



Unsupervised feature selection based on incremental forward iterative Laplacian score

Jiefang Jiang^{1,2} · Xianyong Zhang^{1,2,3} · Jilin Yang^{2,4}

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Feature selection facilitates intelligent information processing, and the unsupervised learning of feature selection has become important. In terms of unsupervised feature selection, the Laplacian score (LS) provides a powerful measurement and optimization method, and good performance has been achieved using the recent forward iterative Laplacian score (FILS) algorithm. However, there is still room for advancement. The aim of this paper is to improve the FILS algorithm, and thus, feature significance (SIG) is mainly introduced to develop a high-quality selection method, i.e., the incremental forward iterative Laplacian score (IFILS) algorithm. Based on the modified LS, the metric difference in the incremental feature process motivates SIG. Therefore, SIG offers a dynamic characterization by considering initial and terminal states, and it promotes the current FILS measurement on only the terminal state. Then, both the modified LS and integrated SIG acquire granulation nonmonotonicity and uncertainty, especially on incremental feature chains, and the corresponding verification is achieved by completing examples and experiments. Furthermore, a SIG-based incremental criterion of minimum selection is designed to choose optimization features, and thus, the IFILS algorithm is naturally formulated to implement unsupervised feature selection. Finally, an in-depth comparison of the IFILS algorithm with the FILS algorithm is achieved using data experiments on multiple datasets, including a nominal dataset of COVID-19 surveillance. As validated by the experimental results, the IFILS algorithm outperforms the FILS algorithm and achieves better classification performance.

Keywords Feature selection · Unsupervised learning · Forward iterative Laplacian score · Incremental forward iterative Laplacian score · Feature significance · Granulation nonmonotonicity and uncertainty

✉ Xianyong Zhang
xianyongzh@sina.com.cn

¹ School of Mathematical Sciences, Sichuan Normal University, Chengdu 610066, China

² Institute of Intelligent Information and Quantum Information, Sichuan Normal University, Chengdu 610066, China

³ Research Center of Sichuan Normal University, National-Local Joint Engineering Laboratory of System Credibility Automatic Verification, Chengdu 610066, China

⁴ College of Computer Science, Sichuan Normal University, Chengdu 610101, China

1 Introduction

Machine learning facilitates the research and application of artificial intelligence (Hai et al 2021; Lim et al 2022; Tham et al 2022; Van et al 2022), and feature selection (FS) is a basic topic in machine learning (Cai et al 2018; Tubishat et al 2021). The aim of FS is validity extraction and dimensionality reduction, and it has been deeply researched in data mining (El-Hasnony et al 2020; Nguyen et al 2020; Sun et al 2021), pattern recognition (Gunal and Edizkan 2008), etc. Moreover, FS has practical application (Kou et al 2020; Remeseiro and Bolon-Canedo 2019). FS is usually classified into three cases based on the decision category information, i.e., supervised FS, semisupervised FS, and unsupervised FS. Supervised and semisupervised cases require objects with completely-labeled and partially-labeled categories, respectively. Currently, unlabeled data exist universally across diverse fields; therefore, unsupervised FS works well for anomaly detection (Amarbayasgalan et al 2020; Yuan et al 2022), clustering learning (Chen et al 2020; Zhou et al 2020), medical analysis (García-Díaz et al 2020; Rostami et al 2020), etc.

Unsupervised FS can be divided into three main methods: “wrapper, filter, hybrid” (Alelyani et al 2018; Solorio-Fernández et al 2020). In terms of the FS strategy, wrapper methods utilize clustering algorithms (Breaban and Luchian 2011) but filter methods do not (Dadaneh et al 2016; Luo et al 2017), while hybrid approaches reach a compromise by combining the above two methods (Solorio-Fernández et al 2016). Filter unsupervised FS depends on the intrinsic properties of data to exhibit good characteristics and applicability, and basic filter learning is motivated by the Laplacian score (LS). He et al (2005) proposed LS to evaluate feature importance, and thus, FS was implemented by the LS algorithm. Then, Zhao et al (2008) used LS for face recognition comparison in semisupervised FS, and Huang et al (2018) presented manifold-based constraint LS for multi-label learning. Moreover, Zhu et al (2012) proposed an LS-based FS algorithm called the iterative LS (ILS) algorithm, and using this approach, the nearest neighborhood graph is iteratively updated by discarding the least relevant feature; thus, better experimental results were achieved using the ILS algorithm compared with the LS algorithm. To enhance both the LS and ILS algorithms, Pang and Zhang (2020a) recently proposed an improved FS algorithm, i.e., the forward iterative LS algorithm. More specifically, LS is modified by both the parameter concretization (on the K nearest neighbors (KNN)) and feature subset assessment, and then, the FILS algorithm introduces the forward strategy and selective criterion by optimally extracting the most important feature in each iteration. Meanwhile, Pang and Zhang (2020b) introduced the modified LS to linearly combine the neighborhood discrimination index, and the corresponding semisupervised FS was further implemented.

Now, the FILS algorithm (Pang and Zhang 2020a) is refocused on, and it utilizes the modified LS assessment and forward iteration strategy to achieve better performance than the LS and ILS algorithms. This algorithm still has room for improvement. In the FILS algorithm, the feature correlation and selection criterion require uncertainty measurements on the subset transfer $A \rightarrow A \cup \{f_k\}$, and this core process actually adheres to dynamic learning with a feature increment; however, only the measurement of the terminal state subset $A \cup \{f_k\}$ is of concern when using the FILS algorithm. However, the initial state A with dynamic change is also worth considering. Thus, in this paper, feature significance (SIG) is introduced through the process of state deference, i.e., $\Delta(A \rightarrow A \cup \{f_k\})$, and the relevant uncertainty characterization can more systematically evaluate and effectively promote forward iterations with dynamics. The in-depth utilization of SIG has already been applied in optimal selection and heuristic reduction of

rough set reasoning (Wang et al 2017; Xu et al 2021; Yuan et al 2021; Zhang et al 2021; Zhang and Yao 2022). In this paper, the modified LS and forward iteration are still utilized. However, the SIG technology is added to implement (filter) unsupervised FS. The relevant algorithm is referred to as the incremental forward iterative LS (IFILS) algorithm, and it truly improves the FILS algorithm, which already improves the LS and ILS algorithms. Thus, better learning performance on classification recognition is achieved using the IFILS algorithm, as validated by data experiments. Moreover, both the basic modified LS and developmental SIG acquire in-depth measurement properties, mainly granulation nonmonotonicity and uncertainty.

In this paper, unsupervised feature selection is implemented, and its purpose is to improve the current FILS algorithm (Pang and Zhang 2020a). Regarding its novelty, SIG and its granulation nonmonotonicity and uncertainty are deeply mined, and thus an improved algorithm, i.e., the IFILS algorithm, is designed to achieve better classification performance. The remainder of this paper is organized as follows. In Sect. 2, unsupervised FS based on FILS is reviewed. In Sect. 3, unsupervised FS based on IFILS algorithm is established. In Sect. 4, an illustrative example is provided for relevant measures and algorithms. In Sect. 5, comparative experiments of the FILS and IFILS algorithms are performed. Finally, in Sect. 6, this paper is concluded.

2 Unsupervised feature selection based on forward iterative Laplacian score (FILS)

LS effectively evaluates feature description abilities. Therefore, it facilitates FS especially in unsupervised learning. In LS-based FS, if samples with similar distances in the original feature space maintain their neighbor relationships in a single feature dimension, then the basic feature is considered to maintain local structures of data; that is, FS utilizes the nearest neighbor graph to explore the local structures. Of course, different LS forms and treatments may cause different algorithms and effects of unsupervised FS. Next, the relevant FILS algorithm is reviewed based on basic LS and improved LS (Pang and Zhang 2020a).

Let $DIS = (X, F, D)$ be a decision information system. Here, universe $X = \{x_1, x_2, \dots, x_u\}$ carries u samples, the condition set $F = \{f_1, f_2, \dots, f_n\} = \{f_r \mid r = 1, 2, \dots, n\}$ contains n features (such as f_r), f_{ri} denotes the value on i -th sample and r -th feature ($i = 1, \dots, u; r = 1, \dots, n$), and D implies the decision set. By separation, information system $IS = (X, F)$ is mainly used for unsupervised FS, and D supports only eventual identification for learning effects. Let $A \subseteq F, \bar{A} = F - A$.

Definition 1 (LS (He et al 2005)) Information system $IS = (X, F)$ has the nearest neighbor graph. If $x_j \in X$ is in KNN of $x_i \in X$, then nodes i and j are connected by an edge. By sample circulation in universe X , weight matrix $S = [S_{ij}]_{u \times u}$ is obtained as follows:

$$S_{ij} = \begin{cases} e^{-td^2(x_i, x_j)}, & \text{if } x_i \in KNN(x_j) \vee x_j \in KNN(x_i), \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Here, $t > 0$ is an adjustable parameter, $d(x_i, x_j)$ is the Euclidean distance for x_i, x_j , and $KNN(x_i)$ denotes the KNN set of x_i . Now, introduce matrix expressions:

$$\mathbf{1} = [1, 1, \dots, 1]_{1 \times u}^T, \mathbf{D} = \text{diag}(\mathbf{S} \times \mathbf{1}), \mathbf{L} = \mathbf{D} - \mathbf{S},$$

$$\mathbf{f}_r = [f_{r1}, f_{r2}, \dots, f_{ru}]^T, \tilde{\mathbf{f}}_r = \mathbf{f}_r - \frac{\mathbf{f}_r^T \times \mathbf{D} \times \mathbf{1}}{\mathbf{1}^T \times \mathbf{D} \times \mathbf{1}} \times \mathbf{1}. \tag{2}$$

Then, LS of feature f_r is defined as

$$J_{LS}(f_r) = \frac{\tilde{\mathbf{f}}_r^T \mathbf{L} \tilde{\mathbf{f}}_r}{\tilde{\mathbf{f}}_r^T \mathbf{D} \tilde{\mathbf{f}}_r} = \frac{\sum_{i,j} (f_{ri} - f_{rj})^2 S_{ij}}{\sum_i (f_{ri} - \mu_r)^2 D_{ii}}, \tag{3}$$

where $\mu_r = \frac{1}{n} \sum_i f_{ri}$ and $D_{ii} = \sum_j S_{ij}$.

Basic LS can be used to guide unsupervised FS, as realized by the LS algorithm (He et al 2005). In this method, the nearest neighborhood graph is mainly iteratively updated, and the importance of a feature is evaluated based on its locality preservation ability. Furthermore, the ILS algorithm is used to remove poor performing features (Zhu et al 2012); therefore, it generally outperforms the LS algorithm. In contrast, selecting higher-ranking features becomes a direct and effective strategy. For example, better effects are achieved using the FILS algorithm with the LS improvement (Pang and Zhang 2020a).

Definition 2 (FILS (Pang and Zhang 2020a)) The original weight matrix in Eq. (1) is defined as

$$S_{ij} = \begin{cases} e^{-\frac{d^2(x_i, x_j)}{\sigma_i \sigma_j}}, & \text{if } x_i \in KNN(x_j) \vee x_j \in KNN(x_i), \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Here, $\sigma_i = d(x_i, x_i^K)$ is given, where x_i^K represents the KNN of x_i , and σ_j has the similar case. By Eqs. (2) (4), \mathbf{S} -based \mathbf{L}, \mathbf{D} is updated to generate matrices:

$$\tilde{\mathbf{Z}}_A = \mathbf{Z}_A - \mathbf{1} \times \left(\frac{\mathbf{Z}_A^T \times \mathbf{D} \times \mathbf{1}}{\mathbf{1}^T \times \mathbf{D} \times \mathbf{1}} \right)^T, \tilde{\mathbf{Z}}_A^T = (\tilde{\mathbf{Z}}_A)^T, \tag{5}$$

where $\mathbf{Z}_A \in R^{(u) \times |A|}$ is the sample matrix of feature subset $A \subseteq F$. Thus, the modified LS of feature subset A is defined as

$$J_{FILS}(A) = \begin{cases} \frac{\text{trace}(\tilde{\mathbf{Z}}_A^T \times \mathbf{L} \times \tilde{\mathbf{Z}}_A)}{\text{trace}(\tilde{\mathbf{Z}}_A^T \times \mathbf{D} \times \tilde{\mathbf{Z}}_A)}, & \text{if } A \neq \emptyset, \\ \infty, & \text{otherwise,} \end{cases} \tag{6}$$

where $\text{trace}(\cdot)$ denotes the sum of the diagonal matrix elements.

The selection criterion (regarding picking up the most important feature) is based on

$$f_{k^*} = \arg \min_{f_k \in A} J_{FILS}(f_k) = \arg \min_{f_k \in A} \left| 1 - \frac{J_{FILS}(A \cup \{f_k\})}{J_{FILS}(F)} \right|, \tag{7}$$

where

$$J_{FILS}(f_k) = \left| 1 - \frac{J_{FILS}(A \cup \{f_k\})}{J_{FILS}(F)} \right|. \quad (8)$$

In Definition 2, the weight matrix is first defined by the modified LS through the assignment of parameter t to the distance production on KNN, i.e., $t = \frac{1}{\sigma_i \sigma_j}$. This strategy is used to introduce relative measurements and statistical information on KNN. Therefore, it largely eliminates similar elements to accurately reflect the intrinsic relationships of the elements. Moreover, feature subsets rather than single features are of concern in variant LS. Therefore, it more completely contains local structures and actual interactions. Briefly, the new LS fully considers mutual influences between elements and features to improve the previous LS, so it is beneficial in unsupervised FS. As a result, the corresponding feature indicator and optimization extraction are presented in Eq. (7) of Definition 2, and they are mainly established by virtue of minimality and inversion. They naturally lead to the FILS algorithm and its validity. Note that abbreviation FILS is used to represent both the relevant mechanism and the feature extraction algorithm (Pang and Zhang 2020a).

3 Unsupervised feature selection based on incremental forward iterative Laplacian score (IFILS)

Based on observation, the FILS algorithm still has room for improvement with respect to unsupervised feature selection. Although each dynamic iteration necessarily contains two states, the beginning and ending states, this algorithm mainly addresses uncertainty measurements for the ending state in an iterative process. Therefore, not describing the beginning state causes both the basic defect of measurement incompleteness and the further possibility of FS improvement. In this section, the incremental technology and complete measurements are introduced to address the underlying measurement issue in the FILS algorithm, and thus, an improved FS algorithm (i.e., the IFILS algorithm) is proposed.

3.1 Feature significance in the incremental process and its granulation nonmonotonicity/uncertainty

Herein, a process measure for feature importance is first mined based on modified LS, and its granulation nonmonotonicity/uncertainty is further revealed.

