International Journal of Approximate Reasoning 187 (2025) 109530

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Fusing fuzzy rough sets and mean shift for anomaly detection

Mengyao Liao [a,b], Zhiyu Chen [a,b], Can Gao [a,b,*], Jie Zhou [a,b], Xiaodong Yue [c]

[a] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China
[b] Guangdong Provincial Key Laboratory of Intelligent Information Processing, Shenzhen, 518060, China
[c] Artificial Intelligence Institute, Shanghai University, Shanghai, 200444, China

ARTICLE INFO

Keywords:
Anomaly detection
Fuzzy rough sets
Fuzzy granule density
Density ratio
Mean shift

ABSTRACT

Outlier detection is a critical but challenging task due to the complex distribution of practical data, and some Fuzzy Rough Sets (FRS)-based methods have been presented to identify outliers from these data. However, these methods still have limitations when facing the co-existence of different types of outliers. In this study, an improved FRS-based unsupervised anomaly detection method is proposed by integrating distance and density information. Specifically, to detect the local outliers, a fuzzy granule density is first defined by introducing a Gaussian kernel similarity to characterize the local density of samples. Then, optimistic and pessimistic fuzzy granule densities are further developed to evaluate the density variation in the local neighborhood. Moreover, a distance measure based on mean shift is introduced to detect global and group outliers. Finally, an outlier detection method that integrates the density and distance measures is designed to effectively identify diverse types of outliers. Extensive experiments on synthetic and public datasets, along with statistical significance analysis, demonstrate the superior performance of the proposed method, achieving an average improvement of at least 12.27% in terms of AUROC.

1. Introduction

In the real world, samples that deviate from the main patterns of the data may represent potential issues and risks, and identifying these samples is of high practical value. Anomaly detection [1], which served as an effective technique to capture these rare samples, has been extensively studied and have been applied in multiple domains, including video surveillance [2], financial fraud detection [3], cybersecurity [4], and medical diagnostics [5].

Anomaly detection can be considered a binary classification task with an imbalanced data distribution. Existing anomaly detection methods can be classified into supervised, semi-supervised, and unsupervised methods in terms of the availability of the label information. Supervised methods assume that outliers exhibit distinct patterns that can be distinguished by supervised models trained on both normal and abnormal samples, while semi-supervised methods aim to construct a model that can effectively identify outliers using normal data or limited anomalous samples. Nevertheless, obtaining outliers is a challenging task [1]. Therefore, unsupervised methods have become a mainstream approach in anomaly detection [6]. Unsupervised anomaly detection methods can be broadly categorized into statistical, distance-based, density-based, clustering-based, and ensemble methods. Statistical methods, such as HBOS [7] and KDE [8], assume specific data distributions and use metrics like mean or variance to detect outliers. Distance-based approaches, exemplified by KNN [9] and ODIN [10], identify outliers by assessing the difference in distance between a sample and its

* Corresponding author at: College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China.
E-mail address: 2005gaocan@163.com (C. Gao).

https://doi.org/10.1016/j.ijar.2025.109530
Received 5 April 2025; Received in revised form 23 June 2025; Accepted 13 July 2025

Available online 18 July 2025
0888-613X/

neighboring samples. Density-based methods, like LOF [11], evaluate local density variations and recognize samples in sparse regions as outliers. Clustering-based methods, including CBLOF [12] and K-Means++ [13], cluster similar samples and consider small-sized groups or isolated samples as outliers. Ensemble methods, such as IForest [14], combine the detection results of multiple models to improve the robustness and accuracy. Nevertheless, most of the methods mentioned above cannot effectively handle data with unordered discrete features, which severely limits their capability and application in anomaly detection.

The rough set theory (RST) proposed by Pawlak [15] is an effective method to handle categorical data with vagueness and uncertainty. Some RST-based anomaly detection methods [16,17] have been presented and have exhibited good anomaly detection performance. However, traditional RST relies on the equivalence relation to form information granules, which limits its ability to deal with continuous or mixed data. To overcome this limitation, fuzzy rough sets (FRS) and neighborhood rough sets (NRS) have been introduced for anomaly detection. For example, Yuan et al. [18] used fuzzy approximation precision to identify outliers within fuzzy rough granules, enabling efficient processing of mixed data. Yuan et al. [19] combined fuzzy rough density and entropy to assess sample compactness and attribute weights, effectively integrating both sample density and fuzziness for anomaly detection. Wu et al. [20] introduced kernelized fuzzy upper and lower approximations to compute anomaly scores of samples. Gao et al. [21] defined a relative granular ratio to reflect the size of the sample neighborhood and introduced an outlier factor to measure the difference between outliers and a granule-based majority set for granule-based majority set. Yuan et al. [22] proposed a hybrid method using multi-granular relative entropy to identify outliers by assessing uncertainty across different granularity levels. Zhang et al. [23] fused multiple neighborhood anomaly factors defined on the three-way neighborhood regions to identify outliers.

Existing FRS-based methods rely on fuzzy similarity relations to detect outliers and inherently have two limitations. On the one hand, the fuzzy similarity relation is essentially a form of distance metric. When data exhibit imbalanced distributions, FRS-based methods may generate poor performance due to their weak ability to capture the local density of samples. On the other hand, existing methods generally use distance or density information to detect outliers, making them challenging to identify different types of outliers from complex distributed data. To address these issues, we introduce a novel anomaly detection method that combines fuzzy granule density and mean shift. The key contributions are as follows.

1) To capture the local neighborhood information of samples, the optimistic and pessimistic fuzzy granule densities are defined to evaluate the density variation, and a fuzzy granule density measure is further introduced to evaluate the outlierness of samples, enhancing the ability to detect local outliers.
2) To evaluate the deviation degree of samples, a distance measure based on mean shift is proposed, which reflects the difference between the maximum and average shift across different attributes, thereby improving the performance in detecting global and group outliers.
3) To cope with the coexistence of different types of outliers, a novel anomaly detection method is presented by fusing fuzzy granule density measure and distance measure. Extensive experiments conducted on artificial and real datasets demonstrate the superior performance of the proposed method over other representative methods, improving over the second-best method by 12.27%.

The remainder of the paper is structured as follows. Section 2 introduces the preliminaries of fuzzy rough sets and mean shift. Section 3 describes the proposed anomaly detection method based on density and distance information. Section 4 validates the effectiveness of the proposed method, highlighting its superiority over other comparative methods. Finally, Section 5 summarizes the paper and suggests potential directions for future research.

2. Preliminaries

In this section, several key concepts related to fuzzy rough sets and mean shifts are introduced. The detailed information can be referred to as [24,25].

2.1. Fuzzy rough sets

In FRS, the given set of samples U is described by a finite set of attributes A, and for any sample $x \in U$, a function f maps a value from the domain of attributes $V = \bigcup_{a\in A} V_a$ to the sample in every attribute $a \in A$, i.e., $f : U \times A \to V$. Thus, a fuzzy information system is generally represented as a quadruple $FIS = (U, A, V, f)$.

Definition 1. Given a fuzzy information system $FIS = (U, A, V, f)$, the fuzzy relation induced by an attribute subset $B \subset A$ can be defined as

$$R_B : U \times U \mapsto [0,1], \tag{1}$$

and satisfies the following properties:

(1) Reflexivity: $R_B(x,x) = 1$;
(2) Symmetry: $R_B(x,y) = R_B(y,x)$;
(3) Transitivity: $R_B(x,z) \geq \sup_{y\in U} T(R_B(x,y), R_B(y,z))$, where T is a T-norm operator and the Min-Max triangular norm [26] is adopted in this study.

Definition 2. Given a fuzzy information system $FIS = (U, A, V, f)$, the collection of fuzzy granules induced by the fuzzy relation R_B can be expressed as

$$U/R_B = \{[x_1]_B, [x_2]_B, ..., [x_i]_B, ..., [x_n]_B\}, \tag{2}$$

where $[x_i]_B = \left(R_B(x_i, x_1), R_B(x_i, x_2), ..., R_B(x_i, x_n)\right)$ and the cardinality of the fuzzy granule $[x_i]_B$ is computed by accumulating all fuzzy similarity degrees of x_i to other samples, with $1 \leq |[x_i]_B| \leq n$.

Definition 3. Given the fuzzy relation R_B, the fuzzy relation matrix is defined as

$$M(R_B) = \begin{bmatrix} R_B(x_1\, x_1) & \cdots & R_B(x_1, x_n) \\ \vdots & \ddots & \vdots \\ R_B(x_n, x_1) & \cdots & R_B(x_n, x_n) \end{bmatrix}. \tag{3}$$

2.2. Mean shift

Mean shift is an effective technique of locating the maxima of a density function or data distribution and has been used in a range of applications such as clustering and tracking. It begins with an initial estimated value and iteratively updates the mean by determining the weight of nearby samples. Some related concepts are presented as follows.

