RESEARCH PAPER

ARDOD: adaptive radius density-based outlier detection

Farshad Rahmati[1] · Reza Heydari Gharaei[1] · Hossein Nezamabadi-pour[1]

Abstract

Outlier detection has garnered considerable attention in recent years due to its wide-ranging applications across various research domains. This surge in interest has led to the development of numerous detection techniques, predominantly based on distance or density metrics. A notable limitation of these existing methods is their reliance on parameter adjustments, significantly affecting the outcome. Additionally, these methods exhibit intrinsic flaws: distance-based approaches struggle with clusters with varying local densities, while density-based methods fail to identify patterns within low-density areas. Moreover, most prior techniques are adept at identifying only one kind of outlier—local, global, or group of outliers. Addressing these challenges, we introduce the Adaptive Radius Density-Based Outlier Detection (ARDOD) method, which departs from the traditional parameter-dependent approach. ARDOD is a novel parameter-free algorithm that dynamically determines the necessary parameters based on the data distribution within the feature space. This innovative method demonstrates robust performance in detecting all three categories of outliers. The efficacy and superior performance of ARDOD are validated through an extensive experimental analysis involving various synthetic and real-world datasets. This analysis showcases ARDOD's advantages over seven established methods: Local Outlier Factor (LOF), Angle-Based Outlier Detection (ABOD), Robust Distance-Based Outlier Score (RDOS), Directed density ratio Changing Rate-based outlier detection (DCROD), Empirical-Cumulative-distribution-based Outlier Detection(ECOD), mean-shift outlier detector(MOD +),and Local–Global Outlier Detection (LGOD), underscoring its potential as a versatile tool in outlier detection research.

Keywords Outlier detection · Anomaly detection · Density-based methods · Distance-based methods · Adaptive radius · Mass sharing method

1 Introduction

The exponential growth in data collection and information processing in recent years underscores the importance of advanced analysis techniques in science and technology. Despite the abundance of data, certain phenomena remain rare and unpredictably deviate from the norm, termed as outliers or anomalies [1]. These anomalies, which starkly contrast to regular observations, often carry critical information valuable for high-level research and applications, highlighting the significance of their detection and analysis. Identifying such anomalies, known as outlier or anomaly detection, is a pivotal aspect of data mining, contributing substantially to fields like intrusion detection, fraud prevention, crime analysis, and traffic management, among others [2].

Outlier detection methods are broadly categorized into supervised and unsupervised approaches, focusing on identifying data points that deviate significantly from most observations [3]. This process is crucial for understanding the underlying structure of data and enhancing the performance of predictive models by identifying and possibly excluding anomalies. In computer science, for instance, detecting unusual patterns can signify security threats. At the same time, in clustering problems, outliers may be treated as noise, affecting the accuracy of the clustering results [4].

The current state of the art in outlier detection has evolved to address these challenges through various methodologies, including distribution-based, clustering-based, density-based, and distance-based methods. Each approach offers unique advantages and faces specific limitations, such as sensitivity to parameter settings or the inherent difficulty in

✉ Reza Heydari Gharaei
   reza.heydari@eng.uk.ac.ir

1 Intelligent Data Processing Laboratory (IDPL), Department
   of Electrical Engineering, Shahid Bahonar University
   of Kerman, Kerman, Iran

handling multi-dimensional data. Despite these advances, the dynamic nature of data and the complexity of modern applications necessitate continuous improvement and innovation in outlier detection techniques [4–9].

This study is motivated by the need to refine outlier detection methods further, aiming to improve their accuracy, efficiency, and applicability across diverse datasets and contexts. By proposing a novel unsupervised outlier detection method, this research seeks to contribute to the field by offering a solution that balances sensitivity and specificity, even in challenging data scenarios. The following sections will delve into the related works [10], present the proposed methodology [11], discuss experimental results on synthetic and real-world datasets [12, 13], and conclude with the implications of the findings [14].

2 Related works

The local outlier factor (LOF) algorithm, one of the pioneering methods in outlier detection, utilizes a density-based approach leveraging the k-nearest Neighbor (k-NN) algorithm [15]. It computes an outlier score for each object based on local reachability density, where a significant discrepancy in neighborhood density increases the outlier score. The essence of LOF lies in calculating each object's outlier score relative to its local clustering structure. However, LOF's efficacy decreases when local distances are not distinct [4]. To improve upon LOF, Tang et al. introduced the Connectivity-based Outlier Factor (COF), which, unlike LOF's reliance on Euclidean distance and k-NN for local density estimation, employs chaining distance better to gauge an object's density [16]. COF distinguishes low density from isolativity—defined as an object's degree of connectivity with others. Nevertheless, COF sometimes inaccurately estimates density due to its indirect data distribution assumptions [4].

Acknowledging the limitations in density estimation by the LOF algorithm, particularly in complex datasets, Gao et al. introduced the Robust Kernel-based Local Outlier Factor (RKOF). This method adapts the kernel and weighted neighborhood to refine local density estimation, addressing the disadvantages of the LOF algorithm, including its dependency on the parameter k for defining the local neighborhood's scale [17].

Papadimitriou et al. proposed the Local Correlation Integral (LOCI) method to identify groups of outliers rather than individual anomalies [18]. LOCI utilizes the Multi-Granularity Deviation Factor (MDEF), marking objects as outliers if their deviation exceeds three times their neighbor's MDEF. This method effectively detects outliers amidst local density variations, identifying both distant clusters and isolated outliers [4].

Jin et al. introduced the Influenced Outlierness (INFLO) method in 2006. This method focuses on abnormal observation detection through a relationship-based density measure. INFLO considers both neighbors and reverse neighbors of an observation to estimate its relative distribution, aiming to overcome LOF's spatial representation limitations [19].

The challenge of dimensional curses significantly hampers the performance of traditional methods reliant on full-dimensional Euclidean space for distance estimation. Addressing this, Kriegel et al. developed the Angle-Based Outlier Detection (ABOD) method, which identifies outliers using the variance in angles between dataset observation vectors [20].

In 2014, Ha et al. proposed a novel approach based on gravity principles from physics, utilizing the center of gravity to denote each observation's geometric stability. Observations with a low center of gravity are deemed more stable and likely inliers, while those with a high center are considered unstable and potential outliers. This method introduces the "instability factor," a new benchmark for outlierness measurement using the k-NN algorithm [21]. Furthermore, in 2015, they developed the Observability Factor (OF), which quantifies the degree of an observation's inlierness, suggesting that objects with lower OF values are likelier to be outliers [22].

Tang et al. offered a local density-based method (RDOS) to estimate an object's density through various means, including k-nearest neighbors, reverse nearest neighbors, and shared neighbors [13]. Subsequently, Ning et al. aimed to refine RDOS by introducing a novel criterion for neighborhood density measurement, presenting the Relative Density-based Outlier Factor (RDOF) to address density-based methods' inability to detect low-density patterns [5].

Zhang et al. proposed a technique to circumvent the need for input parameters in cluster-based methods, introducing a method based on Cluster Outlier Factor and Mutual Density (COF). This technique first identifies an optimal number of neighbors for each observation using the NOF algorithm [23], then calculates mutual density for clustering, treating clusters with few patterns as outliers [24]. Wahid et al. introduced the Relative Kernel Density-based Outlier Score (RKDOS), which employs Weighted Kernel Density Estimation (WKDE) with an adaptive kernel size for density estimation, using both reverse nearest neighbors and k nearest neighbors. RKDOS applies a Gaussian kernel function for measurement smoothness [25].

2020 Henry et al. unveiled the Local-Gravitation-based Method (LGOD) for detecting outliers and boundary points. LGOD assesses each sample's local resultant force (LRF) applied by neighbors, detecting outliers and boundary samples by evaluating variations in LRF. This method asserts that its performance is independent of the k parameter value, addressing a standard limitation among existing techniques [26]. Adding to the innovative landscape of outlier detection, Jinwook Rhyu and his team have developed an automated method that significantly enhances data accuracy in biomanufacturing processes by using various algorithms to estimate missing data,

effectively mitigating the impact of outliers. This method's efficacy, highlighted through its application in monoclonal antibody production, underscores the potential of matrix completion methods in resolving complex data patterns supported by open-source software for broader applicability [27].

Furthermore, addressing the issue of class imbalance in classification tasks, recent advancements in Extreme Learning Machines (ELMs) have been explored for their potential in outlier detection across supervised, unsupervised, and semi-supervised frameworks, emphasizing the importance of methodological diversity in tackling anomalies in fields such as intrusion detection and medical diagnosis [28]. Additionally, exploring outlier detection within multiple circular regression models, particularly utilizing multivariate eye data, showcases the advancement in statistical methods capable of identifying outliers in complex datasets [29].

