symmetry

MDPI

Article

New Online Streaming Feature Selection Based on Neighborhood Rough Set for Medical Data

Dingfei Lei [1], Pei Liang [1], Junhua Hu [1] and Yuan Yuan [2,*]

[1] School of Business, Central South University, Changsha 410083, China; 181611128@csu.edu.cn (D.L.); shirley_lp@csu.edu.cn (P.L.); hujunhua@csu.edu.cn (J.H.)

[2] Third Xiangya Hospital, Central South University, Changsha 410013, China

* Correspondence: kinase200@csu.edu.cn

Received: 7 August 2020; Accepted: 22 September 2020; Published: 3 October 2020

Abstract: Not all features in many real-world applications, such as medical diagnosis and fraud detection, are available from the start. They are formed and individually flow over time. Online streaming feature selection (OSFS) has recently attracted much attention due to its ability to select the best feature subset with growing features. Rough set theory is widely used as an effective tool for feature selection, specifically the neighborhood rough set. However, the two main neighborhood relations, namely k-neighborhood and δ neighborhood, cannot efficiently deal with the uneven distribution of data. The traditional method of dependency calculation does not take into account the structure of neighborhood covering. In this study, a novel neighborhood relation combined with k-neighborhood and δ neighborhood relations is initially defined. Then, we propose a weighted dependency degree computation method considering the structure of the neighborhood relation. In addition, we propose a new OSFS approach named OSFS-KW considering the challenge of learning class imbalanced data. OSFS-KW has no adjustable parameters and pretraining requirements. The experimental results on 19 datasets demonstrate that OSFS-KW not only outperforms traditional methods but, also, exceeds the state-of-the-art OSFS approaches.

Keywords: online stream feature selection; class imbalanced data; neighborhood rough set; weighted dependency degree

1. Introduction

The number of features increases with the growth of data. A large feature space can provide much information that is useful for decision-making [1–3], but such a feature space includes many irrelevant or redundant features that are useless for a given concept. It is of necessity to remove the irrelevant features so that the curse of dimensionality can be relieved. This motivates some sort of research for feature selection methods. Feature selection, as a significant preprocessing step of data mining, can select a small subset, including the most significant and discriminative condition features [4]. Traditional methods are developed based on the assumption that all features are available. Many typical approaches exist, such as ReliefF [5], Fisher Score [6], mutual information (MI) [4], Laplacian Score [7], LASSO [8], and so on [9]. The main benefits of feature selection include speeding up the model training, avoiding overfitting, and reducing the impact of dimensionality during the process of data analysis [4].

However, features in many real-world applications are individually generated one-by-one over time. Traditional feature selection can no longer meet the required efficiency with the growing volume of features. For example, in the medical field, a doctor cannot easily obtain the entire features of a patient. In bioinformatic and clinical medicine situations, acquiring the entire features in a feature space is expensive and inefficient because of high-cost laboratory experiments [10]. In addition, for the task of medical image segmentation, acquiring the entire features is infeasible due to the infinite

number of filters [11]. Furthermore, the symptom of a patient persistently changes over time during the treatment, and judging whether the feature contains useful information is essential for identifying the patient's disease after a new feature has emerged [12]. In these cases, waiting a long time until entire features are available and then performing the feature selection process is the primary method.

Online streaming feature selection (OSFS), presenting a feasible precept to solve feature streaming in an online way, has recently attracted wide concern [13]. The OSFS method must meet the following three criteria [14]: (1) Not all features are available, (2) the efficient incremental updating process for selected features is essential, and (3) accuracy is vital each time.

Many previous studies have proposed some different OSFS methods. For example, a grafting algorithm [15], which employed a stagewise gradient descent approach to feature selection, during which a conjugate gradient procedure was used to carry out its parameters. However, as well as the grafting algorithm, both fast OSFS [16] and a scalable and accurate online approach (SAOLA) [13] need to specify some parameters, which requires the domain information in advance. Rough set (RS) theory [17], which is an effective mathematic tool for features selection, rules extracting, or knowledge acquisition [18], needs no domain knowledge other than the given datasets [19]. In the real world, we usually encounter many numerical features in datasets, such as medical datasets. Under this circumstance, a neighborhood rough set is feasible to analyze discrete and continuous data [20,21]. Nevertheless, all these methods proposed have some adjustable parameters. Considering that selecting unified and optimal values for all different datasets is unrealistic [22], a new OSFS method based on an adapted neighborhood rough set is proposed, in which the number of neighbors for each object is determined by its surrounding instance distribution [22]. Furthermore, in the view of multi-granulation, multi-granulation rough sets is used to compute the neighborhoods of each sample and extract neighborhood size [23]. For the above OSFS methods based on neighborhood relation, dependency degree calculation is a key step. However, very little work has considered the neighborhood structure in the granulation view during this calculation. In additional, the phenomenon of uneven distribution of some data, including medical data, is common, and few works focus on the challenge of the uneven distribution of data.

In this paper, focusing on the strength and weakness of the neighborhood rough set, we proposed a novel neighborhood relation. Further, a weighted dependency degree was developed by considering the neighborhood structure of each object. Finally, our approach, named OSFS-KW, was established. Our contributions were as follows:

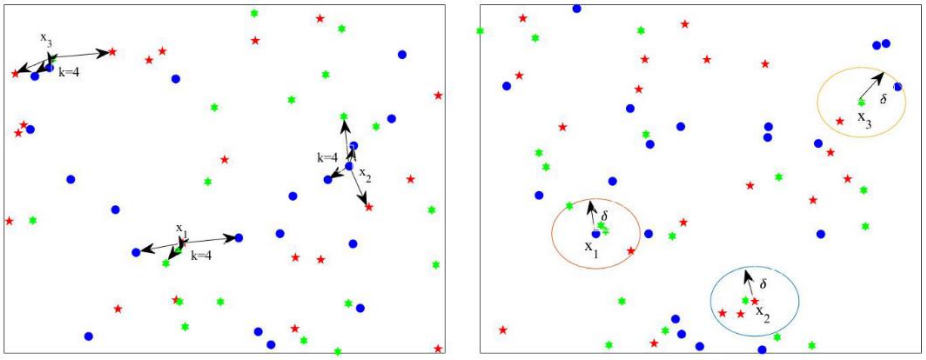
(1) We proposed a novel neighborhood relation, and on this basis, we developed a weighted dependency computation method.
(2) We developed an OSFS framework, named OSFS-KW, which can select a small subset made up of the most significant and discriminative features.
(3) The OSFS-KW was established based on [24] and can deal with the class imbalance problem.
(4) The results indicate that the OSFS-KW cannot only obtain better performance than traditional feature selection methods but, also, better than the state-of-the-art OSFS methods.

The remainder of the paper is organized as follows. In Section 2, we briefly review the main concepts of neighborhood RS theory. Section 3 discusses our new neighborhood relations and proposes the OSFS-KW. Then, Section 4 performs some experiments and discusses the experimental results. Finally, Section 5 concludes the paper.

2. Background

Neighborhood RS has been proposed to deal with numerical data or heterogeneous data. In general, a decision table (DT) for classification problem can be represented as $IS = < U, R, V, f >$ [25], where U is a nonempty finite set of samples. R can be divided condition attributes C and decision attributes D, $C \bigcap D = \varnothing$. $V = \bigcup \{V_r \mid r \in R\}$ is a set of attributes domains, such that V_r denotes the domains of an attribute r. For each $r \in R$ and $x \in U$, a mapping $f: U \times R \to V$ denotes an information function.

There are two main kinds of neighborhood relations: (1) the k-nearest neighborhood relation shown in Figure 1a and (2) the δ neighborhood relation shown in Figure 1b.



(a) k-nearest neighborhood (b) δ neighborhood

Figure 1. Two kinds of neighborhood relations.

Definition 1 [26]. *Given DT, a metric* $\triangle$ *is a distance function, and* $\triangle(x, y)$ *represents the distance between* x *and* y*. Then, for* $\forall x, y, z \in U$*, it must satisfy the following:*

(1) $\triangle(x, y) \geq 0$*, when* x=y *and* $\triangle(x, y)$=0*,*

(2) $\triangle(x, y)$=$\triangle(y, x)$*, and*

(3) $\triangle(x, z) \leq \triangle(x, y) + \triangle(y, z)$.

Definition 2 (δ neighborhood [22]). *Given DT, a feature subset* $B \subseteq C$*, the neighborhood* $\delta_B^r(x)$ *of any object* $x \subseteq U$ *is defined as follows:*

$$\delta_B^r(x) = \{y \mid \triangle_B(x, y) \leq r, y \in U, y \neq x\} \tag{1}$$

where r *is the distance radius, and* $\delta_B^r(x)$ *satisfies:*

(1) $y \in \delta_B^r(x) \Rightarrow x \in \delta_B^r(y)$*,*

(2) $1 \leq card(\delta_B^r(x)) \leq card(U)$*, where* $card(\bullet)$ *denotes the number of elements in the set, and*

(3) $\bigcup_{x \in U} \delta_B^r(x) = U$.

Definition 3 (*k*-nearest neighborhood [22]). *Given DT and* $B \subseteq C$*, the k-nearest neighborhood* $\mathbb{K}_B^k(x)$ *of any object* $x \subseteq U$ *on the feature subset* B *is defined as follows:*

$$\mathbb{K}_B^k(x) = \{y \mid y \in MIN_k\{\Delta_B(x, y)\}, y \in U, y \neq x\} \tag{2}$$

where $MIN_k\{\Delta_B(x, y)\}$ *represents the k neighbors closest to* x *on a subset B, and* $\mathbb{K}_B^k(x)$ *satisfies:*

(1) $\mathbb{K}_B^k(x) \neq \varnothing$*,*

(2) $card(\mathbb{K}_B^k(x)) = k$*, and*

(3) $\bigcup_{x \in U} \mathbb{K}_B^k(x) = U$.

Then, the concepts of the lower and upper approximations of these two neighborhood relations are defined as follows:

Definition 4. *Given DT, for any* $X \subseteq U$*, two subsets of objects, called the lower and upper approximations of* X *with regard to the* δ *neighborhood relation, are defined as follows [27]:*

$$\underline{B_\delta} X = \{x_i \mid \delta(x_i) \subseteq X, x_i \in U\} \tag{3}$$

$$\overline{B_\delta} X = \{x_i \mid \delta(x_i) \bigcap X \neq \varnothing, x_i \in U\} \tag{4}$$

If $x \in \underline{B_\delta} X$, *then* x *certainly belongs to* X, *but if* $x \in \overline{B_\delta} X$, *then it may or may not belong to* X.

Definition 5. *Given DT, for any* $X \subseteq U$, *the lower and upper approximations concerning the k-nearest neighborhood relation are defined as [24]:*

$$\underline{B_{\mathbb{k}}} X = \{x_i \mid \mathbb{k}(x_i) \subseteq X, x_i \in U\} \tag{5}$$

$$\overline{B_{\mathbb{k}}} X = \{x_i \mid \mathbb{k}(x_i) \bigcap X \neq \varnothing, x_i \in U\} \tag{6}$$

Figure 1a shows that the k-nearest neighbor (k = 4) samples of x_1, x_2, *and* x_3 *have different class labels. In detail, the k-nearest neighborhood samples of* x_1 *are from class* C_2 *with the mark "•" and class* C_3 *with the mark "✸"; k-nearest neighborhood samples of* x_2 *are from classes* C_2, C_3, *and* C_1 *with the mark "★"; the k-nearest neighbor samples of* x_3 *are from classes* C_1 *and* C_2. *Figure 1b depicts that all* δ *neighbor samples of* x_1, x_2, *and* x_3 *also come from different class labels. We define the samples of* x_1, x_2, *and* x_3 *as all the boundary objects. The size of the boundary area can increase the uncertain in DT, because it reflects the roughness of* X *in the approximate space.*

By Definition 5, The object space X can be partitioned into positive, boundary, and negative regions [28], which are defined as follows, respectively:

$$POS_B(X) = \underline{B}X \tag{7}$$

$$BOU_B(X) = \overline{B}X - \underline{B}X \tag{8}$$

$$NEG_B(X) = U - \overline{B}X \tag{9}$$

In the data analysis, computing dependencies between attributes is an important issue. We give the definition of the dependency degree as follows:

Definition 6. *Given DT, for any* $B \subseteq C$, *the dependency degree of* B *to decision attribute set* D *is defined as [22]*

$$\gamma_B(D) = \frac{CARD(POS_B(D))}{CARD(U)} \tag{10}$$

The aim of the feature selection is to select a subset B *from* C *and gain the maximal dependency degree of* B *to* D. *Since the features are available one-by-one over time, it is a necessity to measure each feature's importance in the candidate features.*

Definition 7. *Given DT, for* $B \subseteq C$ *and* D, *the significance of a feature* c $(c \in B)$ *to* B *is defined as follows [22]:*

$$\sigma_B(D, c) = \gamma_B(D) - \gamma_{B \setminus \{c\}}(D) \tag{11}$$

In a real application, specifically in the medical field, the instances are often unevenly distributed in the feature space; that is, the distribution around some example points is sparse, while the distribution around others is tight. Neither the *k*-nearest neighborhood relation nor the δ neighborhood relation can portray sample category information well, since the setting of the parameters like *r* and *k* can hardly meet both the sparse and tight distributions. For example, the feature space has two classes, as shown in Figure 2—namely, red and green. Red and green represent two different classes respectively, which have different symbols including pentacle and hexagon as well. Around a sample point x_1, the sample distribution is sparse. The three nearest points to x_1 all have different class from x_1. If applying the *k*-nearest neighborhood relation (*k* = 3), x_1 will be

misclassified. However, if we employ the δ neighborhood relation method, then the category of x_1 is consistent with that of most samples in the neighbors of x_1. On the other hand, the sample distribution around point x_2 is tight, and two class samples are included in its neighborhood, and sample x_2 will be misclassified when applying the δ neighborhood relation denoted by the red circle. In fact, if applying the *k*-nearest neighborhood relation ($k = 3$), x_2 will be classified correctly. Therefore, in Section 3, we proposed a novel neighborhood rough set combining the advantages of the *k*-nearest neighborhood rough set and the δ neighborhood rough set.

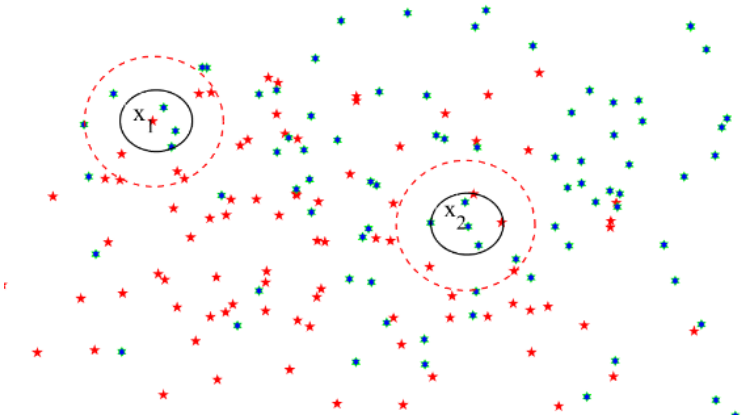


Figure 2. Distribution of the two class examples.

3. Method

In this section, we initially introduce a definition of OSFS. Then, we propose a new neighborhood relation and an approach of a weighted dependency degree. Based on three evaluation criteria—namely, maximal dependency, maximal relevance, and maximal significance—our new method OSFS-KW is presented finally.

