METHODOLOGIES AND APPLICATION

Weighted support vector machine using fuzzy rough set theory

Somaye Moslemnejad[1] · Javad Hamidzadeh[2]

Abstract

The existence of both uncertainty and imprecision has detrimental impact on efficiency of decision-making applications and some machine learning methods, in particular support vector machine in which noisy samples diminish the performance of SVM training. Therefore, it is important to introduce a special method in order to improve this problem. Fuzzy aspects can handle mentioned problem which has been considered in some classification methods. This paper presents a novel weighted support vector machine to improve the noisy sensitivity problem of standard support vector machine for multiclass data classification. The basic idea is considered to add a weighted coefficient to the penalty term Lagrangian formula for optimization problem, which is called entropy degree, using lower and upper approximation for membership function in fuzzy rough set theory. As a result, noisy samples have low degree and important samples have high degree. To evaluate the power of the proposed method WSVM-FRS (Weighted SVM-Fuzzy Rough Set), several experiments have been conducted based on tenfold cross-validation over real-world data sets from UCI repository and MNIST data set. Experimental results show that the proposed method is superior than the other state-of-the-art competing methods regarding accuracy, precision and recall metrics.

Keywords Support vector machine · SVM classifier · Fuzzy rough set theory · Noisy samples · Entropy degree

1 Introduction

Support vector machine (SVM) is supervised machine learning model which has been presented for classification and regression analysis (Vapnik 1995; Yang and Xu 2017; Mao et al. 2014; Santhanama et al. 2016). SVM has been converted into a convex quadratic optimization problem and then solved by a quadratic programming (QP) technique (Vanir 1999). It can also classify nonlinear samples using the kernel function (Cortes and Vapnik 1995; Xue et al. 2018; Moghaddam and Hamidzadeh 2016).

✉ Javad Hamidzadeh
    j_hamidzadeh@sadjad.ac.ir

    Somaye Moslemnejad
    moslemnejad@salman.ac.ir

[1] Department of Computer Engineering, Salman Institute of Higher Education, Mashhad, Iran

[2] Faculty of Computer Engineering and Information Technology, Sadjad University of Technology, Mashhad, Iran

In the real-world applications, there is some complex information under uncertainty which reduces efficiency and accuracy of decision-making systems. Machine learning algorithms would be able to handle decision-making applications via classification algorithms, but it requires a specific algorithm which has high tolerance in dealing with uncertainty aspects, namely noise. The classification hyperplane, which is obtained by SVM, is determined by the support vectors. The presence of noise increases inefficiency of the standard SVM training which makes the decision boundary from the optimal hyperplane. In machine learning, there are many techniques to improve classification (Alcantud et al. 2019; Hamidzadeh and Moradi 2018, 2020; Hamidzadeh et al. 2014, 2017; Hamidzadeh and Namaei 2018; Hamidzadeh and Ghadamyari 2019; Javid and Hamidzadeh 2019). Therefore, there are many methods which have been proposed to improve the SVM classification using identifying uncertainty samples such as noisy and outlier ones in order to discard or delete them (Han et al. 2016; Nguyen et al. 2018; Xu et al. 2016). On the other hand, some solutions have been presented to deal with noisy and outlier samples in weighted support vector machine classification methods in

order to reduce the effect of unimportant samples (Karal 2017; Sheng et al. 2015; Zhou et al. 2016; Yang et al. 2005). The previous methods, in which weighted SVM is based on probabilities, cannot reduce the effect of noisy samples in SVM training accurately.

Fuzzy aspects in the real-world applications play a crucial role due to existence of sophisticated information in real-world applications such as medical diagnosis, pattern recognition and wherever exist big data. With respect to the importance of this aspect, the use of probabilistic and fuzzy methods is effective (Singh et al. 2020; Sivasankar et al. 2020). Also regarding importance of the efficiency and the speed of classification-based model, it is important to introduce a specific method that classifies precisely. In order to realize this aim, it is better to introduce a more precise and quick method. Therefore, the fuzzy rough set theory is redesigned. This paper presents a method to lessen the effect of noise in SVM training with soft margin using fuzzy rough set theory (Dubois and Prade 1990). In the proposed method, a weight coefficient is added to the penalty term Lagrangian formula for optimization problem. This weighted coefficient is called entropy degree, which uses lower and upper approximation for membership function in fuzzy rough set theory. As a result, in the proposed method—WSVM-FRS (Weighted SVM-Fuzzy Rough Set)—noisy samples have low degree, and important samples have high degree. The results have obtained good classification accuracy, precision and recall for SVM training.

The rest of this paper is organized as follows: In Sect. 2, a survey of weighted support vector machine algorithms is presented. In Sect. 3, primary concepts of support vector machine and fuzzy rough set theory are presented. In Sect. 4, the proposed method is introduced. The experimental results are shown in Sect. 5. Finally, Sect. 6 contains conclusions and future works.

2 Related works

Several methods have been proposed for weighted support vector machine. This section surveys some important weighted support vector machine methods to deal with noisy and outlier samples.

Weighted support vector machine (WSVM) (Yang et al. 2005) has been presented to improve the outlier sensitivity problem of SVM. The basic idea is possibility c-means (PCM), which is extended into kernel space to generate different weight values for main training data points and outliers according to their relative importance in the training set. In Du et al. (2017), a fuzzy compensation multiclass SVM method has been introduced to improve the outlier and noise sensitivity problem, which gives the

dual effects to penalty term through treating every data point as both positive and negative classes, but with different memberships. In Sheng et al. (2015), a method has been presented to reduce the noise sensitive issue based on fuzzy least square support vector machine. By applying fuzzy inference and nonlinear correlation measurement, the effects of the samples with low confidence can be reduced. WDRSVM (Li et al. 2016) has been presented a weighted doubly regularized support vector machine to deal with noise by using both the distance information between classes and within each class. Incosh loss (Karal 2017) has been introduced to obtain support vector regression (SVR) models for coping with different noise distributions, which is optimal in the maximum likelihood sense for the hypersecant error distributions. In Ding et al. (2017), the presented method (WLMSVM) is a new classifier for multiclass classification, called weighted linear loss multiple birth support vector machine based on information granulation to enhance the performance of multiple WLTSVM.

BWSVM (Sun et al. 2017) has presented a band-weighted support vector machine, which is to quantify the divergent contributions of different bands when implementing SVM; the BWSVM adopts the L1 norm penalty term of band weights on the original SVM. In Li et al. (2017), a method has been introduced to construct the new weighted mechanisms for both loss and penalty, which has developed the weight partly adaptive elastic net for dealing with the binary classification problem of microarray with noise by using the distances from the sample points to both class centers. In Lu et al. (2017) a probabilistic weighted least squares SVM method has been presented to model these kinds of processes under noise; this method can increase robustness and accuracy even with outliers or non-Gaussian noise. RLS-SVM (Yang et al. 2014) has presented a method based on the truncated least squares loss function for regression and classification with noise. DS-RLSSVM (Zhou et al. 2016) has been developed to model complex systems in the presence of various types of random noise. The integration of the distributed LS-SVM and fuzzy clustering is used to construct the evidence for the LS-SVM parameters. In Zhang et al. (2018), a method has been presented to incorporate prior knowledge into SVM using sample confidence, which is called feature weighted confidence with SVM. This method computes the sample confidence directly from the weights of prior features provided by SVM. In Xu et al. (2015), a new support vector weighted quantile regression approach has been introduced that is closely built upon the idea of support vector machine. It can be estimated by solving a Lagrangian dual problem of quadratic programming and is able to implement the nonlinear quantile regression by introducing a kernel function. In Tang et al. (2019), a new approach of integrating piecewise linear representation and weighted

support vector machine has been introduced to forecast the stock turning points. K-SRLSSVCR (Ma et al. 2019) has been proposed a robust least squares version of K-SVCR (K-RLSSVCR) based on squares ε-insensitive ramp loss and truncated least squares loss, which partially depress the impact of outliers on the new model via its nonconvex ε-insensitive ramp loss and truncated least squares loss.

Overall, in the real-world applications, it is impossible to ignore the importance of productivity and speed of classification-based model due to the existence of complex information. Consequently, it is substantial to introduce a specific method in order to decide and classify precisely. Regarding this, fuzzy aspects and probability distribution can be effective. Although the previous methods have increased the precision and accuracy of SVM classifier, those methods have not been able to handle uncertainty aspects such as noisy and outlier samples based on a probability distribution. In order to realize this aim, it is better to introduce a more precise and quick method. Thus, the fuzzy rough set theory is redesigned. Following the strategy outlined above, this paper proposes a new weighted support vector machine method based on the fuzzy rough set theory in order to decrease uncertainty.

3 Preliminaries

In this section, a brief overview of the constructive concepts of the proposed method is presented. In Sect. 3.1, the concepts regarding the support vector machine are reviewed. In Sect. 3.2, the basic comprehension of fuzzy rough set theory is defined.

3.1 Support vector machine

Suppose that there is a group of training samples $\{(x_i, y_i), x_i \in R^d, y_i \in \{+1, -1\}, i = 1, \ldots, N\}$, where x_i represents the ith sample and y_i represents its corresponding class label. SVM aims to find a hyperplane which separates the positive training samples from those negative ones and maximizing the margin W between both training samples. Also SVM can extend two class problems to multiclass problems (Hsu and Lin 2002) which $y_i \in \{1, \ldots, C_i\}$, where C_i represents the number of classes. There are two common methods for implementation for SVM multiclass classification including one-against-all method and one-against-one method. In this paper, one-against-one method is considered. In order to maximize the margin, thus it needs to minimize $\|W\|$ that converts to primary quadratic programming of SVM (Vapnik 1995; Yang and Xu 2017; Mao et al. 2014; Santhanama et al. 2016) as following:

$$\min \frac{1}{2} \|W\|^2 + C \sum_{i=1}^{N} \xi_i$$
$$s.t \ y_i(w^\tau \phi(x_i) + b) \geq 1 - \xi_i; \quad (1)$$
$$\xi_i \geq 0; \ i = 1, \ldots, N$$

where ξ_i is the error term and $C > 0$ is the regularization parameter. The above optimization can be converted to one problem in the form of (2) by introducing Lagrange parameters

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i[y_i(w^\tau \phi(x_i) + b) - 1 + \xi_i]$$
$$- \sum_{i=1}^{N} \xi_i \mu_i$$
$$s.t \ \alpha_i \geq 0; \ \mu_i \geq 0, i = 1, \ldots, N$$
$$(2)$$

where w is a weight vector, b is bias, α is Lagrangian coefficients, and $\phi(x_i)$ is kernel of training samples.