Lina Zheng, Lijun Chen, and Yini Wang introduced the Information Amount-Based Outlier Factor (IAOF), a new unsupervised method that utilizes information quantity to improve anomaly detection accuracy in categorical data systems, adding to the field's diverse methodological approaches [30].

In 2021, Jiawei Yang, Susanto Rahardja, and Pasi Fränti developed a mean-shift-based outlier detection method(MOD +). This approach, recalculating data points via their k-nearest neighbors, minimizes outlier effects preclustering and has proven more robust and adaptable than existing models across various datasets [6].

2022 saw Kangsheng Li and his team propose (DCROD) a technique based on the rate of change in directed density ratios, offering enhanced detection in complex data scenarios with minimal parameter sensitivity, benefiting sectors like network security and healthcare analytics [31].

That same year, Zheng Li, Yue Zhao, Xiyang Hu, Nicola Botta, Cezar Ionescu, and George H. Chen presented ECOD, a parameter-free method using empirical cumulative distribution functions (ECDF) for outlier identification. ECOD stands out for its accuracy, efficiency, and scalability, representing a leap in unsupervised anomaly detection for large, high-dimensional datasets without the need for intricate hyperparameter adjustments [32].

3 Proposed methodology

3.1 Problem description

In the initial section, we delineated two principal frameworks: supervised and unsupervised learning. Subsequently, we will elaborate on the various anomalies, encompassing local, global, and group of outliers, offering an in-depth analysis of each. Moreover, the advantages and disadvantages of different techniques employed to identify these outliers will be illuminated.

For a more concrete understanding, consider a two-dimensional synthetic dataset as an illustrative example (see Fig. 1), which encapsulates all three outlier types. This dataset is partitioned into four segments: global outliers, local outliers, a group of outliers, and entities resembling one another, termed inliers. The inlier group comprises two clusters, C1 and C2, characterized by the proximity of their members, denoted by the green section. On the other hand, the blue points (P1 & P2) are proximal to the inliers but possess distinct attributes that classify them as local outliers. The red points are significantly distanced from both clusters, categorizing them as global outliers. Conversely, the purple points represent a small cluster (C3), distanced from the inliers yet forming a cluster with a finite number of members, hence identified as a group of outliers. Consequently, entities such as C3, P1, P2, P3, and P4 are recognized as outliers. Over recent years, numerous supervised and unsupervised methodologies have been developed to detect such outliers, broadly categorized into four main types: distribution-based, clustering-based, density-based, and distance-based methods, each with inherent strengths and weaknesses.

- **Distribution-based Methods:** These methods ascertain an object as an outlier if its distribution significantly deviates from a normal distribution. Classified into parametric and non-parametric categories, these methods face limitations such as the unknown distribution of the dataset under study and their ineffectiveness in multi-dimensional datasets, rendering them less appealing to researchers [4].
- **Clustering-based Methods:** These methods identify outliers by attempting to cluster all samples within the dataset. Samples not belonging to any cluster are deemed outliers [4]. Despite primarily focusing on cluster identification rather than outlier detection, these methods have limited efficacy. Moreover, outliers are treated as noise in clustering problems, necessitating their identification and removal from the dataset.
- **Density-based Methods:** These methods detect outliers by contrasting the density of an object's neighborhood with that of others. An object is likely considered an outlier if there is a substantial discrepancy in density compared to its surroundings [5]. This approach is efficient even in datasets comprising multiple clusters of varying densities [14]. However, its efficiency markedly decreases in datasets characterized by low-density patterns, and the effectiveness is contingent upon parameter settings, like neighborhood size.
- **Distance-based Methods:** Primarily, these methods detect outliers by evaluating the distances among observations within the dataset [6]. An observation is presumed to be an outlier if it is significantly distant from its

Fig. 1 Illustrates the delineation of different types of outliers in a two-dimensional synthetic dataset, with clusters C1 and C2 representing inliers, P1 and P2 representing local outliers, P3 and P4 representing global outliers, and cluster C3 representing a collective of outliers
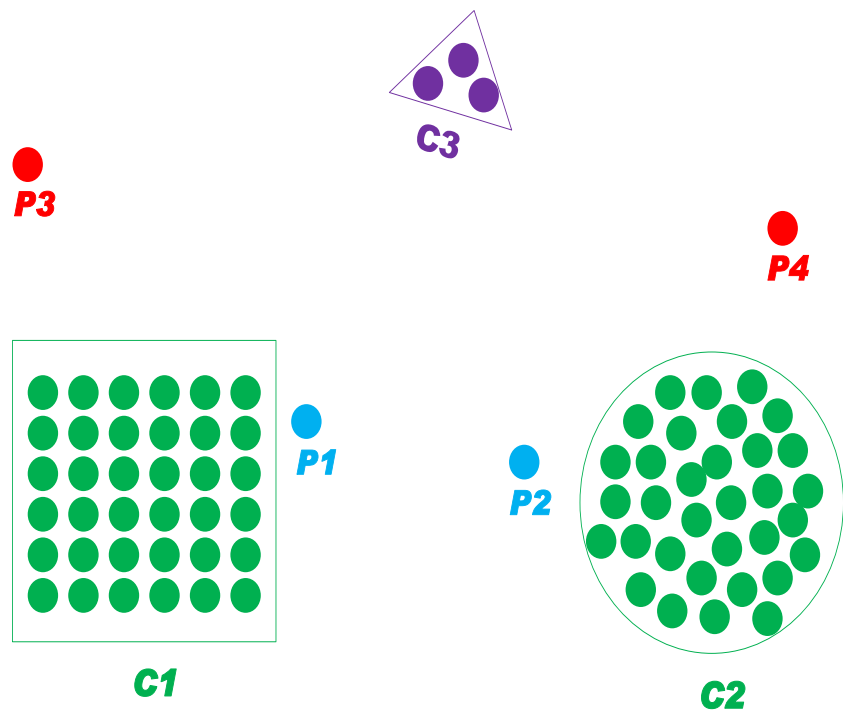
neighbors. These methods are preferred for their simplicity and rapidity, making them suitable for large datasets. Nonetheless, they struggle in datasets featuring multiple clusters of diverse densities, typically only identifying global outliers [7]. Additionally, the computation of distances in high-dimensional datasets may necessitate feature reduction, and selecting optimal parameters remains a challenge dependent on the dataset characteristics.

3.2 Motivation and contribution

As outlined in the preceding section, various methods for detecting outliers come with their challenges, largely dependent on their strategies. Generally, the main limitations of prior approaches can be distilled into three key areas:

a) The necessity of parameter setting in current methodologies poses a common obstacle. Selecting the appropriate parameter is intricate and time-consuming, necessitating a deep understanding of the data type. Moreover, the accuracy of the outcomes heavily relies on this selection, where an incorrect choice may result in unsatisfactory outcomes. Consequently, identifying a solution to mitigate this issue would be significantly beneficial.

b) The second issue concerns the inherent weakness in outlier detection strategies, which were predominantly observed in previous methodologies. This limitation is particularly evident in distance-based and density-based methods when confronted with datasets featuring clusters of varying densities and patterns of low density,

respectively. Thus, methods predicated on these strategies (either distance or density) need to be revised.

c) The third challenge is the previous methods' inability to accurately identify all three outlier types: local, global, and group outliers. In the experimental section, our proposed method is compared against former approaches across these dimensions.

These issues served as the primary motivation behind this article. Given that each methodological group has its pros and cons, the decision-making process involves a compromise, necessitating selecting a method based on the user's understanding of the dataset. Additionally, choosing the correct parameter(s) value poses a challenge, even when the method choice is apt. To address these concerns, we introduce a novel, reliable, parameter-free technique for outlier detection named Adaptive Radius Density-based Outlier Detection (ARDOD). ARDOD eliminates the need for parameter selection, facilitating a swift and straightforward application process. Contrary to most existing techniques that implement the k-NN algorithm, our method adopts the fixed radius nearest neighbor (FRNN) rule due to its heightened sensitivity to the density criterion.