3.1. Problem Statement

$DS_t = (U_t, A_t, V_t)$ denotes the decision system (DS) at time *t*, where $A_t = C_t \bigcup D_t$ is a feature space including the condition feature C_t and decision feature D_t. $U_t = \{x_1, x_2, ..., x_{n_t}\}$ is a nonempty finite set of objects. A new feature arrives individually, while the number of objects in U_t is fixed. OSFS aims to derive a mapping of $f_t^{'}: x_i \to D$ $(x_i \in U)$ at each timestamp *t*, which obtains an optimal subset of features available so far.

Contrary to the traditional feature selection methods, we cannot access the full feature space in the scenarios of the OSFS. However, the two main neighborhood relations cannot make up the shortage caused by the uneven distribution data. Moreover, the class imbalanced issue of medical data is common. For example, abnormal cases attract more attention than the normal ones in the field of medical diagnosis. It is also crucial for the proposed framework to handle the class imbalanced problem.

3.2. Our New Neighborhood Relation

3.2.1. $k-\delta$ Neighborhood Relation

The standard European distance method is applied to eliminate the effect of variance on the distance among the samples. Given any samples x_i and x_j, and an attribute subset $c \in C$, the distance between x_i and x_j is defined as follows:

$$\Delta_B(x_i, x_j) = \sqrt{\sum_{c \in B} \left(\frac{f(x_i, c) - f(x_j, c)}{\sigma_c}\right)^2} \tag{12}$$

where $f(x_i, c)$ denotes the values of x_i relative to the attribute *c*, and σ_c represents standard deviation of attribute *c*.

To overcome the challenge of the uneven distribution of medical data, we proposed a novel neighborhood rough set as follows.

Definition 8. *Given a decision system* $DS = \{U, C, D, f\}$, *where* $U = \{x_1, x_2, x_3, ..., x_n\}$ *is the finite sample set,* $B \in C$ *is a condition attribute subset, and* D *is the decision attribute set, the* $k-\delta$ *neighborhood relation is defined as follows:*

$$\kappa_B(x_i) = \{x_j \in U \mid x_j \in \delta_B^r(x_i) \bigcap \mathbb{k}_B^k(x_i)\} \tag{13}$$

where $\delta_B^r(x_i)$ *is defined as Definition 2, and* $\mathbb{k}_B^k(x)$ *is defined as Definition 3.* $k = 0.15 * n$, *where* n *is the number of instances.*

Meanwhile, based on [22], $r = 1.5 * G_{mean}$ *and* G_{mean} *are defined as follows*:

$$G_{mean} = \frac{G_{\max} - G_{\min}}{n-1} \tag{14}$$

More specifically, $G_{\max}$ *represents the maximum distance from* x_i *to its neighbors, and* $G_{\min}$ *denotes the minimum distance from* x_i *to its neighbors.*

Definition 9. *Given* $DS = \{U, C, D, f\}$ *and its neighborhood relations* R *on* U, *for* $X \subseteq U$, *the lower and supper approximation regions of* X *in terms of the* $k-\delta$ *neighborhood relation are defined as follows:*

$$\underline{B_\kappa X} = \{x_i \mid \kappa_B(x_i) \subseteq X, x_i \in U\} \tag{15}$$

$$\overline{B_\kappa X} = \{x_i \mid \kappa_B(x_i) \bigcap X \neq \varnothing, x_i \in U\} \tag{16}$$

Similar to Definition 4, $\underline{B_\kappa X}$ *is also called the positive region, denoted by* $POS_B^\kappa(D)$.

3.2.2. Weighted Dependency Computation

The traditional dependency degree only considers the samples correctly classified instead of that of neighborhood covering. To solve this problem, we propose an approach of weighted dependency degree, which considers the granular information for features.

Definition 10. *Given* $DS = \{U, C, D, f\}$, *the weighted dependency of* $B \in C$ *is defined as follows:*

$$\tau_B^\kappa(D) = w(B) \bullet \gamma_B^\kappa(D) \tag{17}$$

where

$$w(B) = \log_2(2 - \varphi(B)) \tag{18}$$

$$\gamma_B^\kappa(D) = \frac{CARD(POS_B^\kappa(D))}{CARD(U)} \tag{19}$$

$$\varphi(B) = \sum_{i=1}^{|U|} \frac{CARD(\kappa_B(x_i))}{CARD(U)^2} \tag{20}$$

Theorem 1. *Given* $DS = \{U, C, D, f\}$, $P \subseteq O \subseteq U$. *The weight* $w(\bullet)$ *is monotonic, defined as follows:*

$$w(P) \leq w(O) \tag{21}$$

Proof. According to the monotonicity of neighborhood relation defined in Definition 8, $\kappa_O(x_i) \subseteq \kappa_P(x_i)$ and $\therefore CARD(\kappa_O(x_i)) \le CARD(\kappa_P(x_i))$. According to Equation (20), $\varphi(O) \le \varphi(P)$. Considering Definition 10, $0 < \varphi(O) \le 1$, $0 < \varphi(P) \le 1$, and $\therefore 0 \le w(P) = \log_2(2-\varphi(P)) \le w(O) = \log_2(2-\varphi(O)) < 1$. □

Theorem 2. *Given* $DS = \{U, C, D, f\}$, $B \subseteq C$. *Then,* $\tau_B^\kappa(D) = \tau_C^\kappa(D)$ *is equivalent to* $\gamma_B^\kappa(D) = \gamma_C^\kappa(D)$, *and* $\log_2(2-\varphi(P)) = \log_2(2-\varphi(O))$.

The proof of Theorem 2 is easy according to the monotonicity of the neighborhood relation and Theorem 1.

Definition 11. *Given* $B \subseteq C$ *and a decision attribute set* D, *the significance of feature* $c\,(c \in B)$ *to* B *can be rewritten as follows:*

$$\sigma_B(D,c) = \tau_B^\kappa(D) - \tau_{B\setminus\{c\}}^\kappa(D) \tag{22}$$

3.3. Three Evaluation Criteria

During the OSFS process, many irrelevant and redundant features should be removed for high-dimensional datasets. There are three evaluation criteria used during the process, such as max-dependency, max-relevance, and max-significance.

3.3.1. Max-Dependency

$C = \{c_1, c_2, ..., c_m\}$ denotes the set of *m* condition attributes. The task of OSFS is to find a feature subset $B \subseteq C$, which has the maximal dependency $\mathbb{D}$ on the decision attributes set D. At the moment, the number of features denoted as *d* in the feature space should be as small as possible.

$$\max \mathbb{D}(B,D), \quad \mathbb{D} = \tau_B(D) \tag{23}$$

where $\tau_B(D)$ denotes the weighted dependency between the attribute subset *B* and target class label D. The dependency $\mathbb{D}$ can be rewritten as $\tau_{B_t}(D)$, where $B_t = \{B_{t-1}, c_t\}$. Hence, the increment search algorithm optimizes the following problem for selecting the tth feature from the attribute set $\{C - B_{t-1}\}$:

$$\max_{c_i \in \{C-B_{t-1}\}} \{\tau_{\{B_{t-1},c_i\}}(D)\} \tag{24}$$

which is also equivalent to optimizing the following problem:

$$\max_{c_i \in \{C-B_{t-1}\}} \{\tau_{\{B_{t-1},c_i\}}(D) - \tau_{\{B_{t-1}\}}(D)\} = \max_{c_i \in \{C-B_{t-1}\}} \{\sigma_{B_t}(D,c_i)\} \tag{25}$$

The max-dependency maximizes either the joint dependency between the select feature subset and the decision attribute or the significance of the candidate feature to the already-selected features. However, the high-dimensional space has two limitations that lead to failure in generating the resultant equivalent classes: (1) the number of samples is often insufficient, and (2) during the multivariate density estimation process, computing the inverse of the high-dimensional covariance matric is generally an ill-posed problem [29]. Specifically, these problems are evident for continuous feature variables in real-life applications, such as in the medical field. In addition, the computational speed of max-dependency is slow. Meanwhile, max-dependency is inappropriate for OSFS, because each timestamp can only know one feature instead of the entire feature space in advance.

3.3.2. Max-Relevance

Max-relevance is introduced as an alternative in selecting features, as implementing max-dependency is hard. The max-relevance search feature approximates $\mathbb{D}(B,D)$ in Equation (23) with the mean value of all dependency values between individual feature B_i and the decision attribute D.

$$Max\ \mathbb{R}(B,D), \quad \mathbb{R}=\frac{1}{|B|}\sum_{c_i\in B}\tau_{c_i}(D) \tag{26}$$

where B is the already-selected feature subsets.

A rich redundancy likely exists among the features selected according to max-relevance. For example, if two features c_i and c_j among the large features space highly depend on each other, then after removing any one of them, the class differentiation ability of the other one would not substantially change. Therefore, the following max-significance criterion is added to solve the redundancy problem by selecting mutually exclusive features.

3.3.3. Max-Significance

Based on Equation (22), the importance of each candidate feature can be calculated. The max-significance can select mutually exclusive features as follows:

$$Max\{\mathbb{S}\}\quad \mathbb{S}=\frac{1}{|B|}\sum_{c_i\in B}\{\sigma_B(D,c_i)\} \tag{27}$$

The feature flows individually over time for the OSFS. Testing all combinations of the candidate features and maximizing the dependency of the selected feature set are not appropriate. However, we can initially employ the "max-relevance" criteria to remove the irrelevant features. Then, we employ the "max-significance" criteria to remove the unimportant features in the selected feature set. Finally, the "max-dependency" criteria will be used to select the feature set with the maximal dependency. Based on the three criteria mentioned previously, in the next subsection, a novel online feature selection framework will be proposed.

3.4. OSFS-KW Framework

The proposed weighted dependency computation method based on the $k-\delta$ neighborhood RS in this study is shown in Algorithm 1. First, we calculate the *card* value of each sample x_i and obtain the sum for the final weighted dependency at steps 5–14. The $card(\bullet)\in[0,1]$ denotes the consistency between the decision attribute of x_i and its neighbor's decision attributes. The $k-\delta$ neighborhood relation is used to calculate the dependency of attribute subset B. The value of $\tau_B^{\kappa}(D)$ reveals not only the distribution of labels nearby x_i but, also, the structure granular structure information around x_i.

In the real world, we generally encounter the issue of high-dimension class imbalance, specifically in medical diagnosis. Then, we employ the method proposed in [24], named the class imbalance function, as shown in Algorithm 2. For imbalanced medical data, we apply Algorithm 2 to compute $card(\kappa_B(x_i))$ at step 9 in Algorithm 1.

Algorithm 1 Weighted dependency computation

Require:

B: The target attribute subset;

X_B: Sample values on B;

Ensure:

1: $\tau_B^{\kappa}(D)$: Dependency between B and decision attribute D;

2: $card(B)$: the number of positive samples on B, initial $0\rightarrow card(B)$;

3: $|U|$: the number of instances in universe U;

4: S_B: the number of instance of neighbors on B, initial $0\rightarrow S_B$

5: for each x_i in X_B

6: Calculate the distance from x_i to other instances;

7: Sort the neighbors of x_i from the nearest to the farthest;

8: Find the neighbor sample of x_i as $\kappa_B(x_i)$;
9: Calculate the card value of x_i as $card(\kappa_B(x_i))$;
10: $card(B) = card(B) + card(\kappa_B(x_i))$;
11: Calculate the number of neighbors $\kappa_B(x_i)$ with the same class label of x_i as S_{x_i};
12: $S_B = S_B + S_{x_i}$;
13: end
14: $\tau_B^{\kappa}(D) = \log_2(2 - S_B/|U|^2)) * card(B)/|U|$;
15: return $\tau_B^{\kappa}(D)$

In Algorithm 2, D_{large} denotes the large class, while D_{small} is the small class. The D_{large} sample is different from the D_{small} sample at steps 3–11. For x_i in the large class, if the number of neighbors with the same class label is more than 95% of the number of its total neighbors, then we will set the value of $card(\kappa_B(x_i))$ to 1; otherwise, the value is set to 0. For x_i in the small class, we calculate the ratio of the number of neighbors with D_{small} to the total number of neighbors as the $card(\kappa_B(x_i))$. The method in Algorithm 2 can strengthen the consistency constraints of D_{large} and weaken the consistency constraints of D_{small}, so D_{small} is prevented from being overpowered by the samples in D_{large}.

Algorithm 2 Class imbalance function

Require:
D_{x_i}: The class label of x_i;
B: The target attribute subset;
Ensure:
$card(\kappa_B(x_i))$: the card value of x_i on B;
1: N_B: the number of neighbors $\kappa_B(x_i)$ with the same class label of x_i;
2: N_R: the number of neighbors of x_i on B;
3: if $D_{x_i} == D_{large}$ then
4: if $N_B(x_i) >= 0.95 * N_R$ then
5: $card(\kappa_B(x_i)) = 1$;
6: else
7: $card(\kappa_B(x_i)) = 0$;
8: end
9: else then
10: $card(\kappa_B(x_i)) = N_B / N_R$;
11: end
12: return $card(\kappa_B(x_i))$;

Based on the $k-\delta$ neighborhood relation and the weighted dependency computation method mentioned above, we introduce our novel OSFS method, named "OSFS-KW", as shown in Algorithm 3. The main aim of the OSFS-KW is to maximize $\mathbb{D}(B, D)$ with the minimal number of feature subsets.

Algorithm 3 OSFS-KW

Require:
C: the condition attribute set;
D: the decision attribute;
Ensure:
B: the selected attribute set
1: B initialized to $\varnothing$;
2: $\tau_B^{\kappa}(D)$: the dependency between *B* and *D*, initialized to 0;

3: $Mean_{\tau_B^\kappa(D)}$: the mean dependency of attributes in *B*, initialized to 0;
4: Repeat
5: Get a new attribute c_i of *C* at timestamp *i*;
6: Calculated the dependency of c_i as $\tau_{c_i}^\kappa(D)$ according to Algorithm 1;
7: if $\tau_{c_i}^\kappa(D) < Mean_{\tau_B^\kappa(D)}$ then
8: Discard attribute c_i ; and go to Step 25;
9: end
10: if $\tau_{B\cup c_i}^\kappa(D) > \tau_B^\kappa(D)$ then
11: $B = B \cup c_i$;
12: $\tau_B^\kappa(D) = \tau_{B\cup c_i}^\kappa(D)$;
13: $Mean_{\tau_B^\kappa(D)} = \frac{\sum_{c_i \in B} \tau_{c_i}^\kappa(D)}{card(B)}$;
14: else if $\tau_{B\cup c_i}^\kappa(D) == \tau_B^\kappa(D)$ then
15: $B = B \cup c_i$;
16: random the feature order in B;
17: for each attribute c_j in B
18: calculate the significance of c_j as $\sigma_B(D, c_j)$;
19: if $\sigma_B(D, c_j) == 0$
20: $B = B - \{c_i\}$;
21: $Mean_{\tau_B^\kappa(D)} = \frac{1}{card(B)} \sum_{c_i \in B} \tau_B^\kappa(D)$;
22: end
23: end
24: end
25: Until no attributes are available;
26: return B;

Specifically, in Algorithm 1, we calculate the dependency of B_i when a new attribute c_i arrives at timestamp *i*. Then, the dependency of c_i is compared with the mean dependency of the selected attribute subset B at step 7. If $\gamma_{c_i} > Mean_{\mathbb{D}(B,D)}$, then c_i is added into *B* and goes to step 10. Otherwise, c_i is discarded and goes to step 25 when $\gamma_{c_i} < Mean_{\mathbb{D}(B,D)}$ due to the "max-relevance" constraint.

