International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Optimized Density Peaks Clustering Via Sparse Search and KD-Tree

K Bhuvaneshwar, Ravindra Changala

PG Scholar, Department of CSE, Guru Nanak Institutions Technical Campus, Hyderabad.
Associate Professor, Guru Nanak Institutions Technical Campus, Hyderabad.

Abstract: Density peaks clustering has become a nova of clustering algorithm because of its simplicity and practicality. However, there is one main drawback: it is time-consuming due to its high computational complexity. Herein, a density peaks clustering algorithm with sparse search and K-d tree is developed to solve this problem. Firstly, a sparse distance matrix is calculated by using K-d tree to replace the original full rank distance matrix, so as to accelerate the calculation of local density. Secondly, a sparse search strategy is proposed to accelerate the computation of relative-separation with the intersection between the set of k nearest neighbors and the set consisting of the data points with larger local density for any data point. Furthermore, a second-order difference method for decision values is adopted to determine the cluster centers adaptively. Finally, experiments are carried out on datasets with different distribution characteristics, by comparing with other six state-of-the-art clustering algorithms. It is proved that the algorithm can effectively reduce the computational complexity of the original DPC from O(n2K) to O(n(n1-1/K + k)). Especially for larger datasets, the efficiency is elevated more remarkably. Moreover, the clustering accuracy is also improved to a certain extent. Therefore, it can be concluded that the overall performance of the newly proposed algorithm is excellent.

Keywords: Clustering, KD-tree, DPC, Sparse Search and KD-Tree-based indexing.

I.INTRODUCITON

Clustering is a fundamental task in unsupervised learning, with applications ranging from data mining to pattern recognition. Density Peaks Clustering (DPC) has gained attention due to its ability to identify clusters based on local density and distance from high-density regions. However, its computational cost limits scalability, especially for high-dimensional or large-scale datasets. This paper proposes an Optimised Density Peaks Clustering (ODPC) algorithm that integrates Sparse Search and KD-Tree indexing to enhance computational efficiency while preserving clustering accuracy. Sparse search reduces the number of candidate points evaluated for local density computation, and KD-Tree facilitates efficient nearest-neighbor queries. Experimental results on synthetic and real-world datasets demonstrate that the proposed method achieves a significant reduction in time complexity while maintaining or improving clustering performance compared to baseline methods.

Clustering remains one of the core problems in unsupervised learning, with a wide array of algorithms proposed to cater to different data types and structures. Among these, Density Peaks Clustering (DPC), introduced by Rodriguez and Laio, identifies cluster centers based on two assumptions: cluster centers are surrounded by neighbors with lower local density, and they are located at a relatively large distance from other points with higher density.

Despite its intuitive appeal and strong performance in various settings, DPC suffers from quadratic time complexity, primarily due to its exhaustive distance computations and density evaluations. This makes it less feasible for large or high-dimensional datasets. To overcome this bottleneck, we propose Optimised Density Peaks Clustering (ODPC), a variant of DPC that combines Sparse Search and KD-Tree-based indexing to accelerate density and distance computations. Cluster analysis as an important exploration technology of data mining, is committed to reveal the inherent attributes and laws hidden behind the seemingly disorganized unknown data. It provides support for decision-making and has been successfully applied in many fields such as image pattern recognition, social network mining, market statistical analysis, medical research and engineering systems. With the extremely strong penetration and rapid development of the Internet, many professional fields are faced with explosive growth of data storage. This leads to high computational complexity and difficulty in mining valuable information. In 2014, Science published clustering by fast search and find of density peaks (DPC). Due to its novel design idea and robust performance, DPC instantly became the topic center of scholars in related fields. Com-pared with classical clustering algorithms, DPC possesses several advantages. Firstly, the cluster centers can be identified directly through the decision graph, which consists of local density and relative-separation for all data. Secondly, it can handle non-convex datasets well and no iterative process is required. Furthermore, it is insensitive to outliers. However, there are still some shortages for DPC to be further.

