International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.2 (2025) pp. 2597-2604
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

An Efficient Hybrid Improved Feature Vector Manifold Clustering with Neighbour Search Optimization

L. Dhanapriya[1]*, S. Preetha[2]

[1]Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, Tamil Nadu, India
* **Corresponding Author Email:** dhanapriyabca@gmail.com - **ORCID:** 0000-0002-6151-748X

[2]Associate Professor, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, Tamil Nadu, India
Email: preethacs@srcw.ac.in - **ORCID:** 0000-0002-9005-7259

Article Info:

DOI: 10.22399/ijcesen.1671
Received : 25 January 2025
Accepted : 02 April 2025

Keywords :

Data Mining,
Clustering,
Feature Selection,
Manifold,
Neighbour Search,
Outlier.

Abstract:

In this paper, the IFMCNSO algorithm a novel hybrid Improved Feature Vector Manifold clustering with Neighbour search optimization clustering algorithm —is presented. Many methods for linear or nonlinear manifold clustering have been developed recently. While in many cases they have proven to perform better than classic clustering algorithms, the majority of these approaches have a high complexity. In order to overcome the clustering problem, particularly for high-dimensional datasets, this work provides an effective hybrid method called IFMCNSO. By using this strategy, the domain in which feature vector manifold learning and Neighbor search optimization techniques can be used is greatly expanded, enabling parameterization in real-world data sets. A good or nearly optimal solution is found using the IFMCNSO algorithm in an acceptable amount of time. A comprehensive comparison of the proposed IFMCNSO algorithm with state-of-the-art clustering algorithms, namely DCNaN, RDMN, HFMST, and HFMST-PSO, reveals that IFMCNSO achieves higher Rand Index (RI) and Adjusted Rand Index (ARI) scores, underscoring its exceptional clustering performance and accuracy.

1. Introduction

algorithms for processing massive amounts of data with conventional data analysis techniques. Additionally, it has created intriguing possibilities for investigating and evaluating fresh data kinds as well as for reinterpreting more traditional data types [1]. In data mining, cluster analysis looks for collections of closely related observations, with the goal of making observations inside a cluster more similar to one another than between clusters. Finding observations with features that distinguish them noticeably from the rest of the data is the problem of anomaly detection. To describe these findings as anomalies or outliers. An anomaly detection algorithm's objective is to identify the true anomalies rather than mistakenly classifying regular items as abnormal [2].

Clustering is a fundamental unsupervised machine learning technique that group's data based on similarities and differences, with several applications [3]. Various clustering approaches have been proposed to address data clustering challenges, including fuzzy c-mean, k-means, k-medoids, and simulated annealing. However, these methods are prone to getting trapped in local optima due to their dependence on initial solutions. To overcome this limitation, nature-inspired meta-heuristic optimization techniques have been developed and tested on complex and high-dimensional datasets[4-6]. Traditional clustering algorithms often struggle to manage large and intricate datasets, necessitating the development of evolutionary algorithms like genetic and differential evolution, as well as optimization algorithms like particle swarm optimization, ant colony optimization and firefly optimization. These algorithms have been employed in recent decades to tackle data clustering problems and have shown promising results. The majority of clustering techniques would need an objective function that is well defined. The two validity indices stated above are combined with a Euclidean-based distance measure to calculate the fitness function of every solution that is found. Most metaheuristic algorithms have the ability to automatically divide datasets into the optimal number of clusters and handle noise or outlier identification associated with the datasets. Until the ideal answer is found, these algorithms attempt to optimize a population of randomly generated individuals across a number of generations. The research first focuses on determining the ideal number of clusters, and then it progressively shifts to the globally optimal cluster centers. Based on the existing clustering validation indices and the penalty functions created to reduce noise and regulate the number of clusters, two kinds of continuous fitness functions are created [7].

The fundamental concept for identifying an outlier is to identify abnormal points among the data points. The objects that primarily deviate from a specific data collection are identified by outlier identification. Finding outliers that don't match the rest of the dataset is a significant problem for real-world KDD applications. To determine a point's divergence from other points—which indicates whether or not it is an outlier—a number of metrics are employed. It is not essential to compute these measurements for every point in a data collection because there are very few outliers. Reducing computation time by eliminating points that are most likely not outliers is the goal of outlier detection based on clustering [8].