Feature selection with forward iterations concerns the core process from A to $A \cup \{f_k\}$. Therefore, the additional feature $f_k \in \bar{A}$ should be optimally selected to efficiently increase A . The $A \rightarrow A \cup \{f_k\}$ procedure contains two states: the initial set A and the final extension $A \cup \{f_k\}$. In the FILS algorithm, the evaluation measurement in Eq. (7) only considers the final state $A \cup \{f_k\}$. Based on the broad treatment of the global $A \cup \{f_k\}$, it undoubtedly outperforms the single evaluation of the local $\{f_k\}$. However, this approach represents only the structural absoluteness of the process termination. Therefore, measurement weakness easily occurs. The initial state A inevitably exhibits subsequent iterative renewals; thus, the relevant relativity measurement should be added for integrated reinforcement, based on the concept of double quantification and its advantages (Zhang and Gou 2022). Therefore, a better structural mechanism that addresses both initial and final states is worth introducing to realize systematic extraction and comprehensive assessment. Herein, the process difference $\Delta(A \rightarrow A \cup \{f_k\})$ is considered, and thus, the process $A \rightarrow A \cup \{f_k\}$ is integrally estimated by $|J_{FILS}(A \cup \{f_k\}) - J_{FILS}(A)|$ or its related variants. The contrast difference or

difference operation has been extensively utilized in multiple disciplines, such as physics and mathematics, and its resultant feature significance has already profited the optimization selection and heuristic search in rough set analysis (Wang et al 2017; Zhang et al 2021; Zhang and Yao 2022).

Definition 3 (SIG) In information system $IS = (X, F)$, the significance (SIG) of the additional feature $f_k \in \bar{A} = F - A$ on feature subset $A \subseteq F$ is defined as

$$SIG_A(f_k) = J_{FILS}(A) - J_{FILS}(A \cup \{f_k\}), \tag{9}$$

and it is a new measure from the basic modified LS and related difference operation.

$SIG_A(f_k)$ utilizes the difference of modified LS to express a type of feature importance. Therefore, a better assessment of the increment process $A \rightarrow A \cup \{f_k\}$ with dynamics is obtained. The difference characteristic of SIG implies a general increment, and it has become a common and effective form. In later studies, SIG is utilized to optimize measurements and select features. As a basis and an extension, granulation nonmonotonicity/uncertainty, which is a fundamental property regarding knowledge cognition (Xu et al 2021; Zhang and Jiang 2022), is particularly discussed for the SIG metric. In fact, knowledge-based uncertainty is an inherent quality of intelligent information systems, and its relevant granulation monotonicity and nonmonotonicity have currently gained extensive attention in feature selection and attribute reduction (Stańczyk and Zielosko 2020; Zhang and Yao 2022; Gao et al 2019; Wang et al 2015, 2019). An inclusion relationship of feature subsets, such as $A \subseteq A' \subseteq F$, is theoretically required to discuss the granulation uncertainty and monotonicity/nonmonotonicity of SIG. In practice, researchers may focus more on a usual granulation chain with feature additions, i.e.,

$$A_1 = \{f_1\} \subset A_2 = \{f_1, f_2\} \subset \dots \subset A_n = \{f_1, \dots, f_n\} = F, \tag{10}$$

which concerns $n - 1$ incremental processes:

$$A_r \rightarrow A_{r+1} = A_r \cup \{f_{r+1}\} \quad (r = 1, 2, \dots, n - 1). \tag{11}$$

This totally ordered sequence from feature granulation can be effectively used to probe hierarchical knowledge structures and corresponding measurement manifestations.

Theorem 1 In information system $IS = (X, F)$ containing $A \subseteq A' \subseteq F$, $J_{FILS}(A) \leq J_{FILS}(A')$ does not necessarily hold, and the inequality $J_{FILS}(A) \geq J_{FILS}(A')$ may partially or completely emerge. In other words, the modified LS has granulation nonmonotonicity and uncertainty.

Corollary 1 For $A_1 \subset A_2 \subset \dots \subset A_n$ (Eq. (10)), if $1 \leq r_1 < r_2 \leq n$ or $A_{r_1} \subset A_{r_2}$, then neither $J_{FILS}(A_{r_1}) \leq J_{FILS}(A_{r_2})$ nor $J_{FILS}(A_{r_1}) \geq J_{FILS}(A_{r_2})$ necessarily holds. That is, all modified LS values on the feature-incremental chain, i.e.,

$$J_{FILS}(A_1), J_{FILS}(A_2), \dots, J_{FILS}(A_n), \tag{12}$$

have nonmonotonic and uncertain size changes.

Theorem 2 For $A_1 \subset A_2 \subset \dots \subset A_n$ (Eq. (10)), let $1 \leq r_1 < r_2 \leq n - 1$ so that $A_{r_1} \subset A_{r_2}$. Thus, Eqs. (9), (10) and (11) lead to the following formulas:

$$\begin{aligned} SIG_{A_{r_1}}(f_{r_1+1}) &= J_{FILS}(A_{r_1}) - J_{FILS}(A_{r_1+1}), \\ SIG_{A_{r_2}}(f_{r_2+1}) &= J_{FILS}(A_{r_2}) - J_{FILS}(A_{r_2+1}), \end{aligned} \tag{13}$$

$$\begin{aligned} SIG_{A_{r_1}}(f_k) &= J_{FILS}(A_{r_1}) - J_{FILS}(A_{r_1} \cup \{f_k\}), \\ SIG_{A_{r_2}}(f_k) &= J_{FILS}(A_{r_2}) - J_{FILS}(A_{r_2} \cup \{f_k\}), \end{aligned} \tag{14}$$

where $k \in \{r_2 + 1, \dots, n\}$ so that $f_k \in F - A_{r_2} \subseteq F - A_{r_1}$.

- (1) Based on Eq. (13), neither $SIG_{A_{r_1}}(f_{r_1+1}) \leq SIG_{A_{r_2}}(f_{r_2+1})$ nor $SIG_{A_{r_1}}(f_{r_1+1}) \geq SIG_{A_{r_2}}(f_{r_2+1})$ necessarily holds. This basic size result implies sequential size changes with granulation nonmonotonicity and uncertainty for $n - 1$ values:

$$SIG_{A_1}(f_2), SIG_{A_2}(f_3), \dots, SIG_{A_{n-1}}(f_n). \tag{15}$$

- (2) Based on Eq. (14), neither $SIG_{A_{r_1}}(f_k) \leq SIG_{A_{r_2}}(f_k)$ nor $SIG_{A_{r_1}}(f_k) \geq SIG_{A_{r_2}}(f_k)$ necessarily holds. This basic size result implies sequential size changes with granulation nonmonotonicity and uncertainty for general $r - 1$ values ($r \in \{2, 3, \dots, n\}$):

$$SIG_{A_1}(f_r), \dots, SIG_{A_{r-1}}(f_r), \tag{16}$$

as well as for special $n - 1$ values (where $r = n$) or n values.

$$\begin{aligned} &SIG_{A_1}(f_n), SIG_{A_2}(f_n), \dots, SIG_{A_{n-1}}(f_n), \\ &SIG_{A_1}(f_n), SIG_{A_2}(f_n), \dots, SIG_{A_{n-1}}(f_n), SIG_{A_n}(f_n), \end{aligned} \tag{17}$$

where $SIG_{A_n}(f_n) \equiv 0$ is additionally and reasonably stipulated.

Corollary 2 In information system $IS = (X, F)$, the measure $SIG_A(f_k)$ has nonmonotonic and uncertain size relationships in relation to the variable changes of subset A and feature f_k . In other words, let $A \subseteq A' \subseteq F$ and $1 \leq k \leq k' \leq n$. A pair of metric values may have no necessary sizes for four special pairs:

$$\begin{aligned} &(SIG_A(f_k), SIG_A(f_{k'})) \quad (\text{where } f_k, f_{k'} \in F - A), \\ &(SIG_A(f_k), SIG_{A'}(f_k)) \quad (\text{where } f_k \in F - A' \subseteq F - A), \\ &(SIG_A(f_k), SIG_{A'}(f_{k'})) \quad (\text{where } f_k \in F - A, f_{k'} \in F - A'), \\ &(SIG_A(f_{k'}), SIG_{A'}(f_k)) \quad (\text{where } f_{k'} \in F - A, f_k \in F - A'). \end{aligned} \tag{18}$$

Theorems 1, 2 and Corollaries 1, 2 present the basic nonmonotonicity and further uncertainty for two measures, mainly by the granulation changes of feature subsets. For these metric properties, Theorem 1 and Corollary 1 support modified LS, while Theorem 2 and Corollary 2 support SIG. As a comparison, Theorem 1 and Corollary 2 mainly focus on the direct change of $A \subseteq A'$, while Theorem 1 and Corollary 1 reveal the incremental chain $A_1 \subset A_2 \subset \dots \subset A_n$. The theoretical correctness, especially in terms of the relevant changes of Theorem 1 and Corollary 1 on the feature chain, will be verified through examples and experiments, as shown in Table 3, Fig. 2, Table 6, and Fig. 3. Since SIG is derived from modified LS, the nonmonotonicity and uncertainty of

modified LS induce the similar properties in SIG. In other words, the nonmonotonicity and uncertainty of SIG mainly originate from modified LS.

- In terms of the incremental chain, the chaos in Eq. (12) implies the qualitative difference of positive-negative numbers or the quantitative difference of values in Eqs. (13) and (15). Therefore, the SIG values in Eqs. (13) and (15) also become chaotic. A similar induction or explanation applies for the cases in Eqs. (14) and (16) (17).
- For a direct binary comparison, there are several basic cases of SIG for single and dual changes of the features and subsets, i.e., the four combined types in Eq. (18). Based on Eq. (18), $A \subseteq A'$ represents the knowledge granulation. However, $k \leq k'$ still implies the usual parallel relationship or selection. For a single change, the incomparability is naturally related to the nonmonotonicity/uncertainty in Theorem 1; furthermore, dual changes become more complex and also exhibit the incomparability, and their related derivations can be easily realized.

In summary, Theorem 1 describes the original and basic nonmonotonicity/uncertainty of modified LS, and Corollary 1 is derived from it. Corollary 1 generates Theorem 2, which describes the SIG nonmonotonicity/uncertainty on the incremental feature chain, and the latter is the main conclusion that will be observed and verified. Finally, Corollary 2 can be explained by Theorems 1 or 2. All these properties of granulation nonmonotonicity and uncertainty are in-depth. Therefore, they both enrich theoretical studies and underlie practical applications of modified LS and evolutive SIG for uncertainty characterization.

3.2 Increment-significance-based feature selection criterion and algorithm

In the above, feature significance (SIG) is established via modified LS, and its robust mechanism on the uncertainty measurements firmly underlies further applications, such as feature selection. Next, SIG is utilized to develop the corresponding optimization criterion and implementation algorithm for feature selection.

Definition 4 (IFILS) Eqs. (7) and (9) are combined to offer a new standard of feature selection, i.e.,

$$f_{k^o} = \arg \min_{f_k \in A} J_{IFILS}(f_k) = \arg \min_{f_k \in A} \left| 1 - \frac{SIG_A(f_k)}{J_{FILS}(F)} \right|. \tag{19}$$

Here, the metric integration of feature f_k is obtained by

$$J_{IFILS}(f_k) = \left| 1 - \frac{SIG_A(f_k)}{J_{FILS}(F)} \right|. \tag{20}$$

Algorithm 1 (The IFILS algorithm) Feature selection based on incremental forward iterative Laplacian score

Require: Information system $IS = (X, F)$, nearest neighbor number K , and required feature number r .

Ensure: Selective feature subset A with number r .

- 1: Initialize $A = \emptyset, F = \{f_1, f_2, \dots, f_n\}, \bar{A} = F - A$;
 - 2: Construct the nearest neighbor graph;
 - 3: Use Eq. (4) to compute weight matrix S ;
 - 4: Use Eq. (6) to compute LS of entire set F , i.e., $J_{FILS}(F)$;
 - 5: **while** $(|A| < r \wedge \bar{A} \neq \emptyset)$ **do**
 - 6: **for each** $f_k \in \bar{A}$ **do**
 - 7: Let $A_k = A \cup \{f_k\}$;
 - 8: Use Eq. (6) to compute both $J_{FILS}(A)$ on initial feature subset A and $J_{FILS}(A_k)$ on terminative union subset $A \cup \{f_k\}$;
 - 9: Use Eq. (9) to compute SIG of feature f_k , i.e., $SIG_A(f_k)$;
 - 10: Use Eq. (20) to compute $J_{IFILS}(f_k)$ on feature f_k ;
 - 11: **end for**
 - 12: Use Eq. (19) to sequentially find f_{k° with the minimum $J_{IFILS}(f_k)$;
 - 13: $A \leftarrow A \cup \{f_{k^\circ}\}$;
 - 14: $\bar{A} = F - A$;
 - 15: **end while**
 - 16: Return A .
-

In Definition 4, $SIG_A(f_k)$ is introduced in Eq. (7) to replace $J_{FILS}(A \cup \{f_k\})$. Therefore, an improved selection strategy is constructed through Eq. (19). The new feature index and extraction still use modified LS but carry the difference $SIG_A(f_k)$, and they adhere to the structural relativity and dynamic integrity of the entire incremental process. Hence, Definition 4 (or its Eq. (19)) guides the natural construction of the relevant IFILS algorithm, and corresponding Algorithm 1 has a procedural flowchart that is shown in Fig. 1.

For feature selection, Eqs. (7) and (19) have a large similarity to influence the FILS and IFILS algorithms, respectively. In Eqs. (7) and (19), $J_{FILS}(F)$ informationizes local structures of the original data to act as a contrast criterion; when adding feature f_k , measures $J_{FILS}(A \cup \{f_k\})$ and $SIG_A(f_k)$ imply the process-driven absolute and relative degrees, respectively, for maintaining the local structures of the data. In the FILS algorithm, if $J_{FILS}(A \cup \{f_k\})$ approaches $J_{FILS}(F)$, i.e., $J_{FILS}(f_k) \rightarrow 0$, then feature subset $A \cup \{f_k\}$ can represent the whole feature set F well; in other words, a lower amount of $J_{FILS}(f_k)$ implies more importance of feature f_k . Therefore $J_{FILS}(f_k)$ is minimized to generate the FILS algorithm. Similarly, the IFILS algorithm is designed by minimizing $J_{IFILS}(f_k)$; when facing multiple features, the feature with the smallest label is preferentially utilized. Clearly, the IFILS algorithm and its performance benefit from both the effective uncertainty measurement of SIG and the corresponding convergence treatment of incremental forward iteration.