When c_i satisfies the "max-relevance" constraint, $\gamma_{c_i} > Mean_{\mathbb{D}(B,D)}$, going to step 10 and comparing the dependency of the current attribute subset *B* with $B \cup c_i$. If $\tau_{B\cup c_i}^\kappa(D) > \tau_B^\kappa(D)$, then adding attribute c_i into *B* will increase the dependency of *B*, so c_i is added into *B* with the "max-dependency" constraint; that is, $B = B \cup c_i$. On the other hand, if $\tau_{B\cup c_i}^\kappa(D) = \tau_B^\kappa(D)$, then some redundant attributes exist in $B \cup c_i$. In this condition, we add c_i into *B* firstly. Then, we remove some redundant attributes by steps 16–24. With the "max-significance" constraint, we randomly select an attribute from *B* and compute its significance according to Equation (22). Some attributes with a significance equal to 0 will be removed from *B*. Ultimately, we can obtain the best feature subset for decision-making through the aforementioned three evaluation constraints.

3.5. Time Complexity of OFS-KW

In the process of OSFS-KW, the weighted dependency degree computation, shown in Algorithm 1, is a substantially important step. The number of examples in *DS* is *n*, and the number of attributes *C* is *m*. Table 1 shows the time complexity for different steps of OSFS-KW. In Algorithm 1, we compute the distance between x_i and its neighbors for each sample $x_i \in U$. The time complexity of this process is $O(n)$ ($n = card(U)$). Sorting all neighbors of x_i by instance is essential to find the

neighbors of x_i. The time complex of the quick sorting process is $O(nlogn)$. Thus, the time complexity of Algorithm 1 is $O(n^2logn)$.

Table 1. Time complexity of the online streaming feature selection (OSFS)-KW framework.

Description	Algorithm	Line	Complexity
Compute the distance	1	6	$O(n)$
Sort all the neighbors	1	7	$O(nlogn)$
Repeat loop	1	5–13	$O(n^2logn)$
Compare the dependency	3	10	$O(n^2logn)$
Compare the dependency	3	14–24	$O(card(B)n^2logn)$

At timestamp *i*, as a new attribute c_i is present to the OSFS-KW, the time complexity of steps 6–9 is $O(n^2logn)$. If the dependency of c_i is smaller than $Mean_{\tau_B^\kappa(D)}$, then c_i will be discarded. Otherwise, comparing the dependency of $B \bigcup c_i$ with *B*, the time complexity is also $O(n^2logn)$. If $\tau_{B\bigcup c_i}^\kappa(D) > \tau_B^\kappa(D)$, then c_i can be added into *B*, and step 25 is repeated. However, if $\tau_{B\bigcup c_i}^\kappa(D) = \tau_B^\kappa(D)$, then the time complexity of steps 14–24 is $O(card(B) \bullet n^2logn)$. Thus, the complexity of the OSFS-KW is $O(m^2n^2logn)$. Choosing all features in real-world datasets is impossible. Therefore, the time complexity will be smaller than $O(m^2n^2logn)$.

4. Experiments

4.1. Data and Preprocessing

We use a high-dimensional medical dataset as our test bench to compare the performance of the proposed OSFS-KW with the existing streaming feature selection algorithm. Table 2 summarizes the 19 high-dimensional datasets used in our experiments.

In Table 2, the BREAST CANCER and OVARIAN CANCER datasets are biomedical datasets [30]. LYMPHOMA and SIDO0 datasets are from the WCCI 2008 Performance Prediction Challenges [31]. MADELON and ARCENE are from the NIPS 2003 feature selection challenge [16]. WDBC, HILL, HILL (NOISE), and COLON TUMOR are four UCI datasets, the web can be accessed at https://archive.ics.uci.edu/ml/index.php. And DLBCL, CAR, LUNG-STD, GLIOMA, LEU, LUNG, MLL, PROSTATE, and SRBCT are nine microarray datasets [32,33].

Table 2. Nineteen experimental datasets.

Dataset	Instances	Features	Classes
WDBC	569	30	2
HILL	606	100	2
HILL (NOISE)	606	100	2
COLON TUMOR	60	2000	2
DLBCL	77	6285	2
CAR	174	9182	11
LYMPHOMA	62	4026	3
LUNG-STD	181	5000	2
GLIOMA	50	4433	4
LEU	72	7129	3
LUNG	203	3312	2
MLL	72	5848	3
PROSTATE	102	6033	2
SRBCT	83	2308	4
ARCENE	200	10,000	2

MADELON	500	2600	2
BREAST CANCER	286	17,816	2
OVARIAN CANCER	253	15,154	2
SIDO0	500	999	2

In our experiments, we employ K-nearest neighbor (KNN), support vector machines (SVM), and random forest (RF) as the basic classifiers to evaluate a selected feature subset. The radial basis function is used in SVM, and the Gini coefficient is used to comprehensively measure all variables' importance in RF. Furthermore, a grid search cross-validation is applied to train and optimize these three classifiers to give the best prediction results. Then, search ranges of some adjustable parameters for each basic classifier are shown in Table 3.

Table 3. The search ranges of all adjustable parameters. RF: random forest.

Classifiers	Parameters	Search Ranges
KNN	The number of neighbors	(3,12)
SVM	The kernel coefficient σ for the radial basis function	(0.1,3)
	penalty parameter	(0.1,3)
RF	The number of trees in the forest	(2,15)

As listed below, there are three key metrics employed to evaluate the OSFS-KW with other streaming feature selection methods.

(1) Compactness: the number of selected features,
(2) Time: the running time of each algorithm,
(3) Prediction accuracy: the percentage of the correctly classified test samples.

The results are collected in the MATLAB 2017b platform with Windows 10, Intel(R) Core (TM)i5-8265U,1.8GHz CPU, and 8GB memory. In addition, we applied the Friedman test at a 95% significance level under the null hypothesis to validate whether the OSFS-KW and its rivals have a significant difference in the prediction accuracy, compactness, and running time [34]. Then, accepting the null hypothesis means that the performance of the OSFS-KW has no significant difference with its rivals. However, if the null hypothesis is rejected, then conducting follow-up inspections is necessary. If so, we employed the Nemenyi test [35], with which the performances of the two methods were significantly different if their corresponding average rankings (AR) were greater than the value of the critical difference (CD).

4.2. Experiments and Dicussions

4.2.1. OSFS-KW versus k-Nearest Neighborhood

In this section, we compare OSFS-KW with the *k*-nearest neighborhood relation. We employ the same algorithm framework for both neighborhood relations to reduce the impact of other factors. In addition, for the *k*-nearest neighborhood relation, the value of *k* varies from 3 to 13 in the experiments.

The experiment results are shown in Appendix A. Tables A1 and A2 show the compactness and running time. The *p*-values of the Friedman test are 5.07×10^{-9} and 5.47×10^{-10}, respectively. In addition, Tables A3–A5 show the experimental results about the prediction accuracy on these datasets. The *p*-values on KNN, SVM, and RF are 0.6949, 0.9884, and 0.5388, respectively. Table A6 shows the test results of the OFS-KW versus *k*-nearest neighborhood. Therefore, a significant difference exists among these 19 datasets on compactness and running time. On the contrary, no significant difference is observed on accuracy with KNN, SVM, and RF. According to the Nemenyi test, the value of the CD is 3.8215, and we have the following observations from Tables A1–A5.

In terms of compactness, a significant difference is just observed between OSFS-KW and *k*-nearest neighborhood when $k = 10$, 11, 12, and 13, but OSFS-KW selects the smallest average number of features. According to the running time, there is a significant difference between OSFS-KW and *k*-

nearest neighborhood when k = 3, 4, 5, 6, 7, and 8. In general, the k-nearest neighborhood is faster than OSFS-KW, mainly because the number of neighbors for OSFS-KW is uncertain but fixed for the k-nearest neighborhood. According to the value of AR and CD, there is no significant difference between the OSFS-KW and k-nearest neighborhood, with three basic classifiers' prediction accuracy for the value of k from 3 to 13. On some datasets, such as COLON, TUMOR, DLBCL, CAR, LYMPHOMA, and LUNG-STD, if a proper k is chosen, the k-nearest neighborhood would have a higher prediction accuracy than the OSFS-KW with KNN, SVM, and RF. This finding means that k-nearest neighborhood can perform well with the proper parameter k.

4.2.2. OSFS-KW versus δ Neighborhood

In this section, the OSFS-KW is compared with the δ neighborhood relation. We employ the algorithm framework for both neighborhood relations for equality. In addition, we employ $\delta = r \times D_{\max}$ and conduct experiments with values of r from 0.1 to 0.5 with step 0.05. The experiment results can be seen in Appendix B. Table A7 shows the compactness of different methods on 19 datasets, and the p-value of the Friedman test is 2.92×10^{-15}. Table A8 shows the running time on these datasets, and the p-value is 3.65×10^{-10}. Tables A9–A11 show the results of the prediction accuracy of the OSFS-KW versus δ neighborhood. The p-values on KNN, SVM, and RF are 0.0275, 0.7815, and 0.6683, respectively, shown in Table A12. There is a significant difference among the different algorithms on compactness, running time, and prediction accuracy using KNN, but no significant difference exits in the prediction accuracy of SVM and RF. In addition, the value of CD is 3.1049.

On the number of selected features shown in Table A7, a significant difference is observed between the OSFS-KW and δ neighborhood when r = 0.4, 0.45, and 0.5. Our proposed method OSFS-KW selects the smallest mean number of features. The number of selected features using δ neighborhood increases with the r value increasing. In terms of the running time shown in Table A8, a significant difference exists between the OSFS-KW and δ neighborhood when r = 0.3~0.5, and no significant difference is found when r = 0.1~0.3. The OSFS-KW has the smallest mean running time. On the average ranks shown in Tables A9–A11, for the value of CD, no significant difference is observed with KNN when r = 0.2 and 0.25 and RF when r = 0.1, 0.15, 0.2, 0.25, 0.35, 0.4, and 0.5. Particularly, no significant difference exists with RF under any value of r. On the prediction accuracy, the OSFS-KW has the highest mean of the prediction accuracy δ neighborhood than among these datasets. However, the δ neighborhood can also obtain the highest prediction accuracy with different r values on some datasets, such as DLBCL and LUNG-STD. However, it is impossible for the δ neighborhood relation to uniform the parameters on all different kinds of datasets.

4.2.3. Influence of Feature Stream Order

In this section, we carry out the experiments on the OSFS-KW with three types of feature steam orders, including original, inverse, and random. Figure 3 depicts the results of the compactness of the OSFS-KW on the datasets. Figures 4–6 show the prediction accuracy about KNN, SVM, and RF, respectively.

In addition, we execute the Friedman test at a 95% significance level under the null hypothesis to verify whether there is a significant difference in the compactness, running time, and predictive accuracy. Table A12 in Appendix C shows the calculated p-values. Moreover, it is clear that there is no significant difference, except for the running time, with random order and prediction accuracy, using KNN with the random order. The number of features in the feature space has a remarkable impact on the running time between the original and random orders, specifically when the number of features is very large. For example, the number of features of ARCENE is 10,000, and the difference of the running time on the dataset between the original and random is 157.2334 s.

Figures 3–6 show minor fluctuations in some datasets. However, these three orders have no significant difference with each other on most of the datasets. This result denotes that the feature stream orders have a limited impact on the OSFS-KW.

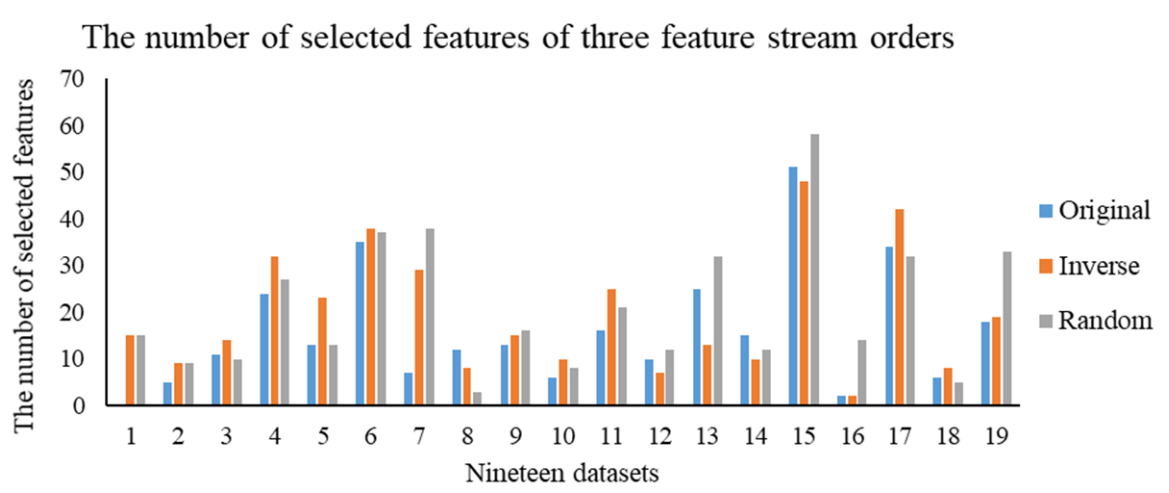


Figure 3. Compactness of the three feature stream orders.

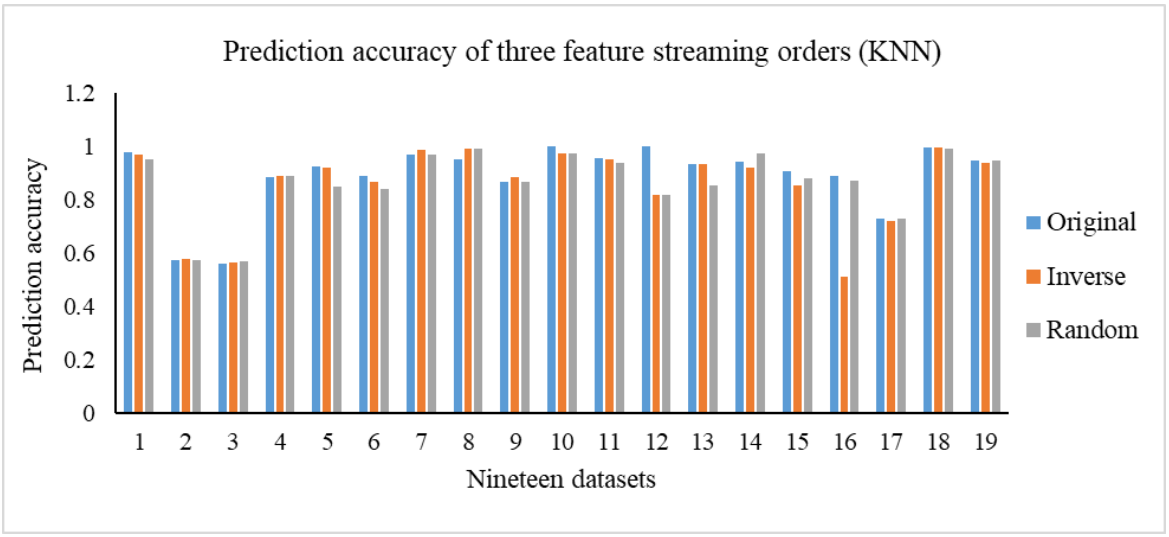


Figure 4. Prediction accuracy of the three feature streaming orders (KNN).

