancer subtype diagnosis indicates that the kernel-free model has good potential for handling real-world problems with data having a vast quantity of features.

The other parts of our article are structured in the following manner. Some well-established support vector machine models are reviewed in Section 2; they are the basis of the presented model in this research. The presented (L1- $v$ -FRQSSVM) model is put forward inside Section 3. The computational tests are depicted inside Section 4, which is performed on several public benchmark datasets and gene expression RNAseq datasets from LUAD/LUSC in the TCGA database. Additionally, the conclusion is described in Section 5.

## 2. Preliminaries

Some notations are introduced in this part, as well as the preconditions used in this paper, along with a short retrospection of several relevant SVM models for the sake of binary classification.

Throughout the research, the lowercase letters are employed for scalars, the lowercase bold letters are employed for vectors; moreover, uppercase bold letters are also employed for matrices.  $\mathbb{R}$  refers to the set of authentic numbers; additionally,  $\mathbb{R}^n$

refers to the  $n$ -dimensional authentic vector.  $\mathbb{R}_+^n$  refers to the  $n$ -dimensional non-negative real vector.  $I_n$  is utilized to denote the  $n \times n$  identity matrix and the set of  $n \times n$  matrices are regard as  $\mathbb{S}_n$ . Also, the binary data set is defined as below:

$$\mathcal{D} = \left\{ (\mathbf{x}^{(i)}, y^{(i)})_{i=1, \dots, N} \mid \mathbf{x}^{(i)} \in \mathbb{R}^n, y^{(i)} \in \{-1, 1\} \right\} \tag{1}$$

Next, a brief retrospection of several corresponding SVM models is offered. As to an assignment of binary classification, the SVM mechanism is used for discovering one best hyperplane that satisfies these classification requirements such that the hyperplane maximizes the separation margin while guaranteeing the classification accuracy (Cortes & Vapnik, 1995). While the data is not separable linearly, to handle training data sets with outliers and noise, the soft-margin is utilized through depicting this slack vector  $\boldsymbol{\xi} = [\xi_1, \dots, \xi_n] \in \mathbb{R}_+^n$ . We figure out the soft-margin linear SVM:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & (b + \mathbf{w}^T \mathbf{x}^{(i)}) y^{(i)} \geq 1 - \xi_i \quad \forall i = 1, \dots, N \\ & b \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n, \boldsymbol{\xi} \in \mathbb{R}_+^n \end{aligned} \tag{SVM}$$

in which  $C > 0$  means a given parameter. Each  $\xi_i$  is utilized to measure the classification error for each  $\mathbf{x}^{(i)}$ . Like the SVM models, the  $\nu$ -SVM model is propounded by Schölkopf et al. for the sake of binary classification inside (Schölkopf et al., 2000). We figure out that model as below:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{v}\|_2^2 - \nu \rho + \frac{1}{N} \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & y^{(i)} (\mathbf{v} \mathbf{x}^{(i)} + d) \geq \rho - \xi_i \quad \forall i = 1, \dots, N \\ & \mathbf{v} \in \mathbb{R}^l, d \in \mathbb{R}, \rho \in \mathbb{R}_+, \boldsymbol{\xi} \in \mathbb{R}_+^n \end{aligned} \tag{\nu-SVM}$$

To address the linearly inseparable case, Schölkopf et al. (2000) similarly propounded the  $\nu$ -SVM models with the kernel. The thought is just to use a nonlinear kernel  $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^l (l > n)$  to map these datasets points in a more elevated dimensional feature space; next, the mapped data is classified. That kind of  $\nu$ -SVM model is figured out as below:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{v}\|_2^2 - \nu \rho + \frac{1}{N} \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & y^{(i)} (\mathbf{v} \phi(\mathbf{x}^{(i)}) + d) \geq \rho - \xi_i \quad \forall i = 1, \dots, N \\ & d \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}_+^n, \mathbf{v} \in \mathbb{R}^l, \rho \in \mathbb{R}_+ \end{aligned} \tag{\nu-KSVM}$$

In this model,  $\nu \in [0, 1]$  is the only parameter, and the role of  $\nu$  is capable of effectively controlling the quantity of these support vectors (Schölkopf et al., 2000). Its dual issue could be figured out as follows:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) y^{(i)} y^{(j)} \alpha_i \alpha_j \\ \text{subject to} \quad & \sum_{i=1}^N \alpha_i y^{(i)} = 0 \\ & \sum_{i=1}^N \alpha_i \geq \nu \\ & 0 \leq \alpha_i \leq \frac{1}{N}, i = 1, \dots, N. \end{aligned} \tag{\nu-DKSVM}$$

in which, the given parameter is denoted by  $\nu$ . Several theoretical qualities have already been studied inside (Schölkopf et al., 2000), such as the fact that the  $\nu$  parameter is bounded and can be utilized to control the number of support vectors. Nevertheless, as to an appointed data set, there exists no general rule of selecting the most appropriate kernel for utilization automatically. Furthermore, the effect of one soft SVM model with a kernel leans chiefly upon the chosen parameter set in the kernel function; moreover, to tune those parameters often takes much effort (Cristianini & Shawe, 2000; Schölkopf & Smola, 2002). To avoid the challenges in using the nonlinear kernel trick, multiple SVM models with no kernel have been put forward and promoted to classify nonlinearly (Dagher, 2008; Luo et al., 2016; Mousavi et al., 2019). Rather than mapping

data points into one more elevated dimensional feature space, those aforementioned SVM models divide data sets through straightforward yielding nonlinear separation surfaces within the source space. One representative kernel-free SVM model listed below is the soft quadratic surface SVM (SQSSVM) model (Luo et al., 2016), maximizing the sum of comparative geometrical margins of data points when penalizing these margin errors:

$$\begin{aligned} \min \quad & \sum_{i=1}^N \|\mathbf{W}\mathbf{x}^{(i)} + \mathbf{b}\|_2^2 + C \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & \\ & y^{(i)} \left( \frac{1}{2} \mathbf{x}^{(i)T} \mathbf{W} \mathbf{x}^{(i)} + \mathbf{x}^{(i)T} \mathbf{b} + c \right) \geq 1 - \xi_i, i = 1, \dots, N, \\ & \mathbf{W} \in \mathbb{S}^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}, \xi \in \mathbb{R}_+^N \end{aligned} \quad (\text{SQSSVM})$$

in which  $C$  denotes an appointed parameter. We could formulate the (SQSSVM) model equivalently again into the model listed below with the aim of implementing more easily:

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{z}^T \mathbf{G} \mathbf{z} + C \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & \\ & y^{(i)} (\mathbf{s}^{(i)T} \mathbf{z} + c) \geq 1 - \xi_i, \quad i = 1, \dots, N, \\ & \mathbf{z} \in \mathbb{R}^{n(n+1)/2}, c \in \mathbb{R}, \xi \in \mathbb{R}_+^N \end{aligned} \quad (\text{SQSSVM}')$$

in which  $\mathbf{G}$  and  $\mathbf{s}^{(i)}$  ( $\forall i = 1, \dots, N$ ) are stated inside (Luo et al., 2016). Furthermore, we figure out the dual issue of (SQSSVM') (Luo et al., 2016) is as stated below:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y^{(i)} y^{(j)} \mathbf{s}^{(i)T} \mathbf{G}^{-1} \mathbf{s}^{(j)} - \sum_{i=1}^N \alpha_i \\ \text{subject to} \quad & \\ & \sum_{i=1}^N \alpha_i y^{(i)} = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, N. \end{aligned} \quad (\text{DSQSSVM}')$$