To apply such trends, the results of partial derivatives are replaced in (2). Therefore, a dual problem is constructed in the form of (3). In fact, it can be solved by quadratic programming.

$$\min \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^\tau x_j - \sum_i \alpha_i$$
$$= \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i, x_j) - \sum_i \alpha_i \quad (3)$$
$$s.t \ \sum_i \alpha_i y_i = 0; \ 0 \leq \alpha_i \leq C, \ \forall i$$

Hence, the solution has the form

$$w = \sum_{i=1}^{n} \alpha_i y_i \phi(x_i) = \sum_{i \in sv} \alpha_i y_i \phi(x_i) \quad (4)$$

where SV is the number of support vectors

$$sv = \{i | 0 \leq \alpha_i \leq C\}$$
$$\forall i \in sv, \ w^\tau \phi(x_i) + b = y_i$$
$$y_i \in \{+1, -1\} \quad (5)$$
$$b_i = y_i - w^\tau \phi(x_i)$$

where x_i is support vector. And the average of all this b_i defines the bias

$$b = \frac{1}{|sv|} \sum_{i \in sv} b_i = \frac{1}{|sv|} \sum_{i \in sv} (y_i - w^\tau \phi(x_i)) \quad (6)$$

Once the optimal pair (w, b) is determined, the decision function is obtained by

$$f = (w, \phi(z)) + b = \mathrm{sign}\left(\sum_{i=1}^{N_{sv}} \alpha_i y_i K(x_i, z) + b\right) \quad (7)$$

where N_{sv} is the number of support vectors.

3.2 Fuzzy rough set

Fuzzy rough set theory (Dubois and Prade 1990) is built upon two other theories, namely fuzzy set theory (Zadeh 1965) and rough set theory (Pawlak 1982). Rough set tries to divide universe of discourse to positive region (lower approximation sets), boundary region, and negative region. Fuzzy rough set for each sample returns a pair of memberships that show lower approximation membership as a degree of certainty and upper approximation membership as a possibility degree of being included in target samples.

The lower and upper approximation memberships are constructed by indiscernibility relationships (IR) between samples and the amount of dependency on the target set, μ_F (Verbiest et al. 2013b). IR as shown in (8) measures the similarity of each pair of samples. When two samples are identical, IR becomes 1, and when the samples are completely different, it shows zero.

$$\mathrm{IR}(x_i, x_j) = \underset{\alpha \in \mathrm{Dimension}}{\tau} \left(1 - \left|x_i(\alpha) - x_j(\alpha)\right|^2\right) \quad (8)$$

where τ is a triangular norm (t-norm), $\tau : [0, 1]^2 \rightarrow [0, 1]$ based on dimension (attribute) α.

According to the above relation, the differences between two samples in every dimension are aggregated by a t-norm. The result is called indiscernibility relation between two samples under decision boundary DB. On the other hand, membership of each sample into target class, μ_F, can be shown differently like binary ones or the others. Therefore, by these two concepts, lower and upper approximation memberships are defined in (9) and (10), respectively:

$$\mu_{\underline{F_{\mathrm{IR},\mu_F}}}(x_i) = \inf_{x_j \in T, x_i \neq x_j} I(\mathrm{IR}(x_i, x_j), \mu_F(x_j)) \quad (9)$$

$$\mu_{\overline{F_{\mathrm{IR},\mu_F}}}(x_i) = \sup_{x_j \in T, x_i \neq x_j} \tau(\mathrm{IR}(x_i, x_j), \mu_F(x_j)). \quad (10)$$

As shown in the above equations, fuzzy operators, implicator (I) and t-norm (τ), combine two basic elements and result in several outputs. Then, "inf" and "sup" select one of these outcomes as the final result. Hence, outlier and noise data can change lower and upper approximation memberships in a wide range. Therefore, Verbiest et al. (2013a) propose adjusted versions of the memberships using order weigh average (OWA) instead of "inf" and "sup." These forms of memberships are presented as follows:

$$\mu_{\underline{F_{\mathrm{IR},\mu_F}}}(x_i) = \underset{x_j \in T, x_i \neq x_j}{\mathrm{OWA}_{\min}} I(\mathrm{IR}(x_i, x_j), \mu_F(x_j)) \quad (11)$$

$$\mu_{\overline{F_{\mathrm{IR},\mu_F}}}(x_i) = \underset{x_j \in T, x_i \neq x_j}{\mathrm{OWA}_{\max}} \tau(\mathrm{IR}(x_i, x_j), \mu_F(x_j)). \quad (12)$$

Many methods like popular fuzzy belief function (Shafer 1976) follow similar trend to handle conflicting, incomplete, and uncertain information (Liu et al. 2011, 2015, 2016). In fact, it tries to explain probability of occurring a subset of universe of discourse based on belief and plausibility functions in the form of (13) and (14), respectively.

$$\forall A \in T : \mathrm{Bel}(A) = \sum_{B:B \in A} P_\alpha(B) \quad (13)$$

$$\forall A \in T : \mathrm{Pl}(A) = \sum_{B:B \cap A \neq \phi} P_\alpha(B) \quad (14)$$

where $P_\alpha(B)$ demonstrates the probability of happening B while the amount of α has the probability of occurrence relevant to α-cut in fuzzy concept (Chen et al. 2008). On the other hand, the relation between these two concepts as explained in (Dubois and Prade 1990; Chen et al. 2008; Yao and Lingras 1998; Wu et al. 2002; Liu et al. 2015) can be gained as follows:

$$\forall A \in T : \mathrm{Bel}(A) = \mu_{\underline{F_{\mathrm{IR},\mu_F}}}(A) \quad (15)$$

$$\forall A \in T : \mathrm{Pl}(A) = \mu_{\overline{F_{\mathrm{IR},\mu_F}}}(A). \quad (16)$$

Consequently, fuzzy rough set is powerful to handle the vagueness and conflict among data and its effectiveness has been proved by remarkable results in fields such as KNN improvement (Verbiest et al. 2013a; Bian and Mazlack 2003; Derrac et al. 2013), fuzzy decision tree expansion (Zhai 2011), and solid multiple traveling salesman problem (Changdar et al. 2016). However, it stands on common fuzzy operators which may be improvable in some issues.

4 Proposed method

In this section, a novel method, namely WSVM-FRS (Weighted SVM-Fuzzy Rough Set), is proposed for SVM training which is one of the novel data characteristic to reduce the effect of noise in SVM training with soft margin toward important samples in contrast to the others. In this method, a weighted coefficient is added to the penalty term Lagrangian formula for optimization problem, using lower and upper approximation for membership function in fuzzy rough set theory. Consequently, in WSVM-FRS noisy samples have low degree. The simplest form of samples is shown by discrete attributes reviewed in the following paragraph.

In Fig. 1, assume a curved line as boundary region dividing universe of discourse. The whole area is sectionalized by indiscernibility relation (IR) shaping equivalent squares. Therefore, these squares contain samples that have similar attributes. According to the rough set theory, the upper approximation represented includes all squares having at least one sample in the boundary region. On the other hand, as shown in Fig. 1, samples in square units located in the boundary region are positive region (lower approximation sets) of the universe segregated by curved boundary region and IR. Consequently, as shown in Fig. 1, simultaneously considering both approximation sets, samples that are certainly in the boundary region gain higher values rather than those on the boundary region. In addition, negative samples are the ones taking zero values.

The approximation memberships constructing ED use (8) and (17) as IR and amount of belonging samples to target set t_s, respectively. t_s demonstrates reverse distance of samples from the center of the class

$$t_s = 1 - \frac{\sum\limits_{\alpha \in \text{Dimension}} \left| \sum_{x_i \in T} x_i(\alpha) - c^* \right|^2}{|c^*|} \quad (17)$$

where T is target set and α dimension (attribute) and $||$ is absolute value. Also c^* is the center of class. The center of each class c^* is defined as follows

$$c^* = \frac{1}{N} \sum_{x_i \in y_i} x_i \quad (18)$$

where N is total number of training sample in each class.

Therefore, the lower and upper approximations memberships are reconstructed according to (9) and (10) by the following

$$\mu_{\underline{F_{\text{IR},\mu_F}}}(x_i) = \inf I(\text{IR}(x_i, c^*), t_s) \quad (19)$$

$$\mu_{\overline{F_{\text{IR},\mu_F}}}(x_i) = \sup \tau(\text{IR}(x_i, c^*), t_s). \quad (20)$$

It is important to introduce a measurement parameter which measures relation among samples based on entropy of samples. Furthermore, WSVM-FRS sets penalty and kernel parameters by using grid search. The new data characteristics discover the certainty of a sample. Each sample, x_i, is recognized as certain one if it is represented by entropy of attributes similar to the others.

In order to satisfy the above explanations, WSVM-FRS creates a map of data importance by computing ED in the form of (21).

$$\text{ED}(x_i) = -\sum_{x_i \in R^d} \mu_{\overline{F_{\text{IR},\mu_F}}}(x_i) \log_2 \mu_{\underline{F_{\text{IR},\mu_F}}}(x_i) \quad (21)$$

Therefore, in this paper, ED concludes that both lower and upper approximation memberships are proposed as certainty index. In fact, ED gains a general perspective of sample's roles by the upper approximation membership. Following this, the lower approximation membership, which computes restricted relation between samples and the decision boundary, is added to ED to improve the level of samples, which are certainly in the decision boundary. Then, it maps to range of [0, 1].

In addition to assigning certain values as ED to each sample, kernel function is also used for mapping data to hyper dimension $\varphi(.)$. The advantages of this mapping appear when inner product is displayed in optimization formula, then it can be replaced with the kernel trick. WSVM-FRS can be optimized as follows to discover boundary:

$$\min \frac{1}{2} \|W\|^2 + C \sum_{i=1}^{N} \text{ED}(x_i)\xi_i$$
$$s.t \; y_i(w^\tau \phi(x_i) + b) \geq 1 - \xi_i; \quad (22)$$
$$\xi_i \geq 0; \; i = 1, \ldots, N.$$

This optimization turns into differentiable form below by adding $\alpha_i \geq 0$ and $\mu_i \geq 0$ as positive Lagrange values:

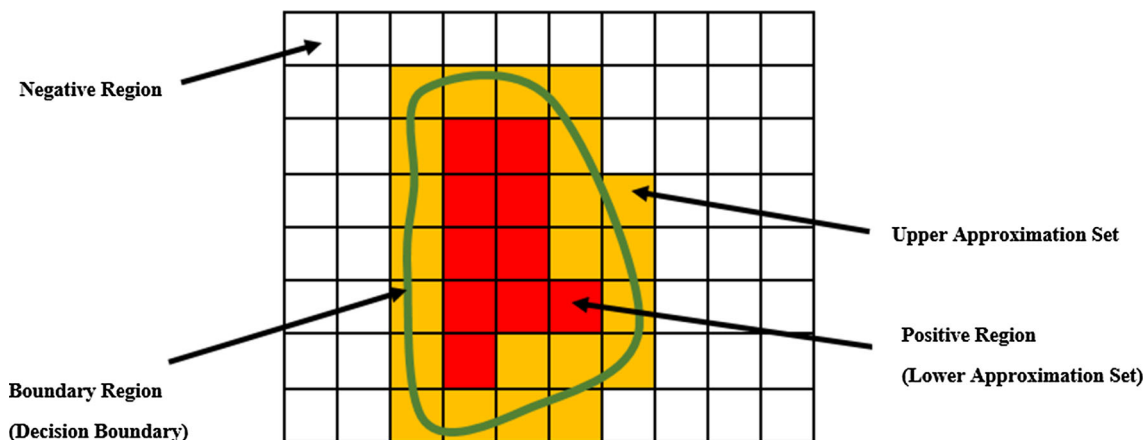**Fig. 1** An example for entropy degree (ED) expression

$$L(w, b, \alpha) = \min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \mathrm{ED}(x_i)\xi_i$$
$$- \sum_{i=1}^{N}\alpha_i[y_i(w^\tau\phi(x_i)+b)-1+\xi_i] - \sum_{i=1}^{N}\xi_i\mu_i \quad (23)$$
$$s.t \; \alpha_i \geq 0; \; \mu_i \geq 0, i = 1,\ldots,N.$$