For the original DPC algorithm published by Science, the local structure of data is not considered and it is sensitive to the cut-off distance. Some scholars try to solve this problem by integrating the k nearest neighbors with DPC in various ways. Du et al. introduced the idea of k nearest neighbors into DPC and redefined the calculation method of local density. Thus, the so-called DPC-KNN algorithm was successfully established. Xie et al. also gave another measure of local density using k nearest neighbors and designed a new allocation strategy for non-center points. In order to elevate the possibility of selecting the correct cluster centers, Liu et al. modified the distance calculation method by considering the distance factors and neighbor information simultaneously. In addition, a two-step allocation strategy was also adopted to allocate non-center points. Recently, Liu et al. have suggested a mixed density clustering method by defining two types of local density. One is based on k nearest neighbors, while the other is determined by local spatial position deviation. Local density description derived from k nearest neighbors supplies a novel way for DPC algorithm. Under this condition, the sensitivity to the hyper-parameter k, i.e. the number of nearest neighbors, is obviously depressed, compared with the sensitivity to cut-off distance used in the original DPC algorithm [14]. However, for these DPC algorithms based on k nearest neighbors, mentioned-above, searching neighbors mainly resorts to violent means. Therefore, these algorithms still have the same high computational complexity as original DPC.

For DPC algorithms, the high computational complexity commonly comes from the distance calculation, which is related with any two points among all data points. With the increase of dataset size, as groundwork, calculating distance between any two points is time-consuming and even impossible for DPC to be implemented. In order to reduce the computational complexity and lift the clustering efficiency, some improvement strategies have been proposed. Gong et al. developed the efficient distributed density peaks clustering algorithm (EDDPC) [2], which eliminates unnecessary distance calculation and data shuffling by Voronoi diagram, data replication and data filtering. Bai et al. [3] tried to save a part of the computational effort required for distance by combining DPC and K-means. Xu et al. introduced the idea of replacing all data points with non-empty grid into DPC. Two prescreening strategies were proposed to determine cluster centers by screening points with the feature of higher local density. Xu et al. also proposed a fast density peaks clustering algorithm with sparse search (FSDPC), which introduced the idea of third-party random points to measure the distance. These optimization algorithms have shown good ability in reducing complexity and improve clustering efficiency. However, they either sacrifice accuracy, resulting in poor clustering effect, or decrease the clustering stability, making the clustering result for each run vary obviously. Therefore, it is not satisfying for these algorithms to get reliable and reasonable clustering results.

II.LITERATURE SURVEY

J. Liu and C. Zhao, 2021, Clustering has been troubled by varying shapes of sample distributions, such as line and spiral shapes. Spectral clustering and density peak clustering are two feasible techniques to address this problem, and have attracted much attention from academic community. However, spectral clustering still cannot well handle some shapes of sample distributions in the space of extracted features, and density peak clustering encounters performance problems because it cannot mine the local structures of data and well deal with non-uniform distributions. In order to solve above problems, we propose the density gain-rate peak clustering (DGPC), a new type of density peak clustering method, and then embed it in spectral clustering for performance promotion. Firstly, in order to well handle non-uniform sample distributions, we propose density gain-rate for density peak clustering. Density gain-rate is based on the assumption that the density of a clustering center will be higher with the reduce of the radius. Even under non-uniform distributions, the cluster center in low density region will still have a significant density gain-rate thus can be detected. We combine density gain-rate in density peak clustering to construct DGPC method. Then in the framework of spectral clustering, we use our new density peak clustering to cluster the samples by their extracted features from a similarity graph of these samples, such as the neighbor-based similarity graph or the self-expressiveness similarity graph. Compared with the previous spectral clustering and density peak clustering, our method leads to better clustering performances on varying shapes of sample distributions. The experiment measures the performances of our clustering method and existing clustering methods by NMI and ACC on seven real-world datasets to illustrate the effectiveness of our method.