Implementing the FRNN rule and an adaptive radius effectively addresses the traditional weaknesses of density-based methods. Furthermore, an automatic radius calculation mechanism, predicated on the sample distribution within the feature space, eradicates the issues associated with parameter selection and the uncertainty of its optimal value. By leveraging the mass-sharing algorithm, we harness the advantages of distance

and density-based methods simultaneously, enhancing the algorithm's efficacy in handling datasets that pose challenges to other density-based methods. Our method's effectiveness and efficiency have been validated through tests on synthetic and real-world datasets across various dimensions. In summary, the contributions of our method are multifaceted:

- We introduced a new density-based outlier detection method that not only enhances performance over previous approaches but also addresses their limitations.
- The principle of limited resources in nature inspired the development of the mass-sharing concept.
- The FRNN rule is adopted over the k-NN algorithm due to the former's difficulty in determining the optimal number of neighbors, which is heavily dependent on the dataset. Additionally, an intuitive approach for determining the most suitable radius value based on the dataset distribution enhances the method's adaptability.
- By utilizing the mass-sharing concept, we effectively navigate the limitations of both density-based and distance-based methods while simultaneously capitalizing on their strengths.

3.3 Description of the proposed method

In recent years, there has been an increasing interest in outlier detection methods, primarily focusing on distance-based and density-based approaches. While these strategies have significantly enhanced performance compared to earlier techniques, they have drawbacks. Distance-based methods often need help with datasets characterized by varying local densities, as illustrated in Fig. 2a. Conversely, density-based methods fail to identify low-density patterns frequently encountered in real-world scenarios, depicted in Fig. 2b [21]. A notable limitation of density-based approaches is the necessity to define a neighborhood size explicitly, which is crucial because the effectiveness of these methods heavily depends on the accuracy of this determination.

Addressing the shortcomings of density-based and distance-based methods necessitates a novel approach that not only facilitates the outcomes of prior techniques but also mitigates their inherent weaknesses. A particular challenge of density-based methods is the requirement to specify a neighborhood size, significantly influencing the results. Traditional methods leave this decision to the user, often leading to

suboptimal outcomes due to improper size selection. Moreover, assigning the same neighborhood size to all samples overlooks that each sample may have a distinct neighborhood density, thereby hindering the accurate identification of outliers.

To illustrate, consider a scenario where a uniform radius is applied across the dataset. Such an approach would mistakenly classify all samples within a less dense cluster as outliers, failing to recognize the true outlier (the red star), as shown in Fig. 2a. This issue is addressed by introducing an adaptive radius, which varies depending on the local density around each sample. For instance, a sample within a sparse cluster (the square) would be assigned a larger adaptive radius, encompassing numerous neighboring samples and correctly identifying it as part of the cluster. Conversely, a sample adjacent to a dense cluster (the star) would have a smaller adaptive radius, ensuring no other sample falls within its range, thereby accurately labeling it as an outlier.

The proposed method, Adaptive Radius for Outlier Detection (ARDOD), aims to refine the detection process by implementing an adaptive neighborhood size for each sample, denoted as, within the dataset D, containing n samples and targeting m outliers. This adaptive approach begins by determining an initial radius r as a fraction (ε) of the average pairwise distance among all samples in D:

$$r = \frac{\varepsilon}{n \times (n-1)} \times \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} d(x_i, x_j), \ (\varepsilon = 0.01) \tag{1}$$

Here, x_i and x_j represent two distinct samples from dataset D. Subsequently, ARDOD identifies the nearest neighbor for each sample, referred to as Nx_i, and calculates an adaptive radius (AR_i) based on the density of samples within a hypersphere centered at Nx_i:

$$AR_i = \frac{1}{n_i' \times (n_i' - 1)} \times \sum_{\substack{s=1 \\ s \neq t}}^{n_i'} \sum_{t=1}^{n_i'} d(x_s, x_t) \tag{2}$$

For samples isolated within their hyperspheres ($n_i' = 0$), the Adaptive Radius (AR_i) is set to zero. The method then determines the number of samples within a hypersphere of radius AR_i around each sample x_i, excluding x_i itself. Samples with no neighboring patterns are considered outliers.

Fig. 2 An example of a dataset with (**a**) different local densities and (**b**) low-density patterns

(b)

(a)

Suppose the initial selection of outliers exceeds the desired number (m). In that case, the epsilon value is adjusted, and the process is repeated until the count of identified outliers matches or falls below m. To allocate the remaining outliers, ARDOD employs a mass-sharing strategy inspired by fitness sharing in evolutionary computing [33], where samples less similar to their neighbors are deemed outliers.

The shared mass calculation for each sample, x_i, within its surrounding hypersphere is as follows:

$$Shared\ Mass(x_i) = \frac{1}{\sum_{j=1}^{n} sh(x_j)} \tag{3}$$

where

$$sh(x_j) = \begin{cases} 1 - \left(\frac{d_{ij}}{r}\right) & if\, d_{ij} < r, \\ 0 & otherwise, \end{cases} \tag{4}$$

with d_{ij} being the Euclidean distance between x_i and x_j. Samples are then ranked by their shared mass, and the ones with the highest values are identified as outliers. ARDOD finalizes the outlier detection by integrating the results from both phases, ensuring a comprehensive and precise identification process. This methodology is succinctly summarized in Algorithm 1, facilitating its implementation and understanding.

Algorithm 1. Adaptive Radius Density-based Outlier Detection (ARDOD) Pseudocode

---

Input: A given d$ataset\ D, number\ of\ outliers\ m$

Output: m samples as predicted outliers

1. Input $\bar{n} = n,\ \varepsilon = 0, D = \{x_i, i = 1,2, \dots, n\}$,
2. Output O
3. While $(\bar{n} \leq m)$ do
   - 3.1. $\varepsilon = \varepsilon + \Delta\varepsilon\ (\Delta\varepsilon = 0.01)$.
   - 3.2. Calculate radius r based on Eq. (1) and given ε.
   - 3.3. Find $1NN$ for all samples in D.
   - 3.4. The subscript base, cap A. d a. p t i. v e, cap R a. d i. u s, end base, sub i. for each sample x sub i. is based on the drawn hypersphere with radius r around their 1NN based on Eq. (2).
   - 3.5. Count the number of samples in the drawn hypersphere, with the radius of subscript base, cap A. cap R, end base, sub i. around each sample x sub i. and set d sub i. equals 0 for those samples alone in their hypersphere.
   - 3.6. Count the number of samples with $d_i = 0$
     $\big(i.e.\ O_{step1} = \{x_i | d_i = 0\},\ and\ \bar{n}$
     $= |O_{step1}|\ where\ |.|\ is\ the\ cardinality\ operator\big)$.
   - End while
4. Specify those samples with $d_i \neq 0\ (R = D - O_{step1})$.
5. Apply the Mass Sharing method on the remaining samples (set R) and select $(\bar{\bar{n}} = m - \bar{n})$ Samples with the highest shared mass (set O_{step2}).
6. $O = O_{step1} \cup O_{step2}$.

Table 1 Description of used datasets

Dataset Name	Dimensions	Number of samples	Number of Outliers	Number of Classes	Percentage of Outliers (%)
Synthetic 1	2	1039	41	4	3.80
Synthetic 2	2	1641	45	3	2.67
Synthetic 3	2	1037	37	1	3.45
Synthetic 4	2	2259	159	1	6.58
Synthetic 5	2	1034	36	1	3.36
Synthetic 6	2	1242	50	3	3.87
Breast	33	198	47	2	19.18
Breast diagnostic	30	569	212	2	27.14
Diabetes	8	768	268	2	25.87
Glass	9	214	9	5	4.04
Lymphography	18	148	6	2	3.90
Musk	166	3062	97	3	3.07
Pima	8	768	268	2	25.87
Satellite	36	6435	2036	6	24.03
Vowels	12	1456	50	3	3.32
WPBC	33	198	47	2	19.18

4 Numerical experiments

This section demonstrates the superiority and effectiveness of the proposed method through a comprehensive analysis conducted on sixteen datasets. These datasets include ten real-world and six synthetic datasets with two dimensions. To evaluate the proposed method's stability and performance, we compared it with seven well-known unsupervised outlier detection methods: LOF [15], ABOD [20], DCROD [31], ECOD [32], MOD+[6], RDOS [13], and LGOD [26]. Table 1 presents the datasets' characteristics, listing each dataset's name, dimensions, size, and the number of outliers identified.

The evaluation of these methods was carried out directly on the entire dataset, without any preprocessing steps. The computational experiments were conducted using an Intel Core i7 2670QM processor with a 2.20 GHz clock speed and 8 GB of DDR3 RAM, running on Microsoft Windows 10 and utilizing the MATLAB environment for processing.