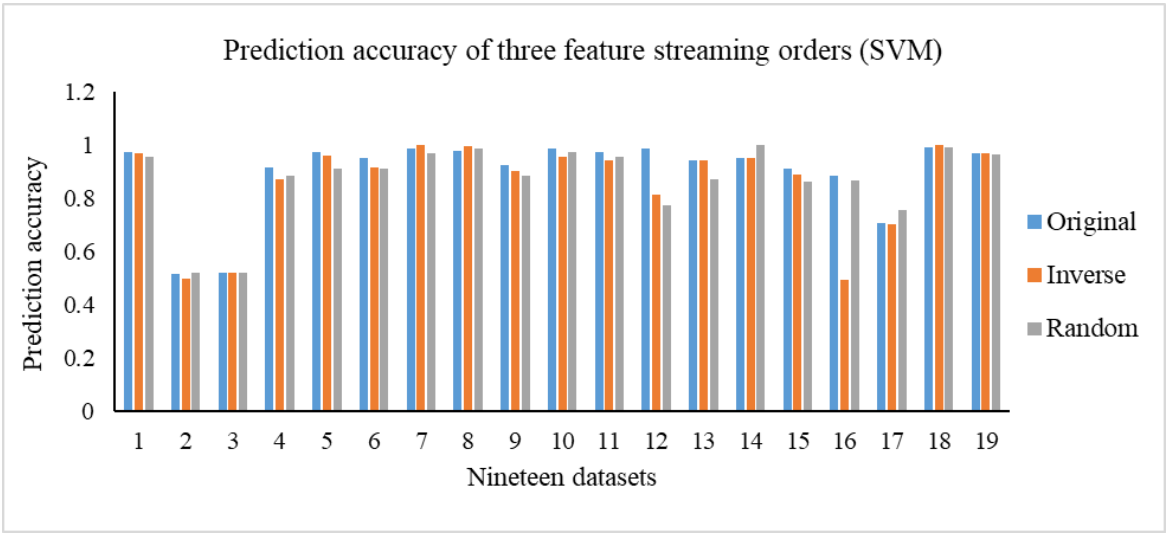


Figure 5. Prediction accuracy of the three different feature streaming orders (SVM).

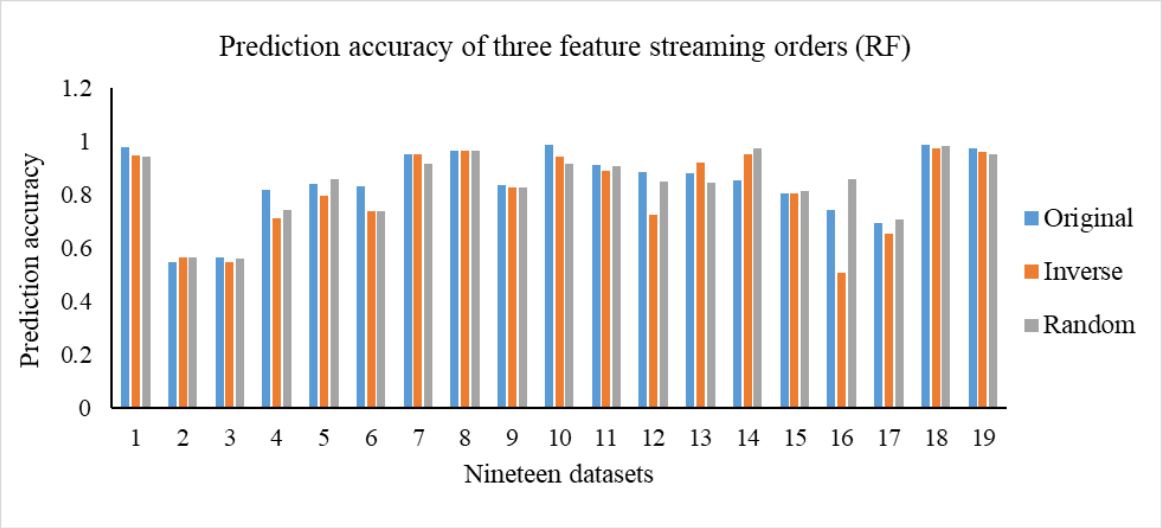


Figure 6. Prediction accuracy of the three different feature streaming orders (RF).

4.2.4. OSFS-KW versus Traditional Feature Selection Methods

In this section, 11 representative traditional feature selection methods are employed to compare with OSFS-KW, including Fisher [36], spectral feature selection [37], Pearson correlation coefficient (PCC) [38], ReliefF [39], Laplacian Score [7], a unsupervised feature selection method with ordinal locality (UFSOL) [40], mutual information (MI) [41], the infinite latent feature selection method (ILFS) [42], lasso regression (Lasso) [43], a fast correlation-based filter method (FCBF) [44], and a correlation based feature selection approach (CFS) [45].

We implement all these algorithms in MATLAB. The k value of ReliefF is set to 7 for the best performance. We rank all features and select the same number of features as the OSFS-KW, considering that all these 11 traditional feature selection methods cannot be applied to the scenario of an OSFS. In addition, we employ three methods as basic classifiers—namely, KNN, SVM, and RF. The results of the prediction accuracy of the three classifiers with five-fold validation are used to evaluate the OSFS-KW and all competing ones.

The experiment results are shown in Appendix D. Tables A14–A16 show the prediction accuracy of the three basic classifiers. The p-values on the accuracy with KNN, SVM, and RF are 1.20×10^{-15}, 3.81×10^{-12}, and 5.99×10^{-13}, respectively. Table A17 shows the test results. Thus, a significant difference is observed between OSFS-KW and the compared algorithms on the prediction accuracy with the three classifiers. According to the value of CD, which is 3.8215, we can observe the following results from Tables A14–A16.

(1) OSFS-KW versus Fisher. According to the values of AR and CD, no significant difference is found between these two methods on the prediction accuracy at a 95% significance level. However, OSFS-KW has a better performance than Fisher on most datasets with the three classifiers.
(2) OSFS-KW versus SPEC. A significant difference exists between these two algorithms on the prediction accuracy with KNN, SVM, and RF. Furthermore, OSFS-KW outperforms SPEC on most of the datasets. On the whole, SPEC cannot handle some datasets well.
(3) OSFS-KW versus PCC. A significant difference is found between OSFS-KW and PCC on the prediction accuracy with KNN and RF but not with SVM. On many datasets, OSFS-KW outperforms PCC.
(4) OSFS-KW versus ReliefF. No significant difference is observed between OSFS-KW and ReliefF on the accuracy with KNN, SVM, and RF. The performance of ReliefF decreases with fewer data, as ReliefF cannot be distinguished among redundant features.
(5) OSFS-KW versus MI. No significant difference is observed between these two algorithms with KNN, SVM, and RF.

(6) OSFS-KW versus Laplacian. A Laplacian score is an unsupervised feature selection algorithm using no class information for each instance. This computes the power of locality, preserving a feature to evaluate the importance of the corresponding feature. A significant difference is found between these two algorithms with KNN, SVM, and RF. OSFS-KW outperforms the Laplacian score on all datasets.
(7) OSFS-KW versus UFSOL. UFSOL is an unsupervised feature selection method that can preserve the topology information, named the ordinal locality. A significant difference is observed between these two algorithms on the prediction accuracy with KNN, SVM, and RF.
(8) OSFS-KW versus ILFS. ILFS is a feature selection algorithm based on a probabilistic latent graph. A significant difference exists between these two algorithms on the prediction accuracy. OSFS-KW outperforms ILFS on most datasets. ILFS has a prediction accuracy of approximately 0.6439, 0.6677, and 0.7152 on dataset OVARIAN CANCER with KNN, SVM, and RF, respectively. By contrast, OSFS-KW has a prediction accuracy of 0.996, 0.9923, and 0.9881. OSFS-KW has a higher ILFS of over 30% on the prediction accuracy among the datasets.
(9) OSFS-KW versus Lasso. Lasso is a regularization method for linear regression and has a weight coefficient of each feature. No significant difference is observed between these two algorithms. On datasets GLIOMA and MADELON, the prediction accuracy of OSFS-KW is more than that of Lasso of nearly 22% on with KNN, SVM, and RF.
(10) OSFS-KW versus FCBF. FCBF, addressing explicitly the correlation between features, ranks the features according to their MI, with the class to be predicted. No significant difference is found between OSFS-KW and FCBF on the prediction accuracy with KNN, SVM, and RF.
(11) OSFS-KW versus CFS. CFS is a correlation-based feature selection method. No significant difference exists between OSFS-KW and CFS. OSFS-KW outperforms CFS on most of the 19 datasets. On some datasets, such as MADELON and ARCENE, the prediction accuracy of CFS is lower than that of OSFS-KW for nearly 30%.

Overall, OSFS-KW not only performs best among the 19 datasets but, also, has the highest average prediction accuracy among KNN, SVM, and RF.

4.2.5. OSFS-KW versus OSFS Methods

In this section, our algorithm is compared to five state-of-the-art OSFS methods—namely, OSFS-A3M [22], OSFS [16], Alpha-investing [46], Fast-OSFS [47], and SAOLA [13].

We implement all aforementioned algorithms in MATLAB [48], and the significant level α is set to 0.05 for the above five algorithms. The threshold a and the wealth w of Alpha-investing are set to 0.5. As shown in Appendix E, Tables A18 and A19 show the compactness and running time of OSFS-KW against the other five algorithms. The p-values of the Friedman test on these three classifiers are 0.0248 and 3.62×10^{-23}. Tables A20–A22 summarize the prediction accuracy on these 19 datasets using the KNN, SVM, and RF classifiers with p-values of 0.0337, 0.0032, and 0.0533, respectively. Table A23 shows the test results. A significant difference is found between the six algorithms on the number of selected features, running time, and the prediction accuracy using KNN and SVM, but no significant difference is observed using RF. According to the value of CD, which is 1.7296, we can observe the following results from Tables A18–A21.

(1) In terms of compactness, no significant difference is observed between OSFS-KW and the other competing algorithms. Fast-OSFS has the smallest mean number of selected features. In addition, for SNAOLA, the number of selected features on some datasets is remarkably large but on some other datasets is zero. This finding demonstrates that SNAOLA cannot handle some types of datasets well.
(2) On the running time, Alpha-investing is the fastest algorithm among all these six algorithms and has the smallest mean running time among these datasets. According to the values of AR and CD, a significant difference exists among OSFS-KW, Alpha-investing, Fast-OSFS, and SAOLA. The difference between OSFS-KW and OFS-A3M on the running time is small.

(3) According to the prediction accuracy, OSFS-KW has the highest mean prediction accuracy on these datasets using all three classifiers. OSFS-KW outperforms the five competing algorithms. No significant difference is observed between OSFS-KW and the other competing methods, except for SAOLA.

In summary, although our method, OSFS-KW, is slower than some competing methods, including Fast-OSFS and SAOLA, OSFS-KW is superior among the six methods in prediction accuracy of the 19 datasets.

5. Conclusions

Most of the exiting OSFS methods cannot deal well with the problem of uneven distribution data. In this study, we defined a new $k-\sigma$ neighborhood relation, combining the advantages of k-neighborhood relation and δ neighborhood relation. Then, we proposed a weighted dependency degree considering the structure of neighborhood covering. Finally, we proposed a new OSFS framework named OSFS-KW, which need not specify any parameters in advance. In addition, this method can also handle the problem of imbalance classes in medical datasets. With three evaluation criteria, this approach can select the optimal feature subset mapping decision attributes. Finally, we used KNN, SVM, and RF as the basic classifiers in conducting the experiments to validate the effectiveness of our method. The results of the Friedman test indicate that a significant difference exists between the OSFS-KW and other neighborhood relations on compactness and running time, but there was no significant difference on the predictive accuracy. Moreover, when comparing with the 11 traditional feature selection methods and five existing OSFS algorithms, the performance of OFS-KW is better than that of the traditional feature selection methods and outperforms that of the state-of-the-art OSFS. However, we only focused on the challenges of medical data and used only medical datasets to verify the validity of our approach. Virtually, our method can be applied into other similar fields, generally. In the future, we will test and evaluate this method using some multidisciplinary datasets.

Author Contributions: Conceptualization, D.L. and P.L.; methodology, D.L.; software, D.L. and P.L.; validation, J.H. and Y.Y.; formal analysis, D.L., P.L. and J.H.; investigation, Y.Y.; resources, Y.Y.; data curation, D.L. and P.L.; writing—original draft preparation, D.L. and Y.Y.; writing—review and editing, P.L. and J.H.; visualization, D.L.; supervision, J.H.; project administration, D.L., P.L., J.H. and Y.Y.; and funding acquisition, J.H. and Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported partially by the Hunan Provincial Natural Science Foundation of China under grant number 2017JJ3472 and the National Natural Science Foundation of China under grant number 71871229.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Results of OSFS-KW versus *k*-Nearest Neighborhood

It is noteworthy that some values in bold are minimum gained by different techniques when analyzing one same data set.

Table A1. OSFS-KW versus *k*-nearest neighborhood (compactness).

Data Set	OSFS-KW	*k* = 3	*k* = 4	*k* = 5	*k* = 6	*k* = 7	*k* = 8	*k* = 9	*k* = 10	*k* = 11	*k* = 12	*k* = 13
WDBC	**18**	23	22	21	24	19	**18**	19	21	22	22	20
HILL	5	6	2	5	2	**1**	**1**	3	5	3	2	6
HILL (NOISE)	11	3	3	3	9	9	15	6	15	12	4	**1**
COLON TUMOR	24	**22**	26	27	33	39	26	38	39	27	33	42
DLBCL	**13**	23	30	32	40	45	43	60	47	71	56	58
CAR	**35**	96	96	96	96	96	96	96	96	96	96	96
LYMPHOMA	**7**	34	34	46	52	47	48	52	70	59	58	61
LUNG-STD	**12**	26	32	42	33	35	58	43	58	77	73	67
GLIOMA	13	14	16	14	9	10	12	8	7	9	**4**	5
LEU	**6**	29	44	60	55	61	54	66	71	107	72	80

LUNG	**16**	62	65	84	91	110	89	115	130	140	137	151
MLL	**10**	25	23	26	20	38	30	28	36	33	33	46
PROSTATE	**25**	32	31	42	42	46	44	50	45	45	71	46
SRBCT	15	10	8	8	10	10	10	9	9	8	7	**6**
ARCENE	51	52	**48**	92	86	97	88	108	94	119	140	132
MADELON	2	2	**1**	12	12	12	12	11	10	14	15	15
BREAST CANCER	34	**1**	**1**	**1**	**1**	**1**	**1**	39	2	78	36	61
OVARIAN CANCER	**6**	57	55	81	115	85	99	128	147	156	148	156
SIDO0	18	**1**	**1**	4	4	**1**	3	4	3	12	32	19
AVERAGE	**16.89**	27.26	28.32	36.63	38.63	40.11	39.32	46.47	47.63	57.26	54.68	56.21
RANKS	**3.89**	4.47	3.76	5.63	6.11	6.53	5.87	7.37	7.87	9.11	8.39	9.00

Table A2. OSFS-KW versus *k*-nearest neighborhood (running time).