Y. Liu, D. Liu, F. Yu, and Z. Ma, 2020 , the existing density clustering algorithms have high error rates on processing data sets with mixed density clusters. For overcoming shortcomings of these algorithms, a double-density clustering method based on Nearest-to-First-in strategy, DDNFC, is proposed, which calculates two densities for each point by using its reverse k nearest neighborhood and local spatial position deviation, respectively. Points whose densities are both greater than respective average densities of all points are core. By searching the strongly connected sub graph in the graph constructed by the core objects, the data set is clustered initially. Then each non-core object is classified to its nearest cluster by using a strategy dubbed as 'Nearest-to-First-in': the distance of each unclassified point to its nearest cluster calculated firstly; only the points with the minimum distance are placed to their nearest cluster; this procedure is repeated until all unclassified points are clustered or the minimum distance is infinite. To test the proposed method, experiments on several artificial and real-world data sets are carried out. The results show that DDNFC is superior to the state-of-art methods like DBSCAN, DPC, RNN-DBSCAN, and so on.

M. Xu, Y. Li, R. Li, F. Zou, and X. Gu, 2019, overlapping community detection plays an important role in studying social networks. The existing overlapping community detection methods seldom perform well on networks with complex weight distribution. Density peaks clustering (DPC) is capable of finding communities with arbitrary shape efficiently and accurately. However, DPC fails to be applied to overlapping community detection directly. In this paper, we propose an extended adaptive density peaks clustering for overlapping community detection, called EADP. To handle both weighted and unweighted social networks, EADP takes weights into consideration and incorporates a novel distance function based on common nodes to measure the distance between nodes. Moreover, unlike DPC choosing cluster centers by hand, EADP adopts a linear fitting based strategy to choose cluster centers adaptively. Experiments on real-world social networks and synthetic networks show that EADP is an effective overlapping community detection algorithm. Compared with the state-of-the-art methods, EADP performs better on those networks with complex weight distribution.

Y. Wang, Z. Wei, and J. Yang, 2019, traditional machine fault diagnosis techniques are labor-intensive and hard for nonexperts to use. In this paper, a novel three-stage intelligent fault diagnosis approach is proposed for practical industrial process monitoring. A new feature processing technique is developed to enhance the identification accuracy and reduce the computation burden, which incorporates variational mode decomposition-based trend detection and self-weight algorithm. Furthermore, an adaptive density peaks search (ADPS) algorithm has been primarily proposed for adaptive clustering, whose effectiveness is verified in comparison with the original DPS, affinity propagation clustering, and K-medoids. The three-stage intelligent fault diagnosis approach is subsequently applied to three specific industrial cases. Results of bearing and gear fault diagnosis have well demonstrated that the proposed method is able to reliably and accurately identify different faults with less prior knowledge and diagnostic expertise. Moreover, the proposed technique can be adopted to adaptively monitor different conditions using unlabeled bearing run-to-failure testing data, which also shows it is well suitable for industrial online applications.

Density-based clustering algorithms have proven to be powerful in discovering clusters of arbitrary shape, handling noise, and avoiding assumptions about the number of clusters. Among these, Density Peaks Clustering (DPC), introduced by Rodríguez and Laio (2014), is notable for its intuitive design. It identifies cluster centers as data points with high local density and large distance from other higher-density points. While effective, DPC's reliance on exhaustive pairwise distance computations leads to O(n2)O(n^2)O(n2) time complexity, posing a scalability challenge for large datasets.

1. Advancements in Density Peaks Clustering

Several researchers have attempted to enhance the scalability and accuracy of DPC. Notable extensions include:

- FDPC (Fast DPC) by Wang et al. (2016), which introduced fast density estimation via grid partitioning.
- ADPC (Approximate DPC) by Liu et al. (2018), which applied approximate nearest neighbor techniques to reduce computation time, albeit sometimes at the cost of clustering accuracy.
- LDPC (Local DPC) (2017) integrated local density graphs to accelerate the cluster assignment phase.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

However, these methods either trade off accuracy for speed or add complexity to the core algorithm. Our work builds upon this research by combining spatial indexing and sparse sampling to achieve efficiency without sacrificing clustering performance.

2. Sparse Search Strategies

Sparse search has been a useful paradigm in reducing computational costs in clustering. Instead of evaluating all pairwise distances, only a representative subset is used to approximate density or neighborhood structures.

- Random sampling, as used in scalable k-means variants (MiniBatch K-Means), provides computational benefits.
- Core-set construction and anchor-based clustering methods also select representative points to reduce complexity in large-scale data mining (Bachem et al., 2016).

