Indonesian Journal of Electrical Engineering and Computer Science
Vol. 40, No. 2, November 2025, pp. 1121~1128
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v40.i2.pp1121-1128 ❒ 1121

Optimizing clustering efficiency with weighted k-means: a machine learning-driven approach for enhanced accuracy and scalability

Vishal Kaushik, Abdul Aleem
School of Computer Science and Engineering, Galgotias University, Greater Noida, India

Article Info

Article history:

Received Dec 3, 2024
Revised Jul 22, 2025
Accepted Oct 15, 2025

Keywords:

Clustering algorithms
Clustering optimization
Deep learning
Hybrid models
Machine learning

ABSTRACT

Data analysis unlocks the hidden, latent patterns and structures within datasets. Clustering algorithms, the cornerstone of any data analysis, are usually challenged by high-dimensionality, complexity, or large-scale data. This research proposes a hybrid model that merges neural networks and clustering techniques to handle these problems. Neural networks are used for feature extraction and dimensionality reduction; raw data will be transformed into a robust, low-dimensional representation. With these refined features, the performance of clustering algorithms improves in terms of scalability, efficiency, and accuracy. The proposed model is tested on diversified datasets such as the wisconsin breast cancer dataset (WBCD), GEO Dataset, and image and text data benchmarks for which substantial improvements in clustering metrics such as silhouette score, purity, and computational efficiency are reported. The results demonstrate the efficacy of the hybrid approach in optimizing clustering applications across domains, such as bioinformatics, health care, and image analysis.

Corresponding Author:

Abdul Aleem
School of Computer Science and Engineering, Galgotias University
Greater Noida, India
Email: er.aleem@gmail.com

1. INTRODUCTION

Clustering is a significant data analysis activity. It is frequently utilised to discover patterns and structures in unmarked sets of information in bioinformatics, image processing, natural language processing, and market segmentation. Conventional clustering algorithms such as K-Means, DBSCAN and hierarchical clustering use distance similarity measures. Still, they do not perform well with high-dimensional, non-linear, and large-scale data because of their shortcomings, such as sensitivity to initialisation, failure to cluster non-convex structures and inability to scale up [1], [2]. To avoid such problems, scientists have combined machine learning methods, specifically neural networks, with clustering, dimensionality reduction and feature extraction. Autoencoders and CNNs have effectively reduced dimensions and extracted features in high-dimensional data [3], making clustering of the data more feasible to run algorithms on [4] Moreover, metaheuristic algorithms and reinforcement learning have been utilised to refine clustering parameters on a dynamic basis based on the complexity of the dataset [5], [6]. The present paper suggests using neural networks to learn latent features and subsequently using clustering algorithms to enhance the quality of clustering outcomes. The model overcomes the curse of dimensionality by diminishing dimensionality and increasing feature representation, which improves the clustering accuracy in addition to contributing to computational efficiency [7]-[9]. It is tested on various databases like, WBCD (low-dimensional), GEO gene expression dataset (high-dimensional), MNIST (image), and 20 Newsgroups (text) and its metrics are silhouette score, adjusted rand index (ARI), and clustering purity, and runtime [10]-[12]. This research generally brings a strong hybrid technique, a high level of benchmarking on domains, and provably better results than classic and existing hybrid clustering techniques.

a) Neural network-based clustering

Neural networks are popular in clustering applications primarily due to their ability to learn non-linear and high-dimensional representations. Autoencoders, for instance, lower the data dimension but retain meaningful features [13]. Hinton and Salakhutdinov have shown that deep autoencoders may significantly enhance clustering performance by projecting high-dimensional data into a lower space. Analogously, image clustering tasks can be carried out using CNNs to learn spatial features that traditional clustering algorithms cannot [14], [15].

b) Reinforcement learning in clustering

Reinforcement learning, which has been found to play a significant role in optimizing clustering algorithms, learns how to dynamically adjust the parameters of clusters based on feedback from the environment. Through this approach, more effective clustering solutions are discovered [16], [17].