The clustering and the outlier detection problem can be related in some situations. Identifying clusters and outliers, for instance, is the primary goal of clustering-based outlier detection algorithms. These are sometimes thought of as noise that needs to be eliminated to produce more dependable grouping. When compared to the data points, some noisy points might be far away while others might be nearby. Given their greater dissimilarity from the data points, the distant noisy points would have a greater impact on the outcome.

2. Related Work

It was introduced a new approach called Particle Swarm Optimization for Density-based Clustering and Classification (PODCC), which aims to address the shortcomings of DBSCAN [9]. One popular Evolutionary and Swarm Algorithm (ESA) that has been employed for optimization issues in a variety of research fields, including data analytics, is Particle Swarm Optimization (PSO). To find the ideal parameters for density-based clustering and classification, PODCC uses SPSO-2011, a PSO version, to explore the parameter space.

A multi-source outlier detection technique was proposed by [10] to accurately identify outliers in several datasets. Based on the correlation between the datasets, three types (Type I–III) of multi-source outliers are explored, based on a number of real-world cases. To find high-score outliers, they created two algorithms: a baseline approach that is an obvious solution and an ideal technique called multiple-data-sources oriented outlier detection (MOD). To expedite the outlier discovery procedure, they also develop the MOD+ approach. We present a new density measure to assess the deviation degrees of multi-source outliers, which combines kNN and RNN.

The Clustering with Outlier Removal (COR) approach was proposed by [11] in response to their consideration of the joint cluster analysis and outlier detection problem. In particular, the process of creating fundamental partitions converts the original space into a binary space. The measure of each cluster's compactness was determined by the authors using holoentropy, which eliminated the need for many outlier candidates. An auxiliary binary matrix is introduced to provide a tidy and fast solution, allowing COR to tackle the difficult problem entirely and effectively using a unified K-means approach with theoretical backing. Extensive experimental findings across multiple domains and data sets show that COR is much more successful and efficient than state-of-the-art techniques for outlier detection and cluster validity.

A unique approach to the detection of mixed attribute outliers was developed by [12], utilizing the neighborhood rough set and multigranulation relative entropy. Optimizing the mixed distance metric and the statistical value's radius first creates the neighborhood system. Furthermore, as a data uncertainty metric, the neighborhood entropy is presented. Moreover, three types of attribute sequences are employed to define three distinct multigranulation relative entropy-based matrices. Subsequently, a matching method is developed, leveraging the proposed outlier detection model as its foundation. This matching method utilizes the outlier degrees and multigranulation relative entropy-based matrices to establish correspondences between objects. A robust rank-constrained sparse learning (RRCSL) technique was presented by [13]. To learn the best graph with resilience, the L2, L1-norm is incorporated into the sparse representation objective function. In order to maintain the data structure, we create an initial graph and search the graph in its vicinity. The trained graph can be used as the cluster indicator straight away by adding a rank constraint, and the results are acquired without the need for further post-processing. Furthermore, the suggested approach is not limited to single-view clustering; it may also be expanded to multi-view clustering. The effectiveness and resilience of the suggested framework have been shown in several tests using both artificial and real-world datasets.

A neighborhood representative (NR) approach was introduced by [14] and allows all of the outlier detectors now in use to effectively identify outliers—including collective outliers—while preserving their computational integrity. It accomplishes this by choosing representative items, assigning a score to them, and then applying the representative objects' score to all of its objects together. NR is compatible with current detectors without requiring changes, and it achieves +8% (0.72 to 0.78 AUC) on average in real-world dataset performance when compared to twelve state-of-the-art outlier detectors. Described an HFMST technique, which is a special HFMST clustering [15]. The HFMST clustering method is a novel clustering technique that combines the EAFFS algorithm with MST. Finding and eliminating redundant characteristics can frequently significantly increase the clustering accuracy, even though it is highly likely that they would show up in the final subset in the existing feature selection procedures. EAFFS and the HFMST technique are two approaches to solving the issue. A new hybrid fuzzy based minimum spanning tree (HFMST) with particle swarm optimization clustering was proposed by [16]. HFMST-PSO is an efficient hybrid method provided by the authors to address the clustering issue, especially with high-dimensional datasets. The HFMST approach may result in an unequal distribution of data, making it difficult to find the perfect solution in an acceptable amount of time when the problem gets large. HFMST is sensitive to startup and is prone to become stuck in local optima. Particle swarm optimization is a common stochastic global optimization technique used to address optimization problems. Within a reasonable period, the PSO algorithm finds an optimal or substantially optimal solution.