Definition 4. Given a dataset U described by the attribute set A, the mean shift vector for a sample $x_i \in U$ is defined as

$$m(x_i) = \frac{\sum_{x_j \in N_A(x_i)} x_j}{|N_A(x_i)|} - x_i, \tag{4}$$

where $N_A(x_i)$ denotes the neighbor set of the sample x_i under the attribute set A. The mean shift vector reflects how a sample shifts toward the weighted centroid of its neighbors.

3. The proposed method

In this section, the overall framework for the proposed anomaly detection method is first introduced. Then, an improved FRS-based density measure is developed to evaluate the density information of samples. Subsequently, a mean-shift-based distance measure is presented to reflect the outlier degree of samples in distance. Finally, a novel anomaly detection method based on these two measures is proposed to identify different types of outliers.

3.1. Overall framework

Existing FRS-based methods generally employ fuzzy relations to group similar samples into fuzzy granules, while samples that have low fuzzy similarities to samples within their fuzzy granules are considered as outliers. Intuitively, FRS-based methods rely on fuzzy similarity between samples and excel in detecting global outliers. However, practical data may exhibit complex distributions and thus contain different types of outliers, such as local, global, and group outliers. Introducing other valuable information has become a key factor for anomaly detection. Essentially, fuzzy granule captures local relationships between samples and can be further exploited to estimate the local density of samples to identify local outliers. On the other hand, the mean shift degree of a sample to the mean of the local neighborhood or the whole data reflects the deviation tendency of the sample, thus facilitating the detection of global or group outliers. Motivated by these facts, a novel anomaly detection method is proposed, and its overall framework is illustrated in Fig. 1.

Specifically, the proposed method first determines the kernel bandwidth for each attribute and computes a fuzzy similarity matrix. Subsequently, the optimistic and pessimistic densities are obtained based on fuzzy granule density, which are then integrated to evaluate the density outlierness of samples. Furthermore, the mean shift degree is derived using the fuzzy distance matrix, and the distance measure of samples is quantified by computing the maximum deviation between the mean shift degree of each attribute and its average value. Ultimately, a novel outlierness measure that combines both density and distance information is proposed to detect various types of outliers.

3.2. Density measure based on fuzzy rough sets

FRS employs fuzzy relations to evaluate the similarity or distance between samples. Although fuzzy granules partially reflect the relationship information between a sample and its neighbors, existing FRS-based anomaly detection methods may be ineffective in identifying outliers from data with uneven distribution due to the lack of accurate estimation of the local relationship of samples. To address this limitation, a density anomaly degree based on fuzzy granules is proposed to capture the local information of samples for anomaly detection. In what follows, some related concepts are first presented.

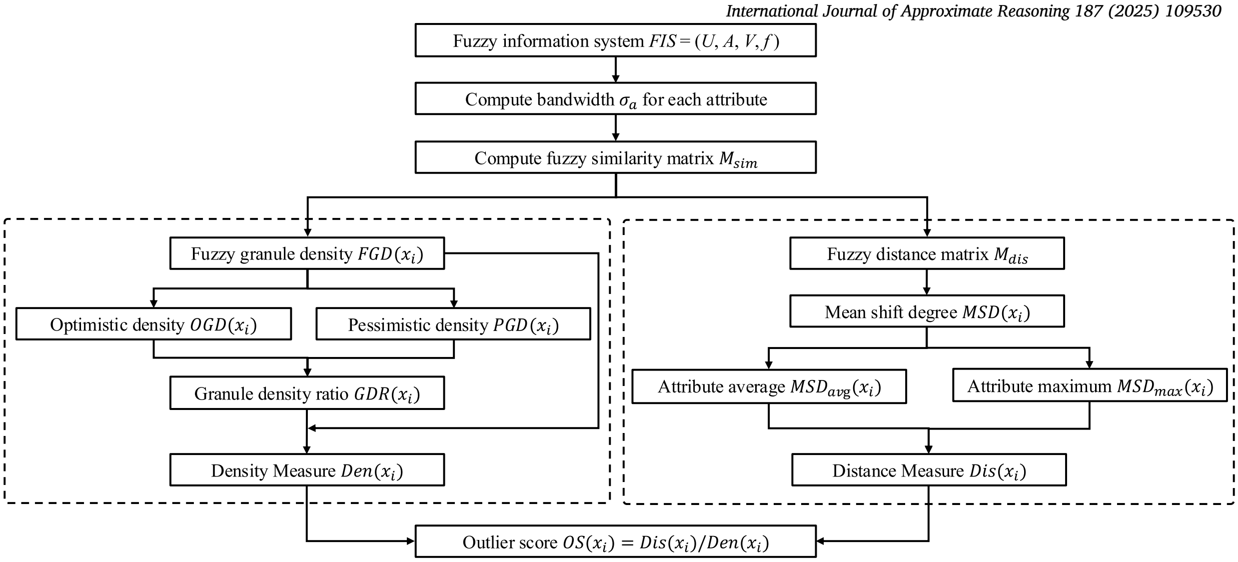


Fig. 1. The framework of the proposed method.

Definition 5. Given a fuzzy information system $FIS = (U, A, V, f)$, the fuzzy similarity between samples x_i and x_j with respect to an attribute $a \in A$ is defined as

$$R_a(x_i, x_j) = \begin{cases} 1 & , f_a(x_i) = f_a(x_j) \text{ and } a \text{ is discrete} \\ 0 & , f_a(x_i) \neq f_a(x_j) \text{ and } a \text{ is discrete} \\ \exp\left(-\frac{\|f_a(x_i) - f_a(x_j)\|^2}{\sigma_a}\right) & , a \text{ is continuous} \end{cases} \tag{5}$$

where $f_a(x_i)$ denotes the value of sample x_i in the attribute a, and the bandwidth for the attribute a is defined as

$$\sigma_a = \mu \cdot \frac{\sum_{x_i \in U} d_a^k(x_i)}{|U|}, \tag{6}$$

where $d_a^k(x_i)$ refers to the k-nearest neighbor distance of sample x_i at the attribute a, and μ is a tuning factor.

For discrete attributes, the similarity of two samples is 1 if their values are identical; otherwise, their similarity is 0. For continuous attributes, the sample similarity is determined by a kernelized exponential function with a scale factor of σ_a.

Definition 6. Given a fuzzy information system $FIS = (U, A, V, f)$, the fuzzy granule density of a sample $x_i \in U$ with respect to the attribute set A is defined as

$$FGD_A(x_i) = \frac{\sum_{a \in A} |[x_i]_a|}{|A|}. \tag{7}$$

The fuzzy granule density is computed by averaging the cardinality of the fuzzy granules of a sample across different features. The higher the fuzzy granule density, the more similar the sample is to its neighbors, and the more likely the sample is to be normal; otherwise, the sample is an outlier.

Definition 7. Given a fuzzy information system $FIS = (U, A, V, f)$, the fuzzy k-neighborhood of a sample $x_i \in U$ with respect to an attribute $a \in A$ is defined as

$$N_a^k(x_i) = \{x_j \in U \mid \left(1 - R_a(x_i, x_j)\right) \leq R_a^k(x_i)\}, \tag{8}$$

where $R_a^k(x_i)$ represents the k-th highest fuzzy distance of the sample x_i to other samples in the attribute a. Traditional KNN neighborhood typically relies on Euclidean distance to evaluate the similarity of samples, while the fuzzy k-neighborhood adopts fuzzy distance between samples to construct the neighborhood of each sample, which implicitly provides fuzzy density information that facilitates the detection of outliers. For a normal sample, its k-th fuzzy distance is relatively small, and its density thus is large. Moreover, a normal sample may also have neighbors with a higher density, and the density difference between the sample and its neighbors is subtle. However, for an abnormal sample, its density is relatively small, and the difference between the sample and its neighbors in density will be large. In view of these observations, the optimistic and pessimistic fuzzy granule densities are introduced to reflect the density variation of a sample and its neighbors.

Definition 8. Given a fuzzy information system $FIS = (U, A, V, f)$, the optimistic and pessimistic fuzzy granule densities of a sample $x_i \in U$ with respect to an attribute $a \in A$ are defined as

$$OGD_a(x_i) = \frac{1}{|N_a^k(x_i)|} \sum_{x_j \in N_a^k(x_i)} \max\left\{FGD_a(x_i), FGD_a(x_j)\right\}, \tag{9}$$

$$PGD_a(x_i) = \frac{1}{|N_a^k(x_i)|} \sum_{x_j \in N_a^k(x_i)} \min\left\{FGD_a(x_i), FGD_a(x_j)\right\}. \tag{10}$$

Definition 9. Given a fuzzy information system $FIS = (U, A, V, f)$, the granule density ratio of a sample $x_i \in U$ with respect to an attribute $a \in A$ is defined as

$$GDR_a(x_i) = \frac{PGD_a(x_i)}{OGD_a(x_i)}. \tag{11}$$

The optimistic and pessimistic fuzzy granule densities reflect the maximum and minimum densities between a sample and its neighbors, respectively, and consider the local density of the sample from different perspectives. While the fuzzy granule ratio captures the difference between the two fuzzy granule densities. When the value of the fuzzy granule ratio is small, the sample is more likely to be normal; otherwise, the sample is an outlier.