Since minimization of (23) is reachable by minimization of w, b, ξ_i and maximization of α_i, μ_i, partial derivations of w, b, ξ_i have been set to zero as follows:

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \Rightarrow w - \sum_{i=1}^{N}\alpha_i y_i \phi(x_i) = 0 \Rightarrow w$$
$$= \sum_{i=1}^{N}\alpha_i y_i \phi(x_i) \quad (24)$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i y_i = 0 \quad (25)$$

$$\frac{\partial L(w, b, \alpha)}{\partial \xi_i} = 0 \Rightarrow \sum_{i=1}^{N} C \times \mathrm{ED}(x_i) - \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\mu_i$$
$$= 0 \Rightarrow \sum_{i=1}^{N}(C - \alpha_i - \mu_i) = 0 \overset{\alpha_i \geq 0, \mu_i \geq 0}{\Rightarrow} C = \alpha_i + \mu_i$$
$$\Rightarrow 0 \leq \alpha_i \leq C, 0 \leq \mu_i \leq C. \quad (26)$$

By replacing equations of (24) and (25) with (26), a form of DB is constructed in the form of (27):

$$\min \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i, x_j) - \sum_i \alpha_i$$
$$s.t \; \sum_i \alpha_i y_i = 0; \; 0 \leq \alpha_i \leq C \times \mathrm{ED}(x_i), \forall i. \quad (27)$$

The above form of optimization can be solved by the well-known quadratic optimization. Positive results of (27) describe support vectors (SVs). These special samples show the boundary and bias toward special ones.

Therefore, if x is a new sample, its decision function in the form of (28):

$$f = (w, \phi(z)) + b = \sum_{i=1}^{N}\alpha_i y_i K(x_i, z) + b$$
$$= \sum_{i \in sv} y_i K(x_i, z) + b. \quad (28)$$

In WSVM-FRS, the boundary is shaped by more effective SVs, the decision boundary becomes more accurate. In the proposed method, grid search is used to find values of penalty and kernel values. Overall, illustration of the proposed method as a step-by step implementation flowchart in Fig. 2 is shown. The next section will demonstrate the superiority of WSVM-FRS over state-of-the-art methods based on experiments conducted on real data sets.

5 Results and discussion

Several experiments have been conducted in terms of accuracy and the value of area under the receiver operating characteristic (ROC) graph to represent the superiority of the proposed method (WSVM-FRS) in comparison with the alternative classification methods. The alternative classification methods involved in this comparison include probabilistic weighted least squares SVM (Lu et al. 2017), DS-RLSSVM (Zhou et al. 2016), Fuzzy-LSSVM (Sheng et al. 2015), RLS-SVM (Yang et al. 2014), PLR-WSVM (Tang et al. 2019) and K-SRLSSVCR (Ma et al. 2019) methods. After describing the implementation details in Subsection 5.1, the results of the experiments are shown in Subsection 5.2.

5.1 Implementation details

In order to validate WSVM-FRS, experiments have been carried out over the real-world data sets taken from the UCI data set repository (Lichman 2013). The case and mechanism of data sets applicability is based on classification under uncertainty due to decision-based models in real-world application because accurate decision is very important such as in medical data, time series data, letter data, etc. which have been described in UCI data set (Lichman 2013).

For all the experiments, the tenfold cross-validation procedure has been used. That is, each data set was divided into ten mutually exclusive blocks, and the proposed method was applied over a training set, built with nine of the ten blocks, and the left one was used as a testing set. Each block was used as the testing set, and the average of the ten tests was reported. The selected data sets and their related parameters are listed in Table 1. In Table 1, #samples, #features, and #classes denote the number of data samples, the number of attributes and the number of classes, respectively.

Experiments have been carried out to evaluate the proposed method against three state-of-the-art weighted SVM methods like probabilistic weighted least squares SVM (Lu et al. 2017), DS-RLSSVM (Zhou et al. 2016), Fuzzy-LSSVM (Sheng et al. 2015), RLS-SVM (Yang et al. 2014), PLR-WSVM (Tang et al. 2019) and K-SRLSSVCR (Ma et al. 2019) methods. In these experiments, grid search has been used for tuning the regularization parameters. A typical soft-margin SVM classifier equipped with an RBF kernel has at least two hyperparameters that need to be tuned for good performance on unseen data. In the grid

Fig. 2 Flowchart of the proposed method
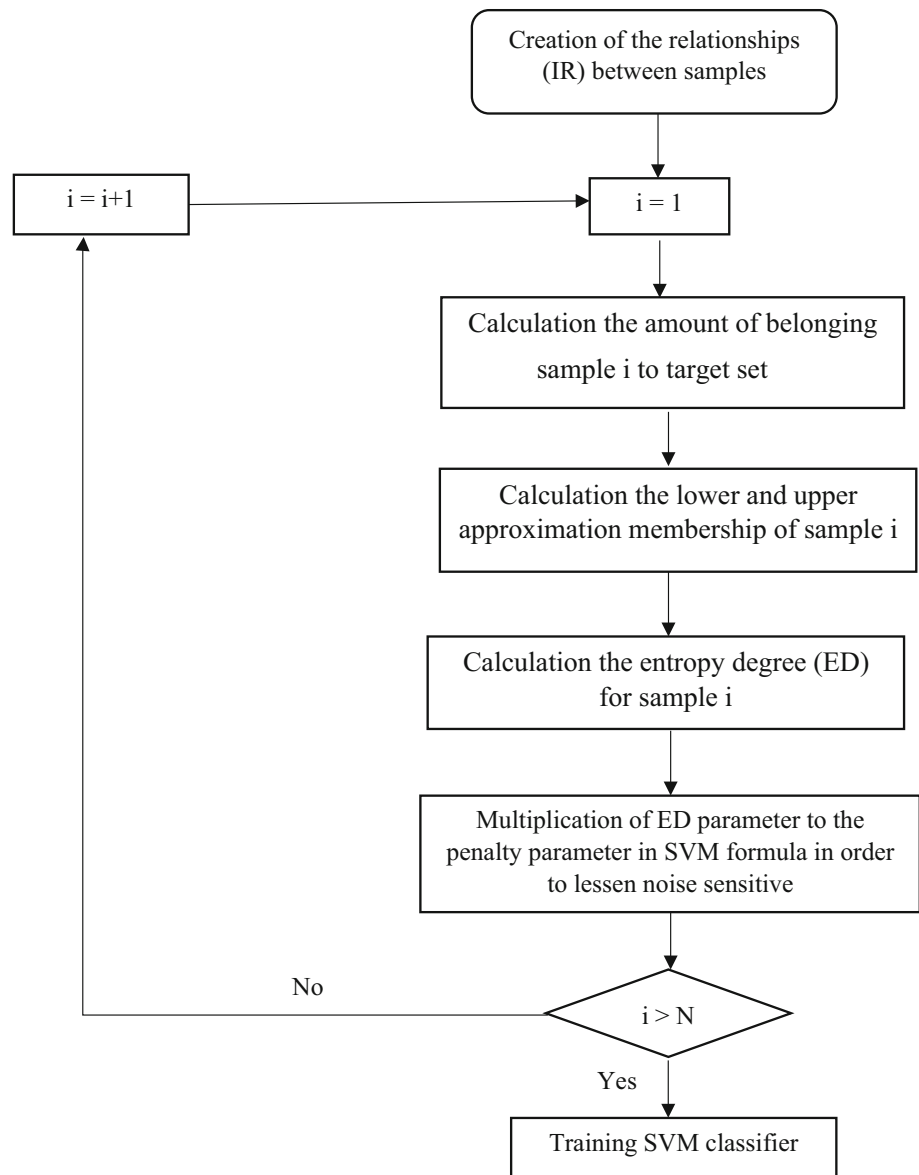
searching scheme, the regularization constant C (penalty parameter) is tuned within the range $\{0, 2^{-1}, 2^0, 2^1, 2^2, 2^3, \ldots\}$, and a kernel parameter γ is tuned within the range $\{2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, \ldots\}$. These experiments have been executed by a computer with an Intel Core i3 2.4 GHz CPU, 8 GB DDR III memory and software of MATLAB R2015b over Microsoft Windows 7 OS. In experiments, libsvm (Chang and Lin 2011) is employed to implement the base classifier of SVM.

The kernel used in all experiments is a radial basis function:

$$f(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right). \tag{29}$$

5.2 Experimental results

The final results of SVM classification can be summarized in four groups: true positive rate is the proportion of positive samples that were correctly identified (TP), false-positive rate is the proportion of negatives samples that were incorrectly classified as positive (FP), true negative rate is defined as the proportion of negatives samples that were classified correctly (TN), and finally false negative rate is the proportion of positives samples that were incorrectly classified as negative (FN). These sample segregations are briefly shown in Table 2.

Accuracy is a popular index displaying the percentages of samples which is truly described as the result. The accuracy can be computed using Eq. (30):

Table 1 Selected data sets of UCI data repository (Lichman 2013) in the experiments

No	Data set	#samples	#features	#classes
1	Iris	150	4	3
2	Sonar	208	60	2
3	Glass	214	9	7
4	Heart	267	44	2
5	Haberman	306	3	2
6	E. Coli	336	7	8
7	Liver	345	6	2
8	Ionosphere	351	34	2
9	Musk	476	167	2
10	Wdbc	569	30	2
11	Transfusion	748	4	2
12	Diabetes	768	8	2
13	Vehicle	846	18	4
14	Vowel	990	10	11
15	Yeast	1484	8	10
16	Segment	2310	19	7
17	Pendigits	3498	16	10
18	Satimage	4435	36	6
19	Letter	15,000	16	26
20	Shuttle	43,500	9	7

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{FN}) + (\text{FP} + \text{TN})}. \tag{30}$$

The results of experiments in the form of accuracy are described in Table 3. In most cases, classification accuracy WSVM-FRS has better results than the others.

A statistical significance analysis was performed by considering the nonparametric Wilcoxon signed-rank test (Demsar 2006) to analyze the results and derive strong conclusions. We used this test to determine whether the improvement in the proposed method (WSVM-FRS) is relevant. The last two rows of Tables 3, 4 and 5 present the results of Wilcoxon test. A significance level 0.05 has been considered for the analysis. Number "1" represents that WSVM-FRS significantly improves over the other competing methods in terms of the average SVM classification accuracy measure.

Table 3 shows the result of average SVM classification accuracy for the competing methods. In some of the data sets, the proposed WSVM-FRS method has a better

classification accuracy than the other methods (Sheng et al. 2015; Zhou et al. 2016; Lu et al. 2017; Yang et al. 2014; Tang et al. 2019; Ma et al. 2019), which is shown in bold. For all data sets, the rank of each method is expressed in Table 3, and the average rank is shown. The best rank in any data set is highlighted. The classification accuracy of SVM training has been shown in terms of percentage (%).