Algorithm 1 mainly simulates but slightly changes the FILS algorithm. The FILS algorithm necessarily enters the loop calculation regardless of A 's condition satisfiability. In contrast, Algorithm 1 adopts the condition judgement to enter the loop, as shown in Step 5. This type of improvement reduces the running costs of a complete

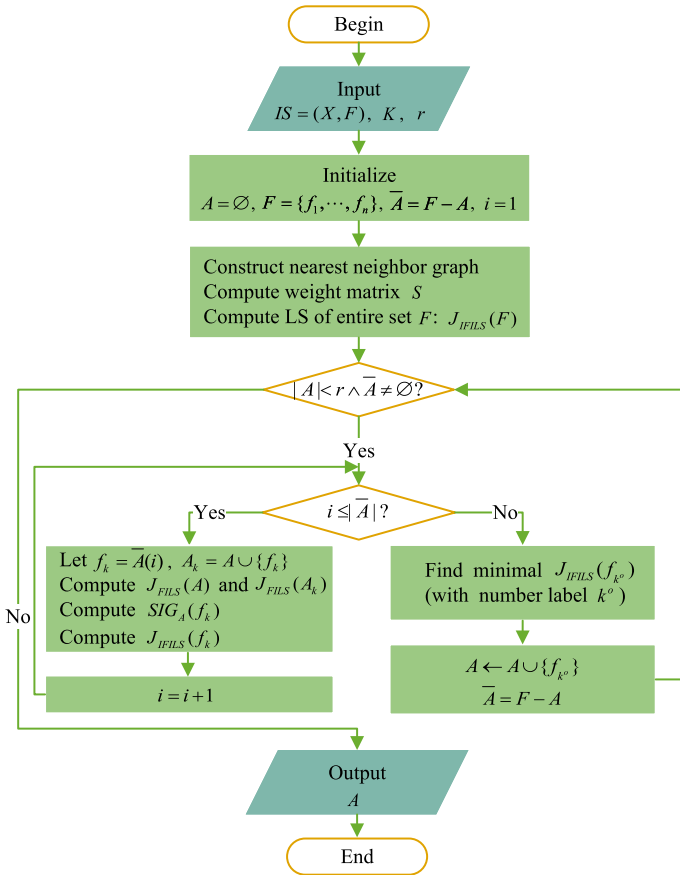


Fig. 1 Procedural flowchart of the IFILS algorithm

circulation. In terms of computational complexity, weight matrix S in Step 3 has a complexity of $O(u^2)$, $J_{FILS}(F)$ in Step 4 has a complexity of $O(nu^2)$, feature iterations on $h \in (0, n]$ in Steps 5–15 have a complexity of $O(2hu^2)$. Therefore, the total complexity is $O(n^2u^2)$. The IFILS algorithm has a partial complexity increase in contrast to the FILS algorithm. This increase mainly comes from the calculation of $SIG_A(f_k)$. In addition, the IFILS algorithm actually concerns the underlying calculation of modified LS only twice, and thus, it becomes feasible.

4 Illustrative example of relevant measures and algorithms

Here, a table example is provided to illustrate relevant measures and algorithms, mainly the granulation nonmonotonicity/uncertainty of the modified LS, SIG and processing contrast of the IFILS and FILS algorithms. For this purpose, a defined part of the University of California at Irvine (UCI) dataset, “Immunotherapy”, is extracted. This numeric dataset and 11 other experimental datasets all come from the UCI repository (<https://archive.ics.uci.edu/ml/index.php>), and their basic information is described in

Table 1 Basic descriptions of experimental ten UCI datasets

No.	Dataset name	Sample number	Feature number	Category number	Data type
(0)	Immunotherapy	90	7	2	Numeric
(1)	Blood	748	4	2	Numeric
(2)	Ccbr (Cervical cancer behavior risk)	72	19	2	Numeric
(3)	Ecoli	336	7	8	Numeric
(4)	Glass	214	9	6	Numeric
(5)	Hayes-roth	132	5	3	Numeric
(6)	Iris	150	4	3	Numeric
(7)	Lung cancer	32	56	3	Numeric
(8)	Wine	178	13	3	Numeric
(9)	COVID-19 (COVID-19 surveillance)	14	7	3	Nominal
(10)	Brain (Multi-view Brain Networks Data Set)	70	70	10	Numeric
(11)	Parkinsons	195	22	2	Numeric

Table 1. On this website, the “view all datasets” link or the “search” function can be directly utilized to acquire the relevant datasets. Moreover, some datasets are saved or opened in other locations. As a result, the data can be received in various manners through searches and downloads. For the Immunotherapy dataset, it can offer

$$DIS = (X = \{x_1, x_2, \dots, x_{90}\}, F = \{f_1, f_2, \dots, f_7\}, D = \{d\}),$$

which has seven selective features.

Max-min normalization is implemented for data preprocessing:

$$f_{ri} \leftarrow \frac{f_{ri} - \min_{i=1}^u f_{ri}}{\max_{i=1}^u f_{ri} - \min_{i=1}^u f_{ri}}. \tag{21}$$

In 10-fold cross-validation, ten random trials are run for performance and statistics assessments, and each trial contains both a 9-subset collection for training and one for testing. In a given case, only one primary treatment is focused on to better illustrate the algorithmic procedures and comparisons. After normalization, the 1st random subset for testing is

Table 2 Decision subsystem on the Immunotherapy dataset’s 1st sample subset

X^*	f_1	f_2	f_3	f_4	f_5	f_6	f_7	d
x_9	1.00	0.10	0.45	0.06	0.00	0.24	0.09	1
x_{10}	1.00	0.41	1.00	0.28	1.00	0.03	0.04	0
x_{14}	1.00	0.00	0.41	0.61	0.00	0.05	0.07	1
x_{19}	1.00	0.00	0.50	1.00	0.00	0.06	0.07	1
x_{50}	0.00	0.56	1.00	0.72	0.00	0.09	0.06	0
x_{58}	1.00	0.93	0.82	0.00	0.50	0.03	0.34	1
x_{64}	0.00	0.07	0.98	0.22	0.50	0.01	0.04	1
x_{78}	1.00	0.76	0.61	0.39	0.00	0.04	0.01	1
x_{90}	1.00	0.20	0.52	0.28	0.00	0.01	0.00	1

extracted, as shown in Table 2, where subuniverse X^* contains nine samples. Next, Table 2 is utilized to measure the observation and demonstrate the algorithm.

4.1 Measure observation

For the nonmonotonicity/uncertainty metric, the relevant results on incremental feature chains, i.e., the modified LS case in Eq. (12) of Corollary 1 and the SIG cases in Eqs. (15) and (17) of Theorem 2, are mainly considered. According to Table 2, the incremental feature chain is defined as $A_1 \subset A_2 \subset \dots \subset A_7$. Therefore, the relevant metric observations mainly concern three groups of measurement values, i.e.,

$$J_{FILS}(A_r), SIG_{A_r}(f_{r+1}), SIG_{A_r}(f_7) \quad (r = 1, 2, \dots, 7). \tag{22}$$

The first group of the modified LS and the two types of SIG are listed in Table 3 and depicted in Fig. 2.

Table 3 and Fig. 2 can be used to illustrate relevant measure definitions, calculations, and properties, and thus, the granulation nonmonotonicity/uncertainty in Eqs. (12) (15) (17) are mainly verified.

- The value sequence $J_{FILS}(A_r)$ ($r = 1, 2, \dots, 7$) shows a general increase. However, there is a decrease in the case when $J_{FILS}(A_3) = 0.1013 > 0.0488 = J_{FILS}(A_4)$. This phenomenon is in accordance with the granulation nonmonotonicity/uncertainty of the modified LS in Corollary 1.
- The sequence of $SIG_{A_r}(f_{r+1})$ ($r = 1, 2, \dots, 6$) contains only the positive value $SIG_{A_3}(f_4) = +0.0525$, which adheres to serial number 3, a middle number. Moreover, there is some fluctuation in the five negative values, such as $SIG_{A_4}(f_5) = -0.1006 < SIG_{A_3}(f_6) = -0.0251 > SIG_{A_6}(f_7) = -0.0468$. These two facts validate the granulation nonmonotonicity/uncertainty of SIG (in item 1) of Theorem 2.
- The sequence of $SIG_{A_r}(f_7)$ ($r = 1, 2, \dots, 7$) contains only negative numbers. However, quantitative differences and fluctuation are observed for these values. Thus, $SIG_{A_1}(f_7) = -0.0199 > SIG_{A_3}(f_7) = -0.0705 < SIG_{A_6}(f_7) = -0.0468$ fully reflects the granulation nonmonotonicity/uncertainty of SIG (in item 2) of Theorem 2.

4.2 Algorithm demonstration

For the Immunotherapy dataset, the training dataset is the complement of the 1st group of subsets, i.e., the opposite of the data in Table 2. Parameter values $K = 5, r = 4$ are

Table 3 Measurement values of the modified LS and SIG on incremental chain of Table 2 (the Immunotherapy dataset)

Measure	Feature-incremental chain $A_1 \subset A_2 \subset \dots \subset A_7$						
	A_1	A_2	A_3	A_4	A_5	A_6	A_7
$J_{FILS}(A_r)$	0.001	0.0054	0.1013	0.0488	0.1494	0.1745	0.2213
$SIG_{A_r}(f_{r+1})$	-0.0044	-0.0959	+0.0525	-0.1006	-0.0251	-0.0468	-
$SIG_{A_r}(f_7)$	-0.0199	-0.0503	-0.0705	-0.0700	-0.0488	-0.0468	0

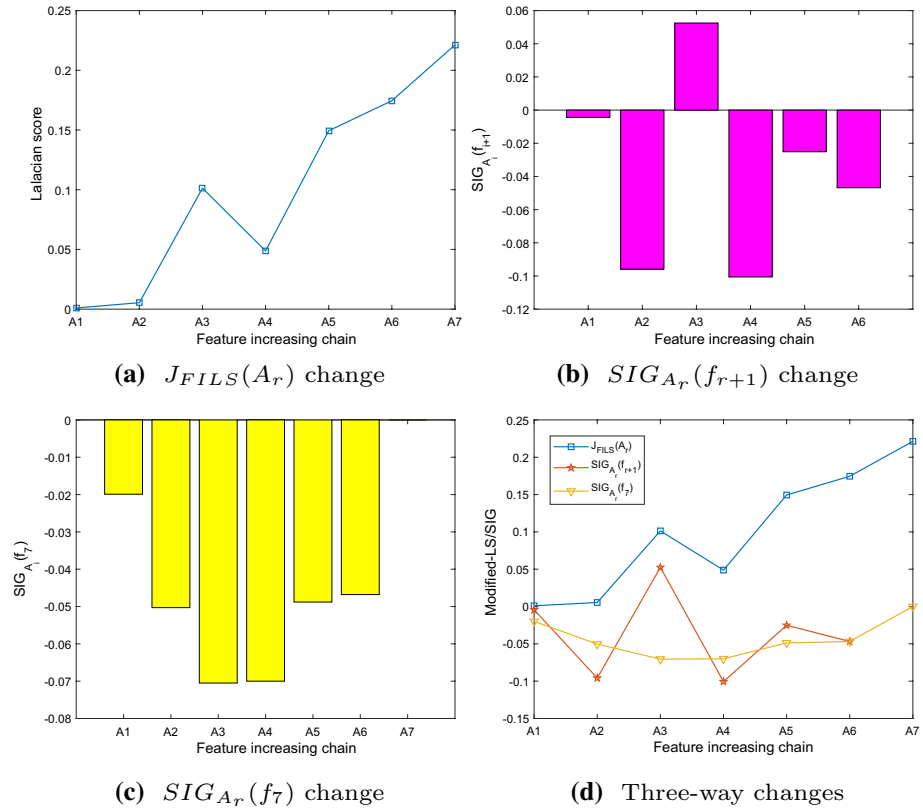


Fig. 2 Three-way non-monotonicity/uncertainty changes of the modified LS and SIG on incremental chain of Table 2 (the Immunotherapy dataset)

set due to algorithmic requirements. The FILS and IFILS algorithms have different selection procedures and results, as shown in Tables 4 and 5 (where $J_{FILS}(F) = 0.2243$), respectively. The FILS algorithm utilizes optimal ordering $f_6 \rightarrow f_7 \rightarrow f_4 \rightarrow f_3$ to obtain a feature subset $\{f_3, f_4, f_6, f_7\}$, while the IFILS algorithm adopts sequence $f_6 \rightarrow f_1 \rightarrow f_5 \rightarrow f_2$ to yield another subset $\{f_1, f_2, f_5, f_6\}$. There are some differences beyond circulation optimization between these two algorithms, e.g., the IFILS algorithm selects f_7 rather than f_1 in the 2nd round.

The subsets selected above, which are based on the unsupervised training, next work on the testing set (i.e., Table 2), and the accuracy of label prediction can be formulated by decision classification. Conclusively, the prediction resorts to the KNN classifier to pursue balanced connections with previous KNN processing approaches, such as Eq. (4), and the classifier that is used requires the same distance and K . In Table 2, $\{f_3, f_4, f_6, f_7\}$ in the FILS algorithm generates predicted labels $[1, 1, 1, 1, 1, 1, 1, 1, 1]$ on X^* , and a classification accuracy of $7/9 \approx 77.78\%$ is achieved when considering the real labels $[1, 0, 1, 1, 0, 1, 1, 1, 1]$; in contrast, a recognition rate of $9/9 = 100\%$ is achieved using $\{f_1, f_2, f_5, f_6\}$ in the IFILS algorithm, which predicts the labels $[1, 0, 1, 1, 0, 1, 1, 1, 1]$. Herein, the classification accuracy (or recognition rate) is defined as the percentage of correct predictions, i.e.,

Table 4 The FILS algorithm's procedure on the Immunotherapy dataset's 1st random sample subset ($K = 5, r = 4$)

No.	A	\bar{A}	$J_{FILS}(f_k) \forall k \in \bar{A}$	Selective feature	Meet loop condition ($ A < r) \wedge (\bar{A} \neq \emptyset$)?	Accuracy
(1)	\emptyset	$\{f_1, f_2, f_3, f_4, f_5, f_6, f_7\}$	$[0.9955, 0.9754, 0.9759, 0.9141, 0.9955, 0.4510, 0.8910]$	f_6	Yes	77.78%
(2)	$\{f_6\}$	$\{f_1, f_2, f_3, f_4, f_5, f_7\}$	$[0.8769, 0.5850, 0.5617, 0.4852, 0.7436, 0.4437]$	f_7	Yes	
(3)	$\{f_6, f_7\}$	$\{f_1, f_2, f_3, f_4, f_5\}$	$[0.7268, 0.3770, 0.3992, 0.2624, 0.6088]$	f_4	Yes	
(4)	$\{f_4, f_6, f_7\}$	$\{f_1, f_2, f_3, f_5\}$	$[0.5531, 0.1142, 0.1026, 0.3755]$	f_3	Yes	
(5)	$\{f_3, f_4, f_6, f_7\}$	$\{f_1, f_2, f_5\}$	$[0.3286, 0.1330, 0.2067]$	f_2	No	

Table 5 The IPILS algorithm's procedure on the Immunotherapy dataset's 1st random sample subset ($K = 5, r = 4$)

No.	A	\bar{A}	Meet loop condition ($ A < r$) \wedge ($\bar{A} \neq \emptyset$)?	$J_{IPILS}(f_k) (f_k \in \bar{A})$	Selective feature	Accuracy
(1)	\emptyset		Yes	[44.1750, 44.1549, 44.1554, 44.0937, 44.1750, 43.6305, 44.0705]	f_6	100%
(2)	$\{f_6\}$	$\{f_1, f_2, f_3, f_4, f_5, f_7\}$	Yes	[0.5741, 0.8660, 0.8892, 0.9657, 0.7073, 1.0072]	f_1	
(3)	$\{f_1, f_6\}$	$\{f_2, f_3, f_4, f_5, f_7\}$	Yes	[1.0814, 1.0928, 1.0789, 1.0355, 1.1500]	f_5	
(4)	$\{f_1, f_5, f_6\}$	$\{f_2, f_3, f_4, f_7\}$	Yes	[1.1643, 1.1698, 1.1687, 1.2348]	f_2	
(5)	$\{f_1, f_2, f_5, f_6\}$	$\{f_3, f_4, f_7\}$	No	–	–	

$$Accuracy = \frac{Object-number(Correct\ prediction)}{Object-number(Testing\ set)}. \quad (23)$$

The contrast accuracies reflect the validity and superiority of the IFILS algorithm, and the root comes from the SIG advancement on the uncertainty measurement.