In real-life applications, there is often a certain amount of unavoidable noise in the datasets, and this noise can reduce the accuracy of the SVM model classification. Therefore, Lin and Wang proposed an SVM model with fuzzy property (Lin & Wang, 2002), and they quoted  $\theta_i \in (0,1]$  as a weight to indicate the importance of each data point  $\mathbf{x}^{(i)}$  for classification. If  $\theta_i$  is closer to 1, then  $\mathbf{x}^{(i)}$  is more important, and conversely if  $\theta_i$  is closer to 0, then  $\mathbf{x}^{(i)}$  is more like an outlier. With the fuzzy attribute, the SVM model helps to classify data sets where noise emerges. Considering one data set  $\mathcal{D}$  as stated by (1), we have a definition of the fuzzy membership  $\theta_i$  of  $\mathbf{x}^{(i)}$  as below:

$$\theta_i = \begin{cases} 1 - \frac{\|\mathbf{x}_+ - \mathbf{x}^{(i)}\|_2}{r_+ + \epsilon} & i \in I_+ \\ 1 - \frac{\|\mathbf{x}_- - \mathbf{x}^{(i)}\|_2}{r_- + \epsilon} & i \in I_- \end{cases} \quad (2)$$

in which a small positive number is used to prevent the fuzzy membership value from being zero. The  $\mathbf{x}_+$  and  $\mathbf{x}_-$  are average of positive data points and negative data points, respectively. The formulations for  $r_+$  and  $r_-$  are as follows, expressed as the radii of two classes of data points, separately.

$$r_+ = \max_{i \in I_+} \{\|\mathbf{x}_+ - \mathbf{x}^{(i)}\|_2\}, \quad r_- = \max_{i \in I_-} \{\|\mathbf{x}_- - \mathbf{x}^{(i)}\|_2\} \quad (3)$$

We figure out the linear fuzzy SVM (Lin & Wang, 2002) as below:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^N \theta_i \xi_i \\ \text{subject to} \quad & \\ & y^{(i)} (\mathbf{b} + \mathbf{w}^T \mathbf{x}^{(i)}) \geq 1 - \xi_i \quad \forall i = 1, \dots, N \\ & \mathbf{b} \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n, \xi \in \mathbb{R}_+^N. \end{aligned} \quad (\text{FSVM})$$

in which  $C > 0$  means the given parameter; additionally,  $\theta_i \in (0,1]$  means the fuzzy membership connected to the data point  $\mathbf{x}^{(i)} (\forall i = 1, \dots, N)$ . Based upon the  $\theta_i$ , the significance of the data point  $\theta_i$  may be handled since a larger value of  $\theta_i$  generates a comparative important data point  $\mathbf{x}^{(i)}$  for classifying. What's more, we calculate the dual issue of (FSVM) as below:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{x}^{(i)T} \mathbf{x}^{(j)} y^{(i)} y^{(j)} \alpha_i \alpha_j - \sum_{i=1}^N \alpha_i \\ \text{subject to} \quad & \\ \sum_{i=1}^N \alpha_i y^{(i)} = 0 & \\ 0 \leq \alpha_i \leq \theta_i C, i = 1, \dots, N. & \end{aligned} \tag{DFSVM}$$

in which  $C$  means the given parameter. Make sure these FSVM models could be offered with kernels to classify nonlinearly. Varieties of fuzzy-based SVM models were presented in literacy works (Tian et al., 2018; Luo et al., 2017), which are employed with the intention of handling disparate categories of the authentic issues.

### 3. The L1 norm regularized v-fuzzy reduced quadratic surface support vector machine model

The (L1-v-FRQSSVM) model is presented within this section for binary classification to improve the computational effectiveness and generalization capability of present approaches.

$$\begin{aligned} \min \quad & \sum_{i=1}^N \|\mathbf{W}\mathbf{x}^{(i)} + \mathbf{b}\|_2^2 - \nu\rho + \frac{1}{N} \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & \\ y^{(i)} \left( \frac{1}{2} \mathbf{x}^{(i)T} \mathbf{W}\mathbf{x}^{(i)} + \mathbf{x}^{(i)T} \mathbf{b} + c \right) \geq \rho - \xi_i, i = 1, \dots, N, & \\ \mathbf{W} \in \mathbb{S}^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}, \rho \in \mathbb{R}_+, \xi \in \mathbb{R}_+^N. & \end{aligned} \tag{(v-SQSSVM)}$$

in which  $\nu$  means an appointed parameter. Also, (v-SQSSVM) is listed (v-SQSSVM') after using an equal reformulation in the formulation (SQSSVM').

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{z}^T \mathbf{G}\mathbf{z} - \nu\rho + \frac{1}{N} \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & \\ (\mathbf{s}^{(i)T} \mathbf{z} + c) y^{(i)} \geq \rho - \xi_i, i = 1, \dots, N, & \\ \mathbf{z} \in \mathbb{R}^{\frac{n(n+1)}{2} + n}, c \in \mathbb{R}, \rho \in \mathbb{R}_+, \xi \in \mathbb{R}_+^N. & \end{aligned} \tag{(v-SQSSVM')}$$

in which  $\nu$  is the given parameter. Mousavi et al. (2019) demonstrated that the introduction of the L1 norm regularization term can bring sparsity to the model. As the penalty parameter of the L1 norm regularization term increases, the decision variables of the quadratic surface become more sparse. With the large penalty parameter of the regularization term, the quadratic surface will be nearly a linear plane. Therefore, we introduce the L1 norm regularization term based on (v-SQSSVM) and propose the (L1-v-SQSSVM) model below.

$$\begin{aligned} \min \quad & \sum_{i=1}^N \|\mathbf{W}\mathbf{x}^{(i)} + \mathbf{b}\|_2^2 - \nu\rho + \frac{1}{N} \sum_{i=1}^N \xi_i + \lambda \sum_{1 \leq i=j \leq n} |W_{ij}| \\ \text{subject to} \quad & \\ y^{(i)} \left( \frac{1}{2} \mathbf{x}^{(i)T} \mathbf{W}\mathbf{x}^{(i)} + \mathbf{x}^{(i)T} \mathbf{b} + c \right) \geq \rho - \xi_i, i = 1, \dots, N, & \\ \mathbf{W} \in \mathbb{S}^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}, \rho \in \mathbb{R}_+, \xi \in \mathbb{R}_+^N. & \end{aligned} \tag{(L1-v-SQSSVM)}$$

where  $\lambda$  and  $\nu$  refer to the given parameters. Similarly, (L1-v-SQSSVM) is (L1-v-SQSSVM') after the equal reformulation is adopted while (SQSSVM') is formulated:

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{z}^T \mathbf{G}\mathbf{z} - \nu\rho + \frac{1}{N} \sum_{i=1}^N \xi_i + \lambda u^T I_1 \\ \text{subject to} \quad & \\ (\mathbf{s}^{(i)T} \mathbf{z} + c) y^{(i)} \geq \rho - \xi_i, i = 1, \dots, N, & \end{aligned} \tag{(L1-v-SQSSVM')}$$

$$\mathbf{z} \in \mathbb{R}^{\frac{n(n+1)}{2}+n}, c \in \mathbb{R}, \rho \in \mathbb{R}_+, \boldsymbol{\xi} \in \mathbb{R}_+^N.$$

where  $\nu$  and  $\lambda$  are the given parameters. We know that the variable  $\mathbf{z}$  in the convex optimization model (L1-v-SQSSVM') becomes a vectorization of the matrix variable  $\mathbf{W}$  inside the (L1-v-SQSSVM), and the quantity of variables connected to these separation surface coefficients is of order of  $O(n^2)$ . Therefore, when the feature dimension of the dataset increases, the computational volume of the model becomes very large, and the computational efficiency decreases rapidly.