The proposed method has the best performance and the top rank in rank average, compared to the other methods; it has, considerably, performed better than the other methods concerning the amount of ranks average. WSVM-FRS method is successful in six out of 20 existing data sets such as Sonar, Heart, E. coli, Diabetes, Yeast, and Letter in in terms of increasing the SVM classification accuracy, compared to the other methods; it has the best performance. The proposed method has the second to third rank among eight data sets. However, it has between the fourth and sixth rank in six of the data sets in increasing the SVM classification accuracy which means that WSVM-FRS method is different from the other methods and chosen as the best.

For some data sets, other methods have better performance in terms of increasing the classification accuracy. For example, in case of Ionosphere, Transfusion, and Vowel data sets, DS-RLSSVM (Zhou et al. 2016) obtained better results. Although DS-RLSSVM method has, among the average ranks, won the second rank in terms of increasing the classification accuracy, in some data sets, it has the second to sixth rank in increasing the classification accuracy. PLR-WSVM (Tang et al. 2019) method has the third rank average in comparison with competing methods which has had the best result in Glass and Wdbc data sets but in others this method has obtained between the second and seventh rank. After that, K-SRLSSVCR (Ma et al. 2019) method has the fourth rank average among other methods and this method has won in Haberman and Segment data sets but in other data sets has earned different ranks. Similarly, in some cases, probabilistic weighted LS-SVM (Lu et al. 2017) method has the best performance, in Musk, Pendigits, and Satimage data sets, in comparison with other methods; it has won the first rank. Probabilistic weighted LS-SVM method has won the fifth rank with considerable difference with the average rank of WSVM-FRS methods. Probabilistic weighted LS-SVM method does not have an acceptable performance in some data sets; it has the worst performance in increasing the classification accuracy, in four data sets, compared to the other methods.

Table 2 Confusion matrix

	Predicted as interested sample	Predicted as unwanted sample
Actually interested sample	True positive (TP)	False negative (FN)
Actually unwanted sample	False positive (FP)	True negative (TN)

Table 3 Result of average SVM classification accuracy

Data set	Probabilistic weighted LS-SVM (Lu et al. 2017)	Fuzzy-LSSVM (Sheng et al. 2015)	DS-RLSSVM (Zhou et al. 2016)	RLS-SVM (Yang et al. 2014)	PLR-WSVM (Tang et al. 2019)	K-SRLSSVCR (Ma et al. 2019)	WSVM-FRS
Iris	83.33 (6)	**90.00 (1)**	85.33 (4)	82.00 (7)	86.23 (3)	83.89 (5)	87.33 (2)
Sonar	80.76 (7)	85.09 (4)	81.25 (6)	83.17 (5)	85.58 (3)	86.12 (2)	**87.50 (1)**
Glass	79.90 (3)	78.50 (4)	75.70 (5)	85.51 (2)	**86.85 (1)**	71.12 (7)	73.17 (6)
Heart	76.40 (5)	75.28 (6)	80.52 (3)	73.40 (7)	79.54 (4)	81.01 (2)	**81.64 (1)**
Haberman	83.98 (6)	88.56 (3)	88.23 (4)	83.33 (7)	88.98 (2)	**89.54 (1)**	86.92 (5)
E. coli	76.48 (5)	79.76 (4)	82.73 (2)	80.65 (3)	72.85 (7)	75.01 (6)	**85.41 (1)**
Liver	78.55 (4)	75.36 (7)	77.68 (6)	**79.71 (1)**	78.98 (2)	78.65 (3)	77.97 (5)
Ionosphere	72.36 (4)	71.50 (5)	**74.35 (1)**	70.08 (7)	70.87 (6)	73.22 (3)	73.50 (2)
Musk	**71.21 (1)**	67.43 (5)	68.90 (3)	67.85 (4)	66.67 (6)	66.02 (7)	70.37 (2)
Wdbc	83.83 (6)	84.00 (5)	85.23 (4)	83.65 (7)	**87.11 (1)**	87.01 (2)	86.99 (3)
Transfusion	73.66 (7)	80.00 (6)	**85.42 (1)**	83.82 (5)	85.02 (2)	84.98 (3)	84.35 (4)
Diabetes	75.26 (7)	76.56 (5)	76.04 (6)	79.68 (2)	77.25 (4)	78.58 (3)	**81.11 (1)**
Vehicle	85.46 (6)	83.21 (7)	86.52 (3)	**87.35 (1)**	85.98 (5)	86.98 (2)	86.40 (4)
Vowel	77.77 (4)	80.10 (2)	**80.60 (1)**	76.96 (5)	73.58 (7)	75.12 (6)	79.89 (3)
Yeast	74.73 (7)	84.23 (6)	84.56 (4)	83.82 (5)	85.58 (2)	84.78 (3)	**86.99 (1)**
Segment	67.79 (5)	67.74 (6)	69.78 (3)	67.53 (7)	70.54 (2)	**71.45 (1)**	69.65 (4)
Pendigits	**70.41 (1)**	67.35 (3)	67.26 (4)	67.12 (5)	66.21 (6)	64.58 (7)	70.18 (2)
Satimage	**81.08 (1)**	78.30 (3)	78.12 (6)	75.12 (7)	78.20 (4)	78.18 (5)	80.92 (2)
Letter	75.81 (3)	67.10 (5)	70.14 (4)	65.50 (6)	76.01 (2)	65.02 (7)	**76.67 (1)**
Shuttle	**88.16 (1)**	83.90 (3)	83.13 (4)	80.45 (5)	78.12 (6)	76.54 (7)	88.02 (2)
Rank average	4.45 (5)	4.50 (6)	3.70 (2)	4.90 (7)	3.75 (3)	4.10 (4)	**2.60 (1)**
z value	− 2.76	− 2.87	− 2.53	− 2.94	− 2.20	− 2.63	
Wilcoxon test result	1	1	1	1	1	1	

Note that the rank of each method is enclosed in parenthesis. The best results are given in bold

Fuzzy-LSSVM (Sheng et al. 2015) method has the sixth rank average. It has the best performance in terms of increasing the classification accuracy in Iris data set compared to the other methods. It stands far away from the proposed method in ranks average for increasing the SVM classification accuracy. RLS-SVM (Yang et al. 2014) has

Table 4 Result of average SVM classification precision

Data set	Probabilistic weighted LS-SVM (Lu et al. 2017)	Fuzzy-LSSVM (Sheng et al. 2015)	DS-RLSSVM (Zhou et al. 2016)	RLS-SVM (Yang et al. 2014)	PLR-WSVM (Tang et al. 2019)	K-SRLSSVCR (Ma et al. 2019)	WSVM-FRS
Iris	83.33 (7)	**90.27 (1)**	84.72 (4)	83.34 (6)	85.65 (3)	84.01 (5)	88.88 (2)
Sonar	84.37 (7)	88.54 (5)	89.58 (4)	86.45 (6)	89.70 (3)	89.92 (2)	**90.62 (1)**
Glass	73.00 (4)	70.00 (6)	72.00 (5)	78.00 (2)	**80.15 (1)**	69.25 (7)	75.00 (3)
Heart	79.48 (3)	77.77 (5)	76.63 (6)	74.35 (7)	79.04 (4)	79.68 (2)	**80.34 (1)**
Haberman	77.14 (6)	84.76 (4)	85.71 (3)	76.19 (7)	86.08 (2)	**87.34(1)**	81.90 (5)
E. coli	78.40 (5)	81.25 (4)	82.32 (2)	82.38 (3)	71.20 (7)	74.21 (6)	**83.52 (1)**
Liver	83.68 (5)	78.42 (7)	80.52 (6)	**84.21 (1)**	84.08 (2)	83.75 (3)	83.69 (4)
Ionosphere	74.02 (2)	70.56 (5)	**74.03 (1)**	69.69 (6)	68.20 (7)	73.66 (3)	73.59 (4)
Musk	**77.49 (1)**	74.16 (5)	74.90 (3)	74.53 (4)	70.11 (6)	66.32 (7)	76.75 (2)
Wdbc	77.82 (4)	75.73 (5)	77.40 (6)	74.05 (7)	**81.36 (1)**	80.31 (2)	79.49 (3)
Transfusion	**97.07 (1)**	86.84 (5)	96.78 (2)	94.15 (4)	85.00 (6)	82.52 (7)	96.49 (3)
Diabetes	74.18 (7)	78.26 (5)	74.72 (6)	79.36 (2)	78.90 (4)	78.88 (3)	**80.15 (1)**
Vehicle	89.33 (4)	89.55 (3)	89.11 (7)	**89.77 (1)**	85.25 (5)	89.75 (2)	89.12 (6)
Vowel	**79.00 (1)**	78.20 (4)	78.60 (2)	78.40 (3)	72.59 (7)	75.00 (6)	77.68 (5)
Yeast	80.38 (7)	87.15 (6)	87.50 (4)	87.38 (5)	94.08 (2)	90.70 (3)	**98.57 (1)**
Segment	60.71 (5)	60.47 (6)	63.57 (2)	60.31 (7)	62.64 (4)	**65.35 (1)**	63.49 (3)
Pendigits	**61.28 (1)**	60.70 (3)	60.65 (4)	60.54 (5)	59.32 (6)	57.99 (7)	61.07 (2)
Satimage	**78.84 (1)**	73.87 (3)	73.82 (5)	72.35 (7)	73.85 (4)	72.88 (6)	78.75 (2)
Letter	76.98 (3)	63.27 (4)	62.50 (5)	62.20 (6)	77.31 (2)	60.02 (7)	**77.51 (1)**
Shuttle	**91.52 (1)**	86.95 (4)	87.45 (3)	86.08 (5)	84.42 (6)	80.64 (7)	91.30 (2)
Rank average	3.75 (2)	4.50 (6)	4.00 (3)	4.70 (7)	4.10 (4)	4.35 (5)	**2.60 (1)**
z value	− 2.12	− 3.21	− 2.80	− 3.28	− 2.57	− 2.72	
Wilcoxon test result	1	1	1	1	1	1	

Note that the rank of each method is enclosed in parenthesis. The best results are given in bold

the seventh rank average in terms of increasing the SVM classification accuracy which is the worst method among ranks average compared to the other methods; it has the best performance in Liver, and Vehicle data sets as well as the first rank. It has the biggest difference with WSVM-FRS method concerning rank average.