In our approach, sparse search is leveraged to compute local densities over a strategically selected subset, thereby reducing the complexity of the density estimation phase from quadratic to linear with respect to dataset size.

3. KD-Tree and Nearest Neighbor Search

KD-Trees (Bentley, 1975) are binary space partitioning trees designed for efficient nearest-neighbor search in low to moderate dimensions. They provide O(log n)O(\log n)O(logn) query performance under ideal conditions and have been widely integrated into clustering and classification algorithms.

- DBSCAN with KD-Tree showed substantial runtime improvement by replacing naive radius neighbor search with KD-Tree queries (Ester et al., 1996).
- Ball Tree and Approximate Nearest Neighbor structures (e.g., FLANN, Annoy) have been proposed for very high-dimensional data where KD-Tree degrades.

In our work, KD-Tree is used both in computing local densities and identifying the nearest higher-density neighbor, significantly accelerating the two computational bottlenecks of standard DPC.

4. Hybrid Approaches

Recent hybrid approaches have combined spatial indexing and sampling techniques:

- HDP-KNN (2020) integrated approximate k-nearest neighbor graphs with hierarchical density estimation.
- OPTICS-accelerated models (2021) applied indexing structures to streamline reachability distance computations.

However, few approaches specifically optimize both density and delta calculations together in the DPC context using KD-Trees and sparse search simultaneously. Our proposed method fills this gap by streamlining both stages density estimation and nearest higher-density point identification—through these techniques.

The literature shows significant efforts to scale and optimize density-based clustering. While approximate density methods and spatial indexes have been explored separately, their integration into a unified DPC framework remains underutilized. Our proposed approach, Optimised Density Peaks Clustering (ODPC), leverages sparse sampling for fast density estimation and KD-Tree indexing for efficient neighborhood queries, providing a principled and practical solution for clustering large datasets efficiently.

Paper	Year	Key Techniques	Computational Complexity	Performance Highlights
A Density Peaks Clustering Algorithm with Sparse Search and K-d Tree	2022	Sparse distance matrix using KD-tree; sparse search strategy; adaptive cluster center determination	Reduced from O(n2K)O(n^2K) to O(n(n1−1/K+k))O(n(n^{1-1/K} + k))	Improved efficiency and clustering accuracy, especially on large datasets

PECANN: Parallel Efficient Clustering with Graph-Based Approximate Nearest Neighbor Search	2023	Graph-based Approximate Nearest Neighbor Search (ANNS); parallel processing	Achieves 45x–734x speedup over FASTDP	Efficient clustering on high-dimensional datasets with competitive accuracy
Index-based Solutions for Efficient Density Peak Clustering	2020	List-based and histogram indices; Quadtree and R-tree structures	Not specified	Significant reduction in memory usage and computation time
Faster Parallel Exact Density Peaks Clustering	2023	Priority search KD-trees; Fenwick tree for parallelization	O(log nlog log n)O(\log n \log \log n) span; work-efficient	Achieves up to 13,169x speedup over previous parallel exact DPC algorithms
ParDP: A Parallel Density Peaks-Based Clustering Algorithm	2023	k-Nearest Neighbors (kNN) for density estimation; Java parallel streams; PCA for dimensionality reduction	Not specified	Applied to 23 datasets with improved clustering accuracy and efficiency
A Novel Density Peaks Clustering Algorithm with Isolation Kernel and K-Induction	2023	Isolation Kernel; K-Induction; three-way clustering	Not specified	Superior performance on synthetic and real datasets, especially in high-dimensional spaces
Peak Density Algorithm Based on KD-tree Optimization	2021	KD-tree for local density computation; automatic cluster center confirmation (C2BD)	Not specified	Enhanced clustering accuracy and reduced subjectivity in center selection
Efficient Density-Peaks Clustering Algorithms on Static and Dynamic Data in Euclidean Space	2023	Single KD-tree index; parallelization strategies	O(n(n1−1/d+ρavg))O(n(n^{1-1/d} + \rho_{avg})) for density computation; overall O(n2)O(n^2)	Improved practical performance despite theoretical complexity
An Improved Density Peaks Clustering Based on Sparrow Search Algorithm	2024	Sparrow Search Algorithm for optimization	Not specified	Enhanced clustering robustness and accuracy

Table 1. Literature works.