Definition 10. Given a fuzzy information system $FIS = (U, A, V, f)$, the fuzzy granule measure of a sample $x_i \in U$ with respect to an attribute $a \in A$ is defined as

$$FGM_a(x_i) = GDR_a(x_i) \cdot FGD_a(x_i). \tag{12}$$

The fuzzy granule measure combines both the fuzzy granule density and the granule density ratio to provide a more comprehensive density measure. The fuzzy granule density captures the density of a sample within its neighborhood, while the granule density ratio reflects the consistency of the sample and its neighbors in density. By multiplying these two factors, a more accurate density estimation can be obtained for anomaly detection. Compared with additive or max-based fusion strategies, the operator of multiplication reflects the consistency trend between the two densities and provides a stricter measure for density, thereby improving the reliability of outlier detection.

Definition 11. Given a fuzzy information system $FIS = (U, A, V, f)$, the density measure of a sample $x_i \in U$ with respect to the attribute set A is defined as

$$Den(x_i) = \frac{1}{|A|} \sum_{a \in A} FGM\left([x_i]_a\right). \tag{13}$$

The density measure characterizes the local density information of a sample across different attributes. A higher density measure indicates that the sample is located in a dense region, and the sample is more likely to be normal; otherwise, the sample is more likely to be an outlier. The density measure effectively exploits the local neighborhood information of samples and provides accurate density estimation for identifying local outlers from uneven distributed data.

3.3. Distance measure based on mean shift

Fuzzy granules can be used to explore the density of samples and facilitate the detection of local outliers. However, practical data may exhibit complex distribution entangled with global and group outliers. Relying solely on density information is thus insufficient for anomaly detection. To address this limitation, a distance-based measure is required to capture sample deviations from different perspectives. Traditional mean shift algorithms, originally designed to locate pattern centers, generate shift vectors during the iterative process. These shift vectors implicitly reflect the relative position of a sample within the distribution, and thus can serve as a meaningful indicator of outlierness. In other words, mean shift evaluates the difference between a sample and the mean of its neighbors, and this deviation can be exploited to assess the outlierness of the sample in distance. Motivated by this fact, a mean shift method is introduced for distance anomaly estimation.

Definition 12. Given a fuzzy information system $FIS = (U, A, V, f)$, the mean shift degree of a sample $x_i \in U$ with respect to an attribute $a \in A$ is defined as

$$MSD_a(x_i) = \frac{\sum_{x_j \in N_k^a(x_i)} \left((1 - R_a(x_i, x_j)\right)}{|N_k^a(x_i)|}. \tag{14}$$

The mean shift degree quantifies the deviation of a sample from its neighbors in a specific attribute. A higher mean shift degree means a greater deviation of the sample in that attribute, which may suggest that the sample is an outlier.

Definition 13. Given a fuzzy information system $FIS = (U, A, V, f)$, the offset distance of a sample $x_i \in U$ with respect to the attribute set A is defined as

$$dist_{off}(x_i) = \frac{1}{|A|} \cdot \sum_{a \in A} MSD_a(x_i). \tag{15}$$

Definition 14. Given a fuzzy information system $FIS = (U, A, V, f)$, the distance measure of a sample $x_i \in U$ with respect to the attribute set A is defined as

$$Dis(x_i) = \max_{a \in A} \left\{ MSD_a(x_i) - dist_{off}(x_i) \right\}. \tag{16}$$

The offset distance reflects the average deviation of a sample across all attributes, providing an overall trend of the sample to its neighbors. While the distance measure captures the maximum deviation of the sample under different attributes from the average offset distance. The smaller the distance measure, the greater the likelihood of the sample being normal; otherwise, the sample is more likely to be an outlier.

3.4. Anomaly detection by combining fuzzy granule density and mean shift distance-based outlier degrees

Real-world data may exhibit complex distribution, and thus different types of outliers may co-exist. To detect local outliers, an improved density measure is proposed, which incorporates neighborhood density information by using fuzzy granules. Meanwhile, a distance measure using mean shift is proposed, which evaluates the distance deviation and contributes to detecting global and group outliers. To enable the identification of various types of outliers, an anomaly measure that fuses density and distance information is developed.

Definition 15. Given a fuzzy information system $FIS = (U, A, V, f)$, the outlier score of a sample $x_i \in U$ is defined as

$$OS\left(x_i\right) = \frac{Dis\left(x_i\right)}{Den\left(x_i\right)}. \tag{17}$$

The outlier score integrates the density and distance information. The density measure effectively enhances the ability to detect local outliers with low density. By further introducing the distance measure, the capability to detect various types of outliers is greatly improved. Specifically, normal samples typically exhibit high densities and small distances, resulting in low outlier scores. Local outliers are located near high-density regions, which gives small distances, but the densities of their neighbors are relatively low, leading to low densities. In contrast, group outliers tend to form small clusters far from the distributions of normal samples. This results in high densities and large distances. Global outliers are usually scattered in positions away from dense regions and thus exhibit low densities and high distances. Besides the above types of outliers, there also exist samples located near the decision boundaries. These samples may not lie in dense regions, nor deviate significantly from the distribution of normal samples. Their outlier scores are very similar to those of local outliers, typically falling between normal samples and typical anomalies. In general, global outliers tend to have the highest scores, normal samples the lowest, while the scores of the local, group, and boundary outliers usually fall in between that of the global outliers and normal samples.

To fuse the distance and density information for outlier detection, different strategies, such as weighted sum, Rank, and Kullback-Leibler (KL) divergence, can be considered. The weighted sum often requires careful tuning of weight parameters, potentially introducing additional hyperparameters. Rank-based fusion is highly sensitive to sample size, which may limit its robustness. KL divergence, as another alternative, primarily captures statistical differences between two measures, but fails to leverage the complementary information provided by the two proposed metrics. In contrast, the ratio-based strategy is adopted, which amplifies the differences between anomalies and normal samples without introducing extra parameters, enabling more effective detection of different types of outliers. Based on the proposed outlier score, a novel anomaly detection measure can be developed, and the overall process of identifying outliers is shown in Algorithm 1.

The algorithm first computes the fuzzy granule density and the fuzzy k-neighborhood for each sample. It then determines the optimistic and pessimistic granule densities, followed by the calculation of the granule density ratio. Simultaneously, the mean shift degree for each attribute is computed, and the distance-based anomaly estimation is obtained after the computation of the offset distance. Finally, the outlier score is derived by integrating these two measures.

The time complexity of the algorithm involves three parts: 1) bandwidth parameter and fuzzy similarity matrix; 2) density measure; and 3) distance measure. Assume that $|A|$ denotes the number of attributes and $|U|$ represents the number of samples. For the first part, the time complexity is $O(|A||U|^2)$, which is mainly due to the calculation of the fuzzy similarity matrix. The calculation of the density measure has a time complexity of $O(|A||U|^2)$. The distance measure computation also takes a time complexity of $O(|A||U|^2)$. Therefore, the total time complexity of the algorithm is $O(|A||U|^2)$, and the space complexity is $O(|A||U|^2)$.

4. Experiment

In this section, the effectiveness of the proposed method is first validated on three artificially generated datasets containing different types of outliers. Subsequently, the proposed method is compared with fourteen benchmark methods on sixteen selected

Algorithm 1 Anomaly detection algorithm by fusing fuzzy granule density and distance shift.

Input: Fuzzy information system $FIS = (U, A, V, f)$, the number of neighbors k, and the tuning factor μ.
Output: Outlier score vector O_S.

```
Initialize outlier score vector O_S = {0, 0, ..., 0};
Compute bandwidth and fuzzy similarity matrix;
for i ← 1 to n do
    for each attribute a ∈ A do
        Calculate fuzzy granule density FGD_a(x_i) and fuzzy k-neighborhood N_a^k(x_i);
        Calculate pessimistic granule density PGD_a(x_i), optimistic granule density OGD_a(x_i) and granule density ratio GDR_a(x_i);
        Compute mean shift degree MSD_a(x_i);
    end for
    Calculate the density measure den(x_i) and distance measure dis(x_i);
    Compute outlier score OS(x_i);
end for
return Outlier score O_S;
```

Table 1
The investigated datasets.