Table 5 Result of average SVM classification recall

Data set	Probabilistic weighted LS-SVM (Lu et al. 2017)	Fuzzy-LSSVM (Sheng et al. 2015)	DS-RLSSVM (Zhou et al. 2016)	RLS-SVM (Yang et al. 2014)	PLR-WSVM (Tang et al. 2019)	K-SRLSSVCR (Ma et al. 2019)	WSVM-FRS
Iris	82.19 (4)	**89.04 (1)**	84.72 (3)	80.00 (5)	78.21 (7)	79.02 (6)	85.33 (2)
Sonar	76.41 (6)	80.95 (4)	74.78 (7)	79.04 (5)	81.75 (3)	81.90 (2)	**83.65 (1)**
Glass	80.21 (5)	81.39 (4)	75.00 (6)	87.64 (2)	**88.45 (1)**	72.30 (7)	85.22 (3)
Heart	70.45 (5)	69.46 (6)	77.31 (3)	67.96 (7)	72.04 (4)	77.78 (2)	**78.33 (1)**
Haberman	76.41 (6)	82.40 (3)	81.08 (4)	75.47 (7)	82.90 (2)	**83.16 (1)**	80.37 (5)
E. coli	77.09 (5)	80.33 (4)	84.30 (2)	81.00 (3)	68.15 (7)	75.02 (6)	**88.02 (1)**
Liver	78.71 (5)	77.20 (7)	79.27 (4)	**82.05 (1)**	80.20 (2)	79.75 (3)	77.94 (6)
Ionosphere	79.90 (6)	83.58 (4)	**85.07 (1)**	82.14 (5)	68.54 (7)	83.70 (3)	84.15 (2)
Musk	**73.42 (1)**	70.27 (5)	71.73 (3)	70.62 (4)	69.02 (6)	67.12 (7)	72.72 (2)
Wdbc	82.66 (7)	84.57 (6)	86.04 (4)	85.09 (5)	**90.30 (1)**	89.80 (2)	88.37 (3)
Transfusion	74.27 (4)	74.06 (5)	**77.15 (1)**	76.12 (2)	73.21 (6)	72.01 (7)	75.86 (3)
Diabetes	74.18 (7)	74.22 (6)	75.13 (5)	77.31 (2)	76.00 (4)	76.50 (3)	**79.11 (1)**
Vehicle	84.27 (6)	80.92 (7)	86.05 (3)	**86.88 (1)**	84.33 (5)	86.22 (2)	85.86 (4)
Vowel	77.45 (4)	81.62 (2)	**82.21 (1)**	76.56 (5)	71.24 (7)	75.02 (6)	81.41 (3)
Yeast	77.45 (7)	86.16 (6)	86.40 (5)	86.98 (4)	88.88 (2)	87.08 (3)	**90.36 (1)**
Segment	75.51 (5)	75.52 (4)	77.09 (2)	75.34 (6)	75.00 (7)	**77.98 (1)**	76.92 (3)
Pendigits	**79.42 (1)**	74.35 (3)	74.24 (4)	74.06 (5)	70.02 (6)	69.54 (7)	79.15 (2)
Satimage	**81.82 (1)**	80.23 (3)	79.85 (5)	75.73 (7)	80.05 (4)	76.30 (6)	81.61 (2)
Letter	77.51 (3)	71.70 (5)	77.17 (4)	71.33 (6)	77.91 (2)	68.14 (7)	78.48 (1)
Shuttle	**88.63 (1)**	85.10 (3)	83.58 (4)	80.48 (5)	79.00 (6)	78.78 (7)	88.57 (2)
Rank average	4.45 (5)	4.40 (4)	3.55 (2)	4.35 (3)	4.45 (5)	4.40 (4)	2.40 (1)
z value	− 3.36	− 3.21	− 2.50	− 3.17	− 2.57	− 2.94	
Wilcoxon test result	1	1	1	1	1	1	

Note that the rank of each method is enclosed in parenthesis. The best results are given in bold

The two last rows of Table 3 show the results of the Wilcoxon test comparing the classification accuracy of WSVM-FRS against the other competing methods. A significance level 0.05 has been considered for the analysis. Number "1" represents that WSVM-FRS significantly

improves over the other competing methods in terms of classification accuracy.

In pattern recognition and information retrieval classification, precision (also called positive predictive value) is the fraction of retrieved samples that are relevant, while recall (also known as sensitivity) is the fraction of relevant samples that are retrieved. Precision and recall can be measured by (31) and (32), respectively. The results of precision comparison are listed in Table 4, while Table 5 shows experimental results of recall.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (31)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \qquad (32)$$

Table 4 shows the result of average SVM classification precision for the competing methods. In some of the data sets, the proposed WSVM-FRS method has a better classification precision than the other methods (Sheng et al. 2015; Zhou et al. 2016; Lu et al. 2017; Yang et al. 2014; Tang et al. 2019; Ma et al. 2019), which is shown in bold. For all data sets, the rank of each method is expressed in Table 4, and the average rank is shown. The best rank in any data set is highlighted. The classification precision of SVM training has been shown in terms of percentage (%).

The proposed method has the best performance and the top rank in rank average, compared to the other methods; it has, considerably, performed better than the other methods concerning the amount of ranks average. WSVM-FRS method is successful in six out of 20 existing data sets such as Sonar, Heart, *E. coli*, Diabetes, Yeast, and Letter in in terms of increasing the SVM classification precision measure, compared to the other methods; it has the best performance. The proposed method has the second to third rank among 9 data sets. However, it has between the fourth and sixth rank in five of the data sets in increasing the SVM classification precision which means that WSVM-FRS method is different from the other methods and chosen as the best.

For some data sets, other methods have better performance in terms of increasing the classification precision. For example, in case of Transfusion, Musk, Pendigits, Satimage, Vowel, and Shuttle data sets, probabilistic weighted LS-SVM (Lu et al. 2017) obtained better results. Although probabilistic weighted LS-SVM method has, among the average ranks, won the second rank in terms of increasing the classification precision, in some data sets, it has the second to seventh rank in increasing the classification precision. Similarly, in some cases, DS-RLSSVM (Zhou et al. 2016) method has the best performance, in Ionosphere data set, in comparison with other methods; it has won the first rank. DS-RLSSVM method has won the

third rank with considerable difference with the average rank of WSVM-FRS methods. DS-RLSSVM method does not have an acceptable performance in some data sets, compared to the other methods. PLR-WSVM (Tang et al. 2019) method has the forth rank average in comparison with competing methods which has had the best result in Glass and Wdbc data sets but in others this method has obtained between the second and seventh rank. After that, K-SRLSSVCR (Ma et al. 2019) method has the fifth rank average among other methods and this method has won in Haberman and Segment data sets but in other data sets has earned different ranks.

Similarity, Fuzzy-LSSVM (Sheng et al. 2015) method has the sixth rank average. It has the best performance in terms of increasing the classification precision in Iris data sets compared to the other methods. It stands far away from the proposed method in ranks average for increasing the SVM classification precision. RLS-SVM (Yang et al. 2014) has the seventh rank average in terms of increasing the SVM classification precision which is the worst method among ranks average compared to the other methods; it has the best performance in Liver and Vehicle data sets as well as the first rank. It has the biggest difference with WSVM-FRS method concerning rank average.

The two last rows of Table 4 show the results of the Wilcoxon test comparing the classification precision of WSVM-FRS against the other competing methods. A significance level 0.05 has been considered for the analysis. Number "1" represents that WSVM-FRS significantly improves over the other competing methods in terms of classification precision.

Table 5 shows the result of average SVM classification recall for the competing methods. In some of the data sets, the proposed WSVM-FRS method has a better classification precision than the other methods (Sheng et al. 2015; Zhou et al. 2016; Lu et al. 2017; Yang et al. 2014; Tang et al. 2019; Ma et al. 2019), which is shown in bold. For all data sets, the rank of each method is expressed in Table 5, and the average rank is shown. The best rank in any data set is highlighted. The classification recall of SVM training has been shown in terms of percentage (%).

The proposed method has the best performance and the top rank in rank average, compared to the other methods; it has, considerably, performed better than the other methods concerning the amount of ranks average. WSVM-FRS method is successful in 6 out of 20 existing data sets such as Sonar, Heart, E. coli, Diabetes, Yeast, and Letter in terms of increasing the SVM classification recall, compared to the other methods; it has the best performance. The proposed method has the second to third rank among 11 data sets. However, it does have between fourth and sixth in three of the data sets in increasing the SVM

classification recall which means that WSVM-FRS method is different from the other methods and chosen as the best.

For some data sets, other methods have better performance in terms of increasing the classification precision. For example, in case of Ionosphere, Transfusion, and Vowel data sets, DS-RLSSVM (Zhou et al. 2016) obtained better results. Although DS-RLSSVM method has, among the average ranks, won the second rank in terms of increasing the classification recall, in some data sets, it has the second to seventh rank in increasing the classification recall. RLS-SVM (Yang et al. 2014) has the third rank average in terms of increasing the SVM classification recall which is the worst method among ranks average compared to the other methods; it has the best performance in Liver, and Vehicle data sets as well as the first rank. After that, K-SRLSSVCR (Ma et al. 2019) method has the fourth rank average among other methods and this method has won in Haberman and Segment data sets but in other data sets has earned different ranks. Similarly, in one case, Fuzzy-LSSVM (Sheng et al. 2015) method has the best performance, in Iris data set, in comparison with other methods; it has won the first rank. Fuzzy-LSSVM method has won the fourth rank with considerable difference with the average rank of WSVM-FRS methods. Fuzzy-LSSVM method does not have an acceptable performance in some data sets, compared to the other methods. Probabilistic weighted LS-SVM (Lu et al. 2017) method has the fifth rank average. It has the best performance in terms of increasing the classification recall in Musk, Pendigits, Satimage, and Shuttle data sets compared to the other methods. It stands far away from the proposed method in ranks average for increasing the SVM classification recall. Finally, PLR-WSVM (Tang et al. 2019) method has the fifth rank average in comparison with competing methods which has had the best result in Glass and Wdbc data sets but in others this method has obtained between the second and seventh rank.

The two last rows of Table 5 show the results of the Wilcoxon test comparing the classification recall of WSVM-FRS against the other competing methods. A significance level 0.05 has been considered for the analysis. Number "1" represents that WSVM-FRS significantly improves over the other competing methods in terms of classification recall.

As the number of positive samples is smaller than the negative ones, the performance of SVM classification in segregating positive samples is more important. Therefore, the value of area under the receiver operating characteristic (ROC) graph, known as area under curve (AUC), is computed as another metric. ROC graph is constructed by plotting TPR in contrast to false-positive rate (FPR).

Hence, more AUC in the form of (33) shows better ability to distinguish positive samples:

$$AUC = (1 + TPR - FPR)/2. \tag{33}$$

Based on Fig. 3, AUC metric represents special power of WSVM-FRS to gain data characteristics of positive samples. This ability causes better results of the proposed method rather than the others. Finally, through analyzing several experiments on data sets taken from the UCI repository, the superiority of WSVM-FRS over state-of-the-art methods (Sheng et al. 2015; Zhou et al. 2016; Lu et al. 2017; Yang et al. 2014; Tang et al. 2019; Ma et al. 2019) is proved.

Overall, in accuracy, precision and recall metrics, WSVM-FRS has had good performance in data sets that have many classes. Also in some data sets, the proposed method has been able to satisfy all metrics to face with large number datasets and in high-dimension data sets WSVM-FRS to some extent has resulted as well.