In (Gao et al., 2022), to circumvent the computing complexity resulting from vast quantities of characteristics and benefit from the kernel-free QSSVM model in the meanwhile, a method of approximately simplifying surfaces is proposed. That is, when designing the variable  $\mathbf{z}$ , only the factors of the diagonal of the matrix  $\mathbf{W}$  are recorded and other upper triangular elements other than that are ignored. We know that for the matrix  $\mathbf{W}$ , the elements on its diagonal correspond to these coefficients of the quadratic terms of the separated surface, and the remaining is just these coefficients of the cross terms. To put it differently, our thought is to divide two classes of data through adopting one reduced quadratic surface; additionally, its quadratic coefficient matrix is diagonal. Based on the above ideas, we propose the corresponding simplified (L1-v-SQSSVM) model:

$$\begin{aligned} & \min \sum_{i=1}^N \|\Sigma \mathbf{x}^{(i)} + \mathbf{b}\|_2^2 - \nu \rho + \frac{1}{N} \sum_{i=1}^N \xi_i + \lambda \sum_{1 \leq i=j \leq n} |\Sigma_{ij}| \\ & \text{subject to} \\ & y^{(i)} \left( \frac{1}{2} \mathbf{x}^{(i)T} \Sigma \mathbf{x}^{(i)} + \mathbf{x}^{(i)T} \mathbf{b} + c \right) \geq \rho - \xi_i, \quad i = 1, \dots, N, \\ & \Sigma \in \mathbb{D}^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}, \rho \in \mathbb{R}_+, \boldsymbol{\xi} \in \mathbb{R}_+^N \end{aligned} \quad (\text{L1-v-RSQSSVM})$$

where  $\nu$  and  $\lambda$  are the given parameters. (L1-v-RSQSSVM) could be formulated equally again as the listed model by means of the identical trick in the formulation. Make  $\mathbf{v} \triangleq \text{dvec}(\Sigma)$ ,  $\forall i = 1, \dots, N$ , and  $\mathbf{r}^{(i)} \triangleq \text{qvec}(\mathbf{x}^{(i)})$ . The reduced (L1-v-SQSSVM') is capable of being expressed:

$$\begin{aligned} & \min \frac{1}{2} \mathbf{v}^T \widehat{\mathbf{G}} \mathbf{v} - \nu \rho + \frac{1}{N} \sum_{i=1}^N \xi_i + \lambda u^T I_1 \\ & \text{subject to} \\ & (\mathbf{r}^{(i)T} \mathbf{v} + c) y^{(i)} \geq \rho - \xi_i, \quad i = 1, \dots, N, \\ & c \in \mathbb{R}, \rho \in \mathbb{R}_+, \mathbf{v} \in \mathbb{R}^{2n}, \boldsymbol{\xi} \in \mathbb{R}_+^N. \end{aligned} \quad (\text{L1-v-RSQSSVM}')$$

where  $\nu$  and  $\lambda$  are the given parameters.

$(\Sigma^*, b^*, c^*, \boldsymbol{\xi}^*, \nu^*)$  denote the optimal solution.  $f^*(x) = \frac{1}{2} \mathbf{x}^{(i)T} \Sigma^* \mathbf{x}^{(i)} + \mathbf{x}^{(i)T} b^* + c^*$  denotes the decision function. If  $f^*(x) \geq 0$ , the data point  $\mathbf{x}$  is the positive class; or it is the negative class. Finally, the fuzzy affiliation is assigned to every one of data points  $\mathbf{x}^{(i)}$ , the following (L1-v-FRQSSVM) model is put forward:

$$\begin{aligned} & \min \sum_{i=1}^N \theta_i \|\Sigma \mathbf{x}^{(i)} + \mathbf{b}\|_2^2 - \nu \rho + \frac{1}{N} \sum_{i=1}^N \theta_i \xi_i + \lambda \sum_{1 \leq i=j \leq n} |\Sigma_{ij}| \\ & \text{subject to} \\ & \left( \frac{1}{2} \mathbf{x}^{(i)T} \Sigma \mathbf{x}^{(i)} + \mathbf{x}^{(i)T} \mathbf{b} + c \right) y^{(i)} \geq \rho - \xi_i, \quad i = 1, \dots, N, \\ & \Sigma \in \mathbb{D}^n, \rho \in \mathbb{R}_+, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}_+^N. \end{aligned} \quad (\text{L1-v-FRQSSVM})$$

where  $\Sigma \mathbf{x}^{(i)}$  is the reduced quadratic surface,  $\nu$  and  $\lambda$  are the given parameters,  $\sum_{1 \leq i=j \leq n} |W_{ij}|$  means reducing the  $\mathbf{W}$  matrix to the form with the cross terms removed,  $\theta_i$  is the fuzzy affiliation, and  $\xi_i$  is the loss term. the optimal solution of the (L1-v-FRQSSVM) model is  $(\Sigma^*, b^*, c^*, \boldsymbol{\xi}^*, \nu^*)$ , and the decision is made in the same way as (L1-v-RSQSSVM).

Its equivalent reconstruction is as follows:

$$\begin{aligned} & \min \frac{1}{2} \mathbf{v}^T \widehat{\mathbf{G}} \mathbf{v} - \nu \rho + \frac{1}{N} \sum_{i=1}^N \theta_i \xi_i + \lambda u^T I_1 \\ & \text{subject to} \\ & (\mathbf{r}^{(i)T} \mathbf{v} + c) y^{(i)} \geq \rho - \xi_i, \quad i = 1, \dots, N, \\ & \mathbf{v} \in \mathbb{R}^{2n}, \rho \in \mathbb{R}_+, c \in \mathbb{R}, \boldsymbol{\xi} \in \mathbb{R}_+^N. \end{aligned} \quad (\text{L1-v-FRQSSVM}')$$

**Remark 1.** The slack variable  $\xi^{(i)}$  gauges this mistake of misclassification; additionally, the fuzzy membership  $\theta_i$  means one comparative in-class significance of data  $\mathbf{x}^{(i)}$ . Recall that the fuzzy membership is placed inside two terms of the objective function. Each of these terms,  $\theta_i \xi^{(i)}$  and  $\theta_i \|\sum \mathbf{x}^{(i)} + \mathbf{b}\|_2^2$  could be considered to be a gauge of  $\xi^{(i)}$  and  $\|\sum \mathbf{x}^{(i)} + \mathbf{b}\|_2^2$  with a weight  $\theta_i$ , separately. While  $\mathbf{x}^{(i)}$  trends to be one outlier, it should be less essential for classification. So, its related weight  $\theta_i$  is anticipated to be small for the sake of lessening the impact of  $\xi^{(i)}$  and  $\|\sum \mathbf{x}^{(i)} + \mathbf{b}\|_2^2$  on the separation surface generated by the presented model.

#### 4. Numeric experiment

In this section, we have done some computer experiments with the objective to detecting efficiency and effectiveness of the presented model (L1-v-FRQSSVM). Firstly, the experimental setup is presented. Next, the presented model and several binary classification models are tested on a few public benchmark data sets.

##### 4.1. Test settings

We compared the presented model to several benchmark models, containing a few developed SVM models, like SVMs with quadratic kernel and RBF kernel, the  $\nu$ -SVM model with quadratic kernel and RBF kernel, Fuzzy SVM (Lin and Wang, 2002), etc. Similarly, we implement the two kernel-free SVM models, such as SQSSVM model (Luo et al., 2016),  $\nu$ -SQSSVM and  $\nu$ -FSQSSVM model (Mousavi et al., 2019). In addition, other typical binary classification methods are experimented with for comparison, such as the decision tree model, the logistic regression model, the artificial neural network model, along with the Gaussian naive Bayes model. Table 1 lists the abbreviations of each tested model and the commercial solvers or packages used for catching out relevant models. Make sure the (L1-v-FRQSSVM) is carried out for (L1-v-FRQSSVM) through adopting the Cplex solver.