In summary, Tables 4 and 5, which come from Table 2 as well as the surplus, thoroughly demonstrate the process and comparison of the FILS and IFILS algorithms. As a result, the measurement mechanism and recognition advantage of the IFILS algorithm are firmly validated.

5 Comparative experimental verification of unsupervised feature selection on the FILS and IFILS algorithms

Finally, data experiments are performed to further demonstrate the effectiveness and superiority of the IFILS algorithm, mainly in contrast to the FILS algorithm. Before this core treatment, the modified LS and SIG are also calculated so that their granulation nonmonotonicity/uncertainty is provided, mainly in terms of incremental feature chains. Three-way values are defined as follows:

$$[J_{FILS}(A_r), SIG_{A_r}(f_{r+1}), SIG_{A_r}(f_n)]. \quad (24)$$

These values are shown in Table 6 and are depicted in Fig. 3. For basic settings, only the 1st sample subset is chosen from the 10-fold cross-validation. The chain number follows $r = \min(7, n)$ for full verification, and the end feature f_n is adopted for the last metric observation of the fixed features. Thus, related results from Table 6 and Fig. 3 effectively validate all theoretical properties, especially the nonmonotonicity/uncertainty in Corollary 1 and Theorem 2.

In the above example, relevant experimental settings and mechanisms are detailed, and they include the dataset description in Table 1, data standardization in Eq. (21), 10-fold cross-validation, KNN classifier prediction, and recognition accuracy in Eq. (23). By deepening and enlarging the above example, more datasets, cross-validation, and parametric optimization are adopted to determine the statistical performance, where a fixed value $K = 5$ is mainly used (Pang and Zhang 2020a). Next, four subsections are respectively formed from the Immunotherapy dataset, eight-numeric datasets, the COVID-19 dataset, and two additional datasets (of medical diagnosis). As described in Table 1, all 12 datasets come from the UCI repository (<https://archive.ics.uci.edu/ml/index.php>). UCI datasets with real-world data have become authoritative and convenient for machine learning, and thus they are also generally utilized for studies on feature selection. Relevant experiments are related to the theoretical validation and practical application, and thus, they adopt more datasets around medical diagnosis, including 8 cases: Immunotherapy, Blood, Ccbr, Ecoli, Lung cancer, COVID-19, Brain, and Parkinsons. Regarding the operational environment, all experiments are performed in MATLAB R2021a and run with an Intel Core i7 CPU at 2.80 GHz, 8 GB RAM and 64-bit operating system.

5.1 The Immunotherapy dataset

The above example is obtained by performing experiments on the Immunotherapy dataset. The 1st group is fixed and $(K, r) = (5, 4)$. These parameters are further extended through

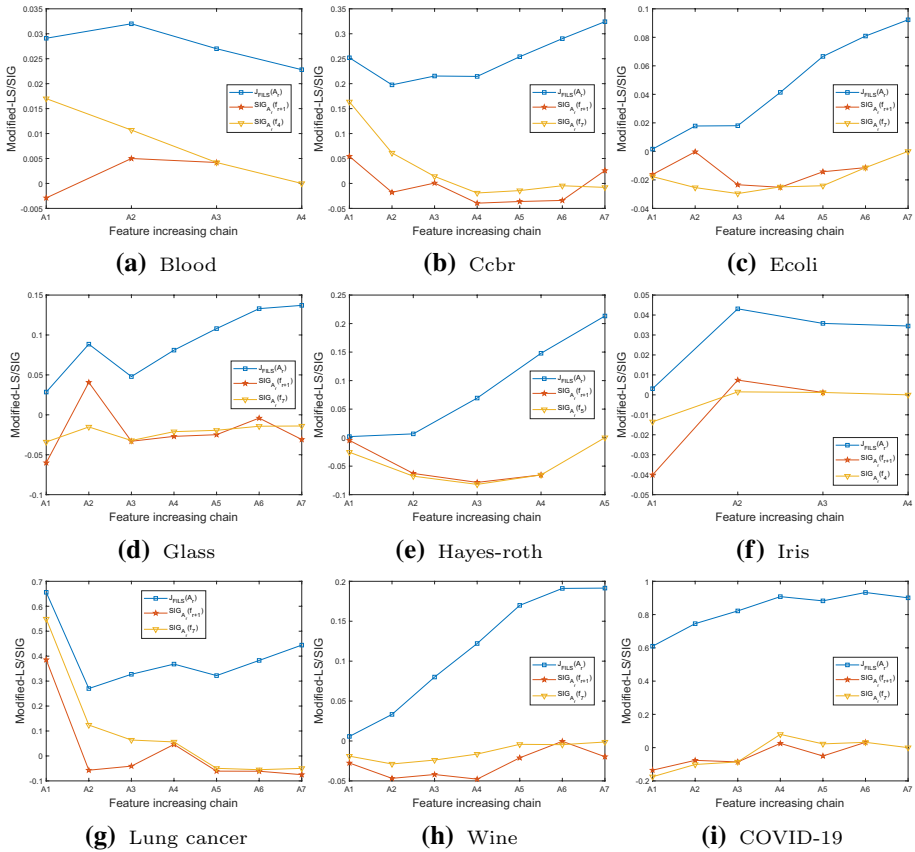


Fig. 3 $[J_{FILS}(A_r), SIG_{A_r}(f_{r+1}), SIG_{A_r}(f_r)]$ -based three-way value figures of the modified LS and SIG on incremental chains of nine UCI datasets

10-fold cross-validation and parameter changes. For the case where $K = 5$, Fig. 4 (a) shows the accuracies obtained on a 10×7 grid, and the IFILS algorithm occupies more dominant points with higher accuracies. Fig. 4 (b) shows the 10-fold average accuracy on feature chain $A_1 \subset A_2 \subset \dots \subset A_7$ with representative subset A_r , and the highest accuracy on each feature number r is achieved using the IFILS algorithm. Hence, the optimal accuracy 86.67% ($r = 4$) achieved by the IFILS algorithm is better than the optimal accuracy 77.78% ($r = 4, 5, 6$) achieved by the FILS algorithm. In terms of the Immunotherapy dataset, the IFILS algorithm outperforms the FILS algorithm, and the optimal accuracies of the two methods are mainly realized by middle feature numbers.

K is the surplus parameter for change analysis, and it is next extended to $\{1, 2, \dots, 10\}$. The K analysis for only the IFILS algorithm is sufficient and reasonable because this new algorithm is improved and more representative. Table 7 shows the 10-fold average accuracies and standard deviations on the (K, r) -grid, and it also shows the edge statistics of K and r optimization. For clarity, Fig. 5(a) shows the average accuracies, and the highest accuracy of 86.67% is obtained when $(K, r) = (5, 4)$. Figure 5(b) shows the cut accuracies (with standard deviations) when $r = 4$. When the K value increases, the mean accuracies first change from the lowest accuracy 57.78% ($K = 1$) to the highest 86.67%

Table 6 [$J_{FHS}(A_p)$, $SI_{G_A}(f_{t+1})$, $SI_{G_A}(f_t)$]-based three-way value table of the modified LS and SIG on incremental chains of nine UCI datasets

No.	Dataset	A_1	A_2	A_3	A_4	A_5	A_6	A_7
(1)	Blood	[0.0291, - 0.0029, + 0.0171]	[0.032, + 0.005, + 0.0107]	[0.027, + 0.0042, + 0.0042]	[0.0228, -, 0]	-	-	-
(2)	Ccbr	[0.2519, + 0.0543, + 0.1636]	[0.1976, - 0.0178, + 0.0611]	[0.2153, + 0.0008, + 0.0141]	[0.2145, - 0.0395, - 0.0191]	[0.254, - 0.0362, - 0.0143]	[0.2902, - 0.034, - 0.0046]	[0.3242, + 0.0258, - 0.0079]
(3)	Ecoli	[0.0016, - 0.0162, - 0.0177]	[0.0178, - 0.0003, - 0.0254]	[0.018, - 0.0234, - 0.0296]	[0.0414, - 0.0252, - 0.0249]	[0.0666, - 0.0143, - 0.0241]	[0.0809, - 0.0114, - 0.0114]	[0.0923, -, 0]
(4)	Glass	[0.0285, - 0.0601, - 0.034]	[0.0885, + 0.0406, - 0.0152]	[0.0479, - 0.0331, - 0.0322]	[0.081, - 0.027, - 0.0211]	[0.108, - 0.0249, - 0.0194]	[0.133, - 0.0041, - 0.0143]	[0.1371, - 0.0309, - 0.014]
(5)	Hayes-roth	[0.0018, - 0.0048, - 0.0256]	[0.0066, - 0.0628, - 0.0676]	[0.0694, - 0.0785, - 0.082]	[0.1479, - 0.0654, - 0.0654]	[0.2133, -, 0]	-	-
(6)	Iris	[0.0031, - 0.0401, - 0.0135]	[0.0431, + 0.0074, + 0.0015]	[0.0358, + 0.0012, + 0.0012]	[0.0345, -, 0]	-	-	-
(7)	Lung cancer	[0.656, + 0.3857, + 0.549]	[0.2703, - 0.0571, + 0.1239]	[0.3274, - 0.0408, + 0.0638]	[0.3682, + 0.0462, + 0.0562]	[0.3221, - 0.0609, - 0.0496]	[0.3829, - 0.0613, - 0.0551]	[0.4443, - 0.075, - 0.0498]
(8)	Wine	[0.0057, - 0.0276, - 0.019]	[0.0333, - 0.0468, - 0.0288]	[0.0801, - 0.042, - 0.0239]	[0.1221, - 0.0478, - 0.0164]	[0.1699, - 0.0211, - 0.0042]	[0.1911, - 0.0004, - 0.0046]	[0.1915, - 0.0197, - 0.0013]
(9)	COVID-19	[0.6099, - 0.1356, - 0.175]	[0.7455, - 0.0765, - 0.1013]	[0.822, - 0.0859, - 0.0859]	[0.9079, + 0.0254, + 0.0798]	[0.8825, - 0.0503, + 0.0225]	[0.9328, + 0.032, + 0.032]	[0.9008, -, 0]

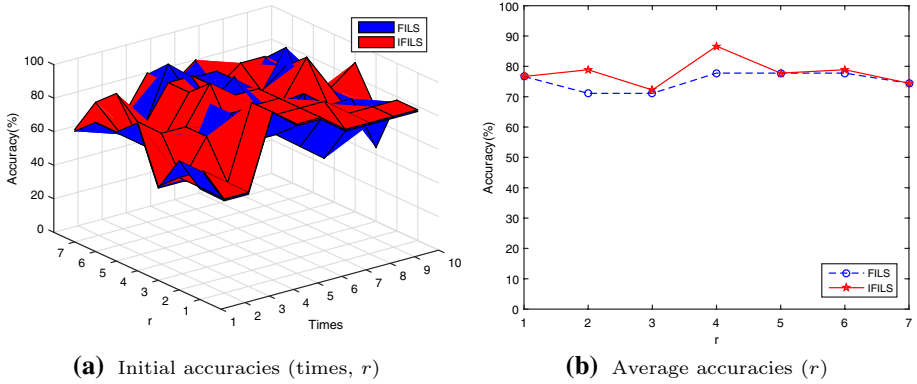


Fig. 4 The FILS and IFILS algorithms' accuracies on the Immunotherapy dataset ($K = 5$)

Table 7 The IFILS algorithm's 10-fold-average accuracies and standard deviations on the Immunotherapy dataset (with K, r)

r	K					
	1	2	3	4	5	6
1	62.22 ± 13.04	52.22 ± 13.91	72.22 ± 12.00	62.22 ± 15.89	76.67 ± 14.30	73.33 ± 11.94
2	63.33 ± 12.88	51.11 ± 11.94	67.78 ± 9.73	55.56 ± 9.07	78.89 ± 13.30	73.33 ± 11.94
3	63.33 ± 17.41	55.56 ± 17.37	66.67 ± 15.71	65.56 ± 15.23	72.22 ± 12.00	65.56 ± 12.23
4	57.78 ± 21.47	64.44 ± 14.63	71.11 ± 15.89	68.89 ± 8.76	↑ 86.67 ± 8.76	71.11 ± 10.73
5	75.56 ± 11.97	68.89 ± 19.12	74.45 ± 12.22	72.23 ± 10.24	77.78 ± 13.15	75.56 ± 12.96
6	↑ 78.89 ± 12.62	↑ 75.56 ± 15.56	↑ 77.78 ± 8.61	↑ 75.56 ± 6.67	78.89 ± 10.48	77.78 ± 13.15
7	74.45 ± 12.22	72.23 ± 10.24	75.56 ± 6.67	74.45 ± 5.09	74.45 ± 8.68	↑ 77.78 ± 11.11
r optimization	↑ 78.89 ± 12.62	↑ 75.56 ± 15.56	↑ 77.78 ± 8.61	↑ 75.56 ± 6.67	↑ 86.67 ± 8.76	↑ 77.78 ± 11.11

r	K				K optimization
	7	8	9	10	
1	77.78 ± 12.83	76.67 ± 12.23	77.78 ± 12.83	76.67 ± 14.30	77.78 ± 12.83
2	78.89 ± 13.30	76.67 ± 13.30	↑ 78.89 ± 13.30	↑ 78.89 ± 13.30	78.89 ± 13.30
3	77.78 ± 13.86	72.22 ± 13.09	76.67 ± 14.30	72.22 ± 17.57	77.78 ± 13.86
4	75.56 ± 12.61	77.78 ± 14.05	77.78 ± 14.91	75.56 ± 14.74	↑ 86.67 ± 8.76
5	↑ 81.11 ± 14.10	↑ 81.11 ± 14.10	77.78 ± 14.05	73.34 ± 14.23	81.11 ± 14.10
6	76.67 ± 14.44	76.67 ± 12.62	74.45 ± 13.19	75.56 ± 12.96	78.89 ± 10.48
7	78.89 ± 13.56	76.67 ± 10.48	76.67 ± 10.48	75.56 ± 12.96	78.89 ± 13.56
r optimization	↑ 81.11 ± 14.10	↑ 81.11 ± 14.10	↑ 78.89 ± 13.30	↑ 78.89 ± 13.30	↑ 86.67 ± 8.76

($K = 5$), and then they decrease to a stabilized value ($K > 5$). Figure 5(c) and (d) show the optimization statistics on K and r , respectively, and the optimal accuracies adhere to the middle values when $K = 5$ and $r = 4$. In summary, $K = 5$ is appropriate for the IFILS algorithm as well as the FILS algorithm (Pang and Zhang 2020a).