No.	Datasets(Abbr.)	$\|A\|(\|A_{dis}\|)$	$\|U\|$	$\|O\|$	Source
1	Annealing(Anne)	38(10)	798	42	https://github.com/Minqi824/ADBench
2	Arrhythmia(Arrhy)	274(66)	452	66	https://github.com/BElloney/Outlier-detection
3	Diabetes(Diab)	16(15)	520	200	https://github.com/BElloney/Outlier-detection
4	Heart(Heart)	12(5)	299	96	https://archive.ics.uci.edu
5	Horse(Horse)	27(20)	256	12	https://github.com/BElloney/Outlier-detection
6	Ionosphere(Iono)	34(1)	249	24	https://github.com/BElloney/Outlier-detection
7	Magic Gamma(Magic)	10(1)	19020	6688	https://github.com/Minqi824/ADBench
8	Monks(Monks)	6(6)	253	25	https://odds.cs.stonybrook.edu
9	Mushroom(Mush)	22(21)	4429	221	https://github.com/BElloney/Outlier-detection
10	PageBlocks(Page)	10(0)	5171	258	https://github.com/BElloney/Outlier-detection
11	PersonGait(Person)	321(0)	48	3	https://archive.ics.uci.edu
12	Sick(Sick)	29(8)	3613	72	https://github.com/BElloney/Outlier-detection
13	Spambase(Spam)	57(1)	2844	56	https://github.com/BElloney/Outlier-detection
14	Thyroid Disease(Thyroid)	28(21)	9172	74	https://github.com/BElloney/Outlier-detection
15	Yeast(Yeast)	8(1)	1484	507	https://github.com/Minqi824/ADBench
16	Zoo(Zoo)	16(15)	101	13	https://archive.ics.uci.edu

datasets, followed by ablation studies and sensitivity analysis. All experiments were implemented in Python 3.9.3 on a computer with the operating system of Windows 11, an Intel Core i9-12900K @3.20 GHz processor, and 128GB RAM.

4.1. Experimental datasets

Sixteen anomaly detection datasets, containing continuous or discrete attributes or both, are selected from publicly available repositories. The details of these datasets are summarized in Table 1, where the third to fifth columns denote the total number of attributes, samples, and outliers, respectively, with the values in the brackets representing the number of discrete attributes. The dataset underwent a unified preprocessing procedure before the experiments. Missing values were imputed based on nearest neighbors. All continuous and discrete features were then scaled to the range $[0, 1]$ using min-max normalization. The UMAP-based visualizations of these datasets are presented in Fig. 2.

4.2. Evaluation metrics

To assess the performance of the proposed method, the precision and recall rates are employed, which are defined as

$$P(t) = \frac{|OS(t) \cap OS^o|}{|OS(t)|}, \tag{18}$$

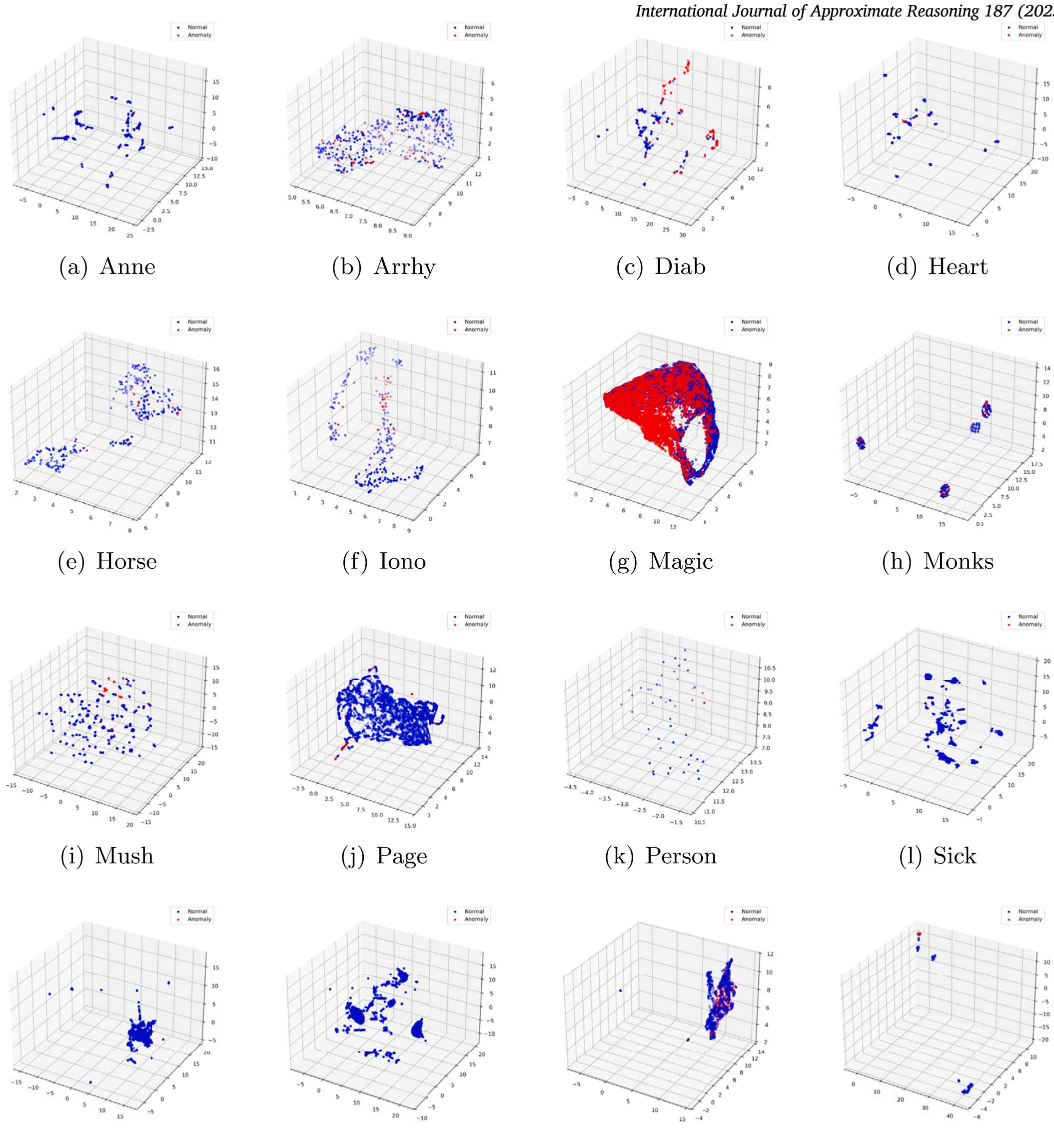
$$R(t) = \frac{|OS(t) \cap OS^o|}{|OS^o|}, \tag{19}$$

where t is a given proportion of outliers, $OS(t)$ denotes the detected outliers with top t scores, and OS^o is the ground-truth outlier set.

In addition, the Area Under the ROC Curve (AUROC) is also adopted, which is defined as

$$\text{AUROC} = \frac{\sum_{x_i \in OS^o} \sum_{x_j \in U \setminus OS^o} \mathbb{I}(S(x_i) > S(x_j))}{|OS^o| \cdot |U \setminus OS^o|}, \tag{20}$$

where $\mathbb{I}(\cdot)$ is the indicator function. AUROC evaluates detection performance by evaluating the probability that an anomaly scores higher than a normal sample. Higher AUROC indicates better detection ability.



(a) Anne (b) Arrhy (c) Diab (d) Heart

(e) Horse (f) Iono (g) Magic (h) Monks

(i) Mush (j) Page (k) Person (l) Sick

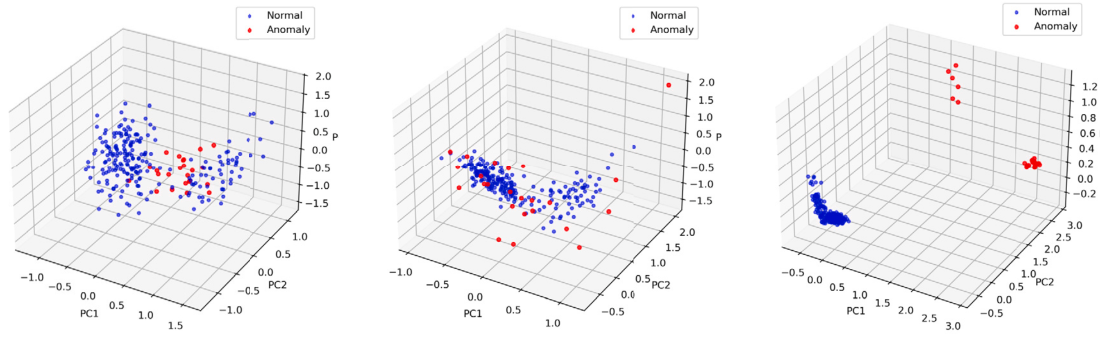
(m) Spam (n) Thyroid (o) Yeast (p) Zoo

Fig. 2. UMAP visualization of experimental datasets. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

4.3. Performance on synthetic datasets

To examine the effectiveness of the proposed method in detecting various types of outliers, three artificially generated datasets containing global, local, and group outliers were synthesized. Particularly, the Heart dataset in Table 1 was considered as the base data, and the generation of different types of outliers followed the method outlined in [27]. Generally, normal samples that follow a distribution of Gaussian Mixture Model (GMM) are selected. Local outliers are drawn from the same GMM with modified covariance. Global outliers follow a uniform distribution, while group outliers are generated from a GMM with shifted means. Note that continuous variables were re-normalized after injecting outliers to ensure consistency. The reduced 3-dimensional data after Principal Component Analysis (PCA) are illustrated in Fig. 3.