3. Research Methodology

All experiments performed on a large-scale real-world dataset are successfully tested by the suggested methodology. This part reports on the research technique, which takes into account an effective clustering process employing data preparation, the Enhanced Adaptive Fuzzy Linkage based Feature Selection (EAFFS) algorithm, Outlier Detection using Feature Ranking Method and Improved Feature Vector Manifold clustering with Neighbor search optimization algorithm (IFMCNSO) Clustering process. Figure 1 displays the overall process flow diagram for the suggested clustering method.

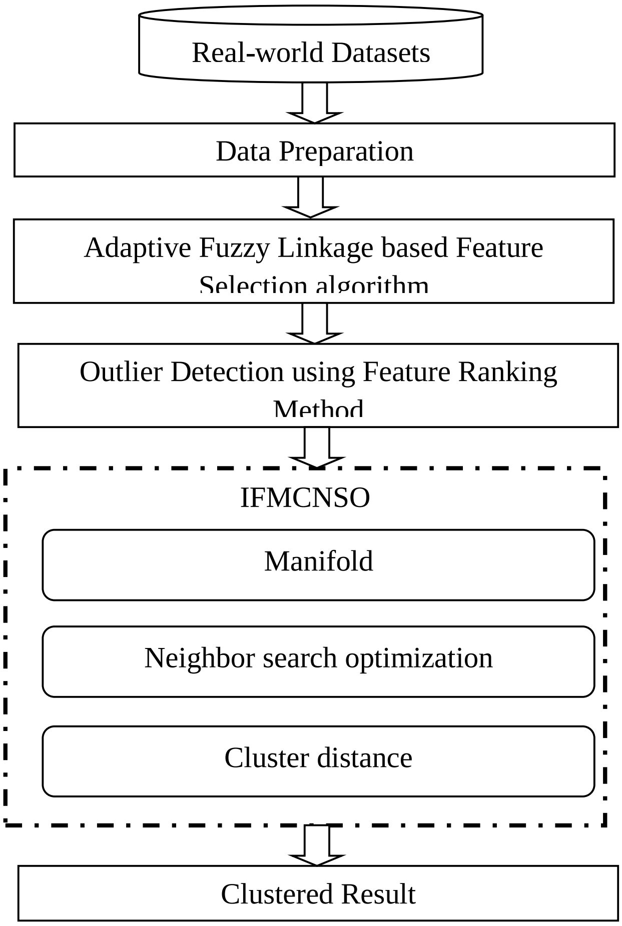

Figure 1. *Proposed process diagram*

3.1 Data Preprocessing

In Data preprocessing process is evaluated in graph formation was already presented (L. Dhanapriya and S. Preetha, 2023). Accuracy in the connected decisions is guaranteed by the perfect data group. Many data sources that can be obtained on different websites are used to collect data. Data on lung cancer, Lukemia, MiniBooNE, Novaratis and Gaussian distributed datasets were included in the datasets [17,18]. Figure 2 provides an overview of the real-world datasets.

3.2 EAFFS: Enhanced Adaptive Fuzzy Linkage based Feature Selection

EAFFS feature selection in real-world datasets was previously presented (L. Dhanapriya and S. Preetha, 2023 [18]). The fuzzy clustering technique is where this technique originated. In order to generate cluster centers and a membership function that fuzzifies every characteristic, fuzzy clustering is employed entropy measure.