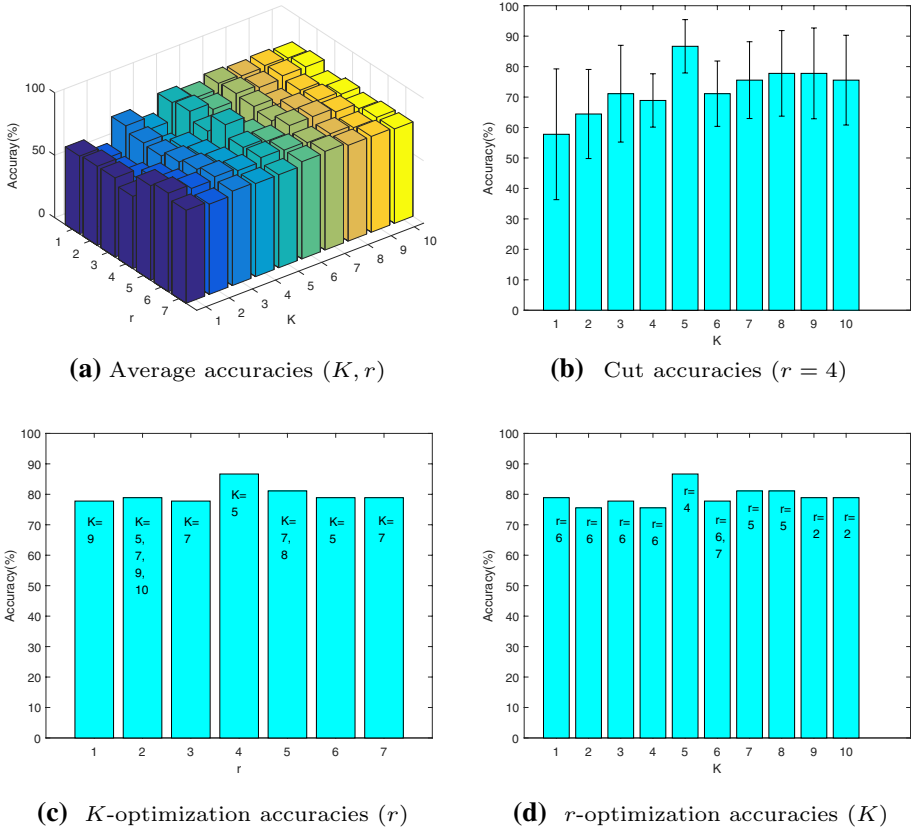


Fig. 5 The FILS algorithm's accuracies on the Immunotherapy dataset (K, r)

5.2 Eight numeric datasets

By referring to the above experiments on the Immunotherapy dataset, eight extra numeric datasets are implemented using the FILS and IFILS algorithms, and they are compared. Except for a 10-fold cross-validation, parameter r concerns an equidistant feature sequence. In terms of classifiers, KNN is still adopted, and $K = 5$ is suitable and fixed; moreover, classifiers CART and SVM are directly added for algorithmic comparison. In comparative experiments, classification accuracies are focused on for the following three dimensions: 10-fold cross-validation, selection number r , and classifier, and the main results are provided after necessary integrated processing.

First, as shown in Fig. 6, the aim is to provide details on 10-fold cross-validation, and r is optimized in the range $|r| < F$ to achieve the highest accuracy. In other words, Fig. 6 shows the r -optimal accuracies on 10-fold cross-validation by using three classifiers. By observation, the IFILS algorithm occupies more 10-fold points with higher accuracies in terms of KNN, CART, SVM. Therefore, it has superiority over the FILS algorithm from the perspective of 10-fold cross-validation.

In general, 10-fold cross-validation is mainly used for integrated statistics. Therefore, the 10-fold dimension is naturally reduced. Based on 10-fold statistics, Tables 8, 9, 10

Table 8 The FILS and IFILS algorithms' 10-fold-average accuracies and r -optimization statistics on eight UCI datasets based on KNN

No.	Dataset	Algorithm	r								Optimization on feature proper subset ($r < F $)	
(1)	Blood	FILS	1	2	3	4	-	-	-	-		
			70.05 ± 3.94	73.53 ± 4.75	75.94 ± 5.72	76.87 ± 5.71	-	-	-	76.87 ± 5.71 (4)	75.94 ± 5.72 (3)	
(2)	Cecbr	IFILS	1	4	7	10	13	16	19	-	-	
			70.05 ± 3.94	75.00 ± 6.73	76.75 ± 5.96	76.87 ± 5.71	-	-	-	-	76.87 ± 5.71 (4)	76.75 ± 5.96 (3)
(3)	Ecoli	FILS	1	2	3	4	5	6	7	-	-	
			29.29 ± 24.57	76.43 ± 16.10	77.68 ± 21.02	79.29 ± 20.00	81.96 ± 16.42	83.39 ± 17.38	87.68 ± 13.79	87.68 ± 13.79	87.68 ± 13.79 (19)	83.39 ± 17.38 (16)
(4)	Glass	IFILS	1	2	3	4	5	6	7	-	-	
			29.29 ± 24.57	65.36 ± 16.56	82.14 ± 14.68	91.96 ± 11.33	89.11 ± 14.30	87.68 ± 13.79	87.68 ± 13.79	-	91.96 ± 11.33 (10)	
(5)	Hayes-roth	IFILS	1	2	3	4	5	6	7	-	-	
			43.15 ± 3.62	49.09 ± 6.66	50.87 ± 8.85	62.75 ± 10.21	80.92 ± 10.24	84.78 ± 6.95	86.28 ± 5.17	86.28 ± 5.17	86.28 ± 5.17 (7)	84.78 ± 6.95 (6)
(6)	Iris	IFILS	1	2	3	4	5	6	7	-	-	
			49.39 ± 7.54	50.56 ± 7.06	63.10 ± 6.23	77.40 ± 5.17	83.32 ± 4.72	84.80 ± 5.20	86.28 ± 5.17	86.28 ± 5.17	86.28 ± 5.17 (7)	84.80 ± 5.20 (6)
(7)	Iris	IFILS	1	2	3	4	5	6	7	-	-	
			35.54 ± 9.80	57.92 ± 10.04	62.14 ± 10.83	65.35 ± 10.17	64.48 ± 8.71	-	-	-	-	65.35 ± 10.17 (7)
(8)	Iris	IFILS	1	2	3	4	5	6	7	-	-	
			34.11 ± 7.94	52.81 ± 11.10	66.77 ± 8.50	67.23 ± 9.34	64.48 ± 8.71	-	-	-	-	67.23 ± 9.34 (7)
(9)	Iris	IFILS	1	2	3	4	5	6	7	-	-	
			42.53 ± 9.99	34.18 ± 11.16	44.67 ± 10.90	45.44 ± 10.51	36.43 ± 10.28	-	-	-	-	45.44 ± 10.51 (4)
(10)	Iris	IFILS	1	2	3	4	5	6	7	-	-	
			42.53 ± 9.99	54.34 ± 14.39	56.87 ± 10.36	71.15 ± 13.42	36.43 ± 10.28	-	-	-	-	71.15 ± 13.42 (4)
(11)	Iris	IFILS	1	2	3	4	5	6	7	-	-	
			52.67 ± 12.75	88.67 ± 8.92	92.00 ± 8.20	96.00 ± 5.62	-	-	-	-	96.00 ± 5.62 (4)	92.00 ± 8.20 (3)
(12)	Iris	IFILS	1	2	3	4	5	6	7	-	-	
			52.67 ± 12.75	96.00 ± 5.62	96.67 ± 4.71	96.00 ± 5.62	-	-	-	-	-	96.67 ± 4.71 (3)

Table 8 (continued)

No.	Dataset	Algorithm	r	1	10	19	28	37	46	56	Optimization realization on feature set ($r = F $)	Optimization on feature proper subset ($r < F $)
(7)	Lung cancer	FILS	1	27.50 ± 27.79	23.33 ± 31.62	27.50 ± 35.15	34.17 ± 28.45	53.33 ± 24.91	47.50 ± 28.34	42.50 ± 27.34	–	53.33 ± 24.91 (37)
			IFILS	27.50 ± 27.79	55.83 ± 22.92	50.00 ± 22.22	54.17 ± 32.22	53.33 ± 24.91	41.67 ± 23.57	42.50 ± 27.34	–	55.83 ± 22.92 (10)
(8)	Wine	FILS	1	49.48 ± 12.10	68.01 ± 11.02	84.35 ± 9.02	95.49 ± 2.38	94.35 ± 3.82	94.90 ± 4.22	96.08 ± 3.77	96.08 ± 3.77 (13)	95.49 ± 2.38 (7)
			IFILS	49.48 ± 12.10	83.76 ± 12.66	91.01 ± 9.52	92.68 ± 7.46	94.90 ± 4.22	95.46 ± 3.67	96.08 ± 3.77	96.08 ± 3.77 (13)	95.46 ± 3.67 (11)

Table 9 The FILS and IFILS algorithms' 10-fold-average accuracies and r -optimization statistics on eight UCI datasets based on CART

No.	Dataset	Algorithm	r	1	2	3	4	—	—	—	—	—	Optimization on feature proper subset ($r < F $)	
(1)	Blood	FILS	75.95 ± 5.76	73.53 ± 4.22	74.87 ± 5.34	74.87 ± 5.34	—	—	—	—	—	—	75.95 ± 5.76 (1)	
		IFILS	75.95 ± 5.76	75.68 ± 5.56	75.68 ± 4.66	74.87 ± 5.34	—	—	—	—	—	—	—	75.95 ± 5.76 (1)
(2)	Cabr	FILS	74.82 ± 17.66	70.71 ± 10.25	68.04 ± 13.37	77.68 ± 12.06	84.82 ± 10.14	83.39 ± 17.37	80.54 ± 16.67	—	—	—	—	84.82 ± 10.14 (13)
		IFILS	74.82 ± 17.66	74.64 ± 17.83	79.11 ± 12.08	83.39 ± 8.67	86.25 ± 13.10	86.25 ± 11.23	80.54 ± 16.67	—	—	—	—	86.25 ± 13.10 (13)
(3)	Ecoli	FILS	43.44 ± 3.56	45.81 ± 4.44	47.93 ± 5.87	58.33 ± 7.86	74.06 ± 10.23	76.13 ± 7.78	80.60 ± 8.67	80.60 ± 8.67	80.60 ± 8.67	80.60 ± 8.67	80.60 ± 8.67	80.60 ± 8.67 (7)
		IFILS	55.34 ± 6.92	46.11 ± 4.41	59.82 ± 6.99	73.51 ± 3.80	78.83 ± 8.20	80.02 ± 9.50	80.60 ± 8.67	80.60 ± 8.67	80.60 ± 8.67	80.60 ± 8.67	80.60 ± 8.67	80.60 ± 8.67 (6)
(4)	Glass	FILS	45.26 ± 16.10	63.98 ± 11.92	68.33 ± 10.32	68.64 ± 12.23	67.36 ± 9.80	—	—	—	—	—	—	68.64 ± 12.23 (7)
		IFILS	45.26 ± 16.10	58.27 ± 10.98	68.64 ± 10.79	68.77 ± 10.91	67.36 ± 9.80	—	—	—	—	—	—	68.77 ± 10.91 (7)
(5)	Hayes-roth	FILS	46.98 ± 12.90	45.44 ± 11.85	46.15 ± 15.28	59.07 ± 9.04	77.15 ± 14.58	—	—	—	—	—	—	77.15 ± 14.58 (5)
		IFILS	46.98 ± 12.90	53.74 ± 11.85	65.17 ± 9.60	77.15 ± 14.58	77.15 ± 14.58	—	—	—	—	—	—	77.15 ± 14.58 (5)
(6)	Iris	FILS	51.33 ± 8.34	84.00 ± 10.04	94.00 ± 2.11	95.33 ± 4.50	—	—	—	—	—	—	—	95.33 ± 4.50 (4)
		IFILS	51.33 ± 8.34	94.00 ± 5.84	96.00 ± 4.66	95.33 ± 4.50	—	—	—	—	—	—	—	96.00 ± 4.66 (3)
(7)	Lung cancer	FILS	45.00 ± 18.92	45.00 ± 31.48	30.83 ± 25.47	35.00 ± 37.02	50.83 ± 39.37	47.50 ± 32.88	49.17 ± 32.50	—	—	—	—	50.83 ± 39.37 (37)
		IFILS	45.00 ± 18.92	45.83 ± 23.33	64.17 ± 24.23	56.67 ± 21.08	56.67 ± 21.08	62.50 ± 29.72	49.17 ± 2.50	—	—	—	—	64.17 ± 24.23 (19)

Table 9 (continued)

No.	Dataset	Algorithm	r						Optimization on feature proper subset ($r < F $)	Optimization on feature realization on feature set ($r = F $)
(8)	Wine		1	3	5	7	9	11	13	
		FILS	45.59 ± 9.12	61.83 ± 10.30	76.99 ± 11.79	89.90 ± 5.09	90.46 ± 10.49	88.72 ± 7.94	88.76 ± 6.93	–
		IFILS	45.59 ± 9.12	83.66 ± 9.35	87.03 ± 9.95	91.57 ± 5.42	90.42 ± 5.93	89.31 ± 6.67	88.76 ± 6.93	–