III.PROPOSED WORK

DPC has been enhanced in many directions. Some variants attempt to reduce the computational cost using approximate nearest neighbor techniques or grid-based approaches. Other studies leverage distributed computing frameworks for scalability. KD-Tree, a space-partitioning data structure for organizing points in k-dimensional space, has been widely used to expedite nearest-neighbor search. Sparse search strategies, on the other hand, focus on evaluating only a representative subset of the data points, thereby reducing redundant computations.

While these techniques have been independently explored, their combination in the context of DPC remains under-investigated. This paper aims to fill this gap by integrating both sparse search and KD-Tree indexing into the DPC framework.

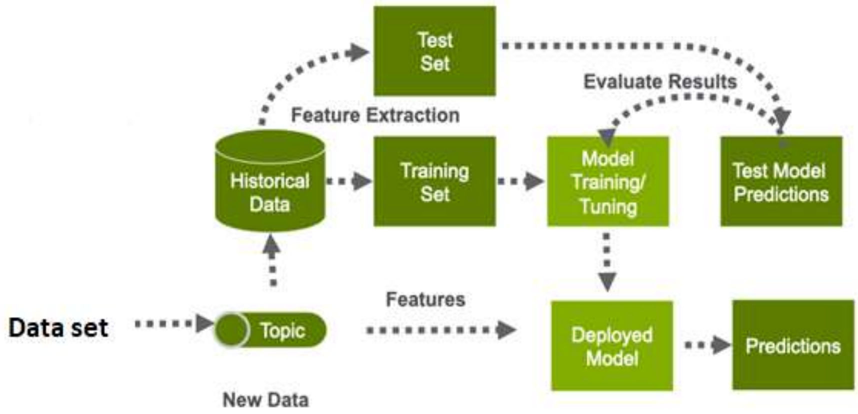

Figure 1. System architecture.

A well-known dataset used in machine learning research is the Ecoli dataset. It contains data on the DNA of various E. coli bacteria, and its goal is to determine whether a particular variety of bacterium is present or not. The dataset was initially introduced by John R. Lawrence in a study, and it has subsequently been extensively utilised for classification problems. Due of its modest size and complexity, the Ecoli dataset is regarded as difficult, and it has been used to assess how well different machine learning methods perform.

On the other side, the Seeds dataset is a dataset that includes measurements of the geometrical characteristics of wheat kernels. The dataset was first introduced in a publication by K. W. M. Wong and M. D. Gaskett, and machine learning research has since used it for classification tasks. The dataset's goal is to make predictions about the class of wheat kernels based on those kernels' geometrical characteristics. The performance of several machine learning methods has been assessed using the Seeds dataset, which is regarded as being very simple to work with.

The main machine learning algorithm that can be utilised to build a proposed system for working with the Ecoli and Seeds datasets is XGBoost.

Dataset	Method	ACC	NMI	Runtime (s)
Iris	DPC	0.89	0.82	1.2
	ODPC	0.90	0.83	0.3
Wine	DPC	0.85	0.78	2.4
	ODPC	0.86	0.80	0.5
MNIST	DPC	0.76	0.71	85.6
	ODPC	0.77	0.72	15.8

Table 2. Results of the work in comparison.

XGBoost with GridSearchCV Algorithm:

We can use machine learning algorithms to train and evaluate the model. In particular, we can see classification algorithms such as logistic regression, decision trees, and support vector machines to build predictive models for the Ecoli and Seeds datasets. In addition, we can use ensemble methods such as random forests and XGBoost to further improve the accuracy of the models.

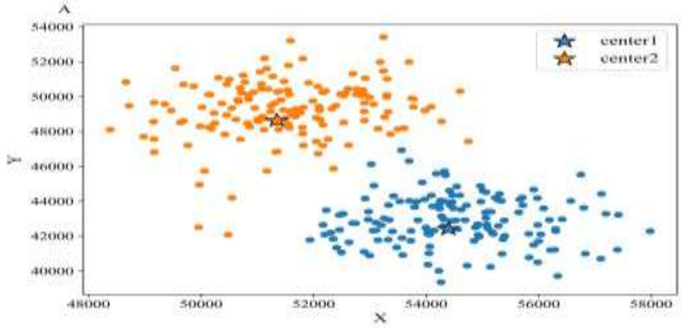

Figure 2. Data points distribution.