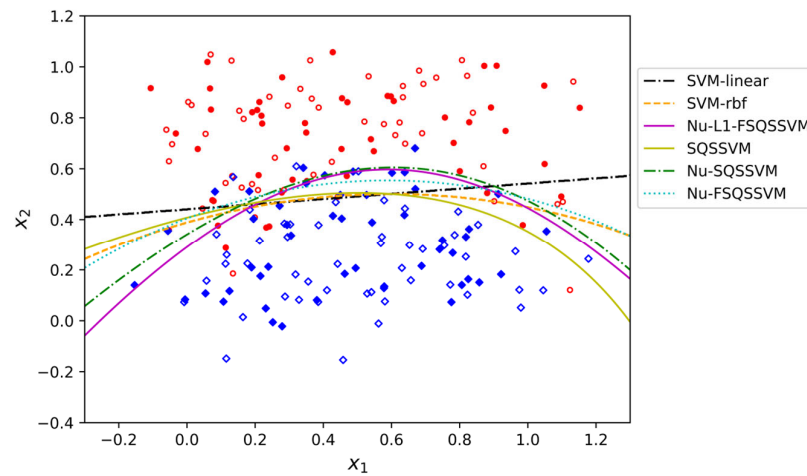
We implement all the computational tests on one computer with two Intel(R) Xeon(R) Gold S117 CPUs @ 2.00 GHz and 64 GB of memory. A procedure for five-fold cross validation adopted for every test. Every one of the tests is duplicated five times with the intention of letting the results to be statistically significant and valid. The standard deviation (SD) and mean of the accuracy scores for each experiment are put down with the aim of verifying every one of the models. What's more, we record the CPU time, which is just the time when conducting a model one time using the fixed parameters.

The parameters are coordinated by a grid method, which is commonly applied to literature (Luo et al., 2016; Gao et al., 2021). Initially, a few tests on artificially designed data sets are implemented. We selected a quadratically separable artificial dataset to pretest the proposed model and compare it with several other selected benchmark SVM models. The classification results are shown in Figure 1. Where the blue and red dots denote two types of data, the hollow dots denote the experiment data, and the solid dots represent the training data. Five-fold cross validation procedure is utilized, where the percentage of the training set is 60%. We present the hyperplane of each model with different lines in the figure and calculate the accuracy of different models.

It is shown in Fig. 1 that the consequences of classification, which is produced by these presented models, are equally great, or even exceeds those of other well-researched models. Indeed, the quadratic separation surfaces are produced by the proposed (L1-v-FRQSSVM) model for binary categorization; additionally, it is anticipated that the quadratically separable data set is classified better. The accuracy of classification is enhanced by the presented model, driving us to implement more computing tests to discern the performance of the presented model in comparison to others.

**Table 1**  
Abbreviations and solvers of experimented models.

Model	Abbreviation	Solver/Package	Parameters
Logistic regression	LR	Scikit-learn	-
Decision tree	DT	Scikit-learn	-
Artificial neural network	ANN	Scikit-learn	-
Gaussian naive Bayes	GNB	Scikit-learn	-
SVM with linear kernel	SVM-linear	Scikit-learn	$(C, r)$
SVM with RBF kernel	SVM-rbf	Scikit-learn	$(C, \gamma)$
$\nu$ -SVM with linear kernel	$\nu$ -SVM-linear	Scikit-learn	$(\nu, r)$
$\nu$ -SVM with RBF kernel	$\nu$ -SVM-rbf	Scikit-learn	$(\nu, \gamma)$
Soft margin quadratic surface SVM	SQSSVM	Cplex	$C$
quadratic surface $\nu$ -SVM	$\nu$ -SQSSVM	Cplex	$C$
Fuzzy SVM	FSVM	Cplex	$C$
Fuzzy quadratic surface $\nu$ -SVM	$\nu$ -FSQSSVM	Cplex	$\nu$
The proposed model	L1-v-FRQSSVM	Cplex	$(\nu, \lambda)$



**Fig. 1.** (L1-v-FRQSSVM) vs. other methods on artificial data sets with outliers.

#### 4.2. Experimented on public benchmark data

(L1-v-FRQSSVM) was experimented on several benchmark datasets, of which the foundational information is as follows inside Table 2.

**Table 2**

Public benchmark data sets

Data Set	n	Sample size (N1 vs, N2)	Data Set	n	Sample size (N1 vs, N2)
Bill	4	610 vs. 762	Pima	8	268 vs. 500
Blood transfusion	4	179 vs. 571	Glass	9	50 vs. 162
Iris	4	50 vs. 50	AusCredit	14	385 vs. 305
Svmguide1	4	2000 vs. 1090	Heart	15	644 vs. 3596
fruit	5	5000 vs. 5000	Seismic-Bumps	15	170 vs. 2414
Careval	6	384 vs. 1210	Brain tumor	17	1178 vs. 97
Ecoli	7	145 vs. 195	CTG	22	1655 vs. 471
Seed	7	50 vs. 50	Parkinson	22	147 vs. 48

As to every one of the data sets, fivefold cross-validation is employed during training for the adjusted parameters. Those parameters are adjusted by means of the grid-search approach. The SD and mean of accuracy scores of each experimented model are put down. The training CPU time of each tested model is also put down for the purpose of comparing the computational effectiveness.

We have these listed observations from the consequences within Tables 3 and 4:

- (L1-v-FRQSSVM) generates the highest or second highest mean accuracy scores on experimented data sets. Make sure there exist large- and small-scale data sets inside Table 2, supporting the efficiency of the presented models on classified data sets in disparate scales.
- The CPU time of the presented (L1-v-FRQSSVM) model is short on masses of the data sets. For the same solver Cplex, the CPU time depleted by (L1-v-FRQSSVM) is less than those by SQSSVM, v-SQSSVM, along with v-FSQSSVM models. Indeed, the model (L1-v-FRQSSVM) has fewer variables than the SQSSVM, the v-SQSSVM, and the v-FSQSSVM models, which leads to a lower computational complexity of the proposed model.
- Make sure that, aside from those kernel-free QSSVM models, all the other experimented models are implemented through adopting Scikit-learn packages supported professionally. Even if the presented model depletes a lengthier CPU time, it's acceptable as before.