Data Set	OSFS-KW	*k* = 3	*k* = 4	*k* = 5	*k* = 6	*k* = 7	*k* = 8	*k* = 9	*k* = 10	*k* = 11	*k* = 12	*k* = 13
WDBC	1.7934	1.1039	1.0673	1.0273	**1.0180**	1.7006	1.7663	1.9128	1.9020	1.8395	1.7437	1.6871
HILL	**5.3830**	5.6782	5.3893	5.9408	6.7386	6.2822	5.6835	5.9713	5.4591	5.9174	5.4391	5.6756
HILL (NOISE)	3.9342	3.4305	3.3936	**3.3310**	3.3427	3.3414	3.3715	3.4386	3.6614	3.6624	3.6308	3.6760
COLON TUMOR	2.0676	**0.95566**	1.0415	1.0433	1.0339	1.0621	1.0534	1.1351	1.1091	1.1115	1.2167	1.1767
DLBCL	10.5498	**4.3684**	4.3931	4.4388	4.7461	4.8728	4.5478	4.7116	4.6009	5.2460	5.1264	4.7404
CAR	200.6726	51.869	53.304	35.3994	**23.8263**	40.2583	35.0849	29.174	33.1931	37.7835	33.357	33.2482
LYMPHOMA	12.4278	5.335	5.262	5.1206	5.2007	**4.8678**	5.2963	5.3348	6.2701	5.6055	5.8868	6.1197
LUNG-STD	55.9968	**36.434**	48.2749	39.9477	45.0579	43.1706	47.4229	45.085	43.5729	51.4130	49.915	48.3410
GLIOMA	4.8191	1.3312	1.3229	1.2728	1.2866	1.2600	1.3111	**1.2588**	1.2640	1.3388	1.2852	1.3367
LEU	4.6364	**2.2495**	2.3131	2.4016	2.3933	2.4948	2.4615	2.5496	3.1112	3.4017	3.2112	3.4583
LUNG	31.9804	**24.4997**	30.035	31.5634	32.7240	34.004	32.1781	34.825	34.7344	35.0237	34.948	31.5197
MLL	11.2642	**4.3727**	5.7569	5.2412	5.2053	4.5691	4.4251	5.2610	5.11078	5.1546	5.1195	5.1029
PROSTATE	24.2364	9.6686	9.0227	10.4484	8.14070	7.9714	7.9191	8.7141	7.8963	7.8965	8.0517	**7.8299**
SRBCT	3.4475	0.9448	**0.8983**	1.0084	1.0722	0.9449	0.9182	0.9020	1.1923	1.2072	1.2102	1.2988
ARCENE	87.9552	**36.4807**	40.048	43.0383	40.5699	44.3189	43.0740	46.626	45.8766	47.3027	48.935	49.1771
MADELON	**538.4690**	563.6911	563.0505	586.0047	9994.3255	562.1598	651.1277	953.0622	1002.0125	1021.341	920.2213	1009.9971
BREAST CANCER	114.8965	110.914	110.857	110.859	**110.336**	111.137	110.530	112.239	111.258	115.893	112.236	112.990
OVARIAN CANCER	80.8324	**70.041**	70.308	72.135	75.095	72.296	73.678	76.297	78.253	78.408	78.025	78.547
SIDO0	64.2822	60.6454	62.642	65.3041	65.3604	63.8099	65.8703	65.270	63.0324	66.0955	63.219	**50.2334**
AVERAGE	66.30	**52.32**	53.60	53.98	548.81	53.19	57.77	73.88	76.50	78.72	72.78	76.64
RANKS	10.00	**3.95**	4.42	4.89	5.53	5.26	5.37	7.16	6.58	9.47	7.79	7.58

Table A3. Predictive accuracy of OSFS-KW versus *k*-nearest neighborhood (KNN).

Data Set	OSFS-KW	*k* = 3	*k* = 4	*k* = 5	*k* = 6	*k* = 7	*k* = 8	*k* = 9	*k* = 10	*k* = 11	*k* = 12	*k* = 13
WDBC	**0.9779**	0.9719	0.9685	0.9544	0.9754	0.9636	0.9737	0.9702	0.9718	0.9719	0.9720	0.9720
HILL	0.5710	0.5644	0.5578	0.5627	**0.5727**	0.5494	0.5494	0.5628	0.56444	0.5628	0.5396	0.5478
HILL (NOISE)	0.5594	0.5675	0.5624	0.5659	0.5495	0.5444	0.5610	0.5527	0.5609	0.5610	**0.5724**	0.5380
COLON TUMOR	0.8833	0.9346	0.9026	**0.9679**	0.9205	0.9013	0.8718	0.8859	0.8705	0.9167	0.8872	0.8859
DLBCL	0.9223	0.9464	0.9875	0.975	**1.0**	0.9875	0.9875	1.0	0.975	1.0	0.9875	0.975
CAR	0.8899	0.8969	0.8744	0.8951	0.9022	0.8854	0.9028	0.8925	**0.9091**	0.8963	0.9024	0.8838
LYMPHOMA	0.970	0.9675	0.9857	0.9818	**1**	0.9818	**1**	**1**	**1**	**1**	**1**	**1**
LUNG-STD	0.9503	0.9944	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
GLIOMA	0.8655	0.9005	0.8873	0.8623	0.8441	**0.9055**	**0.9055**	0.8805	0.8623	0.8441	0.8623	0.8641
LEU	**1**	0.9867	0.9724	0.9857	0.9857	0.9857	0.9857	0.9857	0.9857	**1**	0.9857	0.9857
LUNG	0.9555	0.9606	0.9549	**0.9751**	0.9506	0.9646	0.97	0.9506	0.9647	0.9699	0.9506	0.9597
MLL	**1**	0.9846	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
PROSTATE	0.9314	0.9214	0.951	0.9514	0.951	0.9514	**0.961**	0.931	0.9605	0.9605	0.9605	0.9605
SRBCT	0.9404	0.9882	0.9778	**1**	**1**	0.9889	0.9889	0.9889	0.9889	0.9778	0.9778	0.9408
ARCENE	0.9056	0.8602	0.9004	0.9094	0.8897	**0.9301**	0.8957	0.9006	0.8996	0.9151	0.9104	0.9101
MADELON	0.8885	0.5481	0.5412	0.8958	0.8958	0.8958	0.8958	0.8908	**0.9115**	0.8738	0.8777	0.88
BREAST CANCER	0.7307	0.6685	0.7062	0.6817	0.7064	0.6817	0.7064	0.7167	0.7238	0.7195	0.7302	**0.7656**
OVARIAN CANCER	0.996	**1**	0.996	0.996	1	1	0.996	0.996	0.996	1	1	0.996
SIDO0	0.948	0.972	0.5922	0.968	0.968	0.964	0.97	0.968	0.97	0.9601	0.99	**0.994**
AVERAGE	0.89	0.88	0.86	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**
RANKS	7.55	7.24	8.24	6.34	5.76	6.63	**5.42**	6.92	5.87	5.45	5.71	6.87

Table A4. Predictive accuracy of OSFS-KW versus *k*-nearest neighborhood (SVM).

Data Set	OSFS-KW	*k* = 3	*k* = 4	*k* = 5	*k* = 6	*k* = 7	*k* = 8	*k* = 9	*k* = 10	*k* = 11	*k* = 12	*k* = 13
WDBC	0.9754	0.9701	0.9754	**0.9789**	0.9737	0.9543	0.9632	0.9772	0.9753	0.9754	0.9754	0.9736
HILL	**0.5181**	0.5065	0.5165	0.5131	0.5132	0.5098	0.5098	0.5131	0.5065	0.5099	0.5066	0.5082
HILL (NOISE)	0.5231	0.5132	0.5000	0.5099	0.5248	0.5182	**0.5314**	0.5248	0.5264	0.5231	0.5082	0.5049
COLON TUMOR	**0.9179**	0.8705	0.8692	0.8859	0.9051	0.8705	0.8705	0.8705	0.8692	0.8692	0.8692	0.8846
DLBCL	0.9732	0.9357	0.9775	0.975	0.9875	0.9875	0.975	0.975	0.975	**1.0**	0.975	0.9625
CAR	**0.9534**	0.9229	0.9121	0.9165	0.9224	0.9166	0.9172	0.9118	0.9184	0.9336	0.9398	0.9103
LYMPHOMA	0.9858	0.9532	0.9675	0.9818	**1**	0.9818	**1**	**1**	**1**	0.9818	**1**	**1**
LUNG-STD	0.9778	**1**	0.989	0.9944	**1**	**1**	**1**	0.9944	**1**	0.9944	**1**	**1**
GLIOMA	**0.9255**	0.8986	0.9055	0.8605	0.8041	0.8623	0.8986	0.8623	0.8241	0.7877	0.8805	0.8823
LEU	**0.9857**	0.9857	0.9857	0.9857	0.9857	0.9714	0.9857	0.9857	0.9857	0.9857	0.9857	0.9857
LUNG	0.9751	0.9701	0.9547	**0.9797**	0.9649	0.97	0.9753	0.9454	0.9651	0.9606	0.9454	0.9504
MLL	0.9867	0.9846	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	0.9846	**1.0**	**1.0**	**1.0**
PROSTATE	0.9414	0.941	**0.951**	0.941	0.941	0.951	0.941	0.921	0.951	0.951	0.941	0.951
SRBCT	0.9519	0.9882	0.9889	**1**	0.9889	1	0.9889	0.9889	0.9889	0.9889	0.9778	0.9402
ARCENE	0.9102	0.8954	0.8901	0.9196	0.9102	0.9451	0.9152	0.9203	0.94	**0.9501**	0.94	0.8992
MADELON	**0.8872**	0.5546	0.5565	0.8769	0.8769	0.8769	0.8769	0.8712	0.8815	0.8688	0.875	0.8785
BREAST CANCER	0.7091	0.6749	0.6749	0.6749	0.6749	0.6749	0.6749	0.7514	0.6677	0.7307	**0.7547**	0.7447
OVARIAN CANCER	0.992	**1.0**	0.996	**1.0**	**1.0**	**1.0**	0.996	0.996	**1.0**	0.996	**1.0**	**1.0**
SIDO0	0.972	0.924	0.5715	0.966	0.966	0.882	0.97	0.972	0.97	0.97	**0.99**	**0.99**
AVERAGE	**0.8980**	0.8679	0.8517	0.8926	0.8915	0.8880	0.8942	0.8937	0.8910	0.8935	**0.8980**	0.8930
RANKS	**5.7368**	8.2368	7.8158	6.0789	5.7895	6.3947	6.0526	6.5000	6.3684	6.2368	6.1579	6.6316

Table A5. Predictive accuracy of OSFS-KW versus *k*-nearest neighborhood (RF).

Data Set	OSFS-KW	*k* = 3	*k* = 4	*k* = 5	*k* = 6	*k* = 7	*k* = 8	*k* = 9	*k* = 10	*k* = 11	*k* = 12	*k* = 13
WDBC	**0.9779**	0.9527	0.9544	0.9614	0.9631	0.9719	0.9737	0.9545	0.9545	0.9386	0.9614	0.9544
HILL	0.5461	0.5693	0.5412	0.5577	0.5495	0.5049	0.5099	0.5363	**0.5727**	0.5511	0.5643	0.5528
HILL (NOISE)	0.5642	0.5512	0.5577	0.5595	0.5478	0.5280	**0.5840**	0.5510	0.5278	0.5643	0.5642	0.5017
COLON TUMOR	0.8205	0.7705	0.8705	0.8538	0.7769	0.8218	0.8192	0.8692	0.8231	0.8705	0.8372	**0.8859**

DLBCL	0.8430	0.8723	0.9232	0.9232	0.9357	0.9232	0.9098	0.8973	0.9375	**0.9482**	0.9366	0.9375
CAR	0.8327	0.8	0.8110	**0.8544**	0.7920	0.8062	0.8193	0.8347	0.8226	0.8102	0.8083	0.7905
LYMPHOMA	0.9510	0.9498	0.8088	0.8554	0.9714	0.8556	0.9199	0.9093	**0.9818**	0.8915	0.939	0.9041
LUNG-STD	0.9668	0.9668	0.9614	0.9779	0.9778	0.9557	0.9616	0.9724	0.9889	0.9778	**0.989**	0.9557
GLIOMA	0.8377	0.8291	**0.9073**	0.7877	0.8059	0.8423	0.7659	0.7877	0.7809	0.7714	0.8805	0.8241
LEU	**0.9857**	0.9438	0.9029	0.9286	0.9295	0.9029	0.8752	0.9429	0.9029	0.9162	0.9162	0.9571
LUNG	0.9119	0.8912	0.9201	0.9022	0.9363	0.9347	0.9307	0.9402	0.9407	**0.9505**	0.9358	0.9157
MLL	0.8862	0.96	0.9016	0.9416	0.8729	**0.9733**	0.9416	0.9437	0.9703	0.9026	0.9446	0.9457
PROSTATE	0.8819	0.901	0.901	0.9219	0.8829	0.9119	0.9314	0.9205	0.9019	0.9314	0.8829	**0.941**
SRBCT	0.8539	0.9522	0.9757	**0.9875**	0.9521	0.9515	0.9757	0.9757	0.9764	0.9875	0.9764	0.9513
ARCENE	0.805	0.8258	**0.8438**	0.8397	0.7859	0.8302	0.7801	0.8158	0.8349	0.8298	0.8404	0.7953
MADELON	0.7436	0.5142	0.5188	0.8635	0.8619	0.8712	0.8685	0.8677	**0.8731**	0.8538	0.8546	0.8569
BREAST CANCER	0.6927	0.6475	0.6862	0.6459	0.6543	0.661	0.6682	0.6472	0.6919	0.64	0.6717	**0.7031**
OVARIAN CANCER	0.9881	**1.0**	0.9722	0.9841	0.9881	0.9962	0.9962	0.9762	0.996	0.9921	1	0.996
SIDO0	0.974	**0.99**	0.5645	0.952	0.96	0.9499	0.944	0.96	0.956	0.9659	0.982	0.982
AVERAGE	0.8454	0.8362	0.8170	0.8578	0.8497	0.8522	0.8513	0.8580	**0.8649**	0.8575	0.8676	0.8606
RANKS	6.8684	7.0789	7.6579	6.2105	7.4211	7.0526	7.1316	6.6579	**4.8421**	6.0789	4.6579	6.3421

Table A6. Test results of OFS-KW versus *k*-nearest neighborhood.

Evaluation Criteria	Friedman Test
Compactness	5.07×10^{-9}
Running time	5.47×10^{-10}
Accuracy (KNN)	0.6949
Accuracy (SVM)	0.9884
Accuracy (RF)	0.5388

Appendix B. The Results of OSFS-KW versus δ Neighborhood

Table A7. OFS-KW versus δ neighborhood (compactness).