4.1 Datasets

This section provides an overview of the datasets utilized to evaluate the proposed method across various testing scenarios. The selection of two-dimensional synthetic datasets aims to address common challenges in outlier detection, specifically varying cluster densities and the presence of low-density patterns. These conditions are prevalent obstacles in the field, necessitating datasets encompassing a range of cluster densities and sample sizes. Additionally, numerous studies have previously employed the chosen datasets, offering a basis for comparing and validating our method's effectiveness [22, 26, 31, 34].

Synthetic 1 and Synthetic 2 datasets are mainly designed to highlight issues related to cluster density, a known weakness of distance-based outlier detection methods. Conversely, density-based approaches demonstrate improved performance with these datasets. While Synthetic 2 presents similar challenges to Synthetic 1, it introduces a higher level of complexity. The Synthetic 3 dataset is characterized by a spiral-shaped cluster amidst uniformly distributed outliers, presenting a distinct scenario for outlier detection. In comparison, Synthetic four offers increased complexity over Synthetic 3.

Further diversifying our dataset collection, Synthetic 5 incorporates a sine curve perturbed by Gaussian noise, introducing variability in data distribution. Synthetic six features three nested rectangles that partition the space into distinct regions, each with thin sides. This setup highlights the challenge of detecting low-density patterns, a significant hurdle for density-based detection methods.

Expanding our evaluation to real-world contexts, we selected ten datasets from the UCI repository ("http://www.archive.ics.uci.edu/ml/"). These datasets have been previously utilized in outlier detection research, facilitating a comprehensive assessment of our proposed method's real-world applicability.

Table 1 provides a detailed summary of each dataset, including the number of samples, outliers, and dimensions. Figure 3 illustrates the selection and characteristics of the two-dimensional synthetic datasets, ensuring a thorough examination of the proposed method's performance across a spectrum of testing environments.

Fig. 3 Two-dimension synthetic datasets

Synthetic 1

Synthetic 2

Synthetic 3

Synthetic 4

Synthetic 5

Synthetic 6
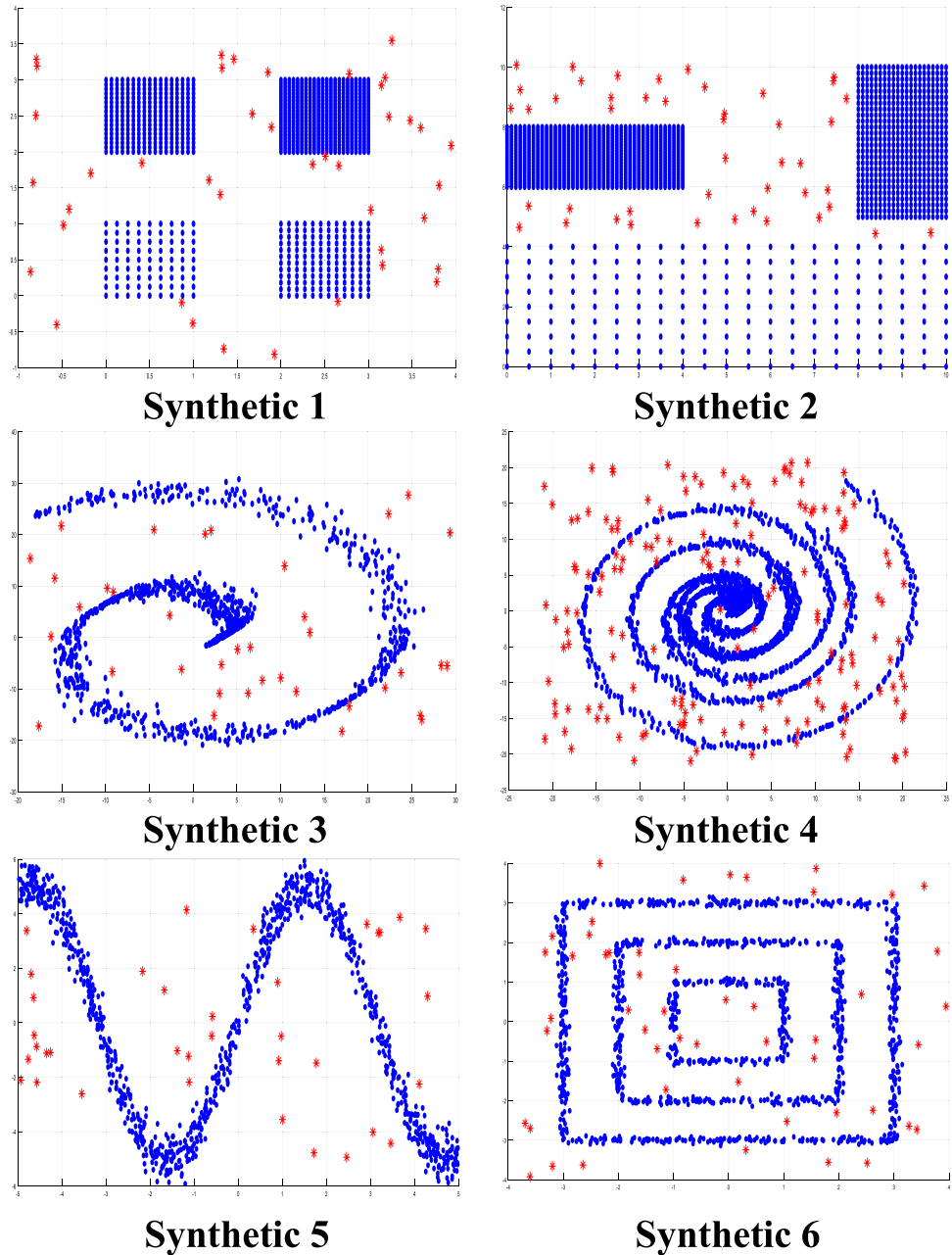
4.2 Metrics

To assess the performance of the proposed method, we employ various metrics, including the Precision, G-Mean, execution time and statistical tests.

- **Precision:** Precision is a metric used to evaluate the accuracy of a model's true predictions. It is beneficial in scenarios with a high cost of false positives. Precision calculates the ratio of true positive predictions to the total number of positive predictions made by the model, including both true positives and false positives. The formula for precision is given by:

$$Precision = \frac{TP}{TP + FP} = \frac{The\ number\ of\ outliers\ that\ are\ correctly\ predicted}{Total\ number\ of\ outliers} \tag{5}$$

This metric does not take into account the true negatives and is, therefore, particularly useful in situations where the focus is on the relevance of the positive predictions [35].

- **G-Mean:** The G-Mean, or Geometric Mean, is a metric that evaluates a model's performance by equally weighing its sensitivity and specificity. This metric is particularly advantageous when dealing with imbalanced class distributions or when the importance of sensitivity and specificity is equivalent. The G-Mean is calculated as the square root of the product of sensitivity and specificity, ensuring that both metrics have an equal impact on the final score. The formula for the G-Mean is:

$$G - Mean = \sqrt{Sensitivity \times Specificity} \tag{6}$$

This measure harmonizes the balance between sensitivity and specificity, providing a singular metric to thoroughly gauge a model's effectiveness.

- **Time:** Evaluating each algorithm's performance also includes assessing the average time required to detect outliers. This metric provides valuable insights into the detection process's efficiency and speed, illustrating each method's practicality in real-world applications. By measuring the time taken for outlier detection, we can compare the computational demands of different algorithms, highlighting those that offer a balance between accuracy and speed. This aspect is crucial for applications where processing time is a limiting factor, ensuring that the chosen method delivers prompt and reliable results.
- **Statistical Tests:** A statistical test is a methodological process utilized for making inferences or decisions about the properties of a population based on sample data. These tests are essential tools in the realm of statistical analysis, employed to assess the validity of hypotheses concerning population parameters. They play a pivotal role in determining whether observed data deviates significantly from what is expected under the null hypothesis, hence facilitating evidence-based conclusions. Within this framework, two notable tests are:
- **The Friedman Test:** This serves as a non-parametric counterpart to the one-way ANOVA with repeated measures, designed for identifying differences across multiple treatment attempts. Its applicability shines in instances where the normality assumption, a prerequisite for parametric tests, is not met. The Friedman test analyzes ordinal or non-normal interval data, offering a robust mechanism for exploring the impacts of varying conditions devoid of the stringent presuppositions associated with parametric statistics. This makes it an indispensable tool

in areas where data often skews from normal distribution patterns [36].