Table 10 The FILS and IFILS algorithms' 10-fold-average accuracies and r -optimization statistics on eight UCI datasets based on SVM

No.	Dataset	Algorithm	r		Accuracy		r -optimization statistics		Optimization real- ization on feature set ($r = F $)	Optimization on feature proper subset ($r < F $)	
(1)	Blood	FILS	1	2	3	4	–	–	–	–	
		IFILS	1	2	3	4	76.21 ± 6.05	76.48 ± 6.85	76.62 ± 6.85	76.62 ± 6.85	76.48 ± 6.85(3)
(2)	Cecbr	FILS	1	4	7	10	76.21 ± 6.05	76.75 ± 6.88	76.62 ± 6.85	–	76.75 ± 6.88(3)
		IFILS	1	2	3	4	73.57 ± 19.30	77.50 ± 25.52	90.54 ± 14.63	89.11 ± 14.30	88.93 ± 13.06(10)
(3)	Ecoli	FILS	1	2	3	4	73.57 ± 19.30	77.50 ± 25.52	90.54 ± 14.63	89.11 ± 14.30	88.93 ± 13.06
		IFILS	1	2	3	4	43.15 ± 3.62	47.60 ± 3.93	53.89 ± 8.66	65.15 ± 4.46	85.69 ± 6.46(7)
(4)	Glass	FILS	1	3	5	7	50.85 ± 4.21	48.78 ± 3.53	64.29 ± 4.43	78.86 ± 7.35	85.71 ± 5.21(6)
		IFILS	1	2	3	4	41.08 ± 13.08	42.60 ± 10.37	56.45 ± 13.64	61.19 ± 9.16	61.64 ± 8.80(9)
(5)	Hayes-roth	FILS	1	2	3	4	38.70 ± 11.95	46.86 ± 13.57	59.37 ± 8.66	65.39 ± 8.79	61.64 ± 8.80
		IFILS	1	2	3	4	34.29 ± 13.47	33.52 ± 10.06	46.04 ± 11.70	58.41 ± 15.58	63.08 ± 13.57(5)
(6)	Iris	FILS	1	2	3	4	34.29 ± 13.47	47.86 ± 15.00	56.87 ± 8.24	73.46 ± 6.55	63.08 ± 13.57(4)
		IFILS	1	2	3	4	54.67 ± 15.01	83.33 ± 13.79	84.67 ± 9.46	90.00 ± 11.44	90.00 ± 11.44(4)
(7)	Lung cancer	FILS	1	10	19	28	54.67 ± 15.01	92.67 ± 5.84	92.00 ± 6.88	90.00 ± 11.44	–
		IFILS	1	2	3	4	45.00 ± 18.92	38.33 ± 21.94	48.33 ± 37.02	44.17 ± 27.79	46.67 ± 24.91
(8)	Wine	FILS	1	3	5	7	45.00 ± 18.92	60.00 ± 31.87	56.67 ± 21.08	50.83 ± 17.77	50.83 ± 28.45
		IFILS	1	2	3	4	45.00 ± 18.92	60.00 ± 31.87	56.67 ± 21.08	50.83 ± 17.77	50.83 ± 28.45
(9)	Wine	FILS	1	3	5	7	44.38 ± 9.45	64.12 ± 9.95	84.38 ± 9.34	94.35 ± 5.32	97.71 ± 4.05
		IFILS	1	2	3	4	44.38 ± 9.45	64.12 ± 9.95	84.38 ± 9.34	94.35 ± 5.32	97.71 ± 4.05

basically summarize the average accuracies and standard deviations of KNN, CART and SVM. Furthermore, optimal or suboptimal accuracies from r statistics are reflected in the last two columns, and the bolded text highlights two-algorithmic comparisons. Moreover, the 10-fold statistical averages of Tables 8, 9, 10 are equivalently depicted in Fig. 7 to vividly clarify the relevant advancement of the IFILS algorithm. Based on comparisons in the table and figure, the IFILS algorithm generally has greater proportions of dominant features with higher accuracies for each dataset and the whole dataset. When r increases, the accuracy first rapidly increases and then gradually increases or decreases. Therefore, the optimal accuracies are usually realized by the final or middle feature numbers. Regarding the entire feature set F , the highest accuracies on the Blood, Ecoli and Wine datasets are achieved using both algorithms. However, the IFILS algorithm can make an additional optimization realization on the Ccbr and Iris datasets, from the perspective of KNN. The surplus observations of the CART and SVM classifiers can lead to similar analyses and results. In terms of appropriate feature subsets, the IFILS algorithm can achieve not only higher optimization accuracies but also less or an equal number of features for the three classifiers on almost all datasets, and the latter results imply practical optimization with dimensionality reduction. The sole exception for internal optimization comes from Wine dataset when using the KNN and SVM classifiers. For the Wine dataset on KNN (Table 8), optimal accuracies of 95.49% on $r = 7$ and 95.46% on $r = 11$ are achieved using the FILS and IFILS algorithms, respectively; although $95.49\% > 95.46\%$, these two values are very close to never become significantly different; moreover, the highest accuracy of 96.08% on the entire feature set F (where $r = 13$) is actually achieved using both algorithms.

Tables 8, 9, 10 mainly come from the 10-fold statistics, and their accuracy values on r -optimal selection (where $r < |F|$) are extracted and summarized in Table 11. Moreover, Table 11 shows a record of the running time of the two algorithms. Therefore, Table 11 provides a comprehensive platform for macroscopically comparing the FILS and IFILS algorithms based on KNN, CART, SVM, and the bottom average results on 10-dataset statistics play an important role. Through observation and analysis, the dominant position of the IFILS algorithm in contrast to the FILS algorithm can be concluded by three assessment indicators, i.e., the reduction length, classification accuracy, and consumption time. In terms of the main indices, the IFILS algorithm exhibits a general advantage in terms of both the length and accuracy. Moreover, its execution time is only slightly greater than that of the FILS algorithm, and the time difference is not large, which implies that these algorithms are on the same level. In other words, the IFILS algorithm has great superiority in pursuing feature simplification and recognition accuracy, and its time is acceptable when compared to the FILS algorithm.

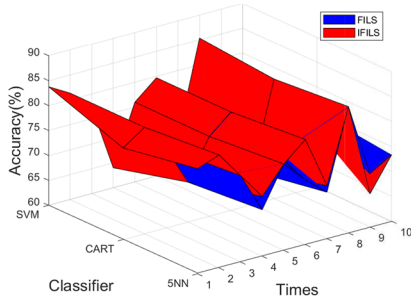
In summary, eight datasets based on the three classifiers provide information on the full algorithmic verification from the correspondence and optimization perspectives, and the IFILS algorithm generally outperforms the FILS algorithm and achieves better classification performance.

5.3 COVID-19 surveillance: a nominal dataset

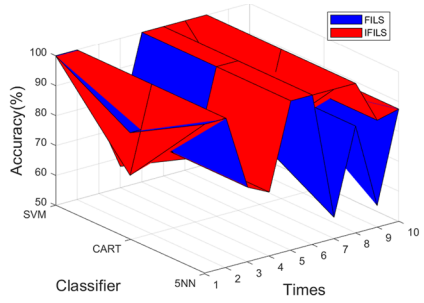
The FILS and IFILS algorithms can also be applied to nominal datasets, and the COVID-19 dataset shown in Table 1 is provided. This important medical case can further reveal the superiority of the new algorithm. There exist only three different operations, i.e., nominal-numeric transformation, cross-validation correction, and parameter K determination. Two qualitative symptoms, i.e., “+,” of seven features naturally

Table 11 The FILS and IFILS algorithms' r -optimal ($r < |F|$) three-way indexes (i.e., number, accuracy, time) on eight UCI datasets based on KNN, CART, SVM

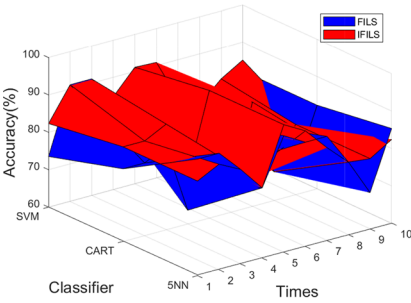
No.	Dataset	Algorithm	KNN			CART			SVM		
			Number	Accuracy (%)	Time (s)	Number	Accuracy (%)	Time (s)	Number	Accuracy (%)	Time (s)
1	Blood	FILS	3	75.94 ± 5.72	2234.34 ± 60.75	1	75.95 ± 5.76	856.29 ± 46.24	3	76.48 ± 6.85	2234.67 ± 60.76
		IFILS	3	76.75 ± 5.96	2270.09 ± 48.89	1	75.95 ± 5.76	878.86 ± 57.67	3	76.75 ± 6.88	2270.47 ± 48.91
2	Cabr	FILS	16	83.39 ± 17.38	67.22 ± 2.09	13	84.82 ± 10.14	75.49 ± 1.54	16	88.93 ± 11.19	67.31 ± 2.09
		IFILS	10	91.96 ± 11.33	87.64 ± 2.89	13	86.25 ± 13.10	103.24 ± 3.19	10	90.54 ± 14.63	87.71 ± 2.88
3	Ecoli	FILS	6	84.78 ± 6.95	455.74 ± 25.63	6	76.13 ± 7.78	455.82 ± 25.63	6	83.01 ± 7.19	455.92 ± 25.63
		IFILS	6	84.80 ± 5.20	655.07 ± 28.10	6	80.02 ± 9.50	655.15 ± 28.09	6	85.71 ± 5.21	655.23 ± 28.09
4	Glass	FILS	7	65.35 ± 10.17	234.02 ± 5.81	7	68.33 ± 10.32	234.11 ± 5.82	7	61.19 ± 9.16	234.16 ± 5.82
		IFILS	7	67.23 ± 9.34	318.13 ± 2.28	7	68.77 ± 10.91	318.23 ± 2.29	7	65.39 ± 8.79	318.27 ± 2.29
5	Hayesroth	FILS	4	45.44 ± 10.51	17.62 ± 0.42	4	59.07 ± 9.04	17.70 ± 0.43	4	58.41 ± 15.58	17.71 ± 0.43
		IFILS	4	71.15 ± 13.42	24.59 ± 0.65	4	77.15 ± 14.58	24.69 ± 0.65	4	73.46 ± 6.55	24.69 ± 0.64
6	Iris	FILS	3	92.00 ± 8.20	14.90 ± 0.02	3	94.00 ± 2.11	14.91 ± 0.03	3	84.67 ± 9.46	14.91 ± 0.004
		IFILS	3	96.67 ± 4.71	21.61 ± 0.01	3	96.00 ± 4.66	21.62 ± 0.01	2	92.67 ± 5.84	16.55 ± 0.02
7	Lungcancer	FILS	37	53.33 ± 24.91	101.06 ± 5.59	37	50.83 ± 39.37	101.08 ± 5.59	46	50.83 ± 28.45	109.36 ± 4.590
		IFILS	10	55.83 ± 22.92	45.10 ± 1.70	19	64.17 ± 24.23	83.65 ± 2.68	10	60.00 ± 31.87	45.19 ± 1.702
8	Wine	FILS	7	95.49 ± 2.38	234.52 ± 3.33	9	90.46 ± 10.49	289.98 ± 9.75	11	97.70 ± 14.05	307.67 ± 10.513
		IFILS	11	95.46 ± 3.67	451.72 ± 21.38	7	91.57 ± 5.42	348.67 ± 5.64	11	97.16 ± 4.85	451.83 ± 21.376
(1)-(8)	Average	FILS	10.38	74.46 ± 20.96	419.93 ± 704.26	10	74.95 ± 20.82	255.67 ± 270.04	12	75.15 ± 20.25	430.21 ± 701.85
		IFILS	6.75	79.88 ± 17.56	484.24 ± 713.86	7.5	79.98 ± 15.79	304.26 ± 299.02	6.63	80.21 ± 18.14	483.74 ± 714.38



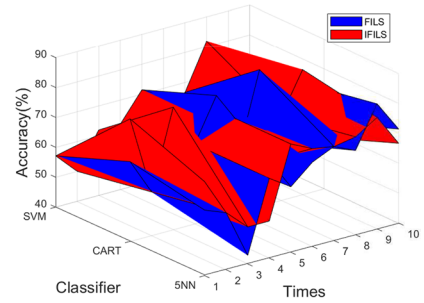
(a) Blood



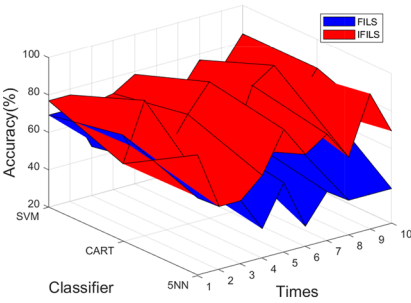
(b) Ccbr



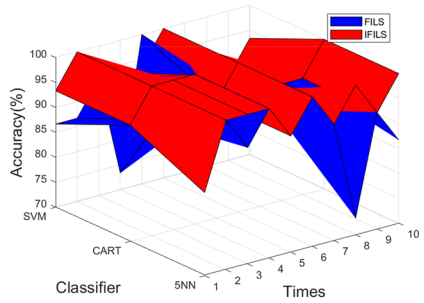
(c) Ecoli



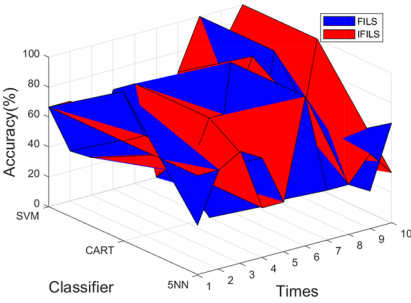
(d) Glass



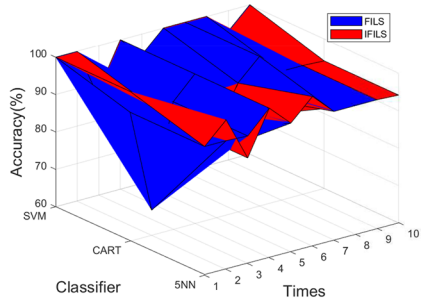
(e) Hayes-roth



(f) Iris



(g) Lung cancer



(h) Wine

Fig. 6 The FILS and IFILS algorithms' r -optimal ($r < |F|$) accuracies on eight UCI datasets (with 10-fold times and three classifiers)