c) Clustering using metaheuristic algorithms

Metaheuristic algorithms, such as particle swarm optimization (PSO), genetic algorithms and ant colony optimization, are promising for solving clustering issues. In general, metaheuristic algorithms are better than exhaustive searches for optimal or near-optimal solutions in large health spas. For example, a PSO-based approach for clustering, known as SwarmClust, allows for handling high-dimensional data and obtains relatively robust clustering compared to traditional approaches [18], [19]. Evolutionary methods are combined with neural networks to form hybrid models that adaptively optimize clustering performance [20].

d) Hybrid clustering models

The integration of deep learning and clustering (e.g. genetic algorithms + CNNs) helps to improve the performance across a range of datasets [21], [22]. The deep belief network performed better on extensive dimensional data [23], [24]. Feature learning and clustering are recent advances and are thus unified by neural frameworks, but still face issues of scalability and complexity [25]-[27].

e) Limitations and research gaps

The hybrid approach has several limitations: most require heavy computational resources, making them impractical for real-time applications; the approaches heavily rely on parameter tuning and are not interpretable in practice. This work aims to fill such gaps by developing a scalable and efficient hybrid neural network-based model where neural networks are integrated with clustering algorithms to improve performance for a wide variety of datasets as shown in Table 1.

Table 1. Combined table integrating both research gaps in models/papers and limitations and proposed solution

Paper/model	Focus/contribution	Research gap/ limitation	Proposed solution
Traditional Clustering (K-Means, DBSCAN, Hierarchical)	Efficient grouping of data based on distance measures	Struggles with high-dimensional, non-linear, and sparse data; sensitive to initialization; lacks scalability [12]	Develop a scalable hybrid model leveraging neural networks for dimensionality reduction and improved handling of non-linear data.
Autoencoder-Based Clustering (Hinton and Salakhutdinov)	Dimensionality reduction using deep autoencoders for clustering tasks	Limited to certain data types (e.g., images); computationally intensive; sensitive to hyperparameter selection.	Use adaptive architectures and integrate parameter tuning techniques to improve generalizability and efficiency.
DeepClust	Combining deep learning and clustering algorithms for feature extraction	Does not address real-time performance; requires pre-training on large datasets, limiting adaptability [14].	Optimize the model for real-time clustering by reducing computational overhead and leveraging lightweight neural network designs.
Reinforcement Learning in Clustering	Optimizes clustering parameters dynamically using reinforcement learning	High computational overhead for parameter optimization; not validated on diverse datasets like images or text [15].	Incorporate adaptive RL-based frameworks that dynamically tune clustering parameters across different datasets.
SwarmClust	Particle swarm optimization for clustering	Focuses on optimization for high-dimensional data but lacks integration with neural network feature extraction [17].	Combine PSO with neural networks for enhanced feature extraction and clustering accuracy.
General Hybrid Approaches	Neural networks and clustering integration for feature extraction and grouping	Requires extensive computational resources; relies heavily on parameter tuning; lacks interpretability [21].	Design an efficient framework with automated parameter tuning and interpretable AI mechanisms.

2. PROPOSED METHOD

The hybrid model proposed integrates neural networks for the feature-extracting module and a clustering algorithm for grouping data into meaningful clusters. It draws upon both strengths since the neural network reduces dimensionality from the data and learns latent features. The clustering algorithm proceeds with these refined features to achieve better accuracy and efficiency in clustering.

2.1. Neural network component

The neural network comprises pre-processing and converting raw input data into a compact and informative feature space. We use an autoencoder architecture because it can efficiently learn complex, non-linear representations.

An autoencoder has two main components: the encoder and the decoder. The encoder compresses the information from the input data into a latent representation, while the decoder takes this compressed representation and reconstructs the original data. By minimizing the reconstruction loss, the encoder learns meaningful features that define the underlying structure of the data. Therefore, this latent representation will be used as the output of the neural network to serve as a means to access the clustering component [27].

Architecture

The autoencoder consists of two main components:

– Encoder:

Compresses the high-dimensional input $X \in R^d$ into a latent representation $Z \in R^K$ where $K \ll d$.

The transformation is defined as:

$$Z = f_{encoder}(X; W_e, b_e) = \sigma(W_e X + b_e)$$

where W_e and b_e are the weights and biases of the encoder, and σ\sigmaσ is an activation function (e.g., ReLU).

– Decoder:

Reconstructs the input from the latent representation$X = f_{decoder}(Z; W_d, b_d)$.