- **The Wilcoxon Test:** This encompasses the Wilcoxon rank-sum test and the Wilcoxon signed-rank test, positioned as non-parametric alternatives to the unpaired and paired t-tests, respectively. It is employed to compare two sample sets to ascertain if their population mean ranks significantly differ. This test is particularly beneficial for examining small sample sizes or data that fails to adhere to a normal distribution. By enabling hypothesis testing without the rigid normality criteria, the Wilcoxon test stands out as a fundamental analytical instrument for researchers working with non-parametric data across various scientific fields [37].

4.3 Numerical results

As delineated in Section 3, it has been observed that all the methods under comparison, with the exception of the proposed method, depend on the parameter k. To shed light on the significance of this parameter, the performance metrics of two specific methods, namely RDOS [13] and LGOD [26], were analyzed across varying k values (20, 50, and 100), as depicted in Fig. 4. The rationale behind the selection of these particular k values is grounded in the guidance provided by several sources [21–23, 25]. It becomes evident that the choice of k exerts a considerable influence on the algorithms' effectiveness, thereby affecting the overall efficiency of the detection methodologies.

Distinctively, the proposed method, which does not rely on any parameter, maintains a uniform performance under diverse conditions. In stark contrast to the methods compared, the performance of the proposed method remains unaffected by changes in k; consequently, metrics such as Precision, G-mean, and execution time were assessed across a broad k value spectrum ranging from 5 to 100. Figure 5 showcases a comparison of the Precision metric of the proposed method against that of seven other methodologies across 16 datasets for k values within the 5 to 100 range, thereby illustrating the proposed method's superior stability attributed to its non-parametric nature.

Figure 5 presents the precision of eight methods under comparison, including the proposed method, across a range of k values from 5 to 100. For datasets such as Synthetic 1 through Synthetic 6, all methods initially demonstrate similar detection performance at lower k values. Nonetheless, as the value of k increases, the superiority of the proposed method becomes increasingly apparent. Specifically, in the cases of Synthetic 4 and Synthetic 6, the ABOD G-mean decreases less than that of other methods with rising k values, indicating a nuanced variation in performance.

Furthermore, the analysis of the Glass, Musk, Pima, and Satellite datasets unequivocally illustrates that the proposed
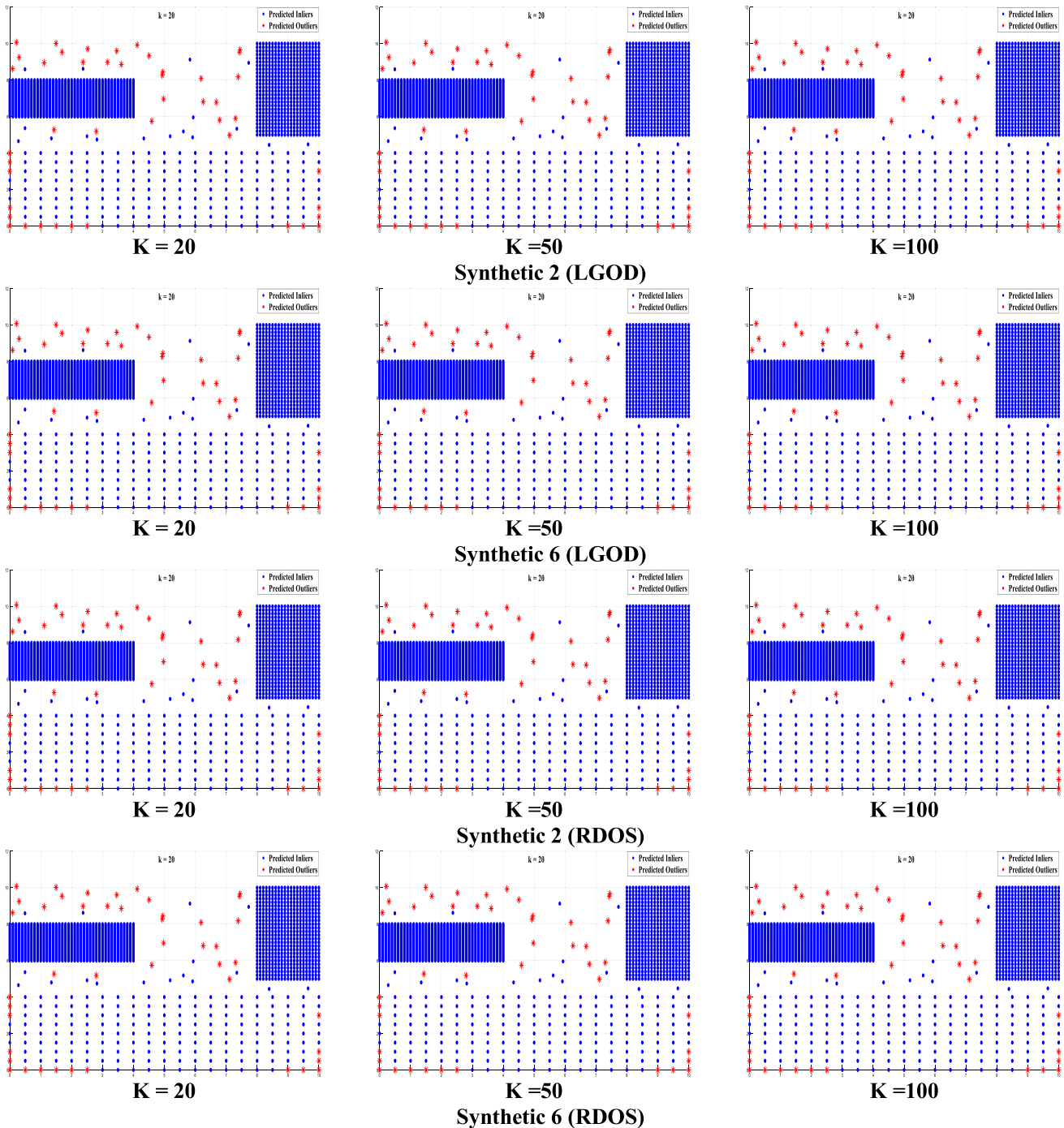
Fig. 4 Outlier detection results of two recently proposed methods (RDOS and LGOD) at varying k values (left: k = 20, middle: k = 50, right: k = 100) on Synthetic two and Synthetic 6 Datasets

method outperforms the existing techniques by a significant margin. This suggests a robust adaptability and efficiency of the proposed method across a variety of data types.

In contrast, for the Breast, Diabetes, Lymphography, and WPBC datasets, while the compared methods generally exhibit similar performances, the proposed method, distinguished by its parameter-free nature, demonstrates enhanced stability. This characteristic suggests its potential for consistent application without the need for intricate parameter tuning.

Particularly in the Breast Diagnostic Dataset, the MOD + method exhibits substantial improvements as the

Fig. 5 Detection performance (G-Mean) of all eight methods
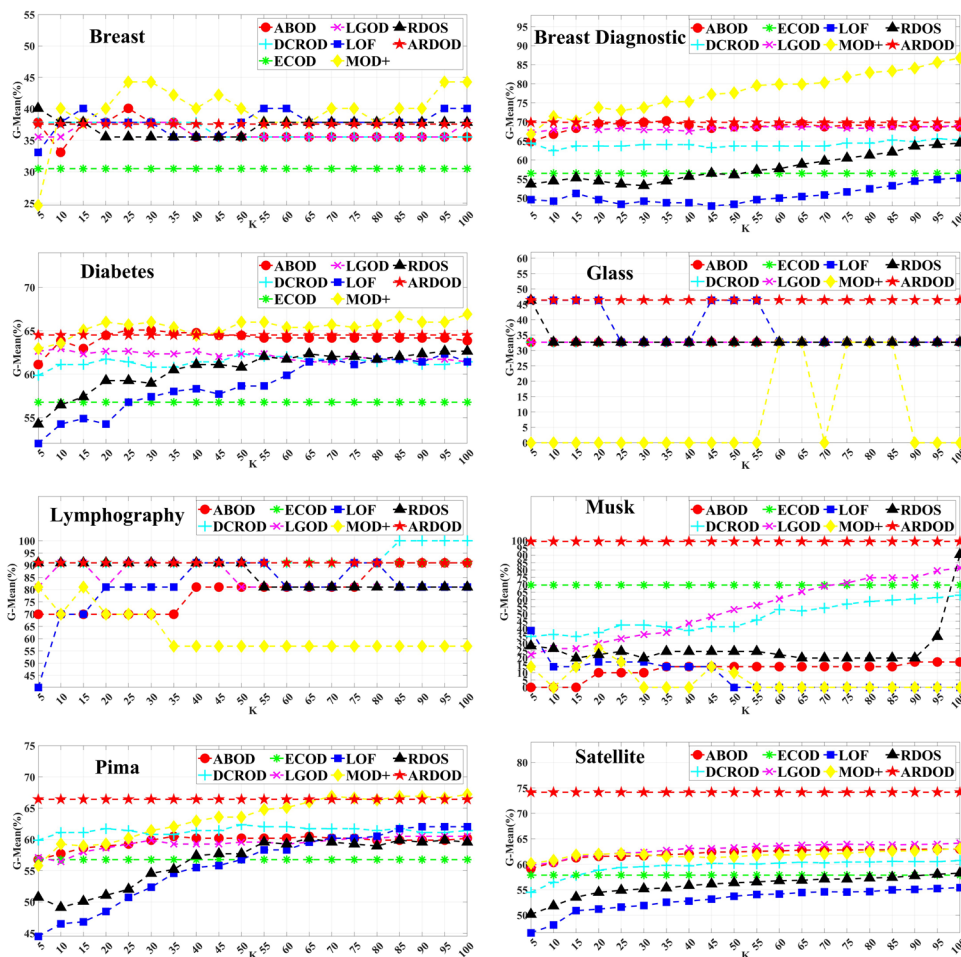
k value progresses from 5 to 100; despite some *k* values favoring the performance of certain methods, the proposed method consistently emerges as more efficient on average. This indicates its capability to maintain a high level of precision across a broad range of conditions.