Dataset	Number of Instances	Cluster	Number of Features
Lukemia	248	6	985
Lung Cancer	32	3	56
MiniBooNE	130064	2	50
Dim-1024	1024	16	32
BioTrain	145751	2	74
Novaratis	103	4	1000
LungA	197	4	1000

Figure 2. *Real-world Dataset*

3.3 Outlier Detection using Feature Ranking Method

The suggested feature ranking algorithm works by removing outliers to reduce dimensionality. Outliers are found using a univariate approach. Let's first visualize the dataset as a matrix. The dataset matrix is reversed, with the samples represented in columns and the features (or attributes) in rows. This is because the original dataset's feature set (samples from the transposed dataset) contains the outliers. The paper proposes an outlier detection-based Feature Ranking (FR) method, leveraging an improved variant of Rank-based Projection Pursuit, termed IRPPFR, for enhanced feature ranking.

The objective is to detect outliers in the feature space, rather than in the data instances themselves. To achieve this, it is necessary to transpose the dataset matrix, such that the features become the objects of interest. This transposition enables the identification of outliers among the features, rather than among the data points. Let D denote the (M×N) dataset matrix, where M represents the number of observations (samples) and N represents the number of features. Consequently, the transposed dataset matrix D' is an (N×M) matrix, where the rows now correspond to the N features and the columns correspond to the M observations (samples). It's important to understand that IRPPFR uses the transposed matrix to choose the best projection for revealing the outliers, or unimportant characteristics. This approach takes into account a one-dimensional projection, denoted by the M-dimensional vector b. The anticipated data coordinates are displayed as an M-dimensional vector mv so that,

$$mv = D'b \tag{1}$$

In the mathematical model, dataset D are divided into k classes, with the n_{th} sample of the m_{th} class making up row $D_{m,n}$. A solution, S_1, is computed from the initial view vw_0, utilizing a basic repulsion-attraction model, parameterized by an initial projection rank Pr^0 and an initial view $vw^0 = DPr^0$. This is achieved through an iterative process, wherein each data point is displaced in the direction of its respective class centroid, while simultaneously; the centroids of each class are repelled from one another with a velocity inversely proportional to their pair-wise distances. Given that each class's centroid within the view is $vw_m{}^0$, the new centroid for the m_{th} class is equal to the mean of $vw_{m,n}{}^0 = D_{m,n}.\ \boldsymbol{P}r^0$, the data of $D_{m,n}$ under $\boldsymbol{P}r^0$ is given by,

$$S_m^1 = vw_m^0 + k\sum_n \frac{(vw_m^0 - vw_n^0)}{|vw_m^0 - vw_n^0|} \tag{2}$$

And the new solution for the n_{th} sample in the m^{th} class

$$S_{m,n}^1 = vw_m^0 + k1(S_m^1 - vw_{m,n}^0) \tag{3}$$

The suggested method for locating outliers based on how closely neighboring data points are to one another. Let a data point p ∈ D and assume that q ∈ Neigk(p) in order to comprehend mutual proximity. Determine the rank of p among all of q's neighbors; that is, compute the group of d(q, e) for each e ∈ D –{q} and ascertain the rank of d(q, p) inside this set. Let this be rankq(p). The outliers of data point p defined by Ok(p),

$$O_k(p) = \frac{\sum_{q\in Neig_k(p)} rank_q(p) + S}{|Neig_k(p)|} \tag{4}$$

P is regarded as an outlier if $O_k(p)$ is 'large'.

3.4 Improved Feature Vector Manifold clustering with Neighbour search optimization algorithm (IFMCNSO)

Given an input dataset, the goal of the IFMCNSO is to partition the data into clusters, such that each cluster contains data points that lie on a single, low-dimensional, and simple manifold. Assuming that the number and dimensions of these low-dimensional manifolds are known. In particular:

Let $\{D_1, D_2, ..., D_n\}$ be a set of n data points, where each data point D_m is an element of R^{dim}, the dim-dimensional Euclidean space. Assume that these data points are derived from the intersection of h non-linear manifolds.