**Table 3**  
The consequences of the experimented models on Benchmark data

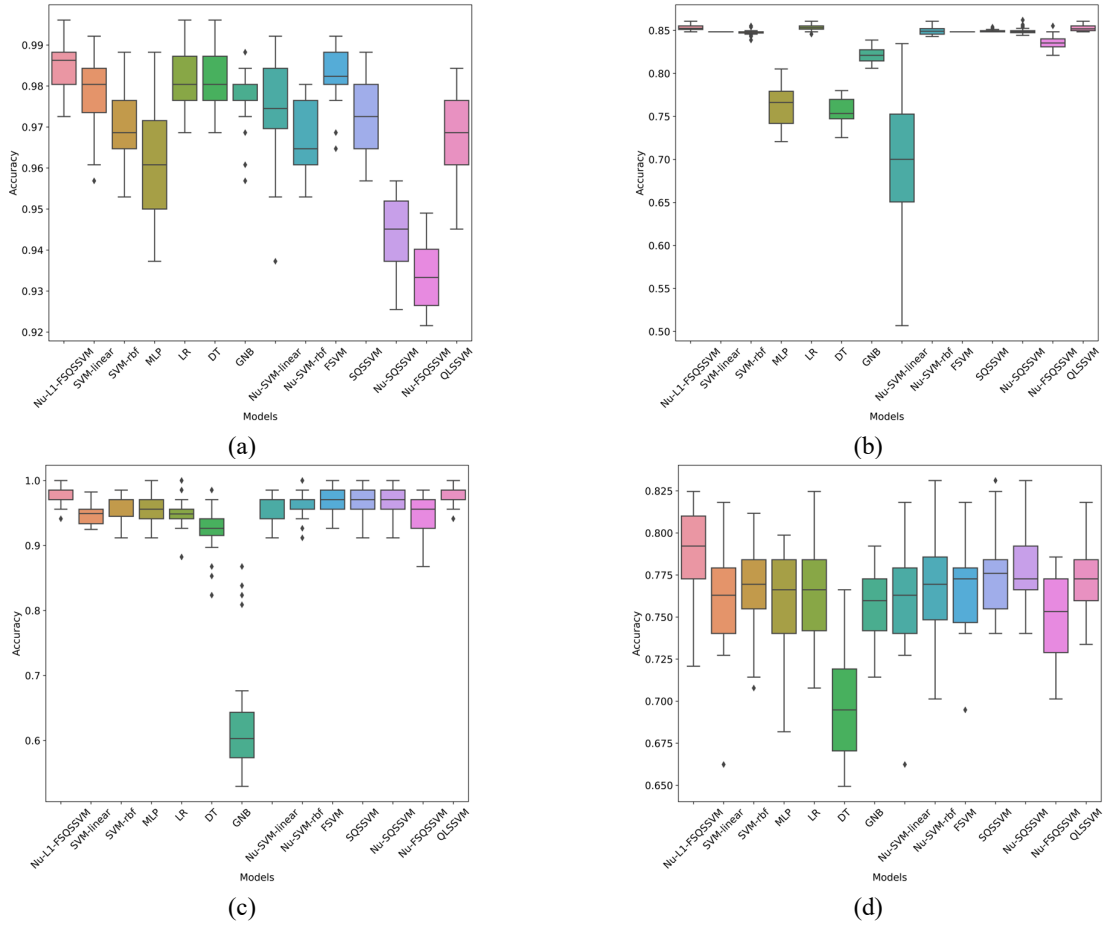
Data set	Model																									
	Accuracy score %																									
	LR		DT		MLP		GNB		SVM-linear		SVM-rbf		v-SVM-linear		v-SVM-rbf		SQSSVM		v-SQSSVM		FSVM		v-FSQSSVM		LI-v-FRQSSVM	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
Bill	95.41	0.78	98.21	0.39	98.4	0.44	83.35	0.79	98.55	0.4	98.41	0.72	98.69	0.44	98.7	0.45	99.85	0.18	99.02	0.27	99.2	0.42	96.73	0.46	100	0
Blood transfusion	77.2	0.5	71.8	2.14	78.2	1	75.27	2.29	76	0	78.53	1.29	59.2	7.56	77.33	3.65	78.8	1.42	80	1.12	76	0	79.6	1.24	79.6	1.24
Iris	96	2	94	2.55	95.5	2.92	95	3.16	96	2	95.5	2.92	96	4.9	97	4	98	4	99	2	99	2	99	2	99	2
Svmguide1	94.56	0.52	95.66	0.6	96.25	0.37	92.15	0.48	95.18	0.48	96.44	0.46	95.76	0.19	96.99	0.3	64.72	0	96.73	0.24	95.5	0.36	95.99	0.37	97.85	0.85
Fruit	92.52	0.26	93.12	0.58	92.58	0.29	92.08	0.24	96.73	0.3	92.79	0.3	96.67	0.25	97.33	0.27	96.17	0.16	95.14	0.26	96.72	0.3	93.18	0.43	97.63	0.66
Careval	85.67	1.05	98.75	0.5	94.29	0.88	85.05	1.2	85.71	1.02	94.29	0.37	84.33	2.9	98.93	0.73	97.24	1.18	96.93	0.83	86.52	1.24	93.98	1.2	99.31	1.28
Ecoli	95.56	0.94	92	2.06	84.89	11.01	66.52	10.02	95.85	0.76	96.47	0.72	96.18	1.18	96.47	0.72	96.18	1.5	96.18	1.5	96.42	2.24	93.24	3.17	96.76	0
Seed	93.93	2.42	95.71	2.42	93.93	3.31	92.5	3.07	97.14	2.67	96.43	3.91	96.43	2.26	97.14	2.67	97.86	1.75	96.43	2.26	100	0	95.71	1.43	100	0
Pima	77.92	1.63	66.88	1.09	78.31	2.34	75.84	1.5	79.09	2.71	78.18	1.86	78.44	2.77	79.22	2.72	80.13	2.92	80.39	2.58	79.09	1.9	75.97	1.97	80.65	1.81
Glass	91.86	1.04	93.72	0.93	93.72	0.93	91.4	1.89	93.02	1.27	94.42	0.87	93.02	4.41	94.42	2.37	95.81	2.71	95.81	2.71	93.49	3.72	93.95	3.15	96.28	2.37
AUScredit	85.8	1.49	80.72	1.11	86.38	2.8	81.96	1.71	85.65	1.25	84.2	1.91	86.09	1.68	85.22	2.18	86.81	2.02	87.54	2.31	85.65	1.25	86.52	2.03	88.41	2
Heart Disease	85.29	0.13	75.87	0.91	84.74	0.47	82.3	0.45	84.77	0	84.74	0.21	71.04	3.81	84.95	0.29	85.27	0.26	84.95	0.29	84.84	0	84.32	0.67	85.38	0.38
Seismic-Bumps	93.31	0.09	88.43	1.01	93.27	0.22	76.85	18.63	93.42	0	93.42	0	66.96	15.11	92.92	0.26	93.42	0	93.38	0.19	93.42	0	93.42	0	93.81	0.2
Brain tumor	93.22	0.32	98.31	0.34	94.51	0.59	98.2	0.34	98.04	0.7	97.1	0.74	98.27	0.73	96.78	1.09	97.8	0.4	94.51	0.78	98.43	0.35	93.8	0.84	98.98	0.43
CTG	95.96	1.39	98	0.36	97.61	1	89.49	3.22	97.14	1.22	98.54	0.27	96.62	1.26	98.59	0.39	92.63	0.78	98.59	0.45	96.24	1.1	89.81	1.61	99.04	0.54
Parkinson	84.1	2.51	87.18	5.85	84.1	2.99	72.82	8.21	86.15	3.48	88.72	4.17	86.15	2.61	91.28	4.76	90.26	2.99	88.72	5.28	89.74	4.29	85.64	3.08	91.28	3.84