Conversely, in the Vowels Dataset, the proposed method does not exhibit strong performance, with ABOD showcasing the highest precision among all the methods compared. This highlights the potential for specific methods to outperform others under certain dataset conditions.

Overall, the absence of parameter adjustment in the proposed method not only ensures better consistency across all compared methods but also leads to the best or at least comparable performance in most of the datasets examined. This underscores the proposed method's versatility and its ability to provide reliable and stable outcomes across diverse analytical scenarios.

Table 2 provides a detailed comparison of the precision of the proposed method against that of other methodologies over a range of *k* values from 5 to 100. Additionally, it outlines the average precision and the ranking of each method based on their performance. It is noteworthy that the proposed method secures the highest precision in 16 out

of the 8 methods evaluated, which may suggest a typo and possibly intends to compare the proposed method across 16 datasets or scenarios. The proposed method distinguishes itself by achieving first place with an overall average precision of 64.35%, surpassing the method in second place by a significant margin of 11%.

This demonstrates not only the effectiveness of the proposed method in achieving high precision across a broad spectrum of *k* values but also its superiority over other evaluated methods in terms of consistent performance. The substantial lead in average precision emphasizes the robustness and reliability of the proposed method, suggesting it as a preferable choice for applications requiring precise anomaly detection or similar tasks. The comparative analysis, underscored by the method's leading position and its significant outperformance of competitors, highlights its potential to serve as a benchmark for future methodological developments in the field.

Table 3 meticulously assesses the speed of performance by cataloging the average duration required by each algorithm to identify outliers across 16 distinct datasets, with k values varying from 5 to 100. The ECOD (Cumulative-distribution-based Outlier Detection) method stands out

Fig. 5 (continued)
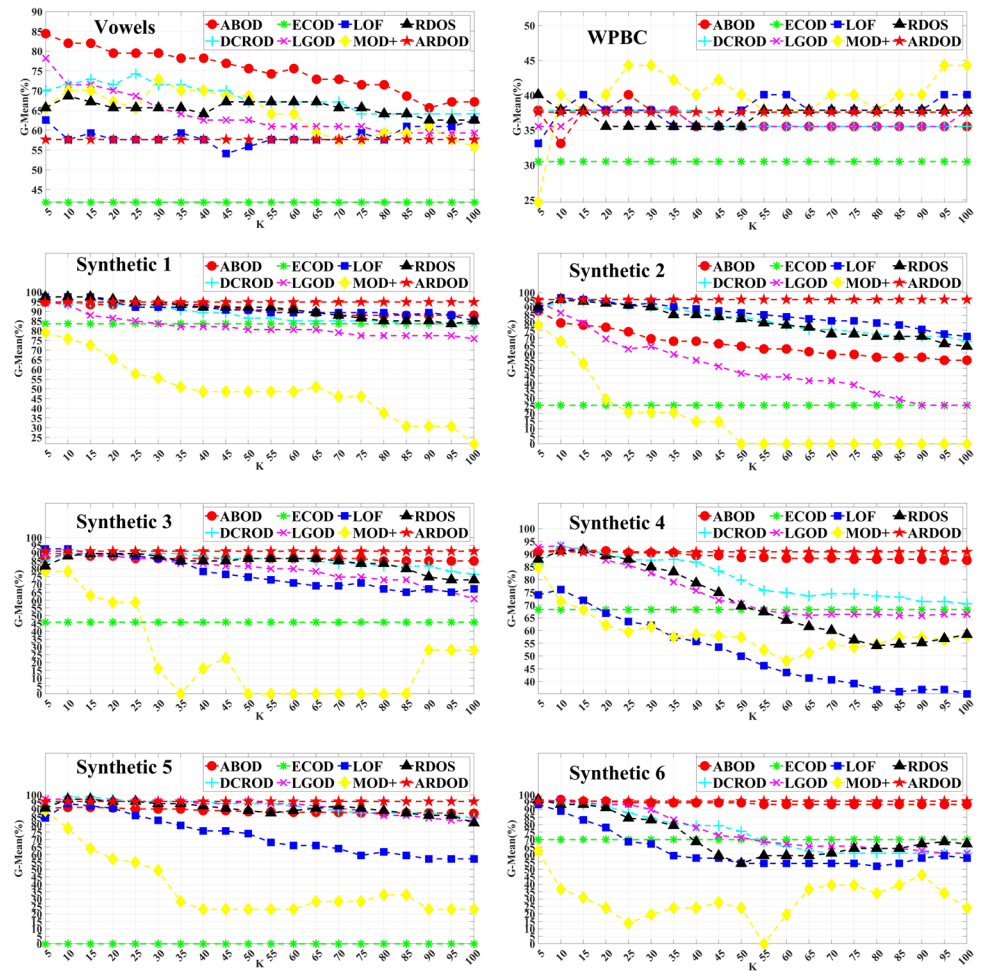
as the fastest among the evaluated techniques. This notable rapidity is primarily attributed to the fact that it does not rely on the KNN (K-Nearest Neighbors) algorithm, which is a significant factor contributing to its superior speed compared to other methods. However, as indicated in Table 2, this advantage in speed comes at the expense of precision. The data suggests a clear trade-off between the method's velocity and its accuracy, highlighting an area of potential compromise for researchers and practitioners considering the ECOD method for outlier detection.

The evaluation of outlier detection methodologies across all 16 datasets, with *k* ranging from 5 to 100, reveals the efficiency of various algorithms in terms of their time consumption. Among these methods, the ARDOD technique stands out for its commendable performance, occupying the 5th rank with an average time consumption of 7.4 s. This positions ARDOD significantly well in comparison to its counterparts, particularly highlighting its superiority over the RDOS and LGOD methods, which are slower, with times of 36.6 and 28.1 s respectively.

ARDOD not only showcases a faster performance but also excels in accuracy, a testament to its sophisticated design that optimizes both speed and precision without the need for parameter adjustments. This dual advantage of speed and accuracy emphasizes ARDOD's effectiveness and efficiency, making it an attractive option for applications that demand rapid and accurate outlier detection.

The Friedman test results in Table 4, elucidate a comparative analysis of algorithmic performance, with ARDOD demonstrating superior efficacy, attaining the lowest average rank of 2.7188. ABOB and DCROD algorithms exhibit parity in performance, sharing an average rank of 3.6562, indicative of moderate efficacy. Conversely, ECOD is discerned as the least effective, with the highest average rank of 6.4375. The rankings are statistically significant, as evidenced by a Friedman statistic of 24.973958 and a notably low p-value of 0.000767, affirming the reliability of the performance differentiation among the evaluated algorithms.

The non-parametric Wilcoxon test was employed for enhanced precision in test analysis. This test conducted

pairwise comparisons between our method and alternative approaches, determining both the magnitude and direction of any differences observed. A p-value below the threshold of α (commonly set at 0.05) indicates the rejection of methodological equivalence, thereby underscoring a significant disparity. The direction of this disparity is reflected in the relationship between R^+ and R^-. According to the Pr criteria presented in Table 5, our method, in all instances, demonstrated statistically significant superiority over the comparative methods, with ARDOD outperforming each alternative as evidenced by the results.