Data Set	OFS-KW	$r = 0.1$	$r = 0.15$	$r = 0.2$	$r = 0.25$	$r = 0.3$	$r = 0.35$	$r = 0.4$	$r = 0.45$	$r = 0.5$
WDBC	18	16	15	16	16	15	14	14	**10**	**10**
HILL	**5**	18	9	6	8	13	5	7	10	13
HILL (NOISE)	11	18	18	11	8	17	**7**	11	8	18
COLON TUMOR	24	17	**9**	24	38	47	70	82	81	72
DLBCL	13	12	**10**	13	14	19	17	14	16	35
CAR	35	**32**	32	37	37	38	32	36	41	47
LYMPHOMA	7	**5**	6	6	9	6	8	8	13	11
LUNG-STD	12	6	**4**	7	9	8	8	9	14	19
GLIOMA	13	14	**13**	10	16	17	15	20	43	37
LEU	6	6	**5**	12	10	9	10	13	12	16
LUNG	16	20	**13**	23	25	18	26	23	33	39
MLL	10	11	11	11	9	**8**	11	12	16	17
PROSTATE	25	**14**	27	24	20	19	106	54	142	170
SRBCT	15	**8**	14	11	15	17	18	20	22	24
ARCENE	51	**30**	30	33	45	56	63	59	138	139
MADELON	**2**	8	13	39	35	40	51	61	56	55
BREAST CANCER	**34**	61	42	50	49	46	79	55	84	81
OVARIAN CANCER	6	6	8	7	**5**	9	11	119	241	364
SIDO0	**18**	25	22	20	68	96	89	98	107	166
AVERAGE	**16.89**	17.21	15.84	18.95	22.95	26.21	33.68	37.63	57.21	70.16
RANKS	**3.71**	3.74	3.26	4.21	4.89	5.39	5.74	6.87	8.13	9.05

Table A8. OFS-KW versus δ neighborhood (running time).

Data Set	OFS-KW	$r = 0.1$	$r = 0.15$	$r = 0.2$	$r = 0.25$	$r = 0.3$	$r = 0.35$	$r = 0.4$	$r = 0.45$	$r = 0.5$
WDBC	1.7934	**1.7735**	1.9180	2.1549	2.0573	2.1659	1.9439	1.9439	2.2182	2.1838
HILL	**5.3830**	7.5729	7.1106	5.6935	7.1577	7.2613	6.1592	10.3283	9.7808	7.6833
HILL (NOISE)	**3.9342**	4.7900	4.8339	4.8170	4.6466	4.8516	4.7399	4.8934	4.9030	5.0067
COLON TUMOR	2.0676	3.7879	6.1031	3.4523	3.4633	**1.8328**	2.0735	1.8638	1.9334	1.8536
DLBCL	10.5498	9.4904	9.2736	10.9896	**8.9530**	12.0197	11.6707	14.1592	13.4430	11.0618
CAR	200.6726	82.0880	**74.8928**	100.9478	134.9345	170.0618	232.9729	230.3482	231.4008	263.2982
LYMPHOMA	**12.4278**	15.3990	15.5832	15.8409	18.8001	22.42501	21.2379	23.1505	17.3421	17.3496
LUNG-STD	**55.9968**	94.1136	110.2884	112.2191	140.1124	128.4688	153.3367	171.3579	184.0983	168.2461
GLIOMA	**4.8191**	5.7425	8.7011	8.2908	7.3862	8.8651	9.7550	10.0932	10.3550	7.9124
LEU	**4.6364**	5.1840	6.5944	7.0113	7.1038	8.1651	8.4208	10.2700	11.8882	13.2705
LUNG	**31.9804**	74.043	68.184	48.724	61.818	83.154	113.622	85.821	105.910	122.239
MLL	**11.2642**	18.4100	19.1743	16.91520	22.4664	24.5996	29.9188	25.8334	31.0735	22.0068
PROSTATE	24.2364	32.8402	25.7984	29.9719	39.5195	40.4784	21.4567	53.0141	**16.6909**	23.4978
SRBCT	**3.4475**	3.6517	6.46288	8.9606	8.8096	10.6264	11.5697	13.7862	14.1859	15.2989
ARCENE	**87.9552**	119.1581	101.0211	107.7780	167.190	191.7981	316.8370	226.6963	143.9225	393.4693
MADELON	538.4690	1218.5929	1056.9129	1040.2067	**487.8704**	551.4037	581.3661	666.5058	692.4043	520.1171
BREAST CANCER	**114.8965**	133.9355	126.5506	130.3485	130.3591	126.8821	139.4627	128.0552	131.3735	131.5025
OVARIAN CANCER	80.8324	100.1605	82.18792	82.43967	83.01642	**78.17236**	97.85803	118.5893	107.6156	101.5973
SIDO0	**64.2822**	118.2797	122.3051	130.5592	86.0615	106.5379	92.5953	74.2628	73.6424	87.7044
AVERAGE	**66.30**	107.84	97.57	98.28	74.83	83.15	97.74	98.47	94.96	100.81
RANKS	**2.00**	4.68	4.32	4.79	4.74	5.89	6.61	7.50	7.42	7.05

Table A9. Predictive accuracy of OFS-KW versus δ neighborhood (KNN).

Data Set	OFS-KW	$r = 0.1$	$r = 0.15$	$r = 0.2$	$r = 0.25$	$r = 0.3$	$r = 0.35$	$r = 0.4$	$r = 0.45$	$r = 0.5$
WDBC	**0.9779**	0.9702	0.9719	0.9684	0.9525	0.9386	0.9667	0.9701	0.9701	0.9701
HILL	**0.5710**	0.5545	0.5528	0.5511	0.5544	0.5627	0.5610	0.5543	0.5593	0.5692
HILL (NOISE)	**0.5594**	0.5575	0.5475	0.5462	0.5281	0.5461	0.5380	0.5545	0.5561	0.5430
COLON TUMOR	0.8833	0.8051	0.8385	0.8231	**0.9038**	0.8397	0.8872	0.841	0.8564	0.8718
DLBCL	0.9223	0.9232	0.9116	0.8875	0.8447	0.9241	0.9233	**0.95**	0.95	0.95
CAR	**0.8899**	0.8763	0.8821	0.8825	0.8204	0.8726	0.8422	0.8509	0.852	0.8535
LYMPHOMA	0.970	**1.0**	**1.0**	0.9526	0.9526	0.9526	**1.0**	**1.0**	0.9818	**1.0**
LUNG-STD	0.9503	0.9946	**1.0**	0.9889	**1.0**	0.9944	**1.0**	0.9889	**1.0**	0.9944
GLIOMA	0.8655	**0.9186**	0.8805	0.9055	0.7295	0.7941	0.8586	0.8673	0.8441	0.8241
LEU	**1.0**	0.9581	0.9714	0.9857	0.9581	0.9867	0.9724	0.9724	0.9857	0.9581
LUNG	0.9555	0.95	0.9502	0.9107	0.9407	0.9607	0.9351	**0.9749**	0.9452	0.9702
MLL	**1.0**	0.9713	0.9446	0.9292	0.9733	0.9467	0.9284	**1.0**	0.9724	0.9857
PROSTATE	0.9314	0.9219	0.7948	0.8238	0.8929	**0.9519**	0.8724	0.9319	0.941	0.9314
SRBCT	0.9404	0.8895	0.8784	0.9162	0.9187	0.9646	0.9403	0.9638	**0.966**	0.9653
ARCENE	**0.9056**	0.8354	0.8708	0.8348	0.8747	0.8804	0.8449	0.8453	0.8448	0.8841
MADELON	**0.8885**	0.5192	0.5265	0.7754	0.5977	0.6177	0.7219	0.71	0.7285	0.5992
BREAST CANCER	**0.7307**	0.7026	0.6818	0.7026	0.6849	0.6856	0.6852	0.727	0.6879	0.7026
OVARIAN CANCER	0.996	**1.0**	**1.0**	0.996	0.996	**1.0**	**1.0**	**1.0**	0.996	0.992
SIDO0	0.948	0.96	0.96	**0.988**	0.9421	0.962	0.9361	0.926	0.916	0.9201
AVERAGE	**0.89**	0.86	0.85	0.86	0.85	0.86	0.86	0.88	0.87	0.87
RANKS	**3.55**	5.55	6.00	6.66	7.26	5.13	6.11	4.61	5.03	5.11

Table A10. Predictive accuracy of OFS-KW versus δ neighborhood (SVM).

Data Set	OFS-KW	$r = 0.1$	$r = 0.15$	$r = 0.2$	$r = 0.25$	$r = 0.3$	$r = 0.35$	$r = 0.4$	$r = 0.45$	$r = 0.5$
WDBC	0.9754	0.9772	**0.9790**	0.9789	0.9771	0.9754	0.9755	0.9737	0.9737	0.9737
HILL	0.5181	**0.5214**	0.4983	0.5066	0.5132	0.5148	0.5049	0.5099	0.5198	0.5000
HILL (NOISE)	0.5231	0.5298	0.5281	0.5231	**0.5314**	0.5248	0.5265	0.5231	0.5116	0.5198
COLON TUMOR	**0.9179**	0.8385	0.8231	0.8551	0.9013	0.8705	0.8538	0.8692	0.8538	0.8692
DLBCL	0.9732	0.9233	0.95	0.9125	0.8857	0.925	**0.9875**	0.95	0.9875	0.9625
CAR	**0.9534**	0.9212	0.9238	0.9098	0.8947	0.8961	0.9136	0.8966	**0.9417**	0.9346
LYMPHOMA	0.9858	0.9857	**1.0**	0.9359	0.956	0.9532	0.969	**1.0**	0.9857	**1.0**
LUNG-STD	0.9778	0.9946	**1.0**	**1.0**	**1.0**	0.9944	**1.0**	**1.0**	**1.0**	**1.0**
GLIOMA	0.9255	0.8823	0.8623	0.9205	**0.9273**	0.8623	0.8736	0.8641	0.8241	0.8605
LEU	0.9857	0.9581	**0.9867**	0.9571	0.9714	0.9867	0.9724	0.9724	0.9714	0.9295
LUNG	0.9751	0.9549	**0.99**	0.9555	0.9406	0.9502	0.95	0.9755	0.9603	0.9504
MLL	0.9867	0.9579	0.9579	0.9579	0.9579	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
PROSTATE	0.9414	0.9505	0.8538	0.8824	0.9214	**0.9514**	0.9214	0.9314	0.941	0.941
SRBCT	0.9519	0.9161	0.9757	0.975	0.9535	0.9889	0.927	**1.0**	0.9417	0.9425
ARCENE	0.9102	**0.9151**	0.8748	0.8699	0.9101	0.9002	0.9001	0.8849	0.8803	0.8248
MADELON	**0.8872**	0.4931	0.5258	0.7846	0.6265	0.6542	0.7342	0.7377	0.7335	0.6246
BREAST CANCER	0.7091	0.6751	**0.7308**	0.6959	0.6778	0.6921	0.6746	0.7271	0.7057	0.6883
OVARIAN CANCER	0.992	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	0.996
SIDO0	0.972	0.9679	0.968	**0.99**	0.976	0.984	0.982	0.986	0.96	0.9699
AVERAGE	**0.8980**	0.8612	0.8646	0.8742	0.8696	0.8750	0.8772	0.8843	0.8785	0.8678
RANKS	**4.2368**	5.9737	5.3421	5.8947	5.7895	5.2105	5.7105	4.5263	5.6842	6.6316

Table A11. Predictive accuracy of OFS-KW versus δ neighborhood (RF).

Data Set	OFS-KW	$r = 0.1$	$r = 0.15$	$r = 0.2$	$r = 0.25$	$r = 0.3$	$r = 0.35$	$r = 0.4$	$r = 0.45$	$r = 0.5$
WDBC	**0.9779**	0.9544	0.9597	0.9526	0.9386	0.9668	0.9439	0.9544	0.9597	0.9509
HILL	0.5461	0.5512	0.5149	0.5329	0.5297	**0.5841**	0.5461	0.5808	0.5379	0.5725
HILL (NOISE)	**0.5642**	0.5380	0.5609	0.5480	0.4998	0.5264	0.5279	0.5544	0.5413	0.5527
COLON TUMOR	0.8205	0.8359	0.8513	0.741	0.7385	0.8218	0.8038	0.8192	0.8346	**0.8526**
DLBCL	0.8430	0.8616	0.9116	0.8848	0.8580	0.8875	0.9125	0.8973	**0.9232**	0.9125

CAR	**0.8327**	0.79017	0.7647	0.7954	0.7176	0.7675	0.8114	0.7208	0.8135	0.7131
LYMPHOMA	0.9510	**0.9675**	0.8926	0.8578	0.9097	0.8716	0.9286	0.9175	0.8452	0.8671
LUNG-STD	0.9668	0.9559	0.9725	0.9722	0.9889	0.9836	0.9778	0.9833	**0.9944**	0.9833
GLIOMA	**0.8377**	0.6977	0.5714	0.7441	0.7845	0.72	0.7359	0.6395	0.6814	0.6077
LEU	**0.9857**	0.9029	0.9171	0.8333	0.8876	0.959	0.9162	0.9038	0.959	0.901
LUNG	0.9119	**0.9408**	0.8873	0.8914	0.9264	0.8958	0.9021	0.8878	0.916	0.9011
MLL	0.8862	0.8605	0.9159	**0.9713**	0.9446	0.9179	0.917	0.9457	0.9067	0.9138
PROSTATE	0.8819	0.8824	0.7738	0.7843	0.8524	0.8533	0.8152	**0.8914**	0.8824	0.8819
SRBCT	0.8539	0.8658	0.8947	0.9151	0.9042	0.9033	0.9055	**0.9298**	0.9028	0.8943
ARCENE	0.805	0.7957	0.7604	0.8154	**0.8498**	0.7958	0.7546	0.7262	0.7951	0.736
MADELON	0.7436	0.5108	0.5104	**0.8015**	0.6246	0.6723	0.7719	0.7819	0.79	0.6535
BREAST CANCER	0.6927	0.6504	0.6681	0.6606	0.6638	0.6784	0.5983	0.6462	0.6642	**0.7062**
OVARIAN CANCER	0.9881	0.9762	0.9763	0.9961	0.9922	**1.0**	0.9921	0.9885	0.992	0.9725
SIDO0	0.974	0.974	0.962	0.982	0.966	**0.984**	0.968	0.954	0.942	0.948
AVERAGE	**0.8454**	0.8164	0.8035	0.8253	0.8198	0.8310	0.8278	0.8275	0.8359	0.8169
RANKS	4.5000	5.9737	6.5526	5.4737	6.0526	**4.3947**	5.5789	5.3684	4.8158	6.2895

Table A12. Comparison results of OFS-KW versus δ neighborhood.

Evaluation Criteria	Friedman Test (p -Values)
Compactness	2.92×10^{-15}
Running time	3.65×10^{-10}
Accuracy (KNN)	0.0275
Accuracy (SVM)	0.7815
Accuracy (RF)	0.6683

Appendix C. The Results of Three Different Feature Stream Orders

Table A13. Comparison results of the three feature stream orders.

	Original	Inverse	Random
Compactness	–	0.1027	0.1027
Running time	–	0.4562	0.0654
Accuracy (KNN)	–	0.4563	0.0631
Accuracy (SVM)	–	0.0554	0.0555
Accuracy (RF)	–	0.0554	0.1027

Appendix D. OSFS-KW versus Traditional Feature Selection Methods

Table A14. Prediction accuracy of OSFS-KW versus traditional feature selection methods (KNN).