- Assign a label k_i to each data point D_m, where k_i is an integer between 1 and t (inclusive), indicating the index of the manifold to which the point belongs.
- Compute the low-dimensional embedding Y_m of each data point D_m on its associated manifold, where Y_m is an element of R^d, the d-dimensional Euclidean space, with d < dim.

The proposed method perform a better Feature in order to choose a small number of neighbors for each data point that cover a low-dimensional affine subspace that passes close to that point, Vector Manifold clustering uses an optimization program based on the neighbor search approach. This means that some of the solution's nonzero elements can be used for clustering because they show which points are on the same manifold. To further aid in dimensionality reduction, the weights assigned to the selected neighbors show how far away they are from the specified data point.

Consider a collection of N data points $\{D_i \in R_d\}_{i=1}^{N}$, where each data point lies in one of n different manifolds $\{M_l\}_{l=1}^{n}$, with intrinsic dimensions $\{d_l\}_{l=1}^{n}$. For each data point x_i in manifold M_l, define the smallest point $SP_i \subset R^d$ that contains the d_l + 1 nearest neighbors of x_i from M_l. Let N_i denote the neighborhood of x_i, which is the set of all data points in D_i excluding x_i itself. In general, this neighborhood Ni contains points from manifold M_l as well as points from other manifolds. Assume that for all i, there exists a non-negative value $\varepsilon \geq 0$ such that the sparsest solution to the system satisfies:

$$\left\| \sum_{j \in N_i} c_{ij}(x_j - x_i) \right\|_2 \leq \epsilon \;\; and \;\; \sum_{j \in N_i} c_{ij} = 1 \qquad (5)$$

The proposed neighbor search optimization method for each data point, to obtain the necessary information for clustering satisfies the below equation,

$$\sum_{j \neq i} \frac{c_{ij}}{\|x_j - x_i\|_2}(x_j - x_i) \approx 0 \qquad (6)$$

Algorithm 1: IFMCNSO CLSUTERING

Input: Input Dataset D, Cluster c, Neighbor search N. Outlier O
Output: Cluster formation
Preparation:
1. Data Preprocessing
2. EAFFS
3. Outlier Detection using Feature Ranking Method
4. Data clustering using IFMCNSO method

Steps:
While (D)
1. D_p ← Data preprocessing // Data noise, NaN removal
2. *EAFFS* ← Feature selection
3. *O*←It will detect the outlier using Ranking deviation method.
4. IFMCNSO ← Hybrid model to form the cluster performance using *Dp*

End While

4. Experimental Result

The proposed IFMCNSO performance was compared to the DCNaN [19], RDMN [20], HFMST and HFMST-PSO algorithms that are currently in use. In this experiment, real-world low- and high-dimensional datasets can be obtained. The accuracy of the dataset was calculated and the requirements for different sorts of experiments were created using the suggested IFMCNSO method.

The Rand Index (RI) measure comparison shown in Figure 3 and Table 1. The definition of the Rand Index is as follows [21]:

$$RandIndex\ (RI) = (a + b)/nC_2) \qquad (7)$$

Were,
a: The number of pair items is part of the cluster of comparable elements..
b: The number pair elements are part of various clusters..
nC_2: A collection of n element's total number of unordered pairings.

A comparative analysis of the Adjusted Rand Index (ARI) measure is presented in Table 2 and illustrated in Figure 4, highlighting the performance evaluation of different clustering algorithms.

Table 1. *Comparison of RI Values*

Dataset	DCNaN	RDMN	HFMST	HFMST-PSO	Proposed IFMCNSO
Lukemia	0.7627	0.5539	0.9264	**0.9421**	**0.9850**
Lung Cancer	0.3468	0.5926	0.7241	**0.8041**	**0.9125**
MiniBooNE	0.5962	0.6033	0.6841	**0.7621**	**0.8436**
Dim-1024	0.9255	0.9547	0.9708	**0.9803**	**0.992**
BioTrain	0.9793	0.9823	0.9912	**0.9985**	**0.9994**
Novaratis	0.6457	0.6587	0.7185	**0.7925**	**0.8602**