5.2.1 Noise analysis

The presence of noise in training data has strong and negative impact on the performance of learning algorithms. Thus, methods should be sufficiently resistant and be able to deal with them. In order to present a deeper discussion and show that the proposed method has better results than competing methods (Sheng et al. 2015; Zhou et al. 2016; Lu et al. 2017; Yang et al. 2014; Tang et al. 2019; Ma et al. 2019), the most common type of artificial noise has been used which is called uniform random addition (Zhu and Wu 2004). Therefore, class noise and attribute noise have been added. The noise level has been raised in an interval of 0% (original datasets) to 30%. To evaluate the impact of noise level in terms of accuracy, the proposed method and all the other comparing methods have been performed on each dataset. The results are demonstrated in Fig. 4 for each dataset. The x-axis indicates the noise level, and the y-axis represents the classification accuracy from different types of classifiers trained.

To keep looking at Fig. 4, when the noise level increases, classification accuracy of all methods has been decreasing dramatically. Furthermore, it is clear that the more noise is added to the datasets, the more proposed method is resistant against the other methods. This is because of a strong and satisfactory classifier based on fuzzy rough set strategy which has been applied. Results indicate the superiority of the proposed method in comparison with the others in dealing with noisy data.
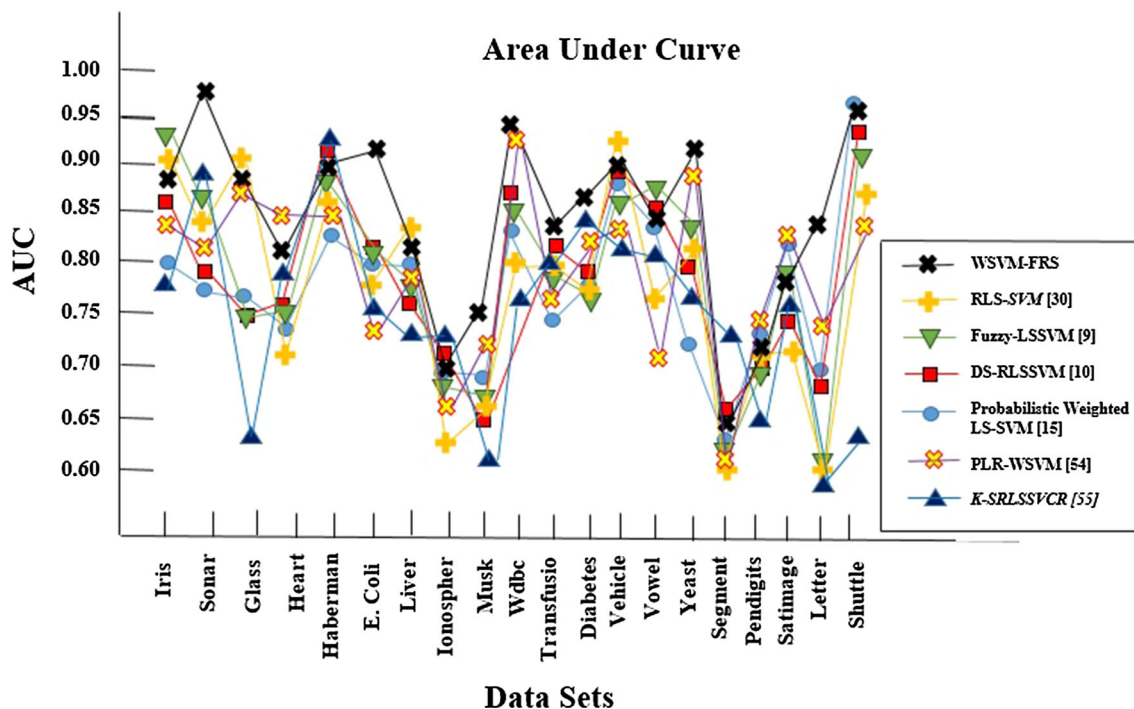
Fig. 3 Result of AUC metric

5.2.2 Real-world data set analysis

In this section, two various real-world data sets are considered to illustrate performance of the proposed method (WSVM-FRS) in comparison with some state-of-the-art methods.

MNIST data set To continue evaluation of the proposed method, specific real-world data set, namely MNIST (LeCun et al. 2010), has been considered, which is related to the handwritten digits date set and is a dataset of simple gray handwritten digits, while ImageNet is a large-scale dataset of labeled high-resolution images. MNIST has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image. Description of MNIST data sat is illustrated in Table 6. Also demonstration of the dataset is shown in Fig. 5.

Experimental results on MNIST dataset Experiment on MNIST data set has been carried out to evaluate the proposed method against three state-of-the-art weighted SVM methods like probabilistic weighted least squares SVM (Lu et al. 2017), DS-RLSSVM (Zhou et al. 2016), Fuzzy-LSSVM (Sheng et al. 2015), RLS-SVM (Yang et al. 2014), PLR-WSVM (Tang et al. 2019) and K-SRLSSVCR (Ma et al. 2019) methods, which are shown in Table 7. In this table, three evaluation measure have been considered such as classification accuracy, recall and precision.

Table 7 shows the result of average SVM classification accuracy, recall and precision for the competing methods. Regarding MNIST data set, the proposed WSVM-FRS method has a better classification performance than the other methods (Sheng et al. 2015; Zhou et al. 2016; Lu et al. 2017; Yang et al. 2014; Tang et al. 2019; Ma et al. 2019), which is shown in bold. For all methods, the rank of each method is expressed in Table 7, and the average rank is shown. The best rank is highlighted, which approves the best performance of the proposed method. The classification of three mentioned evaluations of SVM training has been shown in terms of percentage (%).

Fashion-MNIST dataset Also, other type of real-world data set, which has been used in evaluations, is Fashion-MNIST Dataset (Cohen et al. 2017). Fashion-MNIST is based on the assortment on Zalando's website. Every fashion product on Zalando has a set of pictures shot by professional photographers, demonstrating different aspects of the product, i.e., front and back looks, details, looks with model and in an outfit. The original picture has a light-gray background (hexadecimal color: #fdfdfd) and stored in 762×1000 JPEG format. For efficiently serving different frontend components, the original picture is resampled with multiple resolutions, e.g., large, medium, small, thumbnail and tiny.

The front look thumbnail images of 70, 000 unique products in used to build Fashion-MNIST. Those products come from different gender groups: men, women, kids and neutral. In particular, whitecolor products are not included

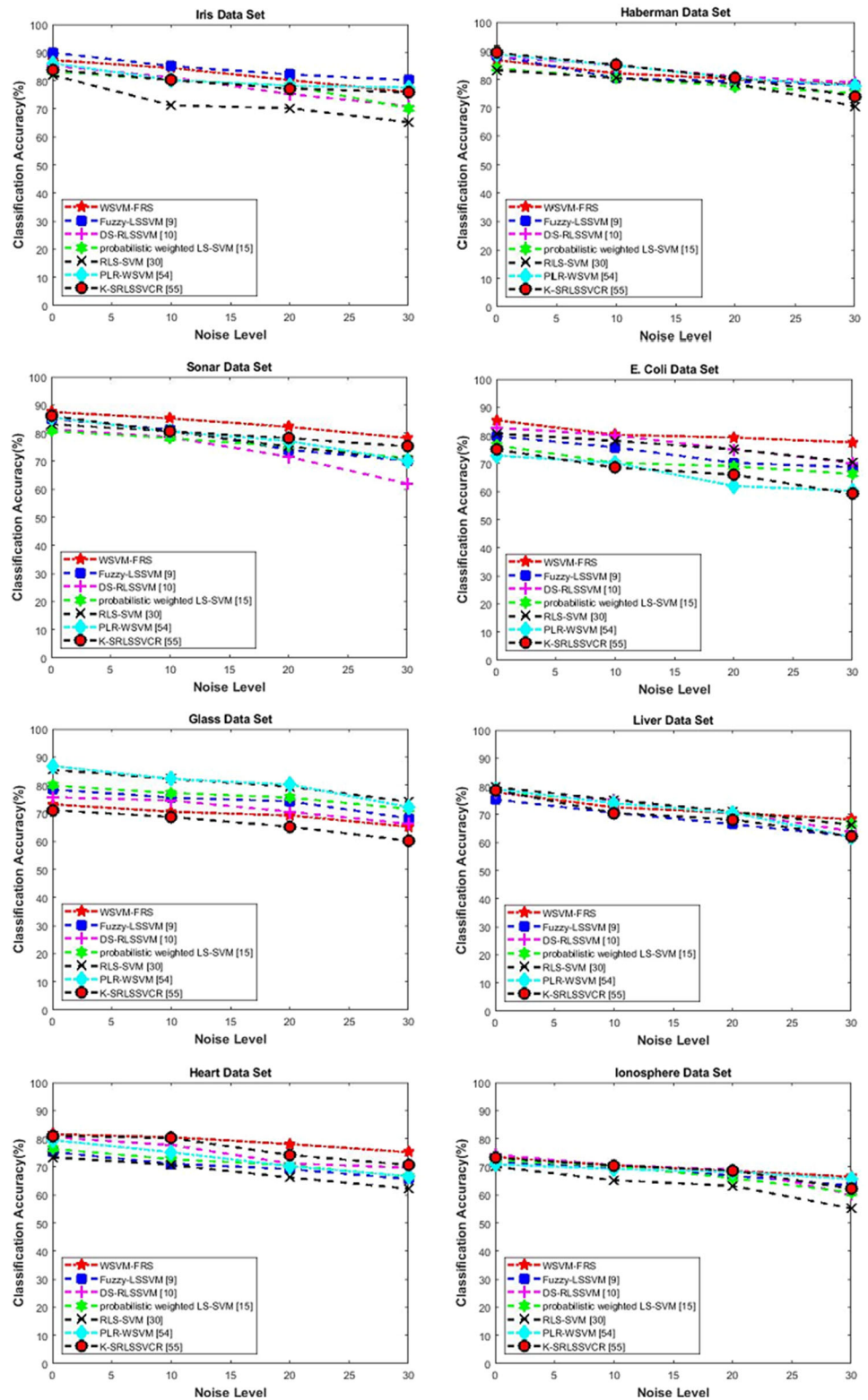**Fig. 4** Experimental results of noise effect on classification accuracy
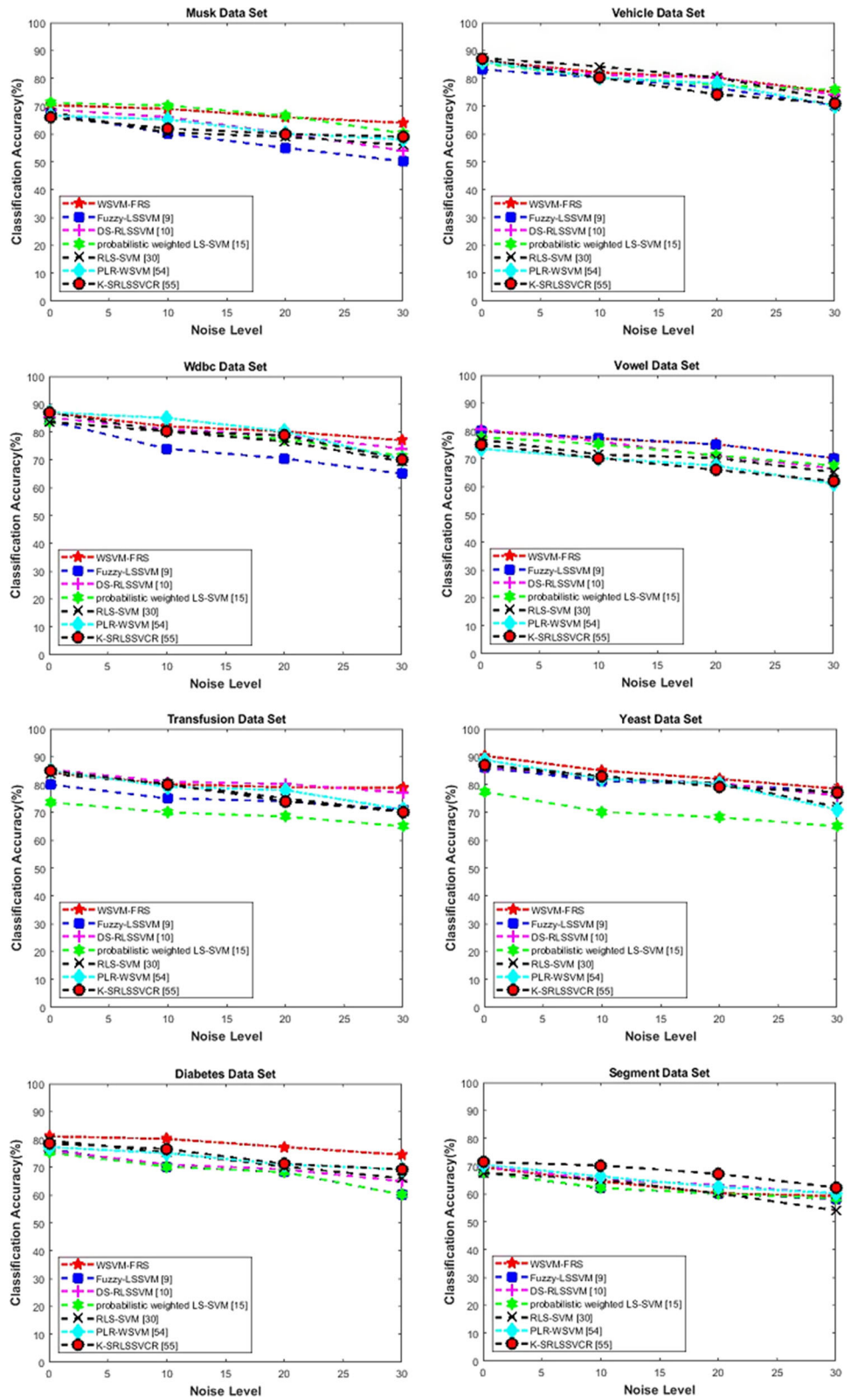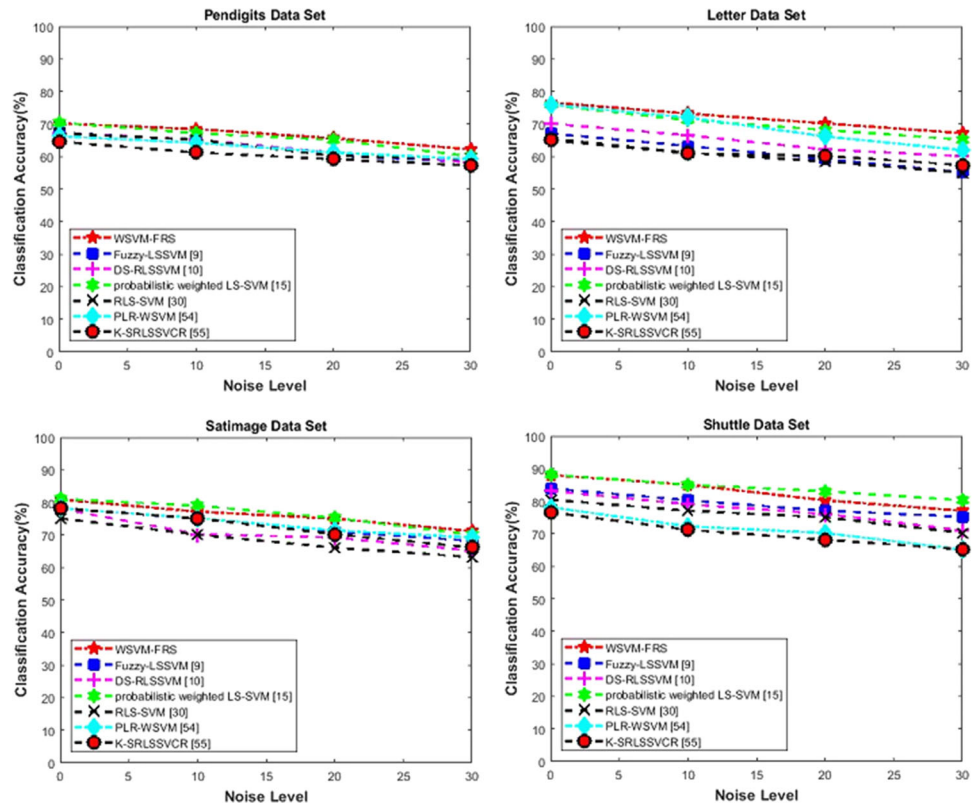
Fig. 4 continued