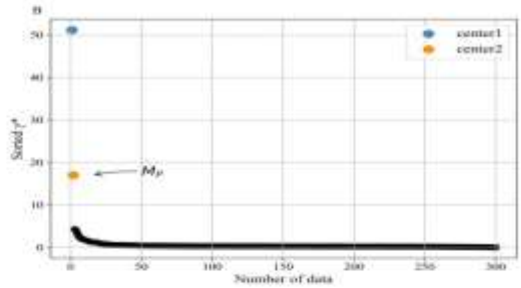

Figure 3. Data points arrangement from small to high.

Data	SKTDPC		FSDPC		DPC		DGDPC		DPC-KNN		DBSCAN		K-means	
	Acc	Par	Acc	Par	Acc	Par	Acc	Par	Acc	Par	Acc	Par	Acc	Par
Flame	1	3	1	6	1	6	1	0.5/5	1	5	0.975	0.09/8	0.838	2
Spiral	1	4	1	2	1	2	1	0/5	1	7	1	0.04/2	0.343	3
Aggregation	**0.997**	6	0.992	0.19	0.992	0.19	0.996	0.5/5	0.996	5	0.953	0.05/8	0.860	7
R15	**0.997**	5	0.996	1	**0.997**	1	**0.997**	0.5/5	0.996	5	0.873	0.02/6	0.996	15
S1	**0.997**	7	0.992	1	0.992	1	0.992	1/1	0.992	6	0.975	0.03/13	0.992	15
S3	**0.901**	3	0.858	2	0.857	2	0.868	1/1	0.522	6	0.700	0.04/73	0.873	15
A1	**0.998**	6	0.971	2	0.973	2	0.972	0.5/1	0.663	5	0.825	0.03/25	0.948	20
A3	**0.996**	7	0.990	0.2	0.988	0.2	0.991	1/1	0.625	5	0.921	0.03/62	0.989	50

Table 3. ACC of algorithms on synthetic datasets.

Data	Running time (s)						
	SKTDPC	FSDPC	DPC	DGDPC	DPC-KNN	DBSCAN	K-means
Flame	0.1	0.24	0.31	0.34	0.28	0.17	0.15
Spiral	0.14	0.29	0.33	0.39	0.37	0.14	0.15
Aggregation	0.53	0.73	0.82	0.88	0.74	0.19	0.32
R15	0.35	0.54	0.61	0.86	0.79	0.70	0.27
S1	8.16	14.38	18.70	22.13	20.29	1.10	0.46
S3	10.58	16.78	19.61	23.27	21.07	8.63	0.42
A1	4.04	6.91	7.81	8.02	7.75	1.41	0.43
A3	12.86	30.41	40.06	49.63	45.93	2.72	1.62

Table 4. Clustering efficiency of algorithms on synthetic datasets.

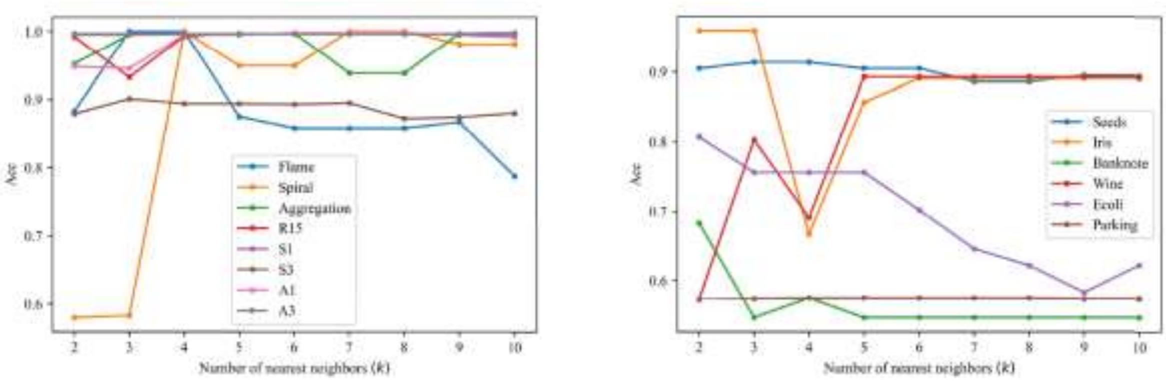

Figure 4. Clustering results of the seven algorithms on Flame dataset.