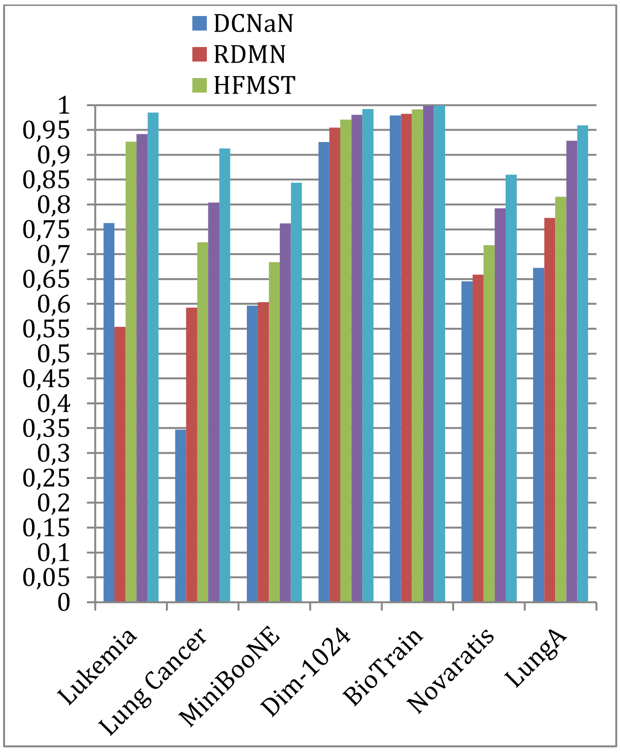

Figure 3. *RI diagram*

5. Conclusions

An enhanced Hybrid Improved Feature Vector Manifold clustering with Neighbor search optimization algorithm (IFMCNSO) method is presented and examined in this work. The proposed IFMCNSO algorithm is designed as a four-stage framework, consisting of Data Preprocessing, EAFFS algorithm, Outlier Detection using Feature Ranking Method, and IFMCNSO Clustering, enabling efficient and effective data analysis and clustering. Data cleaning is employed as a preprocessing step to refine and prepare real-world datasets, encompassing both low and high dimensional data, for further analysis.

Table 2. *Comparison of ARI Values*

Dataset	DCNaN	RDMN	HFMST	HFMST-PSO	Proposed IFMCNSO
Lukemia	0.4011	0.6311	0.7513	**0.8625**	**0.8910**
Lung Cancer	0.1258	0.2288	0.3842	**0.5682**	**0.65272**
MiniBooNE	0.258	0.332	0.4235	**0.6318**	**0.6837**
Dim-1024	0.3111	0.4288	0.4821	**0.5908**	**0.6206**
BioTrain	0.3841	0.4299	0.5233	**0.6824**	**0.7105**
Novaratis	0.3130	0.3300	0.3841	**0.5274**	**0.5942**
LungA	0.3615	0.5344	0.6018	**0.7239**	**0.7801**

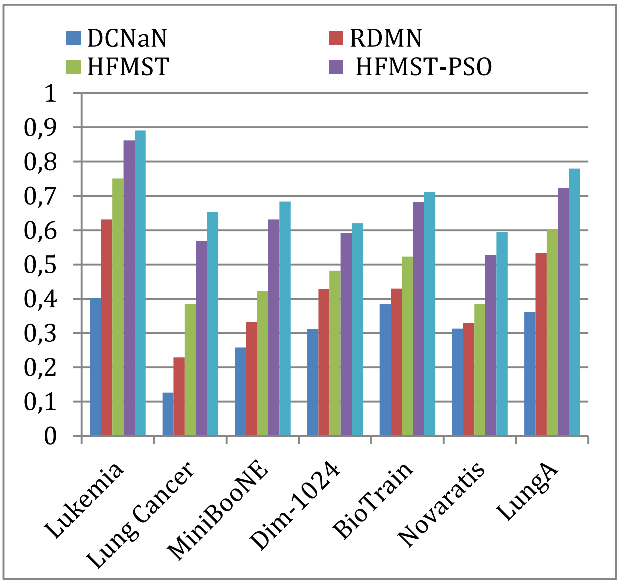

Figure 4. *ARI chart*

Using the neighbor search optimization approach a search optimization program is solved to create a similarity graph that enables low-dimensional embedding. The key to identifying groupings of observations that differ from the majority of the data using this method is the size of the generated clusters. Data mining and AI well studied and reported in the literature [22-31].