As shown in Fig. 3, three types of outliers exhibit distinct distribution patterns. Local outliers lie on the boundaries of regions with high density, global outliers are randomly scattered far from regions with high density, and group outliers form small clusters away from the distribution of normal samples. Table 2 presents the comparative results of the traditional FRS-based method and the proposed method on the three datasets. The results indicate that the proposed method consistently outperforms traditional FRS method in detecting all types of outliers. Although FRS-based methods obtained competitive performance in identifying global and group outliers, it failed to detect some local outliers due to its reliance on the fuzzy similarity metric, which can not well capture the neighborhood structure information. In contrast, the proposed method leveraged fuzzy granule density information on the local neighborhood, enabling a more accurate sample density estimation. Therefore, it attained good detection performance for both local



(a) local outliers (b) global outliers (c) group outliers

Fig. 3. The artificial datasets after PCA dimensionality reduction.

Table 2
Performance of the traditional FRS method and the proposed method on synthetic datasets.

t	Local				Global				Group			
	FRS		Ours		FRS		Ours		FRS		Ours	
	$P(t)$	$R(t)$	$P(t)$	$R(t)$	$P(t)$	$R(t)$	$P(t)$	$R(t)$	$P(t)$	$R(t)$	$P(t)$	$R(t)$
5%	1.0000	0.1458	1.0000	0.1458	1.0000	0.1458	1.0000	0.1458	1.0000	0.1458	1.0000	0.1458
10%	1.0000	0.3021	1.0000	0.3021	1.0000	0.3021	1.0000	0.3021	1.0000	0.3021	1.0000	0.3021
15%	1.0000	0.4583	1.0000	0.4583	1.0000	0.4583	1.0000	0.4583	0.9773	0.4479	1.0000	0.4583
20%	0.9492	0.5833	1.0000	0.6146	0.9492	0.5833	1.0000	0.6146	0.9831	0.6042	1.0000	0.6146
25%	0.8784	0.6771	0.9595	0.7396	0.8649	0.6667	1.0000	0.7708	0.9459	0.7292	1.0000	0.7708
30%	0.8427	0.7813	0.9213	0.8542	0.7978	0.7396	1.0000	0.9271	0.9326	0.8646	1.0000	0.9271
35%	0.7885	0.8542	0.8365	0.9063	0.7115	0.7708	0.9231	1.0000	0.8750	0.9479	0.9231	1.0000
40%	0.7143	0.8854	0.7731	0.9583	0.6639	0.8229	0.8067	1.0000	0.7899	0.9792	0.8067	1.0000
45%	0.6642	0.9271	0.6940	0.9688	0.6045	0.8438	0.7164	1.0000	0.7164	1.0000	0.7164	1.0000
50%	0.6040	0.9375	0.6242	0.9688	0.5705	0.8854	0.6443	1.0000	0.6443	1.0000	0.6443	1.0000
55%	0.5549	0.9479	0.5732	0.9792	0.5305	0.9063	0.5854	1.0000	0.5854	1.0000	0.5854	1.0000
60%	0.5084	0.9479	0.5307	0.9896	0.4860	0.9063	0.5363	1.0000	0.5363	1.0000	0.5363	1.0000
65%	0.4794	0.9688	0.4897	0.9896	0.4639	0.9375	0.4948	1.0000	0.4948	1.0000	0.4948	1.0000
70%	0.4450	0.9688	0.4545	0.9896	0.4354	0.9479	0.4593	1.0000	0.4593	1.0000	0.4593	1.0000
75%	0.4196	0.9792	0.4286	1.0000	0.4152	0.9688	0.4286	1.0000	0.4286	1.0000	0.4286	1.0000
80%	0.3975	0.9896	0.4017	1.0000	0.3933	0.9792	0.4017	1.0000	0.4017	1.0000	0.4017	1.0000
85%	0.3740	0.9896	0.3780	1.0000	0.3701	0.9792	0.3780	1.0000	0.3780	1.0000	0.3780	1.0000
90%	0.3569	1.0000	0.3569	1.0000	0.3532	0.9896	0.3569	1.0000	0.3569	1.0000	0.3569	1.0000
95%	0.3380	1.0000	0.3380	1.0000	0.3380	1.0000	0.3380	1.0000	0.3380	1.0000	0.3380	1.0000
100%	0.3211	1.0000	0.3211	1.0000	0.3211	1.0000	0.3211	1.0000	0.3211	1.0000	0.3211	1.0000
Avg.	0.6318	0.8172	0.6540	0.8432	0.6134	0.7917	0.6695	0.8609	0.6582	0.8510	0.6695	0.8609
AUROC	0.9379		0.9734		0.8805		0.8989		0.9743		1.0000	

outliers and sparsely distributed global outliers. The introduction of the mean shift degree further enhanced its ability to recognize isolated global outliers and group outliers. As a result, the proposed method demonstrated high adaptability to various types of outliers and achieved superior overall detection performance on synthetic datasets.

4.4. Comparison experiments on real-world datasets

To evaluate the effectiveness of the proposed method, comparison experiments with other methods were conducted on real-world datasets, including distance-based k Nearest Neighbors (KNN) [9], density-based methods such as Local Outlier Factor (LOF) [11], Clustering-based Local Outlier Factor (CBLOF) [12], Connectivity-based Outlier Factor (COF) [28], Histogram-based Outlier Score (HBOS) [7], Subspace Outlier Detection (SOD) [29], statistical methods such as Copula-Based Outlier Detection (COPOD) [30], Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions (ECOD) [31], linear models such as Principal Component Analysis (PCA) [32], One-Class Support Vector Machines (OCSVM) [33], and ensemble methods such as Isolation Forest (IForest) [14]. In addition, some FRS-based methods such as Detecting Fuzzy Neighborhood Outliers (DFNO) [34], granular-ball FRS-based anomaly detection (GBFRD) [35], and Anomaly Detection Based on Weighted Fuzzy-Rough Density (WFRDA) [19] were

Table 3
The performance of the selected methods on all datasets.

Datasets	KNN	LOF	CBLOF	COF	HBOS	SOD	COPOD	ECOD	PCA	OCSVM	IForest	DFNO	GBFRD	WFRDA	Ours
Anne	0.7352	0.7294	0.7413	0.6554	0.7479	0.4282	0.7967	0.7874	0.7521	0.7133	0.7500	0.7416	0.7577	0.7387	**0.8207**
Arrhy	0.8053	0.8031	0.7694	0.8086	0.8128	0.7197	0.7998	0.8052	0.7748	0.7872	0.8059	0.7735	0.5274	0.8140	**0.8170**
Diab	0.2176	0.3099	0.2191	0.4891	0.2592	0.2455	0.2492	0.3108	0.2960	0.3518	0.2160	0.3267	0.2452	0.2355	**0.6555**
Heart	0.6670	0.6866	0.5766	0.5709	0.7019	0.6090	0.6570	0.6044	0.6690	0.6610	0.6699	0.6459	0.6540	0.7202	**0.7362**
Horse	0.6554	0.6609	0.5564	0.5792	0.9781	0.7128	0.9857	0.9805	0.9344	0.6230	0.9494	0.9016	0.9624	0.9819	**0.9904**
Iono	0.9978	0.9980	0.9996	0.9781	0.8372	**1.0000**	0.9954	0.9943	0.9993	0.9991	0.9994	0.9969	0.9969	0.9915	**1.0000**
Magic	0.7815	0.7690	0.7547	0.6184	0.7036	0.7584	0.6812	0.6382	0.6675	0.5947	0.7259	0.7476	0.7033	0.7093	**0.8951**
Monks	0.9411	0.9647	0.9223	0.8867	0.5691	0.6724	0.9668	0.9691	0.8961	0.9449	0.9863	0.8953	0.4635	0.9900	**0.9970**
Mush	0.9290	0.9300	0.8500	0.6255	0.9606	0.5632	0.9449	0.9489	0.9243	0.9210	0.8852	0.1267	0.9759	0.9703	**0.9840**
Page	0.9535	0.9719	0.9584	0.6910	0.9285	0.7494	0.9386	0.9376	0.9721	0.9525	0.9641	0.8430	0.9515	0.9428	**0.9810**
Person	0.8370	0.8741	0.1259	0.7704	0.8667	0.5556	0.8444	0.8667	0.8370	0.7852	0.6370	0.8000	0.8667	0.9259	**0.9704**
Sick	0.7406	0.7450	0.7926	0.6604	0.8096	0.7179	0.7804	0.8425	0.7727	0.7723	0.7864	0.8051	0.7816	0.8365	**0.8665**
Spam	0.6641	0.6696	0.6683	0.7639	0.7741	0.8068	0.8146	0.8123	0.7466	0.6704	0.8270	0.8230	0.7036	0.7249	**0.8740**
Thyroid	0.6257	0.6344	0.6659	0.5600	0.6273	0.6274	0.6110	0.5808	0.6716	0.6539	0.6307	0.6507	0.6427	0.5301	**0.8170**
Yeast	0.3801	0.3660	0.4581	0.4449	0.4045	0.4340	0.3813	0.4439	0.4180	0.4141	0.3971	0.3953	0.4048	0.3988	**0.4929**
Zoo	0.6451	0.7893	0.3435	0.0568	0.4003	0.3278	**0.8676**	0.8488	0.7806	0.5656	0.4701	0.3741	0.7972	0.7115	0.8645
Avg.	0.7235	0.7439	0.6501	0.6350	0.7113	0.6205	0.7697	0.7732	0.7570	0.7131	0.7313	0.6779	0.7147	0.7639	**0.8602**