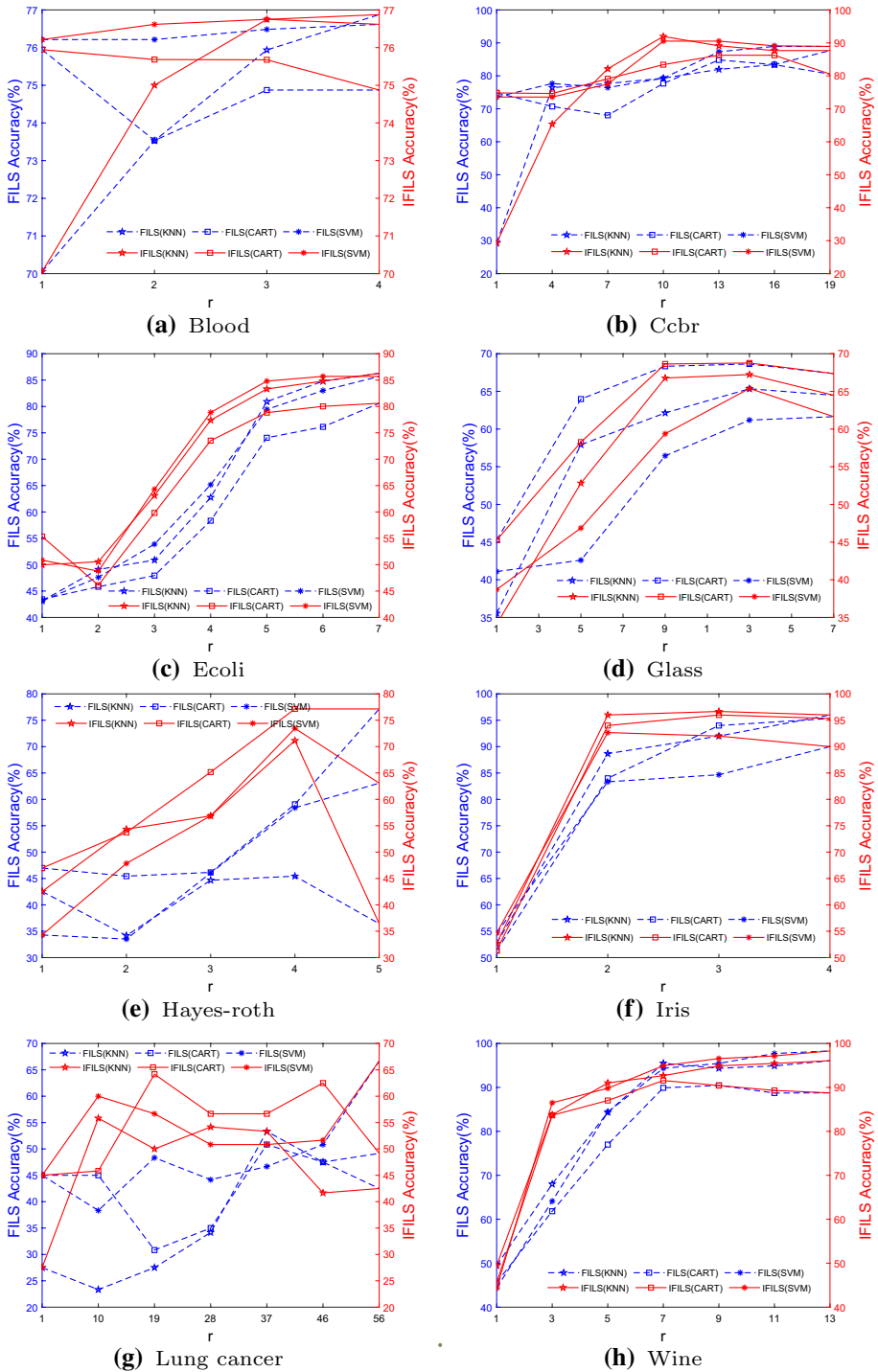
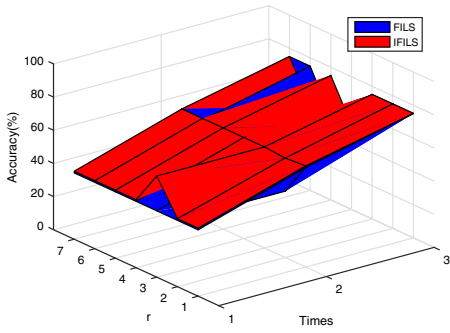


Fig. 7 The FILS and IFILS algorithms' 10-fold-average classification accuracies on eight UCI datasets (with number r and three classifiers)

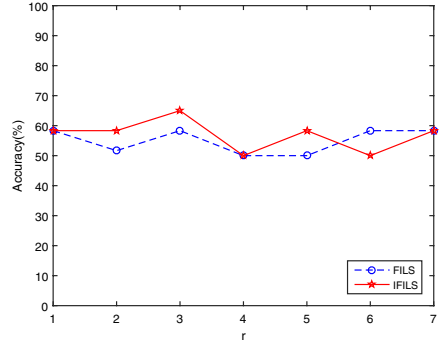
Table 12 The FILS and IFILS algorithms' 3-fold-average accuracies and standard deviations on COVID-19 dataset (with K, r)

r	Algorithm	K			
		1	2	3	4
1	FILS	66.67 ± 24.94	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34
	IFILS	66.67 ± 24.94	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34
2	FILS	56.67 ± 4.71	$\downarrow 65.00 \pm 7.07$	56.67 ± 4.71	58.33 ± 14.34
	IFILS	66.67 ± 24.94	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34
3	FILS	$\downarrow 73.33 \pm 18.86$	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34
	IFILS	$\uparrow 73.33 \pm 18.86$	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$
4	FILS	58.33 ± 14.34	58.33 ± 14.34	50.00 ± 8.16	$\downarrow 58.33 \pm 14.34$
	IFILS	60.00 ± 32.66	58.33 ± 14.34	56.67 ± 4.71	58.33 ± 14.34
5	FILS	41.67 ± 14.34	50.00 ± 8.16	50.00 ± 8.16	50.00 ± 8.16
	IFILS	58.33 ± 14.34	50.00 ± 8.16	63.33 ± 12.47	$\uparrow 65.00 \pm 17.80$
6	FILS	41.67 ± 14.34	50.00 ± 8.16	50.00 ± 24.49	$\downarrow 58.33 \pm 14.34$
	IFILS	58.33 ± 14.34	50.00 ± 8.16	35.00 ± 17.80	50.00 ± 8.16
7	FILS	50.00 ± 8.16	50.00 ± 8.16	35.00 ± 17.80	$\downarrow 58.33 \pm 14.34$
	IFILS	50.00 ± 8.16	50.00 ± 8.16	35.00 ± 17.80	58.33 ± 14.34
r optimization	FILS	$\downarrow 73.33 \pm 18.86$	$\downarrow 65.00 \pm 7.07$	58.33 ± 14.34	58.33 ± 14.34
	IFILS	$\uparrow 73.33 \pm 18.86$	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$
r	K				K optimization
	5	6	7	8	
1	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	66.67 ± 24.94
	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	66.67 ± 24.94
2	51.67 ± 16.50	51.67 ± 16.50	58.33 ± 14.34	50.00 ± 8.16	65.00 ± 7.07
	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	66.67 ± 24.94
3	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	$\downarrow 73.33 \pm 18.86$
	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$	$\uparrow 73.33 \pm 18.86$
4	50.00 ± 8.16	$\downarrow 58.33 \pm 14.34$	$\downarrow 58.33 \pm 14.34$	$\downarrow 58.33 \pm 14.34$	58.33 ± 14.34
	50.00 ± 8.16	50.00 ± 8.16	33.33 ± 24.94	58.33 ± 14.34	60.00 ± 32.66
5	50.00 ± 8.16	50.00 ± 8.16	50.00 ± 8.16	$\downarrow 58.33 \pm 14.34$	58.33 ± 14.34
	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	$\uparrow 65.00 \pm 17.80$
6	$\downarrow 58.33 \pm 14.34$	$\downarrow 58.33 \pm 14.34$	$\downarrow 58.33 \pm 14.34$	$\downarrow 58.33 \pm 14.34$	58.33 ± 14.34
	50.00 ± 8.16	58.33 ± 14.34	50.00 ± 8.16	58.33 ± 14.34	58.33 ± 14.34
7	$\downarrow 58.33 \pm 14.34$	$\downarrow 58.33 \pm 14.34$	50.00 ± 8.16	$\downarrow 58.33 \pm 14.34$	58.33 ± 14.34
	58.33 ± 14.34	58.33 ± 14.34	50.00 ± 8.16	58.33 ± 14.34	58.33 ± 14.34
r optimization	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	58.33 ± 14.34	$\downarrow 73.33 \pm 18.86$
	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$	$\uparrow 65.00 \pm 7.07$	$\uparrow 73.33 \pm 18.86$

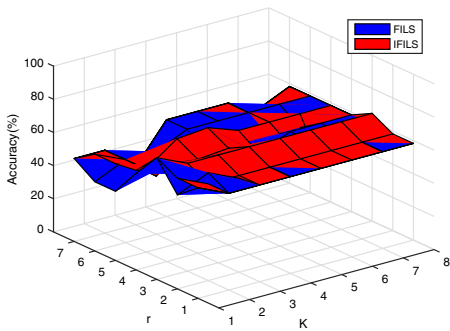
correspond to qualitative numbers +1, -1, and they can be linearly translated to 2, 0 to eliminate the negativity. Since the COVID-19 dataset contains a small number of samples (only 14 samples), 3-fold cross-validation is alternatively adopted. $K = 5$ used in the above experiments may become too large, so K is mainly searched in $\{1, 2, \dots, 8\}$ for full analysis in both algorithms.



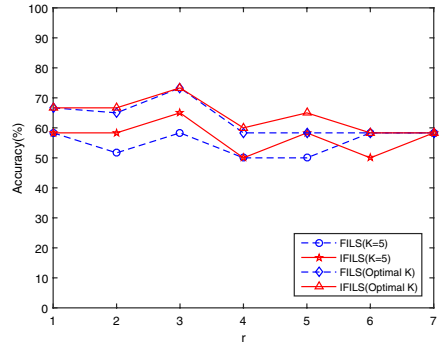
(a) 3 folds and r change ($K = 5$)



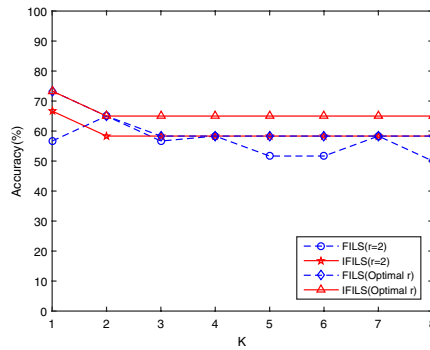
(b) 3-fold average ($K = 5$)



(c) (K, r)-grid on 3-fold average



(d) K cut and optimization



(e) r cut and optimization

Fig. 8 The FLS and IFLS algorithms' accuracies on the COVID-19 dataset

Table 12 shows the accuracy on the (K, r) grid based on the 3-fold statistics, and the 3-fold case and mean are partly reflected in Fig. 8(a) and (b), respectively, when $K = 5$. In Table 12, the main body of the average accuracies are imported from Fig. 8(c), and its cut parts when $K = 5$ and $r = 2$ constitute two broken lines in Fig. 8(c) and (d); furthermore, the right and

lower margins represent the maximal accuracies on K and r optimization, respectively; these form the other two lines in Fig. 8(c) and (d). Using contrast analysis, more proportions on correspondence and higher accuracies on optimization are obtained using the IFILS algorithm. Therefore, the better performance on this nominal medicine dataset is achieved using this algorithm.

5.4 Two additional medical diagnosis datasets

Finally, two additional datasets related to medical diagnosis, the Brain and Parkinsons datasets shown in Table 1, are adopted to comparatively analyze the FILS and IFILS algorithms to further verify the advantage of the IFILS algorithm over the FILS algorithm. Note that both the Brain and Parkinsons datasets are real-world related to medical diagnosis.

- The Brain dataset is a multilayer brain network dataset with resting-state electroencephalography data, and these practical data come from the Department of Otolaryngology of Sun Yat-sen Memorial Hospital, Sun Yat-sen University.
- The Parkinsons dataset is a detection dataset related to Parkinson's disease (PD). In this dataset, a range of biomedical voice measurements were recorded from 31 individuals. Of these individuals, 23 were diagnosed with PD. This dataset can be used to discriminate health patients from patients with PD.

Relevant accuracy results are provided in Table 13 based on changing the feature number r . By using 10-fold statistics, the results in Table 13 are first summarized as average accuracies and standard deviations in terms of the KNN, CART and SVM classifiers. Then, the last columns in Table 13 show a record of the r -optimal statistics, where the bold label highlights the comparative maximum. The algorithm comparison and superiority revelations can be observed in Table 13. When considering the same feature number r on the same classifier, higher recognition rates are often achieved using the IFILS algorithm rather than the FILS algorithm. When $r = |F|$, the same recognition rate is achieved by these two algorithms with the same classifier. These results are generally the highest in relation to feature subsets. However, one exception exists on Parkinsons dataset using the CART classifier. Concretely, a maximum of value of 87.71% is achieved when $r = 10$ using the IFILS algorithm, while a value of 86.18% when $r = |F| = 22$ is achieved using the FILS algorithm, so the IFILS algorithm outperforms the FILS algorithm. Regarding r -optimization with a range of $r < |F|$, the IFILS algorithm has a better effect than the FILS algorithm when observing the two contrast indices for the recognition rate and feature number. For the Brain dataset, when using the three-way classifiers on $r < |F|$, the following results are acquired for the IFILS and FILS algorithms, i.e., (77.14%, 61.43%, 85.71%) > (61.43%, 60.00%, 82.86%) of the recognize accuracy and (34, 23, 23) < (56, 56, 56) of the feature number. As further validated by the two medical diagnosis datasets, the IFILS algorithm not only removes the redundant data but also improves the recognition rate. Therefore, it has true improvements over the FILS algorithm.