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Edelstein, H. A. (1999). Introduction to data mining and knowledge discovery (3rd ed). *Potomac, MD: Two Crows Corp.*

[2] Mirkin, B. Clustering: A Data Recovery Approach, Second Edition. *Chapman & Hall/CRC Computer Science & Data Analysis*.

[3] Alpaydin, E. (2004). Introduction to Machine Learning. *Cambridge, the MIT Press*.

[4] Tan, P. N., Steinbach, M., & Kumar, V. (2005). Introduction to data mining. *Boston, Addison-Wesley*.

[5] Webb, A. (2002). Statistical pattern recognition. *New Jersey, John Wiley & Sons*.

[6] Hathway, R. J., & Bezdek, J. C. (1995). Optimization of clustering criteria by reformulation. *IEEE transactions on Fuzzy Systems*. 3(2), 241–245. https://doi.org/10.1109/91.388178

[7] Rajendra, P., Jatindra, K. D., & Sukumar, N. (2010). Distance based Fast Outlier Detection Method. *Annual IEEE, India Conference (INDICON)*. 1–4. https://doi.org/10.1109/indcon.2010.5712706

[8] Zhang, J., Yang, L., Zhang, Y., Tang, D., & Liu, T. (2022). Non-parameter clustering algorithm based on saturated neighborhood graph. *Applied Soft Computing*. 130, 109647. https://doi.org/10.1016/j.asoc.2022.109647

[9] Guan, C., Yuen, K. K. F., & Coenen, F. (2019). Particle swarm Optimized Density-based Clustering and Classification: Supervised and unsupervised learning approaches. *Swarm and Evolutionary Computation*. 44;876–896. https://doi.org/10.1016/j.swevo.2018.09.008

[10] Ma, Y., Zhao, X., Zhang, C., Zhang, J., & Qin, X. (2021). Outlier detection from multiple data sources. *Information Sciences*. 580, 819–837. https://doi.org/10.1016/j.ins.2021.09.053

[11] Liu, H., Li, J., Wu, Y., & Fu, Y. (2021). Clustering with outlier removal. *IEEE Trans. Knowl. Data Eng*. 33(6);2369-2379. https://doi.org/ 10.1109/TKDE.2019.2954317

[12] Yuan, Z., Chen, H., Li, T., Zhang, X., & Sang, B. (2022). Mult granulation Relative Entropy-Based Mixed Attribute Outlier Detection in Neighborhood Systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 52(8); 5175–5187. https://doi.org/10.1109/tsmc.2021.3119119

[13] Rameshkumar, et al. (2023). Cloud computing and machine learning for analysis of health care data based on neuro fuzzy logistic regression. *Journal of Intelligent & Fuzzy Systems*. 44(6);9955-9964. https://doi.org/10.3233/jifs-223280

[14] Ammal, S. G., et al. (2024). Advanced Cloud-Based Prediction Models for Cardiovascular Disease: Integrating Machine Learning and Feature Selection Techniques. *SN Computer Science*. 5(5);572. https://doi.org/10.1007/s42979-024-02927-w

[15] Ramkumar, M., et al. (2024). Identifying cancer risks using spectral subset feature selection based on multi-layer perception neural network for premature treatment. *Computer Methods in Biomechanics and Biomedical Engineering*. 27(13);1804-1816. https://doi.org/10.1080/10255842.2023.2262662

[16] Dhanapriya, L., & Preetha, S. (2023). HFMST-PSO: An Efficient Hybrid Fuzzy Based Minimum Spanning Tree (HFMST) With Particle Swarm Optimization Clustering Algorithm. *Metszet journal*. 528-533. https://doi.org/10.27896/METSZET8.11/46