selected for comparison, which were performed using the official codes provided by their papers.[1] Deep learning-based methods were not selected because most of the investigated datasets are tabular data, on which deep learning-based methods generally perform relatively poorly. Furthermore, these methods lack interpretability, which is a vital characteristic for outlier detection.

For KNN, LOF, and COF, the parameter k was varied from 1 to $n/2$, and the optimal value was determined using a bisection search approach. Other comparative methods utilized their default parameters as defined in PyOD.[2] For the IForest algorithm, experiments were repeated 20 times to weaken the effect of random initialization, and its final performance was averaged. There are two parameters in the proposed methods. The bandwidth coefficient was varied from 0.05 to 2.5 with an increment of 0.05, while the search strategy for k was kept consistent with the methods of KNN, LOF, and COF. With the aforementioned configurations, the comparative experiments were conducted on the selected datasets, and their results are shown in Table 3, where the best performance of each dataset is in bold.

As seen in Table 3, the selected methods effectively detect outliers on some datasets, but their performance varies significantly. Distance-based method KNN performed well in low-dimensional datasets but worse in high-dimensional datasets like Person due to data sparsity, which seriously affects the accuracy of distance estimation. Density-based method LOF obtained good performance on datasets with continuous attributes, such as Magic, Page, and Person, but performed poorly on datasets containing discrete attributes, such as Horse. This may be due to the limitations in the inaccurate density estimation for samples with discrete attributes. CBLOF showed poor performance on multiple datasets, including Horse, Person, and Zoo. COF ranked worst in overall performance among all compared methods. This may suggest that relying solely on connectivity between samples fails to effectively handle datasets with complex distributions. HBOS exhibited large performance fluctuations across datasets, which may be attributed to the mismatch between the model assumptions and data distribution. SOD demonstrated relatively strong performance on the Iono dataset. However, its weak overall performance may suggest that conducting anomaly detection in low-dimensional subspaces may lose some discriminative information for outliers. Statistical methods such as COPOD and ECOD showed good performance, ranking among the top compared to other selected methods. But their capability of detecting different types of outliers remained limited. IForest achieved a strong result on Monks but showed a poor result on Diab. This inconsistency may be caused by the low quality of the random attribute partition.

For FRS-based methods, DFNO exhibited unsatisfactory performance on datasets such as Arrhy, Mush, and Page. A possible explanation is that fuzzy neighborhood density relies on the local reachable similarity mean, which might be ineffective for datasets with uneven local density distributions. GBFRD demonstrated unstable performance, performing the worst on datasets like Arrhy and Monks. This may be due to the use of k-means for clustering, which is sensitive to initial cluster centers, leading to inconsistent results and potential misclassification of outliers. WFRDA also showed instability, performing well on Arrhy, Heart, and Monks but yielding less favorable results on Iono and Thyroid. The reason for this may be that WFRDA calculates anomaly values by integrating density differences with fixed weights, which makes it difficult to identify outliers with local density changes.

Across the selected datasets, the proposed method significantly outperformed comparative methods. It surpassed KNN, LOF, and CBLOF by 18.89%, 15.63%, and 32.30%, respectively. Compared to COF, the performance gain was 35.46%, highlighting the advantage in handling local density variations. It also outperformed HBOS by 20.92% and SOD by 38.63%, showing the superiority over statistical and subspace-based approaches. Further, the performance improvements over COPOD and ECOD were ranged from 11.76% to 11.24%, while the gains against PCA, OCSVM, and IForest were 13.62%, 20.62%, and 17.63%. In addition, compared to FRS-based methods, the performance improvements were ranged from 12.60% to 26.88%, demonstrating its effectiveness in detecting different types of outliers. This is because the proposed method incorporates a fuzzy granule measure to capture the density of samples, making it more robust in detecting local outliers. Moreover, the measure of outlier scores fuses the density and distance information, providing the ability to identify different types of outliers. As a result, the proposed method achieves good performance across multiple datasets.

[1] https://github.com/BELLoney?tab=repositories.

[2] https://github.com/yzhao062/pyod.

Table 4
The ranking of the selected methods over all datasets.

Datasets	KNN	LOF	CBLOF	COF	HBOS	SOD	COPOD	ECOD	PCA	OCSVM	IForest	DFNO	GBFRD	WFRDA	Ours
Anne	11	12	9	14	7	15	2	3	5	13	6	8	4	10	1
Arrhy	6	8	13	4	3	14	9	7	11	10	5	12	15	2	1
Diab	14	6	13	2	8	10	9	5	7	3	15	4	11	12	1
Heart	7	4	14	15	3	12	9	13	6	8	5	11	10	2	1
Horse	12	11	15	14	5	10	2	4	8	13	7	9	6	3	1
Iono	8	7	3	14	15	1.5	11	12	5	6	4	9.5	9.5	13	1.5
Magic	2	3	5	14	9	4	11	13	12	15	7	6	10	8	1
Monks	8	6	9	12	14	13	5	4	10	7	3	11	15	2	1
Mush	8	7	12	13	4	14	6	5	9	10	11	15	2	3	1
Page	6	3	5	15	12	14	10	11	2	7	4	13	8	9	1
Person	8.5	3	15	12	5	14	7	5	8.5	11	13	10	5	2	1
Sick	13	12	6	15	4	14	9	2	10	11	7	5	8	3	1
Spam	15	13	14	8	7	6	4	5	9	12	2	3	11	10	1
Thyroid	11	7	3	14	10	9	12	13	2	4	8	5	6	15	1
Yeast	14	15	2	3	9	5	13	4	6	7	11	12	8	10	1
Zoo	8	5	13	15	11	14	1	3	6	9	10	12	4	7	2
Avg.	9.4688	7.6250	9.4375	11.5000	7.8750	10.5938	7.5000	6.8125	7.2812	9.1250	7.3750	9.0938	8.2813	6.9375	1.0938

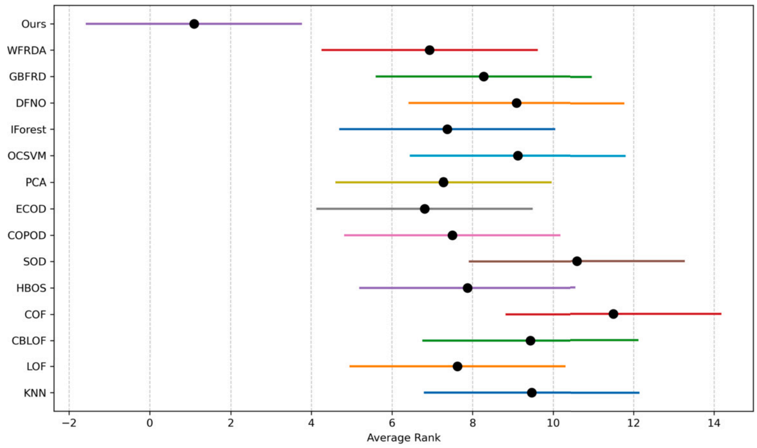


Fig. 4. The Nemenyi test diagram.