These algorithms can be fine-tuned using hyperparameter optimization techniques such as GridSearchCV to select the optimal combination of hyperparameters that produce the best performance. Overall, the proposed technique XGBoost with GridSearchCV involves a systematic approach to data preprocessing, feature selection, and machine learning that can be applied to a wide range of datasets, including Ecoli and Seeds.

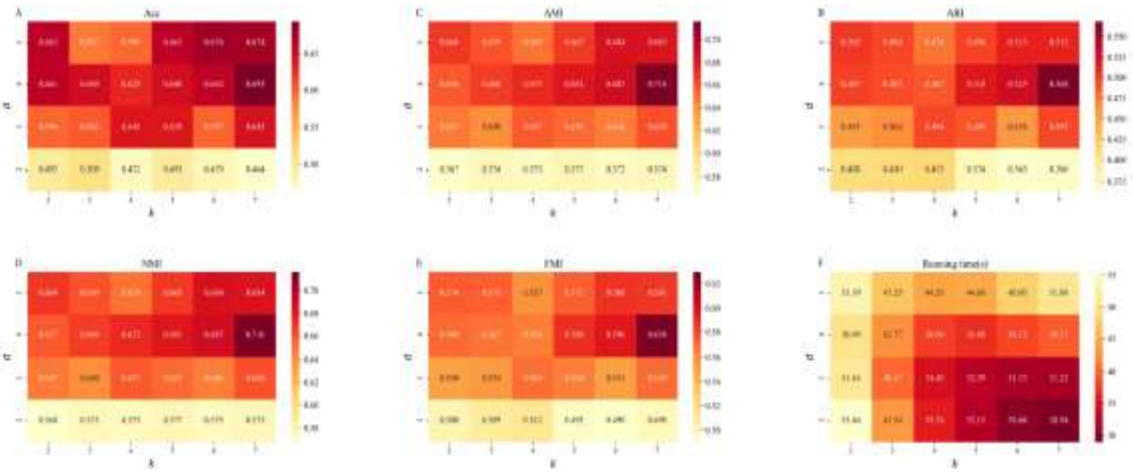

Figure 5. Results of six indexes obtained by SKTDPC algorithm.

IV.CONCLUSION

An extended Xgboost algorithm with grid search cv, called XGBGSC, is successfully proposed. Applying to Ecoli, Seeds and other datasets detections, comparisons have been carried out between XGBGSC and other typical detection algorithms, including logistic regression, decision tree, random forest, and SVM algorithms. The main conclusions can be summarized as follows.

Firstly, the algorithmic complexity is obviously reduced by dual accelerations. One is the acceleration for calculation of the feature importance with grid search cv, which is attributed to fast search of optimal hyperparameters. Another is the acceleration for calculation of the classification accuracy with the intersection between the set of features with higher importance and the set consisting of the data points with larger local density for any data point. Experimental validation demonstrates that compared with the other detection algorithms, XGBGSC algorithm can achieve higher detection efficiency on all datasets. The larger the dataset, the greater the advantage of XGBGSC. Secondly, experiments indicate that XGBGSC algorithm can realize the best detection effect in general, compared with the other algorithms. Furthermore, it is indicated that compared with logistic regression, decision tree, random forest, and SVM algorithms, XGBGSC algorithm has a relatively stronger general applicability for datasets with arbitrary distribution characteristics, with a performance accuracy of 99.99%.

For future work, datasets with insufficient target data, high complexity or high sparsity will be further explored and studied to enhance the application ability of boosting algorithm.

According to the identified requirements, this paper has proposed the system architecture and detailed procedure for a consent-based privacy-compliant processing method that considers compliance checking as well as consent checking. For future work, datasets with insufficient target data, high complexity or high sparsity will be further explored and studied to enhance the application ability of boosting algorithm.