**Table 4**

The results of tested models on Benchmark data

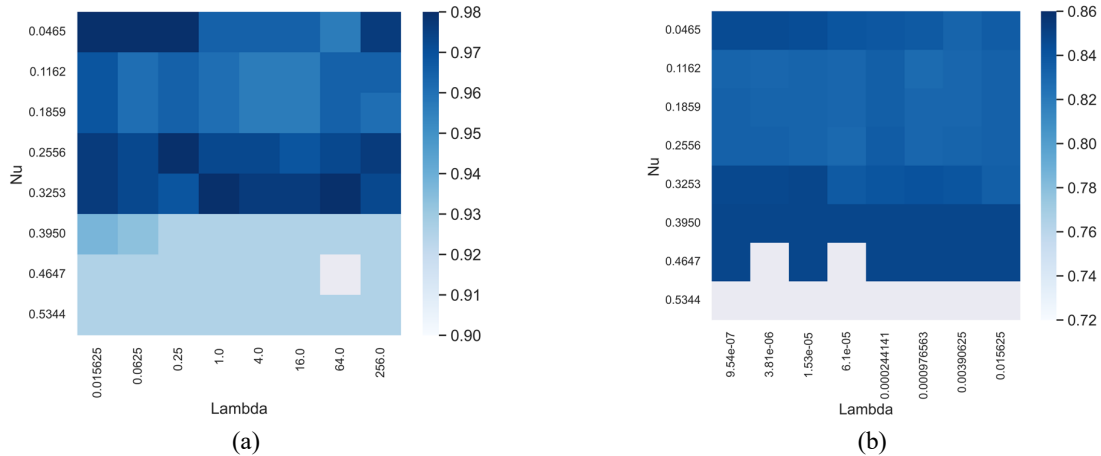
Data set	Model												
	CPU time(s)												
	LR	DT	MLP	GNB	SVM-linear	SVM-rbf	v-SVM-linear	v-SVM-rbf	SQSSVM	v-SQSSVM	FSVM	v-FSQSSVM	L1-v-FRQSSVM
Bill	0.026	0.01	3.158	0.006	0.015	0.05	0.02	0.035	1.13	1.145	0.475	1.45	0.325
Blood transfusion	0.012	0.006	1.894	0.006	0.08	0.04	0.005	0.015	1.485	0.57	0.27	0.82	0.555
Iris	0.008	0.006	0.182	0.006	0.006	0.006	0.005	0.005	0.13	0.135	0.095	0.175	0.125
Svmguide1	0.06	0.01	9.165	0.01	1.26	2.25	0.07	0.145	2.335	1.84	1.235	1.865	0.545
Fruit	0.15	0.024	8.458	0.01	0.645	1.91	0.74	2.46	5.81	5.63	2.77	5.675	1.5
Careval	0.012	0.008	4.196	0.006	0.12	0.03	0.015	0.03	2.97	1.46	0.515	1.755	1.315
Ecoil	0.014	0.006	0.992	0.006	0.006	0.006	0.005	0.005	0.39	0.42	0.185	0.595	0.505
Seed	0.008	0.006	0.194	0.006	0.005	0.005	0.005	0.005	0.35	0.305	0.115	0.36	0.14
Pima	0.014	0.008	1.954	0.004	0.025	0.025	0.01	0.02	0.9	0.9	0.305	1.175	0.68
Glass	0.014	0.006	0.94	0.006	0.006	0.006	0.005	0.005	0.39	0.425	0.14	0.565	0.185
AUScredit	0.035	0.01	3.465	0.01	0.38	0.02	0.01	0.02	2.215	2.12	0.505	5.11	1.443
Heart	0.178	0.024	15.828	0.006	20.338	0.36	0.045	0.315	18.61	10.2	1.46	12.005	0.93
Seismic-Bumps	0.116	0.012	11.426	0.006	6.212	0.08	0.02	0.07	7.27	6.64	0.83	7.46	2.61
Brain tumor	0.168	0.008	5.64	0.006	0.015	0.015	0.01	0.015	5.7	8.07	0.76	5.305	1.265
CTG	0.19	0.015	11.185	0.01	0.08	0.06	0.04	0.07	14.915	14.005	1.51	33.635	1.48
Parkinson	0.012	0.006	1.582	0.004	0.004	0.002	0.005	0.005	2.42	2.995	0.18	2.34	0.445

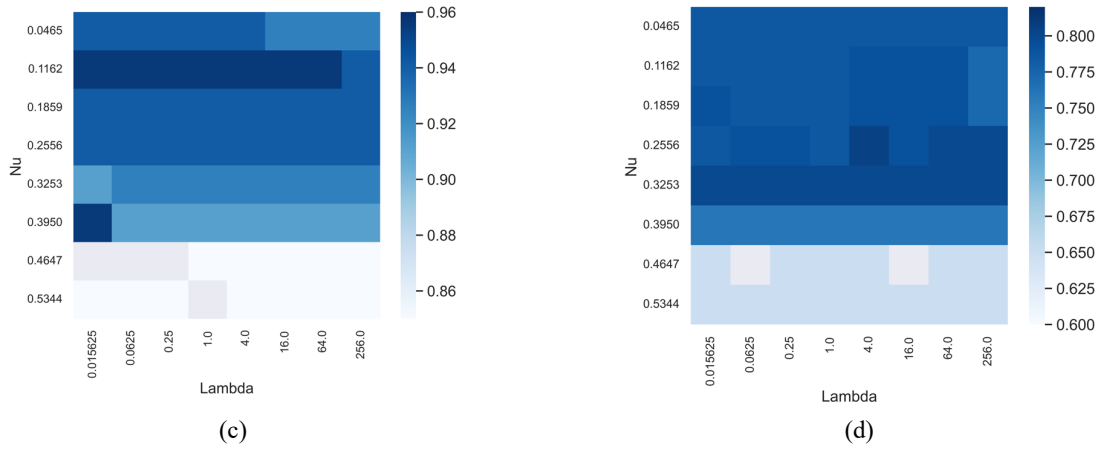
To visualize the difference in performance metrics between our model and other models on different datasets, we selected the Brain tumor, Heart, Ecoli and Pima datasets from the benchmark and made the corresponding box plots Fig. 2 shown. The performance of our model is the best of all models, and the standard deviation of the accuracy is small and fluctuates less, which indicates the good generalization ability of our model.



**Fig. 2.** Boxplots of the consequences of tested models. (a) Results on the Brain tumor; (b) Results on the Heart; (c) Results on the Ecoli; (d) Results on the Pima.

Subsequently, the following figures are plotted with the aim of showing the sensibility of parameters  $\lambda$  and  $\nu$  on the Brain tumor, Heart, Ecoil and Pima datasets from benchmark. It can be noticed that the parameter  $\nu$ , as a parameter with a fixed range, can make the actual parameter tuning process easier and save much effort for training.





**Fig. 3.** The sensibility of parameters  $\nu$  and  $\lambda$ . (a) Results on the Brain tumor; (b) Results on the Heart; (c) Results on the Ecoli; (d) Results on the Pima.

A few statistical tests are also applied with the aim of showing that the presented model outperforms all other experimented models. We firstly figure out the mean rank of accuracy scores from all experimented models on the public benchmark dataset, as depicted inside Table 5.

**Table 5**

Mean rank of the total experimented models on public benchmark data sets.

Model	Mean rank	Model	Mean rank
LR	9.250	$\nu$ -SVM-rbf	5.750
DT	9.000	SQSSVM	5.375
MLP	8.500	$\nu$ -SQSSVM	4.500
GNB	11.188	FSVM	6.000
SVM-linear	7.125	$\nu$ -FSQSSVM	7.750
SVM-rbf	6.750	L1- $\nu$ -FRQSSVM	1.062
$\nu$ -SVM-linear	8.250		

Notice that the proposed L1- $\nu$ -FRQSSVM ranks best in all experimented models. Then, the Friedman tests (Zhou, 2021) on all experimented models are implemented. Make  $k$  be the quantity of models,  $N$  be the quantity of data sets; in addition,  $q_j$  represents the mean rank of the  $j$ th model ( $j = 1, \dots, k$ ). We figure out the statistics  $\lambda_{\chi^2}$  and  $\lambda_F$  as below:

$$\lambda_{\chi^2} = \frac{12N}{k(k+1)} \sum_{j=1}^k \left( q_j - \frac{k+1}{2} \right)^2, \quad \lambda_F = \frac{(N-1)\lambda_{\chi^2}}{(k-1)N - \lambda_{\chi^2}} \quad (4)$$

in which a  $\chi^2$  distribution with  $k-1$  degrees of freedom is followed by  $\lambda_{\chi^2}$ . Moreover, an  $F$ -distribution with degrees of freedom  $k-1$  and  $(k-1)(N-1)$ . is followed by  $\lambda_F$ . With the results in Table 8, we have  $\lambda_{\chi^2} = 82.0409$  and  $\lambda_F = 11.1916$  for  $k = 13$  and  $N = 16$ . The  $F$ -distribution table at the 95% level of significance reaches 2.4589, less than  $\lambda_F = 11.1916$ . As a result, it is confident to refuse the null assumption which the performances of all the experimented models are alike. In another word, there exists an obvious difference in all the experimented models.