Table 13 The FILS and IFILS algorithms' 10-fold-average accuracies and r -optimization statistics on the Brain and Parkinsons datasets based on KNN, CART and SVM

No.	Dataset	Classifier	Algorithm	r	1	2	3	4	7	10	13	16	22	Optimization realization on feature set ($r = F $)	Optimization realization on feature proper subset ($r < F $)
(10)	Brain	KNN	FILS	24.29 ± 15.13	44.29 ± 15.72	57.14 ± 21.3	50.00 ± 24.51	58.57 ± 20.7	61.43 ± 26.13	85.71 ± 11.66	85.71 ± 11.66(70)	61.43 ± 26.13(56)			
			IFILS	24.29 ± 15.13	67.14 ± 26.98	72.86 ± 20.7	77.14 ± 23.52	77.14 ± 20.43	72.86 ± 19.58	85.71 ± 11.66	85.71 ± 11.66(70)	77.14 ± 23.52(34)			
		CART	FILS	11.43 ± 11.27	54.29 ± 24.09	54.29 ± 26.77	52.86 ± 25.24	54.29 ± 22.13	60.00 ± 24.09	62.86 ± 18.07	62.86 ± 18.07	60.00 ± 24.09(56)			
			IFILS	11.43 ± 11.27	44.29 ± 14.21	61.43 ± 19.11	54.29 ± 18.81	58.57 ± 19.58	54.29 ± 16.22	62.86 ± 18.07	62.86 ± 18.07(70)	61.43 ± 19.11(23)			
		SVM	FILS	5.71 ± 12.05	45.71 ± 13.13	68.57 ± 19.98	67.14 ± 21.35	72.86 ± 20.70	82.86 ± 16.22	88.57 ± 13.13	88.57 ± 13.13(70)	82.86 ± 16.22(56)			
			IFILS	5.71 ± 12.05	78.57 ± 27.15	85.71 ± 16.50	82.86 ± 14.75	85.71 ± 13.47	84.29 ± 12.51	88.57 ± 13.13	88.57 ± 13.13(70)	85.71 ± 16.50(23)			
No.	Dataset	Classifier	Algorithm	r	1	4	7	10	13	16	22	Optimization realization on feature proper subset ($r < F $)			
(11)	Parkinsons	KNN	FILS	80.00 ± 8.09	76.39 ± 7.77	76.89 ± 8.73	79.03 ± 8.98	81.05 ± 10.5	91.34 ± 6.64	92.82 ± 6.45	92.82 ± 6.45(22)	91.34 ± 6.64(16)			
			IFILS	80.00 ± 8.09	85.68 ± 6.65	90.29 ± 8.8	89.21 ± 5.72	90.24 ± 7.09	91.84 ± 4.97	92.82 ± 6.45	92.82 ± 6.45(22)	91.84 ± 4.97(16)			
		CART	FILS	76.47 ± 7.09	78.37 ± 11.62	78.89 ± 8.11	78.00 ± 6.22	79.03 ± 4.76	83.61 ± 9.4	86.18 ± 6.26	86.18 ± 6.26	83.61 ± 9.40(16)			
			IFILS	76.47 ± 7.09	82.58 ± 6.43	83.53 ± 7.00	87.71 ± 7.17	84.13 ± 7.96	84.21 ± 6.33	86.18 ± 6.26	-	87.71 ± 7.17(10)			
		SVM	FILS	74.84 ± 9.72	76.34 ± 10.04	77.87 ± 9.98	76.87 ± 7.77	76.87 ± 9.39	83.05 ± 10.48	84.61 ± 8.68	84.61 ± 8.68(22)	83.05 ± 10.48(16)			
			IFILS	74.84 ± 9.72	76.87 ± 8.12	84.05 ± 5.79	84.05 ± 6.35	83.58 ± 7.28	84.00 ± 8.09	84.61 ± 8.68	84.61 ± 8.68(22)	84.05 ± 5.79(7)			

6 Conclusions

This paper devotes to an unsupervised feature selection method based on modified LS, and thus, the IFILS algorithm, whose flowchart is shown in Fig. 1, is proposed as an improvement over the FILS algorithm. The algorithmic validity and superiority come from the introduction and perfection of the systematic measure SIG based on the modified LS (Definition 3), and they are completely supported through examples and data experiments. In terms of the root of uncertainty in measurements, both the main SIG and underlying modified LS have granulation uncertainty/nonmonotonicity, and their characterizations (especially SIG's informatization) effectively function in unsupervised feature selection.

As a result, the IFILS algorithm outperforms the FILS algorithm, and it achieves better classification performance. This conclusion can be generally validated by the experimental results, especially on the main ten datasets (Sects. 5.2 and 5.4), and the eight datasets and their results in Sect. 5.2 are further summarized. In terms of the 8 datasets, Tables 8, 9, and 10 provide information on the chain search and optimization determination. From this detailed information, the comparative advantages of the IFILS algorithm are revealed by using the KNN, CART and SVM algorithms; furthermore, Table 11 shows the optimal situations of feature selection, and they comprehensively reflect the superiority of the IFILS algorithm in terms of selection number, prediction accuracy, and running time, using the three classifiers. In terms of the 8 dataset figures, Fig. 6 shows the 10-fold cross-validation to determine the r -optimal accuracies (where $r < |F|$) based on the KNN, CART and SVM classifiers, while Fig. 7 shows the feature number r to reflect the 10-fold average accuracies based on the three classifiers; these two figures also reveal the advantages of the IFILS algorithm from different yet clear perspectives.

The IFILS algorithm has become a new learning method for feature selection, and there are two possible limitations and corresponding development opportunities. Since the underlying modified LS is dependent on the KNN calculations of all elements, the IFILS algorithm is concerned with relevant matrix processing, which is very time-consuming. Meanwhile, the IFILS algorithm adheres to a specific approach to uncertainty measurements, and more robust measures and algorithms need to be constructed by comprehensively considering additional perspectives. In the future, the time efficiency of the IFILS algorithm should be improved by utilizing some optimization strategies, while its recognition effect can be further reinforced by combining other practical measures such as the dependency degree and information entropy. Furthermore, the IFILS algorithm can be further considered and utilized from the unsupervised FS to semisupervised learning perspective, and its data mining capability is worth studying from numeric and nominal datasets to hybrid datasets. Moreover, in-depth experiments based on broader datasets are required for the IFILS algorithm and subsequent improvements, thus better facilitating real-world intelligent applications of machine learning.

Acknowledgements The authors thank both the reviewers and editors for their valuable suggestions, which substantially improve this paper. The work was supported by National Natural Science Foundation of China (61673285, 11671284), Sichuan Science and Technology Program of China (2021YJ0085), National Science Foundation of Sichuan Province (2022NSFSC0929), and Joint Research Project of Laurent Mathematics Center of Sichuan Normal University and National-Local Joint Engineering Laboratory of System Credibility Automatic Verification (ZD20220101).

Declarations

Conflict of Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Alelyani S, Tang JL, Liu H (2018) Feature selection for clustering: a review. *Data Clustering*, pp 29–60
- Amarbayasgalan T, Pham VH, Theera-Umpon N et al (2020) Unsupervised anomaly detection approach for time-series in multi-domains using deep reconstruction error. *Symmetry* 12(8):1251. <https://doi.org/10.3390/sym12081251>
- Breaban M, Luchian H (2011) A unifying criterion for unsupervised clustering and feature selection. *Pattern Recogn* 44(4):854–865. <https://doi.org/10.1016/j.patcog.2010.10.006>
- Cai J, Luo JW, Wang SL et al (2018) Feature selection in machine learning: a new perspective. *Neurocomputing* 300:70–79. <https://doi.org/10.1016/j.neucom.2017.11.077>
- Chen JC, Zeng YJ, Li Y et al (2020) Unsupervised feature selection based extreme learning machine for clustering. *Neurocomputing* 386:198–207. <https://doi.org/10.1016/j.neucom.2019.12.065>
- Dadaneh BZ, Markid HY, Zakerolhosseini A (2016) Unsupervised probabilistic feature selection using ant colony optimization. *Expert Syst Appl* 53:27–42. <https://doi.org/10.1016/j.eswa.2016.01.021>
- El-Hasnony IM, Barakat SI, Elhoseny M et al (2020) Improved feature selection model for big data analytics. *IEEE Access* 8:66989–67004. <https://doi.org/10.1109/ACCESS.2020.2986.232>
- Gao C, Lai ZH, Zhou J et al (2019) Granular maximum decision entropy-based monotonic uncertainty measure for attribute reduction. *Int J Approx Reason* 104:9–24. <https://doi.org/10.1016/j.ijar.2018.10.014>
- García-Díaz P, Sánchez-Berriel I, Martínez-Rojas JA et al (2020) Unsupervised feature selection algorithm for multiclass cancer classification of gene expression rna-seq data. *Genomics* 112(2):1916–1925. <https://doi.org/10.1016/j.ygeno.2019.11.004>
- Gunal S, Edizkan R (2008) Subspace based feature selection for pattern recognition. *Inform Sci* 178(19):3716–3726. <https://doi.org/10.1016/j.ins.2008.06.001>
- Hai A, Bharath G, Daud M et al (2021) Valorization of groundnut shell via pyrolysis: product distribution, thermodynamic analysis, kinetic estimation, and artificial neural network modeling. *Chemosphere* 283:131162. <https://doi.org/10.1016/j.chemosphere.2021.131162>
- He XF, Cai D, Niyogi P (2005) Laplacian score for feature selection. In: *International conference on neural information processing systems*, pp 507–541
- Huang R, Jiang WD, Sun GL (2018) Manifold-based constraint Laplacian score for multi-label feature selection. *Pattern Recogn Lett* 112:346–352. <https://doi.org/10.1016/j.patrec.2018.08.021>
- Kou G, Yang P, Peng Y et al (2020) Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Appl Soft Comput* 86:105836. <https://doi.org/10.1016/j.asoc.2019.105836>
- Lim HR, Khoo KS, Chia WY, et al (2022) Smart microalgae farming with internet-of-things for sustainable agriculture. *Biotechnol Adv*. <https://doi.org/10.1016/j.biotechadv.2022.107931>
- Luo MN, Nie FP, Chang XJ et al (2017) Adaptive unsupervised feature selection with structure regularization. *IEEE T Neur Net Lear* 29(4):944–956. <https://doi.org/10.1109/TNNLS.2017.2650.978>
- Nguyen BH, Xue B, Zhang MJ (2020) A survey on swarm intelligence approaches to feature selection in data mining. *Swarm Evol Comput* 54:100663. <https://doi.org/10.1016/j.swevo.2020.100663>
- Pang QQ, Zhang L (2020a) Forward iterative feature selection based on laplacian score. In: *International conference on neural information processing*. Springer, New York, pp 381–392. https://doi.org/10.1007/978-3-030-63833-7_32
- Pang QQ, Zhang L (2020) Semi-supervised neighborhood discrimination index for feature selection. *Knowl-Based Syst* 204:106224. <https://doi.org/10.1016/j.knosys.2020.106224>
- Remeseiro B, Bolon-Canedo V (2019) A review of feature selection methods in medical applications. *Comput Biol Med* 112:103375. <https://doi.org/10.1016/j.combiomed.2019.103375>

- Rostami M, Forouzandeh S, Berahmand K et al (2020) Integration of multi-objective pso based feature selection and node centrality for medical datasets. *Genomics* 112(6):4370–4384. <https://doi.org/10.1016/j.ygeno.2020.07.027>
- Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF (2016) A new hybrid filter-wrapper feature selection method for clustering based on ranking. *Neurocomputing* 214:866–880. <https://doi.org/10.1016/j.neucom.2016.07.026>
- Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF (2020) A review of unsupervised feature selection methods. *Artif Intell Rev* 53(2):907–948. <https://doi.org/10.1007/s10462-019-09682-y>
- Starczyk U, Zielosko B (2020) Heuristic-based feature selection for rough set approach. *Int J Approx Reason* 125:187–202. <https://doi.org/10.1016/j.ijar.2020.07.005>
- Sun L, Yin TY, Ding WP et al (2021) Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy. *IEEE T Fuzzy Syst* 30(5):1197–1211. <https://doi.org/10.1109/TFUZZ.2021.3053844>
- Tham PE, Ng YJ, Vadivelu N et al (2022) Sustainable smart photobioreactor for continuous cultivation of microalgae embedded with internet of things. *Bioresource Technol* 346:126558. <https://doi.org/10.1016/j.biortech.2021.126558>
- Tubishat M, Ja'afar S, Alswaitti M et al (2021) Dynamic salp swarm algorithm for feature selection. *Expert Syst Appl* 164:113873. <https://doi.org/10.1016/j.eswa.2020.113873>
- Van JCF, Tham PE, Lim HR, et al (2022) Integration of internet-of-things as sustainable smart farming technology for the rearing of black soldier fly to mitigate food waste. *J Taiwan Inst Chem E*. <https://doi.org/10.1016/j.jtice.2022.104235>
- Wang CZ, Hu QH, Wang XZ et al (2017) Feature selection based on neighborhood discrimination index. *IEEE T Neur Net Lear* 29(7):2986–2999. <https://doi.org/10.1109/TNNLS.2017.2710422>
- Wang CZ, Shi YP, Fan XD et al (2019) Attribute reduction based on k-nearest neighborhood rough sets. *Int J Approx Reason* 106:18–31. <https://doi.org/10.1016/j.ijar.2018.12.013>
- Wang GY, Yu H et al (2015) Monotonic uncertainty measures for attribute reduction in probabilistic rough set model. *Int J Approx Reason* 59:41–67. <https://doi.org/10.1016/j.ijar.2015.01.005>
- Xu JC, Qu KL, Yuan M et al (2021) Feature selection combining information theory view and algebraic view in the neighborhood decision system. *Entropy* 23(6):704. <https://doi.org/10.3390/e23060704>
- Yuan Z, Chen HM, Li TR et al (2021) Unsupervised attribute reduction for mixed data based on fuzzy rough sets. *Inform Sci* 572:67–87. <https://doi.org/10.1016/j.ins.2021.04.083>
- Yuan Z, Chen HM, Li TR, et al (2022) Multigranulation relative entropy-based mixed attribute outlier detection in neighborhood systems. *IEEE T Syst Man Cyber A* 52(8):5175–5187. <https://doi.org/10.1109/TSMC.2021.3119119>
- Zhang XY, Gou HY (2022) Statistical-mean double-quantitative k-nearest neighbor classification learning based on neighborhood distance measurement. *Knowl-Based Syst* 250:109018. <https://doi.org/10.1016/j.knosys.2022.109018>
- Zhang XY, Jiang JF (2022) Measurement, modeling, reduction of decision-theoretic multigranulation fuzzy rough sets based on three-way decisions. *Inform Sci* 607:1550–1582. <https://doi.org/10.1016/j.ins.2022.05.122>
- Zhang XY, Yao YY (2022) Tri-level attribute reduction in rough set theory. *Expert Syst Appl* 190:116187. <https://doi.org/10.1016/j.eswa.2021.116187>
- Zhang XY, Fan YR, Yang JL (2021) Feature selection based on fuzzy-neighborhood relative decision entropy. *Pattern Recogn Lett* 146:100–107. <https://doi.org/10.1016/j.patrec.2021.03.001>
- Zhao JD, Lu K, He XF (2008) Locality sensitive semi-supervised feature selection. *Neurocomputing* 71(10–12):1842–1849. <https://doi.org/10.1016/j.neucom.2007.06.014>
- Zhou P, Chen JY, Fan MY et al (2020) Unsupervised feature selection for balanced clustering. *Knowl-Based Syst* 193:105417. <https://doi.org/10.1016/j.knosys.2019.105417>
- Zhu LL, Miao LS, Zhang DQ (2012) Iterative Laplacian score for feature selection. *Commun Comput Inform Sci* 321:80–87. https://doi.org/10.1007/978-3-642-33506-8_11

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.