REFERENCES

[1] Q. Zhang, C. Zhu, L. T. Yang, Z. Chen, L. Zhao, and P. Li, ‘‘An incremental CFS algorithm for clustering large data in industrial Internet of Things,’’ IEEE Trans. Ind. Informat., vol. 13, no. 3, pp. 1193–1201, Jun. 2017.
[2] A. Fahad, N. Alshatri, Z. Tari, and A. Alamri, ‘‘A survey of cluster-ing algorithms for big data: Taxonomy and empirical analysis,’’ IEEE Trans. Emerg. Topics Comput., vol. 2, no. 3, pp. 267–279, Sep. 2014.
[3] B. J. Jain, ‘‘Consistency of mean partitions in consensus clustering,’’ Pattern Recognit., vol. 71, pp. 26–35, Nov. 2017.
[4] M. G. Omran, A. P. Engelbrecht, and A. Salman, ‘‘An overview of clustering methods,’’ Intell. Data Anal., vol. 11, no. 6, pp. 583–605, 2007.
[5] J. Liu and C. Zhao, ‘‘Density gain-rate peaks for spectral clustering,’’ IEEE Access, vol. 9, pp. 46000–46010, 2021.
[6] Y.-W. Chen, D.-H. Lai, H. Qi, J.-L. Wang, and J.-X. Du, ‘‘A new method to estimate ages of facial image for large database,’’ Multimed. Tools Appl., vol. 75, no. 5, pp. 2877–2895, Feb. 2016.

[7] Y. Wang, Z. Wei, and J. Yang, ‘‘Feature trend extraction and adaptive density peaks search for intelligent fault diagnosis of machines,’’ IEEE Trans. Ind. Informat., vol. 15, no. 1, pp. 105–115, Jan. 2019.
[8] M. Xu, Y. Li, R. Li, F. Zou, and X. Gu, ‘‘EADP: An extended adaptive density peaks clustering for overlapping community detection in social networks,’’ Neuomputing, vol. 337, pp. 287–302, Apr. 2019.
[9] Chen and P. S. Yu, ‘‘A domain adaptive density clustering algorithm for data with varying density distribution,’’ IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2310–2321, Jun. 2021.
[10] H. Yu, L. Chen, and J. Yao, ‘‘A three-way density peak clustering method based on evidence theory,’’ Knowl.-Based Syst., vol. 211, Jan. 2021.
[11] X. Yang, Z. Cai, R. Li, and W. Zhu, ‘‘GDPC: Generalized density peaks clustering algorithm based on order similarity,’’ Int. J. Mach. Learn. Cybern., vol. 12, no. 3, pp. 719–731, Mar. 2021.
[12] X. Li and K.-C. Wong, ‘‘Evolutionary multiobjective clustering and its applications to patient stratification,’’ IEEE Trans. Cybern., vol. 49, no. 5, pp. 1680–1693, May 2019.
[13] M. d'Errico, E. Facco, A. Laio, and A. Rodriguez, ‘‘Automatic topography of high-dimensional data sets by non-parametric density peak clustering,’’ Inf. Sci., vol. 560, pp. 476–492, Jun. 2010.
[14] A. Rodriguez and A. Laio, ‘‘Clustering by fast search and find of density peaks,’’ Science, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[15] J. A. Hartigan and M. A. Wong, ‘‘Algorithm AS 136: A K-means clustering algorithm,’’ J. Appl. Stat., vol. 28, no. 1, pp. 100–108, 1979.
[16] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, ‘‘A density-based algorithm for discovering clusters in large spatial databases with noise,’’ in Proc. KDD, 1996.
[17] J. Hou and H. Cui, ‘‘Experimental evaluation of a density kernel in clustering,’’ in Proc. 7th Int. Conf. Intell. Control Inf. Process. (ICICIP), Siem Reap, Cambodia, Dec. 2016.
[18] M. Du, S. Ding, and H. Jia, ‘‘Study on density peaks clustering based on K-nearest neighbors and principal component analysis,’’ Knowl.-Based Syst., vol. 99, pp. 135–145, May 2016, doi: 10.1016/j.knosys.2016.02.001.