Fig. 4 continued

Table 6 Files contained in the MNIST dataset

Name	Description	#Examples	Size
Train-images-idx3-ubyte.gz	Training set images	60,000	9,912,422 bytes
Train-labels-idx1-ubyte.gz	Training set labels	60,000	28,881 bytes
t10k-images-idx3-ubyte.gz	Test set images	10,000	1,648,877 bytes
t10k-labels-idx1-ubyte.gz	Test set labels	10,000	4542 bytes

Fig. 5 Demonstration of the MNIST data set

in the dataset as they have low contrast to the background. The thumbnails (51×73) are then fed into the following conversion pipeline, which is visualized in Fig. 6.

1. Converting the input to a PNG image.

2. Trimming any edges that are close to the color of the corner pixels. The "closeness" is defined by the distance within 5% of the maximum possible intensity in RGB space.

3. Resizing the longest edge of the image to 28 by subsampling the pixels, i.e., some rows and columns are skipped over.

4. Sharpening pixels using a Gaussian operator of the radius and standard deviation of 1.0, with increasing effect near outlines.

5. Extending the shortest edge to 28 and put the image to the center of the canvas.

6. Negating the intensities of the image.

7. Converting the image to 8-bit grayscale pixels.

The dataset is divided into a training and a test set. The training set receives a randomly selected 6000 examples from each class. Images and labels are stored in the same file format as the MNIST data set, which is designed for storing vectors and multidimensional matrices. The result

Table 7 Result of the proposed method performance compared to other competitors regarding MNIST data set

Evaluation metrics	Probabilistic weighted LS-SVM (Lu et al. 2017)	Fuzzy-LSSVM (Sheng et al. 2015)	DS-RLSSVM (Zhou et al. 2016)	RLS-SVM (Yang et al. 2014)	PLR-WSVM (Tang et al. 2019)	K-SRLSSVCR (Ma et al. 2019)	WSVM-FRS
Accuracy	88.50 (2)	77.98 (7)	80.54 (6)	85.51 (4)	82.10 (5)	86.41 (3)	**90.01 (1)**
Precision	80.15 (4)	78.45 (5)	86.48 (2)	75.10 (6)	84.45 (3)	75.01 (7)	**88.50 (1)**
Recall	82.15 (4)	86.45 (2)	72.45 (7)	85.20 (3)	76.00 (6)	80.08 (5)	**89.51 (1)**
Average rank	3.33 (2)	5.66 (6)	5.00 (5)	4.33 (3)	4.66 (4)	5.00 (5)	**1.00 (1)**

Note that the rank of each method is enclosed in parenthesis. The best results are given in bold
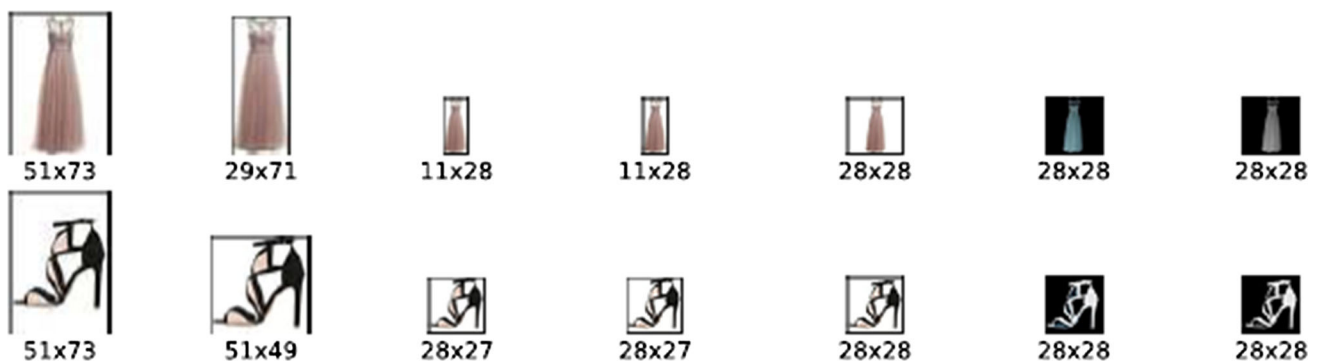
Fig. 6 Diagram of the conversion process used to generate Fashion-MNIST dataset. Two examples from dress and sandals categories are depicted, respectively. Each column represents a step described previously

Table 8 Files contained in the Fashion-MNIST dataset

Name	Description	#Examples	Size
Train-images-idx3-ubyte.gz	Training set images	60,000	25 MBytes
Train-labels-idx1-ubyte.gz	Training set labels	60,000	140 Bytes
t10k-images-idx3-ubyte.gz	Test set images	10,000	4.2 MBytes
t10k-labels-idx1-ubyte.gz	Test set labels	10,000	92 Bytes

files are listed in Table 8. Examples have been sorted by their labels while storing, resulting in smaller label files after compression comparing to the MNIST. It is also easier to retrieve examples with a certain class label.

For the class labels, the silhouette code of the product has been used. The silhouette code is manually labeled by the in-house fashion experts and reviewed by a separate team at Zalando. Each product contains only one silhouette code. Table 9 gives a summary of all class labels in Fashion-MNIST with examples for each class.

Experimental results on fashion-MNIST dataset Experiment on Fashion-MNIST Dataset has been conducted out to evaluate the proposed method compared to three state-of-the-art weighted SVM methods like probabilistic weighted least squares SVM (Lu et al. 2017), DS-

RLSSVM (Zhou et al. 2016), Fuzzy-LSSVM (Sheng et al. 2015), RLS-SVM (Yang et al. 2014), PLR-WSVM (Tang et al. 2019) and K-SRLSSVCR (Ma et al. 2019) methods, which are demonstrated in Table 10. In this table, three evaluation measure have been considered such as classification accuracy, recall and precision.