With the aim of inspecting the differences between two models, the Nemenyi post hoc experiment is implemented (Gao et al., 2022; Zhou, 2021). The critical difference  $C$  is figured out as below:

$$C = p.05(k(k+1)/6N)^{1/2} \quad (5)$$

in which  $p.05$  means the critical value of the Tukey distribution (Gao et al., 2022; Zhou, 2021). Two models become disparate on condition that the difference of their mean ranks surpasses  $CD$ . Inside our case,  $p.05 = 3.354$  and  $C = 4.5933$ . Accordingly, the Nemenyi experiment investigates the obvious distinctions between the presented (L1- $\nu$ -FRQSSVM) model and the other

experimented models except SQSSVM and v-SQSSVM models. However, it's not difficult to observe that the mean rank of the presented model exceeds that of SQSSVM and v-SQSSVM models.

**Remark 2.** *With the consequences inside Tables 3 and 4, alongside consequences from the statistical experiments, it is surely inferred that the presented (L1-v-FRQSSVM) model becomes powerful competitive in all tested models. It's significant to utilize the presented model in authentic problems.*

#### 4.3. Application to diagnose the lung cancer subtypes

The presented (L1-v-FRQSSVM) model is employed in the diagnosis of lung carcinoma subtypes (LUAD/LUSC) in this part. The background and dataset of AD are introduced briefly.

##### 4.3.1. Background and data

Lung carcinoma is one category of malignant tumor that usually originates in cells of the lung tissue and can invade surrounding tissues and spread to other sites. Lung carcinoma becomes one of the most common carcinomas throughout the world; additionally, it is also one of the dominating reasons for deaths related to cancers. The survival rate for patients with lung carcinoma is merely 17.8%, however, is more elevated significantly on condition that the disease is investigated while it is localized in lungs as before. Nevertheless, merely 15% of cases are investigated during an early phase (Cabrera et al., 2015). It means that early diagnosis is vital to the therapy of lung carcinoma. There are several types of lung cancer, two common types of which are lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC), which occupy major categories of lung cancer, respectively (Yu et al., 2020). LUAD and LUSC may differ in terms of clinical presentation, diagnosis, treatment, and prognosis. Clinicians will develop individualized treatment plans based on the patient's condition, pathological diagnosis, and stage to maximize patient outcomes and survival (Hsu and Si, 2018). Therefore, it is very important to make accurate pathological diagnosis and classification for subtypes of lung cancer patients.

In fact, machine learning methods are now frequently used in the medical community to subtype cancer based on data such as gene expression RNAseq (Yu et al., 2020; Hsu and Si, 2018). In this subsection, the present (L1-v-FRQSSVM) model is promoted and employed for lung cancer subtype prediction in LUAD and LUSC, as well as all other tested models. The gene expression RNAseq dataset for LUAD/LUSC used in this study was obtained from the TCGA database. The final acquired LUAD/LUSC gene expression RNAseq dataset contains a total of 1129 data points, including 576 data points in the LUAD class and 553 data points in the LUSC class. The data points have 20531 features showing the transcriptional estimates for each gene level. In order to avoid larger values for some input features and smaller values for preferences, we also performed maximum-minimum normalization for all data points after transforming the raw gene expression values using the  $\log_2^{(2x+1)}$  method.

##### 4.3.2. Numerical tests on lung cancer subtype data sets

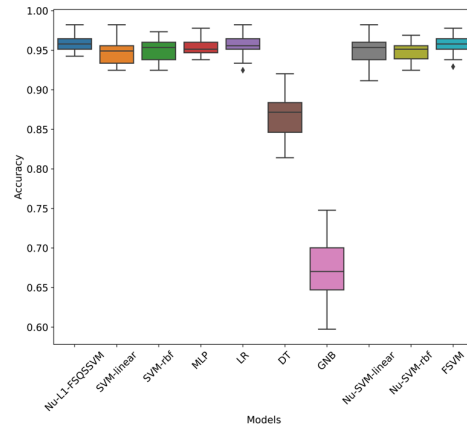
Indeed, the number of features in the lung cancer subtype (LUAD/LUSC) data set is  $n=20531$ . Due to the enormous dimensionality, we apply Principal Component Analysis (PCA) to this data set to reduce the feature dimensionality. We finally choose to retain 90% of the information and obtain a reduced dimensionality data set with feature number  $n=497$ . For the SQSSVM models, even if the feature number  $n=497$ , it still has  $n(n+1)/2=123753$  variables, which exceeds the memory and arithmetic power of the computer, so we do not test this dataset using the SQSSVM models. Instead, the proposed (L1-v-FRQSSVM) model encountered no computational problems. All the results are listed in Table 6. These boxplots of the consequences are depicted as follows (in Fig. 3). The listed observations originate from the consequences:

- In accordance with these results in Table 6 and the box plots, it can be seen that the presented (L1-v-FRQSSVM) model is implemented better, compared with the other experimented models on the part of classification accuracy. Additionally, SD of accuracy scores from the presented model is the smallest in those of the experimented models. To sum up, the (L1-v-FRQSSVM) model offers the steadiest and most accurate classifier in the total experimented models to diagnose the lung cancer subtypes.
- Make sure the representative kernel-free QSSVM models, including SQSSVM, v-SQSSVM or v-FSQSSVM, aren't capable of dealing with the LUAD/LUSC data set in the computing circumstance in our article, on account of the vast quantity of characteristics. Nevertheless, the presented model is capable of coping with this issue; additionally, it is implemented better, compared to the other experimented models on the part of accuracy. The time for CPU, which is depleted by the presented model, is receivable, despite that it's longer compared with several of the experimented models that perform by means of commonly used solvers. Capitalizing on the reduced quadratic surface for separating, the presented (L1-v-FRQSSVM) might cope with authentic issues more effectively compared with other QSSVM models with no kernel.

**Table 6**

The accuracy score consequences of the experimented models on lung cancer subtype figures.

Model	Accuracy score %	
	Mean	Std
LR	95.18	0.72
DT	84.87	1.97
MLP	95.09	0.77
GNB	65.97	5.30
SVM-linear	94.82	0.80
SVM-rbf	93.85	0.88
v-SVM-linear	95.49	0.90
v-SVM-rbf	94.78	0.95
FSVM	95.04	0.86
L1-v-FRQSSVM	98.84	0.72

**Fig. 4.** Boxplots of the consequences of the experimental models on lung cancer subtype figures

## 5. Conclusions

To sum up, one advanced nonlinear binary classification kernel-free (L1-v-FRQSSVM) model has been presented by us. The (L1-v-FRQSSVM) model has been derived and investigated rigorously. The classification efficiency and effectiveness of the presented model have already been verified using numerical tests on a few public benchmark data sets. What's more, the presented model has been employed for the sake of diagnosing lung carcinoma subtypes within the LUAD/LUSC dataset based on gene expression RNAseq in the database of TCGA. The main discoveries of this research are summed up as follows:

- The presented (L1-v-FRQSSVM) is implemented more greatly compared with all the other experimented well-known models on the part of the accuracy of classification. differing from the nonlinear SVM models with the kernel, the presented (L1-v-FRQSSVM) model demands no kernels and tunes no kernel parameters, saving multitudes of efforts in reality.
- Through introducing the L1 norm regularization term, the sparsity of the model is promoted, so that the computational complexity and overfitting can be avoided. At the same time, it enables the model to infinitely approximate the linear plane when facing the linearly divisible case of the split interface. The proposed model combines the v-SVM idea. As a bounded parameter, the parameter  $v$  is selected in a fixed range, which makes the model more convenient for parameter tuning.
- Numerical tests have already been implemented in an attempt to illustrate the impact of the above-described parameters on classification accuracy. Furthermore, the hopeful numerical consequences on several artificial and several public benchmark data sets show the efficiency of the presented models in dealing with the authentic binary classification issues. In the end, the elevated efficiency of the presented algorithm for the (L1-v-FRQSSVM) model has been validated by means of the short training CPU time.
- Especially, the presented model has shown its anticipated effectiveness and acceptable efficiency when adopted in the lung cancer dataset. The gratifying performance of the presented model for lung cancer subtype diagnosis has offered another dependable machine tool in the domain of lung cancer subtype diagnosis research.

## Author Contributions

Junyuan Xiao: Conceptualization, Methodology, Software, Writing – original draft. Guoyi Liu: Validation, Software, Writing – review & editing. Min Huang: Resources, Writing – review & editing. Zheming Gao: Conceptualization, Writing – review & editing. Zhihua Yin: Data curation. The first two authors contributed equally.

## Funding

This work has been supported by the National Natural Science Foundation of China under Grant 72201052 and the Fundamental Research Funds for the Central Universities under Grant N2204017.

**Data Availability Statement:** Data will be made available on request.

## Acknowledgments

Data used in the preparation of this article were obtained from the TCGA database (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>). As such, the investigators within the TCGA contributed to the design and implementation of LUAD/LUSC data and/or provided data but did not participate in analysis or writing of this report.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

- Bai, Y., Han, X., Chen, T., & Yu, H. (2015). Quadratic kernel-free least squares support vector machine for target diseases classification. *Journal of Combinatorial Optimization*, 30, 850-870.
- Cabrera, J., Dionisio, A., & Solano, G. (2015, July). Lung cancer classification tool using microarray data and support vector machines. In 2015 6th International Conference on Information, Intelligence, *Systems and Applications (IISA)* (pp. 1-6). IEEE.
- Cholette, M. E., Borghesani, P., Di Galleonardo, E., & Braghin, F. (2017). Using support vector machines for the computationally efficient identification of acceptable design parameters in computer-aided engineering applications. *Expert Systems with Applications*, 81, 39-52.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
- Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. *Cambridge university press*.
- Dagher, I. (2008). Quadratic kernel-free non-linear support vector machine. *Journal of Global Optimization*, 41(1), 15-30.
- Danenas, P., & Garsva, G. (2015). Selection of support vector machines based classifiers for credit risk domain. *Expert systems with applications*, 42(6), 3194-3204.
- Di, M., & Joo, E. M. (2007, December). A survey of machine learning in wireless sensor networks from networking and application perspectives. In 2007 6th international conference on information, communications & signal processing (pp. 1-5). IEEE.
- Gao, Q. Q., Bai, Y. Q., & Zhan, Y. R. (2019). Quadratic kernel-free least square twin support vector machine for binary classification problems. *Journal of the Operations Research Society of China*, 7, 539-559.
- Gao, Z., Fang, S. C., Luo, J., & Medhin, N. (2021). A kernel-free double well potential support vector machine with applications. *European Journal of Operational Research*, 290(1), 248-262.
- Gao, Z., Wang, Y., Huang, M., Luo, J., & Tang, S. (2022). A kernel-free fuzzy reduced quadratic surface v-support vector machine with applications. *Applied Soft Computing*, 127, 109390.
- Hsu, Y. H., & Si, D. (2018, July). Cancer type prediction and classification based on rna-sequencing data. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 5374-5377). IEEE.
- Lin, C. F., & Wang, S. D. (2002). Fuzzy support vector machines. *IEEE transactions on neural networks*, 13(2), 464-471.
- Luo, J., Fang, S. C., Bai, Y., & Deng, Z. (2016). Fuzzy quadratic surface support vector machine based on fisher discriminant analysis. *Journal of Industrial and Management Optimization*, 12(1), 357-373.
- Luo, J., Fang, S. C., Deng, Z., & Guo, X. (2016). Soft quadratic surface support vector machine for binary classification. *Asia-Pacific Journal of Operational Research*, 33(06), 1650046.
- Luo, J., Tian, Y., & Yan, X. (2017). Clustering via fuzzy one-class quadratic surface support vector machine. *Soft Computing*, 21, 5859-5865.
- Mainali, S., Darsie, M. E., & Smetana, K. S. (2021). Machine learning in action: stroke diagnosis and outcome prediction. *Frontiers in Neurology*, 12, 734345.
- Malik, A. K., Ganaie, M. A., Tanveer, M., Suganthan, P. N., & Alzheimer's Disease Neuroimaging Initiative Initiative. (2022). Alzheimer's disease diagnosis via intuitionistic fuzzy random vector functional link network. *IEEE Transactions on Computational Social Systems*.
- Mousavi, A., Gao, Z., Han, L., & Lim, A. (2019). Quadratic surface support vector machine with L1 norm regularization. *arXiv preprint arXiv:1908.08616*.
- Mousavi, A., Rezaee, M., & Ayanzadeh, R. (2019). A survey on compressive sensing: Classical results and recent advancements. *arXiv preprint arXiv:1908.01014*.

- Mousavi, S., & Shen, J. (2019). Solution uniqueness of convex piecewise affine functions based optimization with applications to constrained  $\ell_1$  minimization. *ESAIM: Control, Optimisation and Calculus of Variations*, 25, 56.
- Schölkopf, B., & Smola, A. J. (2002). Learning with kernels: support vector machines, regularization, optimization, and beyond. *MIT press*.
- Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural computation*, 12(5), 1207-1245.
- Shen, J., & Mousavi, S. (2018). Least Sparsity of p-Norm Based Optimization Problems with  $p > 1$ . *SIAM Journal on Optimization*, 28(3), 2721-2751.
- Tang, S., Cao, P., Huang, M., Liu, X., & Zaiane, O. (2022). Dual feature correlation guided multi-task learning for Alzheimer's disease prediction. *Computers in Biology and Medicine*, 140, 105090.
- Tian, Y., Yong, Z., & Luo, J. (2018). A new approach for reject inference in credit scoring using kernel-free fuzzy quadratic surface support vector machines. *Applied Soft Computing*, 73, 96-105.
- Xie, Y., Meng, W. Y., Li, R. Z., Wang, Y. W., Qian, X., Chan, C., ... & Leung, E. L. H. (2021). Early lung cancer diagnostic biomarker discovery by machine learning methods. *Translational oncology*, 14(1), 100907.
- Yan, H. S., & Xu, D. (2007). An Approach to Estimating Product Design Time Based on Fuzzy  $\nu$ -Support Vector Machine. *IEEE transactions on Neural Networks*, 18(3), 721-731.
- Yan, X., Bai, Y., Fang, S. C., & Luo, J. (2018). A proximal quadratic surface support vector machine for semi-supervised binary classification. *Soft Computing*, 22, 6905-6919.
- Yu, Z., Wang, Z., Yu, X., & Zhang, Z. (2020). RNA-seq-based breast cancer subtypes classification using machine learning approaches. *Computational intelligence*
- Zhou, Z. H. (2021). Machine learning. *Springer Nature*.



© 2023 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).