In summary, the proposed method consistently outperforms other techniques in various scenarios, overcoming the limitations typically associated with distance-based and density-based methods. Its parameter-free nature not only facilitates stable performance but also simplifies usage, distinguishing it as a superior choice for outlier detection across both synthetic and real-world datasets.

5 Challenges and future directions of ARDOD

5.1 Recognizing the boundaries

While ARDOD represents a significant advancement in outlier detection methodology, we acknowledge certain limitations inherent to our approach:

1. **Data Dependency:** The effectiveness of ARDOD is contingent upon the underlying data distribution. In scenarios where the data is extremely sparse or highly uniform, the adaptability of the radius might be less effective, potentially impacting outlier detection performance.
2. **Computational Complexity:** The adaptive nature of the radius calculation and mass-sharing algorithm can lead to increased computational demands compared to simpler methods, especially as the size of the dataset grows.

Table 2 Detection performance (average precision) of eight methods (k in The Range of 5 to 100)

Dataset Name	ABOD	DCROD	ECOD	LGOD	LOF	MOD+	RDOS	ARDOD
Synthetic 1	83.41	78.78	70.73	68.54	83.78	27.68	83.41	**90.00**
Synthetic 2	45.00	68.22	6.67	30.33	73.78	8.22	66.78	**91.11**
Synthetic 3	75.54	73.92	21.62	64.46	60.95	13.51	71.62	**83.78**
Synthetic 4	80.85	66.76	48.43	58.46	55.00	37.20	54.40	**83.65**
Synthetic 5	87.92	85.69	0.00	83.61	53.33	18.61	83.47	**91.67**
Synthetic 6	89.40	57.90	50.00	59.90	42.00	11.00	54.30	**92.00**
Breast	17.66	17.87	12.77	17.66	19.15	**21.17**	18.51	17.02
Breast diagnostic	61.34	55.68	46.70	60.87	39.95	**72.59**	48.33	58.49
Diabetes	54.42	51.10	45.52	51.96	47.84	**56.04**	50.06	50.37
Glass	11.11	11.11	11.11	11.11	15.00	2.78	11.67	**22.22**
Lymphography	65.00	**86.67**	83.33	72.50	67.50	40.00	75.00	83.33
Musk	1.75	24.28	49.48	32.68	1.75	0.88	9.54	**98.97**
Pima	48.99	51.10	45.52	48.64	44.51	**53.69**	45.32	52.99
Satellite	50.17	46.96	44.94	51.18	39.11	49.88	42.51	**63.02**
Vowels	**57.60**	47.50	18.00	42.00	35.20	42.40	43.90	34.00
WPBC	17.66	17.87	12.77	17.66	19.15	**21.17**	18.51	17.02
Average	**52.99**	52.59	35.47	48.22	43.62	29.80	48.58	**64.35**
Rank	**2**	**3**	**7**	**5**	**6**	**8**	**4**	**1**

Table 3 Average time-consumption of eight methods on all 16 datasets (k in The Range of 5 to 100)

#	ABOD	DCROD	ECOD	LGOD	LOF	MOD+	RDOS	ARDOD
Time (Sec)	**138.1**	**1.3**	**0.05**	**28.1**	**0.4**	**2.6**	**36.6**	**7.4**
Rank	**8**	**3**	**1**	**6**	**2**	**4**	**7**	**5**

Table 4 Average ranking of Friedman test

ABOD	DCROD	ECOD	LGOD	LOF	MOD+	RDOS	ARDOD
3.6562	3.6562	6.4375	4.4688	5.2188	5.1875	4.6562	**2.7188**

3. **Domain-specific Adaptations:** While ARDOD is designed to be versatile, its performance can benefit from domain-specific adaptations, such as fine-tuning the ε parameter or integrating additional features relevant to the specific application context.

5.2 Expanding the horizon of ARDOD

While the ARDOD method has demonstrated promising results in outlier detection across various datasets, its evolution presents numerous opportunities for further research and improvement. We envision the following areas as potential avenues for future work:

a) **Scalability to High-dimensional Data:** Although ARDOD shows superior performance in the current settings, its scalability and efficiency in handling high-dimensional datasets could be further investigated. The adaptation of the algorithm to effectively manage the "curse of dimensionality" will be a critical area of research. Techniques such as dimensionality reduction or feature selection could be incorporated into the ARDOD framework to enhance its applicability to complex datasets.

b) **Real-time Outlier Detection:** The development of a real-time variant of ARDOD could significantly impact fields requiring immediate anomaly detection, such as cybersecurity, financial fraud detection, and real-time monitoring systems. Future work will explore modifications to the algorithm that reduce computational complexity and allow for incremental learning from streaming data.

c) **Integration with Supervised Learning Models:** The current implementation of ARDOD is unsupervised. An interesting extension could involve its integration with supervised learning models to leverage labeled data, potentially improving the detection of outliers in semi-supervised or fully supervised settings. This hybrid

approach could refine the algorithm's sensitivity to subtle anomalies.

d) **Adaptation to Distributed Computing Environments:** With the exponential growth of data, the need for distributed computing solutions has become more pronounced. Adapting ARDOD to work efficiently within distributed computing frameworks, such as Apache Hadoop or Spark, could address scalability issues and enhance its suitability for big data applications.

e) **Deep Learning-based Enhancements:** Incorporating deep learning techniques to automate the feature learning process in ARDOD could provide a significant boost in performance, especially in datasets where relevant features for outlier detection are not readily apparent. Convolutional neural networks (CNNs) or autoencoders could be used to extract meaningful features automatically, which could then be fed into the ARDOD algorithm for outlier detection.

By pursuing these directions, we aim to refine and extend the ARDOD method, further contributing to the field of outlier detection. We believe that addressing these future work areas will not only enhance the robustness and applicability of ARDOD but also open new avenues for research and innovation in anomaly detection.

6 Conclusion

This paper proposes a new outlier detection technique called adaptive radius density-based outlier detection (ARDOD) based on the fitness-sharing concept. The proposed method is parameter-free and capable of detecting all kinds of outliers, including local, global, and group of outliers. The ARDOD takes advantage of both distance and density concepts to deal with the weaknesses of the previous methods. Also, no need for parameter adjustment makes the proposed method a suitable option for any given dataset. Moreover, unlike density-based and distance-based methods, the ARDOD works fine with datasets with different clusters' densities and low-density patterns. The proposed method is compared with five previous techniques, including LOF, ABOD, INS, RDOS, and LGOD, on six synthetic datasets. Also, we employed the proposed method on ten real-world datasets. The results of both artificial and real-world datasets show the superiority and effectiveness of the proposed method.

Table 5 The Wilcoxon signed-ranks Pr test results on datasets

	R^+	R^-	p-value	Hyphothesis (α=0.05)
ARDOD vs. ABOD	106.0	30.0	0.045098	**Rejected**
ARDOD vs. DCROD	113.0	23.0	0.017882	**Rejected**
ARDOD vs. ECOD	120.0	0.0	0.000539	**Rejected**
ARDOD vs. LGOD	120.0	16.0	0.006288	**Rejected**
ARDOD vs. LOF	130.0	6.0	0.001143	**Rejected**
ARDOD vs. MOD+	114.0	22.0	0.015516	**Rejected**
ARDOD vs. RDOS	123.0	13.0	0.003871	**Rejected**

Reza Heydari Gharaei: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing- Original draft preparation, Visualization.

Hossein Nezamabadi-pour: Conceptualization, Investigation, Writing- Reviewing and Editing, Supervision, Project Administration.

Data availability The multidimensional datasets utilized in this study are publicly accessible at the following URL: http://odds.cs.stonybrook.edu/ Furthermore, the two-dimensional synthetic dataset featured in this research was obtained via email from Jihyun Ha, the author of "Accurate Ranking Method for Detecting Outliers." These datasets have also been employed in numerous other studies, including:

- "Natural Neighbor: a self-adaptive neighborhood method without parameter K [38]."
- "A precise ranking method for outlier detection [22]."
- "A non-parameter outlier detection algorithm based on Natural Neighbor [23]."
- "Relative Density-Based Outlier Detection Algorithm [5]."
- "A novel outlier cluster detection algorithm without top-n parameter [39]."
- "ADD: a new average divergence difference-based outlier detection method with a skewed distribution of data objects [40]."
- "NaNOD: A natural neighbor-based outlier detection algorithm [41]."
- "Robust outlier detection based on the changing rate of directed density ratio [31]."