Table 10 shows the result of average SVM classification accuracy, recall and precision for the competing methods. Regarding Fashion-MNIST data set, the proposed WSVM-FRS method has a better classification performance than the other methods (Sheng et al. 2015; Zhou et al. 2016; Lu et al. 2017; Yang et al. 2014; Tang et al. 2019; Ma et al. 2019), which is shown in bold. For all methods, the rank of each method is expressed in Table 10, and the average rank is shown. The best rank is highlighted, which approves the

Table 9 Class names and example images in Fashion-MNIST dataset

Label	Description	Example
0	T-Shirt/Top	
1	Trouser	
2	Pullover	
3	Dress	
4	Coat	
5	Sandals	
6	Shirt	
7	Sneaker	
8	Bag	
9	Ankle Boots	

Table 10 Result of the proposed method performance compared to other competitors regarding Fashion-MNIST data set

Evaluation metrics	Probabilistic weighted LS-SVM (Lu et al. 2017)	Fuzzy-LSSVM (Sheng et al. 2015)	DS-RLSSVM (Zhou et al. 2016)	RLS-SVM (Yang et al. 2014)	PLR-WSVM (Tang et al. 2019)	K-SRLSSVCR (Ma et al. 2019)	WSVM-FRS
Accuracy	90.02 (2)	86.02 (5)	88.99 (3)	87.50 (4)	85.54 (6)	84.48 (7)	**91.21 (1)**
Precision	86.41 (7)	89.75 (5)	87.87 (6)	90.20 (3)	88.21 (4)	91.91 (2)	**92.55 (1)**
Recall	86.41 (3)	87.18 (2)	83.06 (5)	82.29 (6)	81.59 (7)	83.14 (4)	**88.06 (1)**
Average rank	4.00 (2)	4.00 (2)	4.66 (4)	4.33 (3)	5.66 (5)	4.33 (3)	**1.00 (1)**

Note that the rank of each method is enclosed in parenthesis. The best results are given in bold

best performance of the proposed method. The classification of three mentioned evaluations of SVM training has been shown in terms of percentage (%).

6 Conclusions and future works

In this paper, a novel method, namely WSVM-FRS, has been introduced to reduce uncertainty effect, say, noise in SVM training. The primary reason of introducing the method is consideration of sophisticated information in the real-world applications under uncertainty so that it would retain training effectiveness against of challenging and large-scale datasets, without losing satisfactory speed.

WSVM-FRS has introduced a novel weighted support vector machine to improve the noisy sensitivity problem of standard support vector machine for multiclass data classification. To keep basic idea, weighted coefficient has been added to the penalty term Lagrangian formula for optimization problem, which is called entropy degree, using lower and upper approximation for membership function in fuzzy rough set theory. Consequently, noisy

samples have low degree, and important samples have high degree. The performance of WSVM-FRS has been examined on 20 data sets taken from the UCI repository and real-world data sets so that the results have been compared to six other algorithms in recent literature. Experimental results demonstrate that the proposed method has good classification accuracy, precision and recall due to consideration and handling uncertainty aspect including noisy sample. The Wilcoxon test proves that the methods is more statistically different, in terms of appropriated performance, regarding accuracy, precision and recall metrics.

In the future, we not only would endeavor to enhance effectiveness of WSVM-FRS in dealing with data stream, but also would introduce more accurate kernel using ED concept when the method would face with more challenging real-world data sets. Furthermore, we are fascinated to seek other aspects of weighted SVM in order to reduce the noise more effectively and quickly.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with animals performed by any of the authors.

References

Alcantud JCR, Díaz S, Montes S (2019) Liberalism and dictatorship in the problem of fuzzy classification. Int J Approx Reason 110:82–95

Bian H, Mazlack L (2003) Fuzzy-rough nearest-neighbor classification approach. In: IEEE 22nd international conference of the North American Fuzzy Information Processing Society, pp 500–505

Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):27

Changdar C, Pal RK, Mahapatra GS (2016) A genetic ant colony optimization based algorithm for solid multiple travelling salesmen problem in fuzzy rough environment. Soft Comput 21:1–15

Chen D, Yang W, Li F (2008) Measures of general fuzzy rough sets on a probabilistic space. Inf Sci 178:3177–3187

Cohen S Afshar JT, van Schaik A (2017) Emnist: an extension of mnist to handwritten letters. arXiv preprint http://arxiv.org/abs/1702.05373

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Demsar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

Derrac J, Verbiest N, García S, Cornelis C, Herrera F (2013) On the use of evolutionary feature selection for improving fuzzy rough set based prototype selection. Soft Comput 17:223–238

Ding S, Zhang X, An Y, Xue Y (2017) Weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification. Pattern Recognit 67:32–46

Du W, Cao Z, Song T, Li Y, Liang Y (2017) A feature selection method based on multiple kernel learning with expression profiles of different types. BioData Min 10:1–16

Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. Int J Gen Syst 17:191–209

Hamidzadeh J, Ghadamyari R (2019) Clustering data stream with uncertainty using belief function theory and fading function. Soft Comput 24:127–138

Hamidzadeh J, Moradi M (2018) Improved one-class classification using filled function. Appl Intell 48:1–17

Hamidzadeh J, Moradi M (2020) Enhancing data analysis: uncertainty-resistance method for handling incomplete data. Appl Intell 50:74–86

Hamidzadeh J, Namaei N (2018) Belief-based chaotic algorithm for support vector data description. Soft Comput 23:1–26

Hamidzadeh J, Monsefi R, Yazdi HS (2014) LMIRA: large margin instance reduction algorithm. Neurocomputing 145:477–487

Hamidzadeh J, Sadeghi R, Namaei N (2017) Weighted support vector data description based on Chaotic bat algorithm. Appl Soft Comput 60:540–551

Han D, Liu W, Dezert J, Yang Y (2016) A novel approach to pre-extracting support vectors based on the theory of belief functions. Knowl Based Syst 110:210–223

Hsu CW, Lin CJ (2002) A comparison of methods for multi-class support vector machines. IEEE Trans Neural Netw 13:415–425

Javid M, Hamidzadeh J (2019) An active multi-class classification using privileged information and belief function. Int J Mach Learn Cybern 11:1–14

Karal O (2017) Maximum likelihood optimal and robust Support Vector Regression with lncosh loss function. Neural Netw 94:1–12

LeCun Y, Cortes C, Burges CJC (2010) The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist

Li J, Wang Y, Cao Y, Xu C (2016) Weighted doubly regularized support vector machine and its application to microarray classification with noise. Neurocomputing 173:595–605

Li J, Wang J, Zheng Y, Xiao H (2017) Microarray classification with noise via weighted adaptive elastic net. In: IEEE data driven control and learning systems, pp 26–27

Lichman M (2013) UCI machine learning repository. http://archive.ics.uci.edu/ml.

Liu Z, Dezert J, Pan Q, Mercier G (2011) Combination of sources of evidence with different discounting factors based on a new dissimilarity measure. Decis Support Syst 52:133–141

Liu Z, Pan Q, Dezert J, Mercier G (2015) Credal c-means clustering method based on belief functions. Knowl Based Syst 74:119–132

Liu Z, Pan Q, Dezert J, Martin A (2016) Adaptive imputation of missing values for incomplete pattern classification. Pattern Recognit 52:85–957

Lu X, Liu W, Zhou C, Huang M (2017) Probabilistic weighted support vector machine for robust modeling with application to hydraulic actuator. IEEE Trans Ind Inform 13(4):1723–1733

Ma J, Zhou S, Li. Chen, W. Wang, Z. Zhang, (2019) A sparse robust model for large scale multi-class classification based on K-SVCR. Pattern Recognit Lett 117:16–23

Mao WT, Xu JC, Wang C et al (2014) A fast and robust model selection algorithm for multi-input multi-output support vector machine. Neurocomputing 130:10–19

Moghaddam VH, Hamidzadeh J (2016) New Hermite orthogonal polynomial kernel and combined kernels in Support Vector Machine classifier. Pattern Recognit 60:921–935

Nguyen VL, Destercke S, Masson MH (2018) Partial data querying through racing algorithms. Int J Approx Reason 96:36–55

Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11:341–356

Santhanama V, Morariua VI, Harwooda D, Davisa LS (2016) A non-parametric approach to extending generic binary classifiers for multi-classification. Pattern Recogn 58:149–158

Shafer G (1976) A mathematical theory of evidence. Princeton University Press

Sheng H, Xiao J, Wang Z, Li F (2015) Electric vehicle state of charge estimation: nonlinear correlation and fuzzy support vector machine. J Power Sources 281:131–137

Singh S, Shreevastava S, Som T, Somani G (2020) A fuzzy similarity-based rough set approach for attribute selection in set-valued information systems. Soft Comput 24:4675–4691

Sivasankar E, Selvi C, Mahalakshmi S (2020) Rough set-based feature selection for credit risk prediction using weight-adjusted boosting ensemble method. Soft Comput 24:3975–3988

Sun W, Liu C, Xu Y, Tian L, Li W (2017) A band-weighted support vector machine method for hyperspectral imagery classification. IEEE Geosci Remote Sens Soc 14:1710–1714

Tang H, Dong P, Shi Y (2019) A new approach of integrating piecewise linear representation and weighted support vector machine for forecasting stock turning points. Appl Soft Comput 78:685–696

Vanir V (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10(5):988–999

Vapnik V (1995) The nature of statistical learning theory. IEEE Trans Neural Netw 10(5):988–999

Verbiest N, Cornelis C, Herrera F (2013a) FRPS: a fuzzy rough prototype selection method. Pattern Recognit 46:2770–2782

Verbiest N, Cornelis C, Herrera F (2013b) FRPS: a fuzzy rough prototype selection method. Pattern Recognit 46(10):2770–2782

Wu WZ, Leung Y, Zhang WX (2002) Connections between rough set theory and Dempster–Shafer theory of evidence. Int J Gen Syst 31:405–430

Xu Q, Zhang J, Jiang C, Huang X, He Y (2015) Weighted quantile regression via support vector machine. Expert Syst Appl 42:5441–5451

Xu P, Davoine F, Zha H, Denœuxa T (2016) Evidential calibration of binary SVM classifiers. Int J Approx Reason 72:55–70

Xue Y, Zhang L, Wang B, Zhang Z, Li F (2018) Nonlinear feature selection using Gaussian kernel SVM-RFE for fault diagnosis. Appl Intell 48:1–26

Yang L, Xu Z (2017) Feature extraction by PCA and diagnosis of breast tumors using SVM with DE-based parameter tuning. Int J Mach Learn Cyber 10:1–11

Yang X, Song Q, Cao A (2005) Weighted support vector machine for data classification. IEEE Int Joint Conf Neural Netw 2:859–864

Yang X, Tan L, He L (2014) A robust least squares support vector machine for regression and classification with noise. Neurocomputing 140:41–52

Yao Y, Lingras P (1998) Interpretations of belief functions in the theory of rough sets. Inf Sci 104:81–106

Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353

Zhai J (2011) Fuzzy decision tree based on fuzzy-rough technique. Soft Comput 15:1087–1096

Zhang W, Yu L, Yoshida T, Wang Q (2018) Feature weighted confidence to incorporate prior knowledge into support vector machines for classification. Knowl Inf Syst 58:1–27

Zhou C, Lu X, Huang M (2016) Dempster–Shafer theory-based robust least squares support vector machine for stochastic modelling. Neurocomputing 182:145–153

Zhu X, Wu X (2004) Class noise vs. attribute noise: a quantitative study of their impacts. Artif Intell Rev 22:177–210