4.5. Statistical significance analysis

To further compare the proposed method and other selected methods, statistical significance tests were employed to evaluate their performance difference. The first is the Friedman test, which is used to determine whether the difference in performance between all methods is statistically significant. The null hypothesis of the Friedman test is that “all methods have the same performance”. If the proposed method passes the initial test, a Nemenyi post-hoc test is required to further determine whether the proposed method is statistically superior to other methods. The selected methods were first ranked based on their performance, and then the Friedman test was applied to their ranked values.

In the experiments, there are 15 methods and 16 datasets, and the critical value of $F(14, 210)$ at a significance level $\alpha = 0.05$ is 1.7391. If the performance of the selected methods is equivalent, the statistical value of the Friedman test should not exceed the critical value of 1.7391; otherwise, there is a significant difference in performance between the selected methods. The rank information of all selected methods is presented in Table 4.

According to the performance rank results in Table 4, the statistical value of the Friedman test is 5.7194, which is greater than the critical value of 1.7391. The null hypothesis was rejected, and the performance difference among the methods is statistically significant. The results of the Friedman test indicate that a post hoc test is needed to further examine the performance differences between the selected methods. Thus, the Nemenyi post hoc test was performed. At a significance level of $\alpha = 0.05$, the critical value of the Nemenyi test is 5.3620, and the performance differences between the selected methods are shown in Fig. 4.

As shown in Fig. 4, if the difference between the average rank values of two methods is greater than the critical value 5.3620, it means a significant difference in their performance. By observing the results in Fig. 4, it is evident that the average rank value of the proposed method is significantly better than that of other methods, and its differences in rank values from other methods are greater than the critical value. Thus, there is a significant difference in performance between the proposed method and other methods. These results further highlight the superiority of the proposed method in detecting different types of outliers.

4.6. Ablation experiment

To evaluate the significance of each component in the proposed method, ablation experiments were conducted on the selected datasets, and the results are shown in Table 5, where the second and third columns report the results with only distance anomaly

Table 5
The results of ablation experiments.

Dataset	Distance anomaly degree	Density anomaly degree	The proposed method
Anne	0.7304	0.7727	**0.8207**
Arrhy	0.2772	0.8103	**0.8170**
Diab	0.6462	0.2436	**0.6555**
Heart	0.7242	0.7053	**0.7362**
Horse	0.2145	0.9815	**0.9904**
Iono	**1.0000**	0.9974	**1.0000**
Magic	0.2469	0.8944	**0.8951**
Monks	0.9779	0.9939	**0.9970**
Mush	0.9090	0.9513	**0.9840**
Page	0.9705	0.9675	**0.9810**
Person	0.4926	0.3481	**0.9704**
Sick	0.8647	0.7993	**0.8665**
Spam	0.6813	0.8138	**0.8740**
Thyroid	0.2772	0.8103	**0.8170**
Yeast	**0.4929**	0.4604	**0.4929**
Zoo	0.8601	0.7946	**0.8645**
Avg.	0.6478	0.7715	**0.8602**

degree and density anomaly degree, respectively, and the last column indicate the proposed method with both distance and density information.

As shown in Table 5, using distance or density only exhibits unstable performance across different datasets. The former often excels on datasets dominated by global outliers, while the latter is more effective for local outliers. The proposed method integrates the advantages of both components, effectively capturing different types of outliers across different datasets. For instance, on the Anne dataset, the proposed method improved by 12.36% and 6.21% compared to its components, respectively. On the Horse dataset, the performance of the distance component was 0.2145, and the performance of the density component was 0.9815. After fusing these information, the proposed method achieved a performance of 0.9904. By averaging the results on all datasets, the proposed method outperforms the distance and density components by 32.79% and 11.50%, respectively, indicating its effectiveness on anomaly detection.

4.7. Parameter sensitivity analysis

There are two parameters in the proposed method, namely the bandwidth coefficient and the number of nearest neighbors to compute the fuzzy similarity of samples. To investigate the effect of these two parameters, a grid search was conducted. The bandwidth coefficient μ was changed from 0.1 to 2.0 with a step size of 0.1, and k was varied from 1 to $n/4$ with a step size of $n/40$, where n is the number of samples. The results of the grid search are shown in Fig. 5.

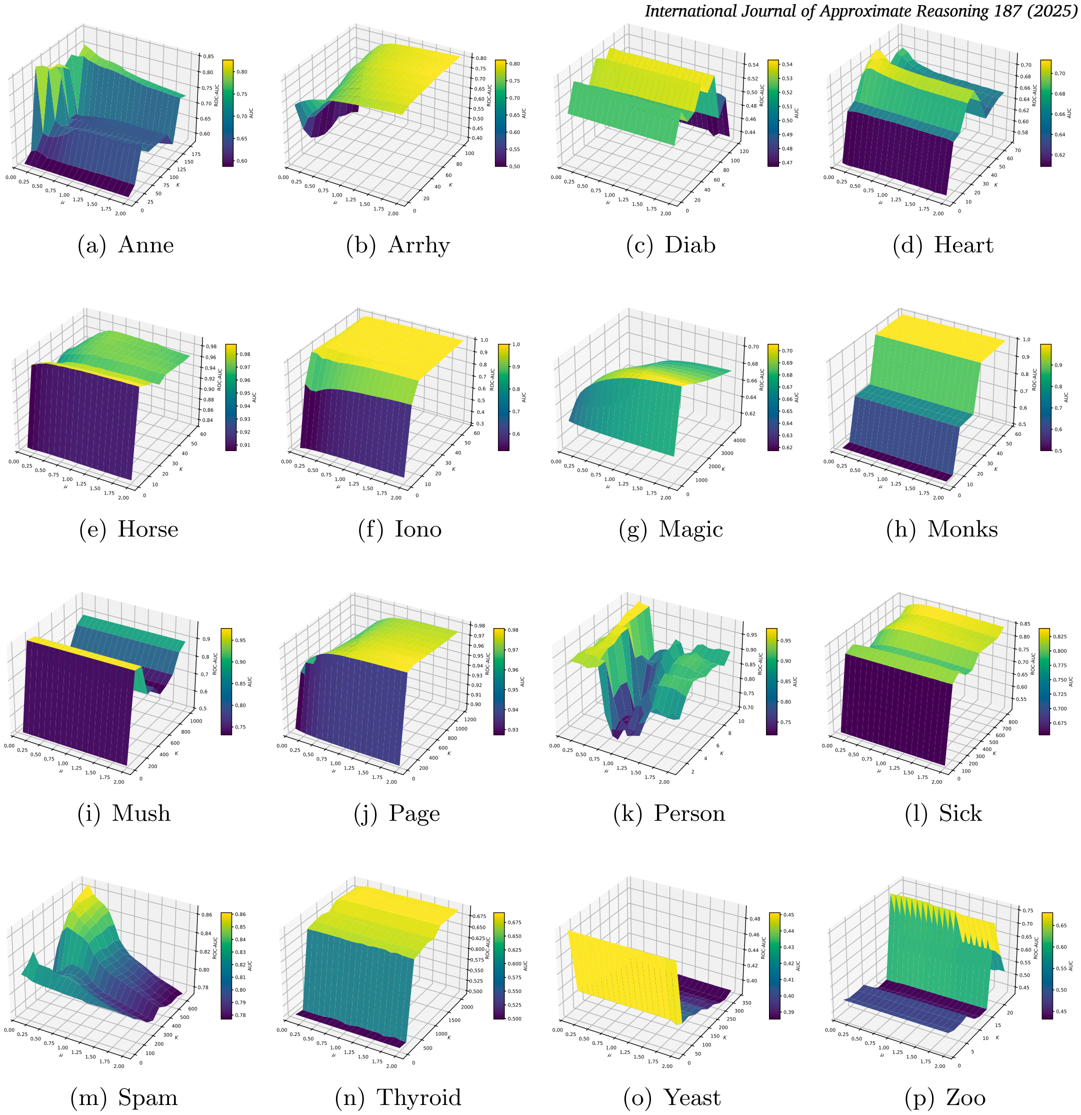
From the results in Fig. 5, the following conclusions could be drawn. Firstly, for some datasets with discrete attributes, the bandwidth coefficient μ has a limited impact on performance, whereas the parameter k plays a more significant role. For example, on the Sick and Thyroid datasets, the performance greatly increases when varying the value of k from 1 to n/80, but remains relatively stable when changing the value of μ from 0.1 to 2.0. The reason for this is that when computing fuzzy similarity, discrete attributes are essentially treated with a binary distance of 0 and 1, which makes k to be the main factor affecting the bandwidth.

Secondly, for some datasets dominated by continuous attributes, the overall performance improves as k increases and μ decreases. For example, on the PersonGait and Spambase datasets, performance generally fluctuates upward as μ decreases and k increases. A possible explanation is that reducing μ and increasing k enhances the ability to detect subtle variations when evaluating the outlier degree of samples.