[17] Pasi, F. et al. (2015). *Clustering datasets*. Available: http://cs.uef.fi/sipu/datasets/

[18] Yeh, C. C., & Yang, M. S. (2017). Evaluation measures for cluster ensembles based on a fuzzy generalized Rand index. *Applied Soft Computing*. 57, 225–234. https://doi.org/10.1016/j.asoc.2017.03.030

[19] Xie, J., Xiong, Z.-Y., Zhang, Y.-F., Feng, Y., & Ma, J. (2018). Density corebased clustering algorithm with dynamic scanning radius. *Knowledge-Based Systems.* 142;58–70. https://doi.org/10.1016/j.knosys.2017.11.025

[20] Mishra, G., & Mohanty, S. K. (2022). RDMN: A Relative Density Measure Based on MST Neighborhood for Clustering Multi-Scale Datasets. *IEEE Transactions on Knowledge and Data Engineering*. 34(1). https://doi.org/10.1109/tkde.2020.2982400

[21] Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*. 17(2–3);107–145. https://web.itu.edu.tr/sgunduz/courses/verimaden/paper/validity_survey.pdf

[22] ARSLANKAYA, S., & Saltan YAŞLI, G. (2024). Predicting Diabetes Disease Using Data Mining Classification Algorithms and Comparison of Algorithm Performances. *International Journal of Computational and Experimental Science and Engineering,* 10(2). https://doi.org/10.22399/ijcesen.233

[23] M. Revathy Meenal, & S. Mary Vennila. (2025). Renyi Entropy Predictive Data Mining And Weighted Xavier Deep Neural Classifier For Heart Disease Prediction. *International Journal of Computational and Experimental Science and Engineering,* 11(1). https://doi.org/10.22399/ijcesen.1000

[24] García, R., Carlos Garzon, & Juan Estrella. (2025). Generative Artificial Intelligence to Optimize Lifting Lugs: Weight Reduction and Sustainability in AISI 304 Steel. *International Journal of Applied Sciences and Radiation Research* , 2(1). https://doi.org/10.22399/ijasrar.22

[25] Arya P. Pillai, & N.V. Chinnasamy. (2025). Prediction of Postpartum Depression With Dataset Using Hybrid Data Mining Classification Technique. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.750

[26] Hafez, I. Y., & El-Mageed, A. A. A. (2025). Enhancing Digital Finance Security: AI-Based Approaches for Credit Card and Cryptocurrency Fraud Detection. *International Journal of Applied Sciences and Radiation Research*, 2(1). https://doi.org/10.22399/ijasrar.21

[27] Fowowe, O. O., & Agboluaje, R. (2025). Leveraging Predictive Analytics for Customer Churn: A Cross-Industry Approach in the US Market. *International Journal of Applied Sciences*

and Radiation Research, 2(1). https://doi.org/10.22399/ijasrar.20

[28] S. Menaka, & V. Selvam. (2025). Bibliometric Analysis of Artificial Intelligence on Consumer Purchase Intention in E-Retailing. *International Journal of Computational and Experimental Science and Engineering,* 11(1). https://doi.org/10.22399/ijcesen.1007

[29] Ibeh, C. V., & Adegbola, A. (2025). AI and Machine Learning for Sustainable Energy: Predictive Modelling, Optimization and Socioeconomic Impact In The USA. *International Journal of Applied Sciences and Radiation Research,* 2(1). https://doi.org/10.22399/ijasrar.19

[30] Taşkaya, S., Wu, D., Kurt, M., Liao, Y., Xu, J., & Liao, W. (2024). Exploring the Application of Building Information Modeling (BIM) in Town Planning: Key Roles in the Relationship Between Buildings and Parcels. *International Journal of Computational and Experimental Science and Engineering, 10*(4). https://doi.org/10.22399/ijcesen.459

[31] Olola, T. M., & Olatunde, T. I. (2025). Artificial Intelligence in Financial and Supply Chain Optimization: Predictive Analytics for Business Growth and Market Stability in The USA. *International Journal of Applied Sciences and Radiation Research,* 2(1). https://doi.org/10.22399/ijasrar.18