Declarations

Competing interests The authors declare no competing interests.

References

1. Edwin MK (1998) Algorithms for mining distance-based outliers in large datasets. In: Proceedings of 24th International Conference on Very Large DataBases (VLDB'98), pp 392–403
2. Sharify R, Gharaei RH, Mahmoud Taheri S (2022) Improved LOF algorithm using random point. In: 2022 9th Iranian joint congress on fuzzy and intelligent systems (CFIS). IEEE, pp 1–6
3. Angiulli F, Ben-Eliyahu - Zohary R, Palopoli L (2008) Outlier detection using default reasoning. Artif Intell 172:1837–1872. https://doi.org/10.1016/j.artint.2008.07.004
4. Wang H, Bah MJ, Hammad M (2019) Progress in outlier detection techniques: a survey. IEEE Access 7:107964–108000. https://doi.org/10.1109/ACCESS.2019.2932769
5. Ning J, Chen L, Chen J (2018) Relative density-based outlier detection algorithm. In: Proceedings of the 2018 2nd international conference on computer science and artificial intelligence. ACM, New York, NY, USA, pp 227–231
6. Yang J, Rahardja S, Fränti P (2021) Mean-shift outlier detection and filtering. Pattern Recognit 115:107874. https://doi.org/10.1016/j.patcog.2021.107874
7. Gharaei RH, Sharify R, Nezamabadi-Pour H (2022) An efficient outlier detection method based on distance ratio of k-nearest neighbors. In: 2022 9th Iranian joint congress on fuzzy and intelligent systems (CFIS). IEEE, pp 1–5
8. Mirzaei B, Rahmati F, Nezamabadi-pour H (2022) A score-based preprocessing technique for class imbalance problems. Pattern Anal Appl 25:913–931. https://doi.org/10.1007/s10044-022-01084-1
9. Rahmati F, Nezamabadi-pour H, Nikpour B (2020) A gravitational density-based mass sharing method for imbalanced data classification. SN Appl Sci 2:260. https://doi.org/10.1007/s42452-020-2039-2
10. Hawkins DM (1980) Identification of outliers. Springer, Netherlands, Dordrecht
11. Zimek A, Campello RJGB, Sander J (2014) Ensembles for unsupervised outlier detection. ACM SIGKDD Explor Newsl 15:11–22. https://doi.org/10.1145/2594473.2594476
12. Aggarwal CC (2017) Outlier analysis. Springer International Publishing, Cham
13. Tang B, He H (2017) A local density-based approach for outlier detection. Neurocomputing 241:171–180. https://doi.org/10.1016/j.neucom.2017.02.039
14. Gao X, Yu J, Zha S et al (2022) An ensemble-based outlier detection method for clustered and local outliers with differential potential spread loss. Knowl Based Syst 258:110003. https://doi.org/10.1016/j.knosys.2022.110003
15. Breuniq MM, Kriegel HP, Ng RT, Sander J (2000) LOF: identifying density-based local outliers. SIGMOD record (ACM special interest group on management of data). ACM Press, New York, New York, USA, pp 93–104
16. Tang J, Chen Z, Fu AWC, Cheung DW (2002) 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 pp 535–548
17. Gao J, Hu W, Zhang Z, Zhang X, Wu O (2011) RKOF: robust kernel-based local outlier detection. In: Pacific-Asia conference on knowledge discovery and data mining, pp 270–283
18. Papadimitriou S, Kitagawa H, Gibbons PB, Faloutsos C (2003) LOCI: fast outlier detection using the local correlation integral. In: proceedings - international conference on data engineering. IEEE, pp 315–326
19. Jin W, Tung AK, Han J, Wang W (2006) Ranking outliers using symmetric neighborhood relationship. In: Advances in Knowledge Discovery and Data Mining: 10th Pacific-Asia Conference, PAKDD 2006, Singapore, April 9–12, 2006. Proceedings 10, pp 577–593
20. Kriegel H-P, Schubert M, Zimek A (2008) Angle-based outlier detection in high-dimensional data. In: proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, NY, USA, pp 444–452
21. Ha J, Seok S, Lee J-S (2014) Robust outlier detection using the instability factor. Knowl Based Syst 63:15–23. https://doi.org/10.1016/j.knosys.2014.03.001
22. Ha J, Seok S, Lee J-S (2015) A precise ranking method for outlier detection. Inf Sci (N Y) 324:88–107. https://doi.org/10.1016/j.ins.2015.06.030
23. Huang J, Zhu Q, Yang L, Feng J (2016) A non-parameter outlier detection algorithm based on natural neighbor. Knowl Based Syst 92:71–77. https://doi.org/10.1016/j.knosys.2015.10.014
24. Zhang Z, Zhu M, Qiu J et al (2019) Outlier detection based on cluster outlier factor and mutual density. Commun Comput Inform Sci 986:319–329. https://doi.org/10.1007/978-981-13-6473-0_28
25. Wahid A, Rao ACS (2019) RKDOS: A relative kernel density-based outlier score. IETE Tech Rev (Inst Electron TelecommunEng, India) 1–12. https://doi.org/10.1080/02564602.2019.1647804
26. Xie J, Xiong Z, Dai Q et al (2020) A local-gravitation-based method for the detection of outliers and boundary points. Knowl Based Syst 192:105331. https://doi.org/10.1016/j.knosys.2019.105331
27. Rhyu J, Bozinovski D, Dubs AB et al (2024) Automated outlier detection and estimation of missing data. Comput Chem Eng 180:108448. https://doi.org/10.1016/j.compchemeng.2023.108448
28. Kiani R, Jin W, Sheng VS (2024) Survey on extreme learning machines for outlier detection. Mach Learn. https://doi.org/10.1007/s10994-023-06375-0

29. Ibrahim S, Alkasadi NA, Yusoff MI, Zhe LW, Ramli IM (2024) Comparative study of outlier detection methods on multivariate eye data via multiple circular regression model. In: AIP Conference Proceedings, vol 2905, no 1. AIP Publishing

30. Zheng L, Chen L, Wang Y (2024) A new unsupervised outlier detection method. J Intell Fuzzy Syst 46:1713–1734. https://doi.org/10.3233/JIFS-236518

31. Li K, Gao X, Fu S et al (2022) Robust outlier detection based on the changing rate of directed density ratio. Expert Syst Appl 207:117988. https://doi.org/10.1016/j.eswa.2022.117988

32. Li Z, Zhao Y, Hu X, et al (2022) ECOD: unsupervised outlier detection using empirical cumulative distribution functions. https://doi.org/10.1109/TKDE.2022.3159580

33. Sareni B, Krahenbuhl L (1998) Fitness sharing and niching methods revisited. IEEE Trans Evol Comput 2:97–106. https://doi.org/10.1109/4235.735432

34. Gharaei RH, Nezamabadi-Pour H (2022) RDOD: a robust distance-based technique for outlier detection. In: 2022 30th international conference on electrical engineering (ICEE). IEEE, pp 885–890

35. Campos GO, Zimek A, Sander J et al (2016) On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. Data Min Knowl Discov 30:891–927. https://doi.org/10.1007/s10618-015-0444-8

36. Motulsky HJ, Brown RE (2006) Detecting outliers when fitting data with nonlinear regression – a new method based on robust nonlinear regression and the false discovery rate. BMC Bioinformatics 7:123. https://doi.org/10.1186/1471-2105-7-123

37. Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1:80. https://doi.org/10.2307/3001968

38. Zhu Q, Feng J, Huang J (2016) Natural neighbor: a self-adaptive neighborhood method without parameter K. Pattern Recognit Lett 80:30–36. https://doi.org/10.1016/j.patrec.2016.05.007

39. Huang J, Zhu Q, Yang L et al (2017) A novel outlier cluster detection algorithm without top-n parameter. Knowl Based Syst 121:32–40. https://doi.org/10.1016/j.knosys.2017.01.013

40. Xiong Z-Y, Gao Q-Q, Gao Q et al (2022) ADD: a new average divergence difference-based outlier detection method with skewed distribution of data objects. Appl Intell 52:5100–5124. https://doi.org/10.1007/s10489-021-02399-y

41. Wahid A, Annavarapu CSR (2021) NaNOD: a natural neighbour-based outlier detection algorithm. Neural Comput Appl 33:2107–2123. https://doi.org/10.1007/s00521-020-05068-2