In summary, the optimal parameters are data-specific. To balance different types of datasets, the parameter μ is suggested to be selected from the range of [0.3, 1.8], and the parameter k is picked from the range of [n/10, n/5].

5. Conclusions

In real-world scenarios, practical data may exhibit complex distributions entangled with different types of outliers, posing great challenges for anomaly detection. This study first proposes a density measure based on the fuzzy granule density to capture the density variation of samples in the neighborhood, facilitating the detection of local outliers. Then, a distance measure based on the concept of mean shift is developed to identify global and group outliers. Finally, a novel anomaly detection method is presented by integrating the density and distance measures. Extensive comparative experiments demonstrate the superior performance of the proposed method over other representative methods. Exploring more effective strategies to fuse different information for anomaly detection deserves further investigation. In addition, identifying outliers from data with missing values is another direction for future research.



(a) Anne (b) Arrhy (c) Diab (d) Heart
(e) Horse (f) Iono (g) Magic (h) Monks
(i) Mush (j) Page (k) Person (l) Sick
(m) Spam (n) Thyroid (o) Yeast (p) Zoo

Fig. 5. Performance under different parameters.

CRediT authorship contribution statement

Mengyao Liao: Writing – original draft, Visualization, Validation, Software, Methodology, Data curation. **Zhiyu Chen:** Writing – original draft. **Can Gao:** Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Jie Zhou:** Writing – review & editing, Data curation. **Xiaodong Yue:** Writing – review & editing, Data curation.

Declaration of competing 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.

Acknowledgements

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation (Grant Nos. 2024A1515011367 and 2025A1515011511), the National Natural Science Foundation of China (Grant Nos. 62476171 and 62476172), and the Guangdong Provincial Key Laboratory (Grant No. 2023B1212060076).

Data availability

Data will be made available on request.

References

[1] A. Boukerche, L. Zheng, O. Alfandi, Outlier detection: methods, models, and classification, ACM Comput. Surv. 53 (3) (2020) 1–37.
[2] J.T. Zhou, J. Du, H. Zhu, X. Peng, Y. Liu, R.S.M. Goh, Anomalynet: an anomaly detection network for video surveillance, IEEE Trans. Inf. Forensics Secur. 14 (10) (2019) 2537–2550.
[3] M. Ahmed, A.N. Mahmood, M.R. Islam, A survey of anomaly detection techniques in financial domain, Future Gener. Comput. Syst. 55 (2016) 278–288.
[4] C.W. Ten, J. Hong, C.C. Liu, Anomaly detection for cybersecurity of the substations, IEEE Trans. Smart Grid 2 (4) (2011) 865–873.
[5] M.E. Tschuchnig, M. Gadermayr, Anomaly detection in medical imaging-a mini review, in: International Data Science Conference, Springer Fachmedien, Wiesbaden, 2021, pp. 33–38.
[6] S. Ahmad, A. Lavin, S. Purdy, Z. Agha, Unsupervised real-time anomaly detection for streaming data, Neurocomputing 262 (2017) 134–147.
[7] M. Goldstein, A. Dengel, Histogram-based outlier score (HBOS): a fast unsupervised anomaly detection algorithm, in: 35th Annual German Conference on AI (KI-2012), 2012, pp. 59–63.
[8] L.J. Latecki, A. Lazarevic, D. Pokrajac, Outlier detection with kernel density functions, in: International Workshop on Machine Learning and Data Mining in Pattern Recognition, Springer, 2007, pp. 61–75.
[9] S. Ramaswamy, R. Rastogi, K. Shim, Efficient algorithms for mining outliers from large data sets, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 427–438.
[10] V. Hautamaki, I. Karkkainen, P. Franti, Outlier detection using k-nearest neighbour graph, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004, pp. 430–433.
[11] M.M. Breunig, H.-P. Kriegel, R.T. Ng, J. Sander, LOF: identifying density-based local outliers, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 93–104.
[12] Z. He, X. Xu, S. Deng, Discovering cluster-based local outliers, Pattern Recognit. Lett. 24 (9–10) (2003) 1641–1650.
[13] S. Chawla, A. Gionis, k-means: a unified approach to clustering and outlier detection, in: Proceedings of the 2013 SIAM International Conference on Data Mining, 2013, pp. 189–197.
[14] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation forest, in: 2008 Eighth IEEE International Conference on Data Mining, IEEE, 2008, pp. 413–422.
[15] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, W. Ziarko, Rough sets, Commun. ACM 38 (11) (1995) 88–95.
[16] F. Jiang, Y. Sui, C. Cao, A rough set approach to outlier detection, Int. J. Gen. Syst. 37 (5) (2008) 519–536.
[17] F. Maciá-Pérez, J.V. Berna-Martinez, A.F. Oliva, M.A.A. Ortega, Algorithm for the detection of outliers based on the theory of rough sets, Decis. Support Syst. 75 (2015) 63–75.
[18] Z. Yuan, H. Chen, T. Li, B. Sang, S. Wang, Outlier detection based on fuzzy rough granules in mixed attribute data, IEEE Trans. Cybern. 52 (8) (2021) 8399–8412.
[19] Z. Yuan, B. Chen, J. Liu, H. Chen, D. Peng, P. Li, Anomaly detection based on weighted fuzzy-rough density, Appl. Soft Comput. 134 (2023) 109995.
[20] Y. Wu, S. Wang, H. Chen, D. Peng, Z. Yuan, Kernelized fuzzy-rough anomaly detection, IEEE Trans. Fuzzy Syst. 32 (8) (2024) 4285–4296.
[21] L. Gao, M. Cai, Q. Li, A relative granular ratio-based outlier detection method in heterogeneous data, Inf. Sci. 622 (2023) 710–731.
[22] Z. Yuan, H. Chen, T. Li, X. Zhang, B. Sang, Multigranulation relative entropy-based mixed attribute outlier detection in neighborhood systems, IEEE Trans. Syst. Man Cybern. Syst. 52 (8) (2022) 5175–5187.
[23] X. Zhang, Z. Yuan, D. Miao, Outlier detection using three-way neighborhood characteristic regions and corresponding fusion measurement, IEEE Trans. Knowl. Data Eng. 36 (5) (2024) 2082–2095.
[24] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (1990) 191–209.
[25] Y. Cheng, Mean shift, mode seeking, and clustering, IEEE Trans. Pattern Anal. Mach. Intell. 17 (8) (1995) 790–799.
[26] Z. Yuan, H. Chen, P. Xie, P. Zhang, J. Liu, T. Li, Attribute reduction methods in fuzzy rough set theory: an overview, comparative experiments, and new directions, Appl. Soft Comput. 107 (2021) 107353.
[27] S. Han, X. Hu, H. Huang, M. Jiang, Y. Zhao, Adbench: anomaly detection benchmark, Adv. Neural Inf. Process. Syst. 35 (2022) 32142–32159.
[28] J. Tang, Z. Chen, A.W.-C. Fu, D.W. Cheung, Enhancing effectiveness of outlier detections for low density patterns, in: Advances in Knowledge Discovery and Data Mining: 6th Pacific-Asia Conference, PAKDD 2002 Taipei, Taiwan, May 6–8, 2002 Proceedings 6, Springer, 2002, pp. 535–548.
[29] H.P. Kriegel, P. Kröger, E. Schubert, A. Zimek, Outlier detection in axis-parallel subspaces of high dimensional data, in: Advances in Knowledge Discovery and Data Mining: 13th Pacific-Asia Conference, PAKDD 2009 Bangkok, Thailand, April 27-30, 2009 Proceedings 13, Springer Berlin Heidelberg, 2009, pp. 831–838.
[30] Z. Li, Y. Zhao, N. Botta, C. Ionescu, X. Hu, COPOD: copula-based outlier detection, in: 2020 IEEE International Conference on Data Mining (ICDM), IEEE, 2020, pp. 1118–1123.
[31] Z. Li, Y. Zhao, X. Hu, N. Botta, C. Ionescu, G.H. Chen, Ecod: unsupervised outlier detection using empirical cumulative distribution functions, IEEE Trans. Knowl. Data Eng. 35 (12) (2022) 12181–12193.
[32] M.L. Shyu, S.C. Chen, K. Sarinnapakorn, L. Chang, A novel anomaly detection scheme based on principal component classifier, in: Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop, 2003, pp. 172–179.
[33] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution, Neural Comput. 13 (7) (2001) 1443–1471.
[34] Z. Yuan, P. Hu, H. Chen, Y. Chen, Q. Li, DFNO: detecting fuzzy neighborhood outliers, IEEE Trans. Knowl. Data Eng. 37 (2024) 200–209.
[35] X. Su, Z. Yuan, B. Chen, D. Peng, H. Chen, Y. Chen, Detecting anomalies with granular-ball fuzzy rough sets, Inf. Sci. 678 (2024) 121016.