Data Set	OSFS-KW	Fisher	SPEC	PCC	ReliefF	MI	Laplacian	UFSOL	ILFS	Lasso	FCBF	CFS
WDBC	**0.9779**	0.9702	0.9472	0.9313	0.9701	0.9526	0.9737	0.9649	0.9614	0.9507	0.9577	0.9701
HILL	0.5710	0.5249	0.5214	0.5610	0.5495	0.5379	0.5379	0.5512	0.5412	**0.5759**	0.5397	0.5297
HILL (NOISE)	**0.5594**	0.5000	0.5345	0.5116	0.5495	0.5429	0.5411	0.5394	0.5446	0.5248	0.5166	0.5560
COLON TUMOR	0.8833	0.8538	0.7769	0.8218	0.8538	0.8244	0.7282	0.6654	0.6128	0.7885	**0.9026**	0.8859
DLBCL	0.9223	0.9116	0.7538	0.9080	0.95	0.9375	0.8830	0.7546	0.7270	0.9732	0.95	**1.0**
CAR	0.8899	0.74388	0.2180	0.8018	**0.9277**	0.8521	0.5621	0.5335	0.5400	0.8034	0.9139	0.8350
LYMPHOMA	0.970	0.9857	0.6887	0.9521	0.9108	**1.0**	0.9303	0.8194	0.9652	0.9857	**1.0**	0.8578
LUNG-STD	0.9503	0.9667	0.9556	0.9889	0.9835	0.9889	0.9944	0.906	0.9833	0.9944	0.9889	**1.0**
GLIOMA	0.8655	0.7623	0.2791	0.3668	0.7591	0.6882	0.6014	0.5068	0.8259	0.66	**0.8805**	0.7827
LEU	**1.0**	0.9714	0.7086	0.9714	0.9438	0.9286	0.9581	0.6695	0.6362	0.9857	0.9571	0.9162
LUNG	**0.9555**	0.8522	0.7789	0.8815	0.8815	0.9254	0.8281	0.8031	0.9012	0.8913	0.9553	0.8861
MLL	**1.0**	0.9579	0.3617	0.8903	0.9713	0.8596	0.9313	0.6619	0.9579	0.9016	0.9857	0.9303
PROSTATE	0.9314	0.931	0.51	0.8924	0.941	0.9505	0.639	0.6481	0.501	**0.961**	0.9505	0.9324
SRBCT	0.9404	**1.0**	0.7074	0.8593	0.9011	0.783	0.6879	0.6948	0.8791	0.9055	0.9875	0.966
ARCENE	**0.9056**	0.7499	0.56	0.7197	0.7808	0.5549	0.7649	0.83	0.6503	0.7057	0.8452	0
MADELON	**0.8885**	0.5719	0.5185	0.5727	0.5946	0.6404	0.5119	0.6335	0.5019	0.5604	0.5804	0.5638
BREAST CANCER	0.7307	0.7271	0.6781	0.6818	0.7023	0.6994	0.6573	0.6365	0.668	**0.8176**	0.7201	0.6648
OVARIAN CANCER	**0.996**	0.9722	0.8269	0.9528	0.9722	0.9722	0.7466	0.6724	0.6439	0.9488	0.9921	0.9686
SIDO0	0.948	0.99	0.924	0.936	0.98	0.992	0.882	0.882	0.9581	0.93	**0.996**	**0.996**
AVERAGE	**0.8887**	0.8391	0.6447	0.8001	0.8486	0.8227	0.7557	0.7038	0.7368	0.8350	0.8747	0.8022
RANKS	**3.1053**	5.8947	10.4737	7.4737	4.8947	5.7632	8.1842	9.2368	8.1842	5.5789	3.5789	5.6316

Table A15. Prediction accuracy of OSFS-KW versus traditional feature selection methods (SVM).

Data Set	OSFS-KW	Fisher	SPEC	PCC	ReliefF	MI	Laplacian	UFSOL	ILFS	lasso	FCBF	CFS
WDBC	0.9754	**0.9789**	0.9596	0.9489	0.9719	0.9648	0.9772	0.9491	0.9737	0.9578	0.9683	0.9719
HILL	**0.5181**	0.5115	0.5132	0.5132	0.5033	0.5016	0.5132	0.5115	0.5132	0.5115	0.5099	0.5065
HILL (NOISE)	0.5231	0.5083	0.5215	0.5132	**0.5313**	0.5198	0.5182	0.5264	0.5099	0.5082	0.4983	0.4852
COLON TUMOR	**0.9179**	0.8692	0.7077	0.8051	0.8385	0.8244	0.7103	0.6321	0.6462	0.8526	0.8859	0.8846
DLBCL	0.9732	0.8991	0.7538	0.925	0.9125	0.9625	0.8483	0.7663	0.7252	0.9607	0.95	**1.0**
CAR	**0.9534**	0.7489	0.2259	0.8192	0.9252	0.9066	0.5949	0.6029	0.5657	0.8722	**0.9285**	0.9095

LYMPHOMA	0.9858	**1.0**	0.7434	0.9379	0.9418	**1.0**	0.9121	0.8194	0.9818	0.9857	**1.0**	0.8669
LUNG-STD	0.9778	0.9778	0.9722	**1.0**	**1.0**	0.9944	0.9889	0.9444	**1.0**	**1.0**	0.9944	0.989
GLIOMA	**0.9255**	0.7441	0.3386	0.4968	0.7827	0.7114	0.6614	0.4936	0.7877	0.7	0.8805	0.8291
LEU	**0.9857**	0.9714	0.7476	0.9571	0.9438	0.9438	0.9448	0.6524	0.6524	0.9857	0.9438	0.9457
LUNG	**0.9751**	0.8268	0.7832	0.8825	0.9107	0.9301	0.8326	0.7942	0.9012	0.911	0.9504	0.9065
MLL	**0.9867**	0.9426	0.3884	0.9037	0.9426	0.8596	0.96	0.627	0.9426	0.8884	0.9724	0.917
PROSTATE	0.9414	0.9305	0.5986	0.9214	0.931	0.9314	0.6967	0.6876	0.5095	**0.9605**	0.941	0.9514
SRBCT	0.9519	0.9889	0.7023	0.9411	0.9764	0.9262	0.7904	0.8186	0.8444	0.9631	**1.0**	0.9771
ARCENE	0.9102	0.74	0.5752	0.7399	0.8362	0.56	0.7648	0.7954	0.5957	**0.9505**	0.8946	0
MADELON	**0.8872**	0.6177	0.5462	0.6177	0.6231	0.6088	0.4819	0.6269	0.5146	0.5538	0.6204	0.5712
BREAST CANCER	0.7091	0.727	0.6749	0.73	0.7367	0.699	0.6749	0.6749	0.6749	**0.8564**	0.7655	0.6749
OVARIAN CANCER	**0.9923**	0.9722	0.8501	0.9448	0.9722	0.9722	0.7153	0.6686	0.6677	0.9526	0.9921	0.9565
SIDO0	0.972	**0.996**	0.932	0.944	0.992	0.994	0.882	0.882	0.974	0.9461	0.992	0.992
AVERAGE	**0.8980**	0.8395	0.6597	0.8180	0.8564	0.8321	0.7615	0.7091	0.7358	0.8588	0.8783	0.8071
RANKS	**2.8947**	5.4737	10.0263	7.0263	4.9474	6.2895	8.0526	9.3684	8.1316	5.4737	4.0789	6.2368

Table A16. Prediction accuracy of OSFS-KW versus traditional feature selection methods (RF).

Data Set	OSFS-KW	Fisher	SPEC	PCC	ReliefF	MI	Laplacian	UFSOL	ILFS	lasso	FCBF	CFS
WDBC	**0.9779**	0.9579	0.9439	0.9174	0.9632	0.9264	0.9492	0.9439	0.9474	0.9421	0.9544	0.9438
HILL	0.5461	0.5263	0.5215	0.5165	0.5016	0.4901	0.4883	0.5346	0.5248	**0.5478**	0.5197	0.5032
HILL (NOISE)	**0.5642**	0.4867	0.5146	0.4834	0.5132	0.5081	0.5280	0.5526	0.4933	0.4933	0.4604	0.5131
COLON TUMOR	0.8205	0.8679	0.7423	0.741	0.7897	0.8538	0.6628	0.6949	0.5154	0.7218	0.8859	**0.9013**
DLBCL	0.8430	0.8857	0.7538	0.9116	0.8866	**0.9375**	0.7690	0.7029	0.6627	0.8322	0.9375	0.9107
CAR	0.8327	0.9675	0.6947	0.9015	0.9381	**0.9714**	0.8848	0.7729	0.9357	0.934	0.9714	0.7753
LYMPHOMA	0.9510	0.9779	0.9611	0.9833	0.9779	0.9502	0.9778	0.8949	0.9889	0.9889	0.9889	**0.989**
LUNG-STD	0.9668	0.9889	0.9835	0.9667	0.9779	**0.9944**	0.9778	0.9225	0.9833	0.9449	0.9778	0.9889
GLIOMA	0.8377	0.6014	0.3223	0.4082	0.7095	0.6732	0.5382	0.4836	0.6732	0.5882	**0.9455**	0.7427
LEU	**0.9857**	0.9438	0.64	0.9162	0.9295	0.8867	0.9457	0.5838	0.6095	0.9571	0.9438	0.8876
LUNG	0.9119	0.8314	0.7733	0.8582	0.8802	0.9058	0.7869	0.7624	0.8466	0.8921	**0.9201**	0.8483
MLL	0.8862	0.9426	0.3063	0.8627	0.9118	0.874	0.8227	0.5923	0.9426	0.8903	**0.9457**	0.9016
PROSTATE	0.8819	0.9114	0.5586	0.9019	0.9214	0.8829	0.6176	0.5776	0.6671	0.9029	**0.941**	0.9219
SRBCT	0.8539	0.9284	0.7478	0.8554	0.8925	0.8568	0.568	0.7921	0.6992	0.8973	**0.9889**	0.9165
ARCENE	0.805	0.6742	0.56	0.595	0.7851	0.5549	0.7549	0.7282	0.5602	0.8202	**0.8746**	0
MADELON	**0.7436**	0.5246	0.5135	0.5258	0.5696	0.6127	0.5077	0.6027	0.4919	0.5377	0.5296	0.5369
BREAST CANCER	0.6927	0.6635	0.6647	0.6883	0.7091	0.6893	0.6041	0.6326	0.6368	0.7585	**0.7659**	0.6817
OVARIAN CANCER	**0.9881**	0.9682	0.8107	0.9489	0.9682	0.9682	0.7313	0.629	0.7152	0.9608	0.9251	0.8732
SIDO0	0.974	0.982	0.9341	0.9341	0.986	0.984	0.874	0.8721	0.962	0.9241	0.988	**0.9939**
AVERAGE	0.8454	0.8226	0.6814	0.7851	0.8322	0.8169	0.7363	0.6987	0.7293	0.8176	**0.8665**	0.7805
RANKS	4.5789	5.2632	9.0526	7.7105	4.6579	5.8684	8.7632	9.3421	8.0263	5.7632	**3.4737**	5.5000

Table A17. Test results of OFS-KW versus traditional feature selection methods.

Evaluation Criteria	Friedman Test (p -Values)
Accuracy (KNN)	1.20×10^{-15}
Accuracy (SVM)	3.38×10^{-13}
Accuracy (RF)	1.22×10^{-10}

Appendix E. OSFS-KW versus OSFS Methods

Table A18. Compactness.

Data Sets	OSFS-KW	OFS-A3M	OSFS	Alpha-Investing	Fast-OSFS	SAOLA
WDBC	18	17	10	19	**4**	24
HILL	5	12	**1**	5	5	78
HILL (NOISE)	11	12	**2**	3	7	76
COLON TUMOR	24	26	7	4	4	**0**
DLBCL	13	16	**6**	11	8	148
CAR	35	38	**10**	30	14	58
LYMPHOMA	**7**	7	13	18	8	69
LUNG-STD	12	8	17	77	11	**0**
GLIOMA	13	18	15	**6**	7	1224
LEU	6	14	11	23	**1**	478
LUNG	16	25	35	39	**2**	**2**
MLL	10	13	23	12	**7**	13
PROSTATE	25	30	6	19	5	**1**
SRBCT	15	16	4	32	8	**0**
ARCENE	51	55	51	29	10	**0**
MADELON	2	2	2	7	14	**0**
BREAST CANCER	34	40	1	13	16	**0**
OVARIAN CANCER	6	9	10	68	6	**2**
SIDO0	18	49	**2**	75	13	7
AVERAGE	16.8947	21.4211	11.8947	25.7895	**7.8947**	114.7368
RANKS	3.5526	4.4737	3.0263	4.0789	**2.5000**	3.3684

Table A19. Running time (seconds).

Data Set	OSFS-KW	OFS-A3M	OSFS	Alpha-Investing	Fast-OSFS	SAOLA
WDBC	1.7934	2.8427	0.8745	0.2117	0.4246	**0.1438**
HILL	5.3830	6.6946	3.7473	**0.0180**	0.0911	0.0778
HILL (NOISE)	3.9342	4.5308	2.3738	**0.0052**	0.4670	0.0750
COLON TUMOR	2.0676	1.8749	0.9296	0.2778	0.7427	**0.0916**
DLBCL	10.5498	17.1439	4.7319	**0.8760**	2.6201	11.1175
CAR	200.6726	68.5552	13.5982	**1.1405**	15.5793	5.1563
LYMPHOMA	12.4278	6.7790	1.7897	**0.5756**	5.7105	5.9966
LUNG-STD	55.9968	54.6860	17.6993	2.2994	16.9619	**0.1351**
GLIOMA	4.8191	6.2993	1.7851	**0.4096**	2.6653	9.6878
LEU	4.6364	4.4316	1.2731	**0.2926**	0.3908	1.5318
LUNG	31.9804	71.2026	13.6662	**0.8823**	1.2404	0.3039
MLL	11.2642	12.5412	3.1975	**0.5738**	3.7881	13
PROSTATE	24.2364	7.5123	3.5531	0.6009	1.5547	**0.0987**
SRBCT	3.4475	4.3123	1.1155	0.1655	1.0674	**0.0394**
ARCENE	87.9552	153.5089	65.9577	5.1478	8.4269	**0.5413**
MADELON	538.4690	1098.8598	822.9196	**0.08597**	2.2586	0.1640
BREAST CANCER	114.8965	117.0676	71.1601	2.8190	7.865126	**0.3178**
OVARIAN CANCER	80.8324	81.2105	33.7745	5.0253	4.431689	**0.3477**

SIDO0	64.2822	74.8208	20.6689	0.6273	1.9891	**0.0982**
AVERAGE	66.2971	94.4671	57.0956	**1.1597**	4.1198	2.5751
RANKS	5.1053	5.5789	3.5789	**1.5789**	2.8947	2.2632

Table A20. Predictive accuracy using KNN.

Data Sets	OSFS-KW	OFS-A3M	OSFS	Alpha-Investing	Fast-OSFS	SAOLA
WDBC	**0.9779**	0.9685	0.9491	0.9632	0.9736	0.9666
HILL	**0.5710**	0.5561	0.5446	0.5627	0.5561	0.5593
HILL (NOISE)	0.5594	0.5594	0.5395	0.5577	**0.5644**	0.5510
COLON TUMOR	0.8833	0.8692	0.8846	0.6154	**0.8859**	0
DLBCL	0.9223	0.8840	**0.9875**	0.9375	0.975	0.8072
CAR	**0.8899**	0.86717	0.71485	0.7388	0.7496	0.8010
LYMPHOMA	0.970	0.9857	**1.0**	**1.0**	**1.0**	**1.0**
LUNG-STD	0.9503	0.9833	**0.9944**	0.9724	0.9944	0
GLIOMA	0.8655	0.8473	0.7295	0.6164	**0.8673**	0.8441
LEU	**1.0**	0.959	0.9724	0.9019	0.6267	0.9857
LUNG	0.9555	0.956	**0.9562**	0.9501	0.6856	0.725
MLL	**1.0**	0.9867	0.9713	0.957	1	0.8892
PROSTATE	0.9314	**0.9324**	0.9324	0.9324	0.9324	0.9324
SRBCT	0.9404	**0.966**	0.966	0.966	0.966	0
ARCENE	**0.9056**	0.8552	0.8708	0.84	0.8298	0
MADELON	**0.8885**	0.5196	0.5196	0.6023	0.5831	0
BREAST CANCER	**0.7307**	0.7307	0.6682	0.7275	0.7304	0
OVARIAN CANCER	0.996	0.9722	**1.0**	0.9922	0.9922	0.8936
SIDO0	0.948	0.952	0.942	0.962	**0.994**	0.958
AVERAGE	**0.8887**	0.8606	0.8496	0.8313	0.8372	0.5744
RANKS	**2.7632**	3.3947	3.5789	3.7632	2.8421	4.6579

Table A21. Predictive accuracy using SVM.

Data Set	OSFS-KW	OFS-A3M	OSFS	Alpha-Investing	Fast-OSFS	SAOLA
WDBC	0.9754	0.9772	0.9526	**0.9806**	0.9683	0.9736
HILL	0.5181	0.5148	0.5148	0.5098	0.5082	**0.5412**
HILL (NOISE)	0.5231	0.5231	0.5065	0.4934	0.5248	**0.5579**
COLON TUMOR	**0.9179**	0.8718	0.8526	0.6462	0.8526	0
DLBCL	0.9732	0.9625	**0.9875**	0.9375	0.9875	0.8580
CAR	**0.9534**	0.92737	0.7557	0.8212	0.79820	0.6453
LYMPHOMA	0.9858	0.9675	**1.0**	0.9857	**1.0**	0.6792
LUNG-STD	0.9778	0.9944	0.9944	0.9944	**1.0**	0
GLIOMA	0.9255	**0.9636**	0.8127	0.6514	0.8491	0.4391
LEU	0.9857	**0.9867**	0.9571	0.9857	0.6524	0.7505
LUNG	**0.9751**	0.9603	0.9601	0.9452	0.6853	0.7151
MLL	0.9867	0.9857	0.9426	0.9857	**1.0**	0.9016
PROSTATE	0.9414	**0.9514**	0.9514	0.9514	0.9514	0.9514
SRBCT	0.9519	**0.9771**	0.9771	0.9771	0.9771	0
ARCENE	**0.9102**	0.8896	0.9049	0.8949	0.8197	0
MADELON	**0.8872**	0.4758	0.4758	0.615	0.6265	0
BREAST CANCER	0.7091	0.7091	0.6749	0.717	**0.7756**	0
OVARIAN CANCER	0.992	0.9842	**1.0**	0.9962	0.996	0.8893
SIDO0	0.972	0.976	0.912	0.99	**0.994**	0.958
AVERAGE	**0.8980**	0.8736	0.8491	0.8462	0.8404	0.5190
RANKS	**2.8158**	3.0526	3.6316	3.3947	3.0526	5.0526

Table A22. Predictive accuracy using RF.

Data Set	OSFS-KW	OFS-A3M	OSFS	Alpha-Investing	Fast-OSFS	SAOLA
WDBC	**0.9779**	0.9578	0.9403	0.9509	0.9491	0.9526
HILL	0.5461	0.5577	0.5280	**0.5675**	0.5577	0.5610
HILL (NOISE)	0.5642	**0.5825**	0.5594	0.5348	0.5347	0.5346
COLON TUMOR	0.8205	0.7577	**0.8679**	0.5808	0.8218	0
DLBCL	0.8430	0.9241	0.9241	0.8848	**0.9357**	0.8064
CAR	**0.8327**	0.76051	0.68523	0.6778	0.71930	0.7486
LYMPHOMA	0.9510	0.9366	**0.9857**	0.9188	0.9002	0.9197
LUNG-STD	0.9668	0.9389	0.9833	0.9835	**0.9889**	0
GLIOMA	**0.8377**	0.655	0.5745	0.5086	0.7741	0.7191
LEU	**0.9857**	0.959	0.9581	0.86	0.5848	0.7895
LUNG	0.9119	0.901	**0.9454**	0.876	0.6608	0.6553
MLL	0.8862	0.8226	0.9559	0.9019	**0.9579**	0.917
PROSTATE	0.8819	0.9114	0.9405	0.9405	0.9119	**0.96**
SRBCT	0.8539	0.9424	0.9071	**0.9425**	0.9306	0
ARCENE	0.805	0.7849	**0.83**	0.7948	0.815	0
MADELON	**0.7436**	0.5204	0.5169	0.6062	0.5788	0
BREAST CANCER	0.6927	0.6537	0.5802	0.689	**0.7688**	0
OVARIAN CANCER	0.9881	0.98	0.992	0.9842	**1.0**	0.8973
SIDO0	0.974	0.968	0.9679	0.9879	**0.988**	0.962
AVERAGE	**0.8454**	0.8165	0.8233	0.7995	0.8094	0.5486
RANKS	**2.8947**	3.5263	3.3158	3.5526	2.9737	4.7368

Table A23. Comparison results.

Evaluation Criteria	Friedman Test (p -Values)
Compactness	0.0248
Running time	3.62×10^{-23}
Accuracy (KNN)	0.0337
Accuracy (SVM)	0.0032
Accuracy (RF)	0.0533

References

1. Zhou, H.; Wang, J.; Zhang, H. Stochastic multicriteria decision-making approach based on SMAA-ELECTRE with extended gray numbers. *Int. Trans. Oper. Res.* **2019**, *26*, 2032–2052, doi:10.1111/itor.12380.
2. Tian, Z.-P.; Wang, J.; Wang, J.-Q.; Chen, X.-H. Multicriteria decision-making approach based on gray linguistic weighted Bonferroni mean operator. *Int. Trans. Oper. Res.* **2018**, *25*, 1635–1658, doi:10.1111/itor.12220.
3. Tian, Z.-P.; Wang, J.; Wang, J.; Zhang, H.-Y. Simplified Neutrosophic Linguistic Multi-criteria Group Decision-Making Approach to Green Product Development. *Group Decis. Negot.* **2017**, *26*, 597–627, doi:10.1007/s10726-016-9479-5.
4. Cang, S.; Yu, H. Mutual information based input feature selection for classification problems. *Decis. Support Syst.* **2012**, *54*, 691–698, doi:10.1016/j.dss.2012.08.014.
5. Wang, Z.; Zhang, Y.; Chen, Z.; Yang, H.; Sun, Y.; Kang, J.; Yang, Y.; Liang, X. Application of ReliefF algorithm to selecting feature sets for classification of high resolution remote sensing image. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–

15 July 2016; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2016; pp. 755–758.
6. Saqlain, S.M.; Sher, M.; Shah, F.A.; Khan, I.; Ashraf, M.U.; Awais, M.; Ghani, A. Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. *Knowl. Inf. Syst.* **2018**, *58*, 139–167, doi:10.1007/s10115-018-1185-y.
7. Benabdeslem, K.; Elghazel, H.; Hindawi, M. Ensemble constrained Laplacian score for efficient and robust semi-supervised feature selection. *Knowl. Inf. Syst.* **2015**, *49*, 1161–1185, doi:10.1007/s10115-015-0901-0.
8. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B (Statal Methodol.)* **2011**, *73*, 273–282, doi:10.2307/41262671.
9. Kumar, V.; Minz, S. Multi-view ensemble learning: An optimal feature set partitioning for high-dimensional data classification. *Knowl. Inf. Syst.* **2015**, *49*, 1–59, doi:10.1007/s10115-015-0875-y.
10. Wang, J.; Zhao, P.; Hoi, S.C.H.; Jin, R. Online Feature Selection and Its Applications. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 698–710, doi:10.1109/TKDE.2013.32.
11. Glocer, K.; Eads, D.; Theiler, J. Online feature selection for pixel classification. In Proceedings of the 22nd International Conference on Software Engineering: ICSE 2000, the New Millennium, Limerick, Ireland, 4–11 June 2000; Association for Computing Machinery (ACM): New York, NY, USA, 2005; pp. 249–256.
12. Javidi, M.M.; Eskandari, S. Online streaming feature selection: A minimum redundancy, maximum significance approach. *Pattern Anal. Appl.* **2018**, *22*, 949–963, doi:10.1007/s10044-018-0690-7.
13. Yu, K.; Wu, X.; Ding, W.; Pei, J. Scalable and Accurate Online Feature Selection for Big Data. *ACM Trans. Knowl. Discov. Data* **2016**, *11*, 1–39, doi:10.1145/2976744.
14. Eskandari, S.; Javidi, M.M. Online streaming feature selection using rough sets. *Int. J. Approx. Reason.* **2016**, *69*, 35–57, doi:10.1016/j.ijar.2015.11.006.
15. Perkins, S.; Theiler, J.; Online feature selection using grafting. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Los Alamos, NM, USA, 21–24 August 2003, pp. 592–599.
16. Wu, X.; Yu, K.; Ding, W.; Wang, H.; Zhu, X. Online Feature Selection with Streaming Features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 1178–1192, doi:10.1109/tpami.2012.197.
17. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 341–356.
18. Yao, Y.; She, Y. Rough set models in multigranulation spaces. *Inf. Sci.* **2016**, *327*, 40–56, doi:10.1016/j.ins.2015.08.011.
19. Javidi, M.M.; Eskandari, S. Streamwise feature selection: A rough set method. *Int. J. Mach. Learn. Cybern.* **2016**, *9*, 667–676, doi:10.1007/s13042-016-0595-y.
20. Kumar, S.U.; Inbarani, H.H. PSO-based feature selection and neighborhood rough set-based classification for BCI multiclass motor imagery task. *Neural Comput. Appl.* **2016**, *28*, 3239–3258, doi:10.1007/s00521-016-2236-5.
21. Zhang, J.; Li, T.; Ruan, D.; Liu, D. Neighborhood rough sets for dynamic data mining. *Int. J. Intell. Syst.* **2012**, *27*, 317–342, doi:10.1002/int.21523.
22. Zhou, P.; Hu, X.-G.; Li, P.; Wu, X. Online streaming feature selection using adapted Neighborhood Rough Set. *Inf. Sci.* **2019**, *481*, 258–279, doi:10.1016/j.ins.2018.12.074.
23. Lin, Y.; Li, J.; Lin, P.; Lin, G.; Chen, J. Feature selection via neighborhood multi-granulation fusion. *Knowl. Based Syst.* **2014**, *67*, 162–168, doi:10.1016/j.knosys.2014.05.019.
24. Zhou, P.; Hu, X.; Li, P.; Wu, X. Online feature selection for high-dimensional class-imbalanced data. *Knowl. Based Syst.* **2017**, *136*, 187–199, doi:10.1016/j.knosys.2017.09.006.
25. Pawlak, Z. Rough sets and intelligent data analysis. *Inf. Sci.* **2002**, *147*, 1–12, doi:10.1016/s0020-0255(02)00197-4.
26. Hu, Q.; Yu, D.; Liu, J.; Wu, C. Neighborhood rough set based heterogeneous feature subset selection. *Inf. Sci.* **2008**, *178*, 3577–3594, doi:10.1016/j.ins.2008.05.024.
27. Mac Parthalain, N.; Shen, Q.; Jensen, R. A Distance Measure Approach to Exploring the Rough Set Boundary Region for Attribute Reduction. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 305–317, doi:10.1109/tkde.2009.119.
28. Maciá-Pérez, F.; Berna-Martinez, J.V.; Oliva, A.F.; Ortega, M.A.A. Algorithm for the detection of outliers based on the theory of rough sets. *Decis. Support Syst.* **2015**, *75*, 63–75, doi:10.1016/j.dss.2015.05.002.
29. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238, doi:10.1109/tpami.2005.159.

30. Wang, Y.; Klijn, J.G.; Zhang, Y.A.; Sieuwerts, M.; Look, M.P.; Yang, F.; Talantov, D.; Timmermans, M.; Meijer-van Gelder, M.E.; Yu, J. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **2005**, *365*, 671–679.
31. Rosenwald, A.; Wright, G.; Chan, W.C.; Connors, J.M.; Campo, E.; Fisher, R.I.; Gascoyne, R.D.; Müller-Hermelink, H.K.; Smeland, E.B.; Giltnane, J.M.; et al. The Use of Molecular Profiling to Predict Survival after Chemotherapy for Diffuse Large-B-Cell Lymphoma. *N. Engl. J. Med.* **2002**, *346*, 1937–1947, doi:10.1056/nejmoa012914.
32. Yu, L.; Ding, C.; Loscalzo, S. Stable feature selection via dense feature groups. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08), Las Vegas, NV, USA, 24–27 August 2008; Association for Computing Machinery (ACM): New York, NY, USA, 2008; 803p.
33. Yang, K.; Cai, Z.; Li, J.; Lin, G. A stable gene selection in microarray data analysis. *BMC Bioinform.* **2006**, *7*, 228, doi:10.1186/1471-2105-7-228.
34. Richardson, A.M. Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach by Gregory W. Corder, Dale I. Foreman. *Int. Stat. Rev.* **2010**, *78*, 451–452, doi:10.1111/j.1751-5823.2010.00122_6.x.
35. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
36. Gu, Q.; Li, Z.; Han, J. Generalized Fisher Score for Feature Selection. *arXiv* **2012**, arXiv:1202.3725.
37. Zhao, Z.; Liu, H. Spectral feature selection for supervised and unsupervised learning. In Proceedings of the 24th International Conference on Real-Time Networks and Systems (RTNS '16), Brest, France, 19–21 October 2016; Association for Computing Machinery (ACM): New York, NY, USA, 2017; pp. 1151–1157.
38. Wasikowski, M.; Chen, X.-W. Combating the Small Sample Class Imbalance Problem Using Feature Selection. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1388–1400, doi:10.1109/tkde.2009.187.
39. Robnik-Šikonja, M.; Kononenko, I. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Mach. Learn.* **2003**, *53*, 23–69, doi:10.1023/a:1025667309714.
40. Guo, J.; Guo, Y.; Kong, X.; He, R.; Quo, Y. Unsupervised feature selection with ordinal locality. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017; pp. 1213–1218.
41. Li, F.; Miao, D.; Pedrycz, W. Granular multi-label feature selection based on mutual information. *Pattern Recognit.* **2017**, *67*, 410–423, doi:10.1016/j.patcog.2017.02.025.
42. Roffo, G.; Melzi, S.; Castellani, U.; Vinciarelli, A. Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1407–1415.
43. Fonti, V.; Belitser, E.; Feature selection using lasso. *VU Amst. Res. Pap. Bus. Anal.* **2017**, *30*, 1–25.
44. Yu, L.; Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003, pp. 856–863.
45. Chutia, D.; Bhattacharyya, D.K.; Sarma, J.; Raju, P.N.L. An effective ensemble classification framework using random forests and a correlation based feature selection technique. *Trans. GIS* **2017**, *21*, 1165–1178, doi:10.1111/tgis.12268.
46. Zhou, J.; Foster, D.P.; Stine, R.A.; Ungar, L.H. Streamwise feature selection, *J. Mach. Learn. Res.* **2006**, *7*, 1861–1885.
47. Wu, X.; Yu, K.; Wang, H.; Ding, W. Online streaming feature selection. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Citeseer, Haifa, Israel, 21–24 June 2010, pp. 1159-1166.
48. Yu, K.; Ding, W.; Wu, X. LOFS: A library of online streaming feature selection. *Knowl. Based Syst.* **2016**, *113*, 1–3, doi:10.1016/j.knosys.2016.08.026.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article 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/).