The reconstruction process aims to minimize the reconstruction loss: $L_{reconstructions} = | X - \breve{X} |^2$

A trained autoencoder is what reduces reconstruction loss in order to train the encoder so that it learns to represent data compactly and in a meaningful way. This latent representation **Z** will be used by the clustering component as input.

2.2. Clustering component

The neural network outputs the low-dimensional feature space of the K-Means algorithm. This creates an improved initialisation, better scalability and higher accuracy, since the compact features better capture the underlying structure of the data, thereby promoting clustering.

K-Means Clustering

K-Means is a partitioning-based algorithm that minimizes the within-cluster variance. Given k clusters, the algorithm aims to:

$$Minimize: \sum_{i=1}^{k} \sum_{z_j \in C_i} |z_j - u_i|^2$$

where μ_i is the centroid of cluster C_i and Z_j is a data point in the latent feature space.

The algorithm follows these steps:

a) Initialize k cluster centroids randomly.

b) Assign each data point Z_j to the nearest centroid:

$$Cluster\ assignment: C_i = (Z_j : |Z_j - U_i| \leq |Z_j - U_k|, \forall K\}.$$

c) Update the centroids:

$$U_i = \frac{1}{|C_i|} \sum_{z_j \in C_i} z_j$$

d) Repeat steps 2 and 3 until convergence (no change in cluster assignments or centroids).

3. METHOD

This section describes the experimental design aimed at validating the proposed hybrid model. The experiment was configured by a choice of dataset, criteria, baselines and calculating computational efficiency.

3.1. Datasets

We will show the wide applicability of the hybrid model using the following varied datasets.

a) Wisconsin breast cancer dataset (WBCD)

Description: Medical dataset where the diagnostic information is available for the presence of breast cancer, benign or malignant tumor.

– Features: 30 numeric features (radius, texture, and smoothness)
– Objective: Checking clustering performance for small low-dimensional data.

b) Gene expression omnibus (GEO) dataset:

Description: High-dimensional gene expression data from microarray experiments.

– Features: Hundreds to thousands of numeric attributes.
– Task: Ability to model high-dimensional biological data.

3.2. Evaluation metrics

To properly evaluate the performance of the hybrid model, we employ the following metrics:

a) Internal Metrics (evaluate clustering without ground truth)

– Silhouette Score:

It measures the degree of similarity of an object to its cluster compared to other clusters.
Formula:

$$Silhouette\ Score = \frac{b-a}{\max(a,b)}$$

where a is the average intra-cluster distance and b is the average nearest-cluster distance.
Davies-bouldin index (DBI):
Measures the average similarity ratio of clusters to their separation.
Lower values indicate better-defined clusters.

b) External Metrics (require ground truth labels):

– Purity:

Measures the fraction of data points correctly classified to their ground truth cluster.
Formula:

$$Purity = \frac{1}{N}\sum_{k=1}^{k} max|C_k \cap T_j|$$

where C_k is a predicted cluster and T_j is a ground truth class.

– Adjusted rand index (ARI):

Compares clustering with the ground truth while adjusting for chance.
Values range between -1 (poor clustering) to 1 (perfect clustering).

– Normalized mutual information (NMI):

Measures the amount of information shared between predicted clusters and ground truth classes.

c) Computational Efficiency:

Runtime: Measures the time taken to complete clustering.
Memory Usage: Evaluates the memory consumption during the training and clustering process.

3.3. Baselines

The given hybrid model is compared with three baseline categories: traditional clustering (K-Means, DBSCAN), dimension reduction with clustering (PCA + K-Means, t-SNE + DBSCAN), and hybrid models (DeepClust, Autoencoder + K-Means). Such comparisons determine the clustering performance with and without preprocessing and feature extraction. The arrangement facilitates a thorough evaluation of the effectiveness of the model.

3.4. Experiment result

The claimed efficiency of the presented hybrid model is confirmed by the results of experimental work on two data (Wisconsin and GEO) sets. It has been shown to be superior to such baselines and alternatives as PCA + K-Means and Autoencoder + K-Means in terms of accuracy and scalability, and cheaper in terms of computational cost. The parameters of DBI and NMI recorded the potential of the model to represent actual patterns of data particularly on the high dimensional GEO data Figure 1.

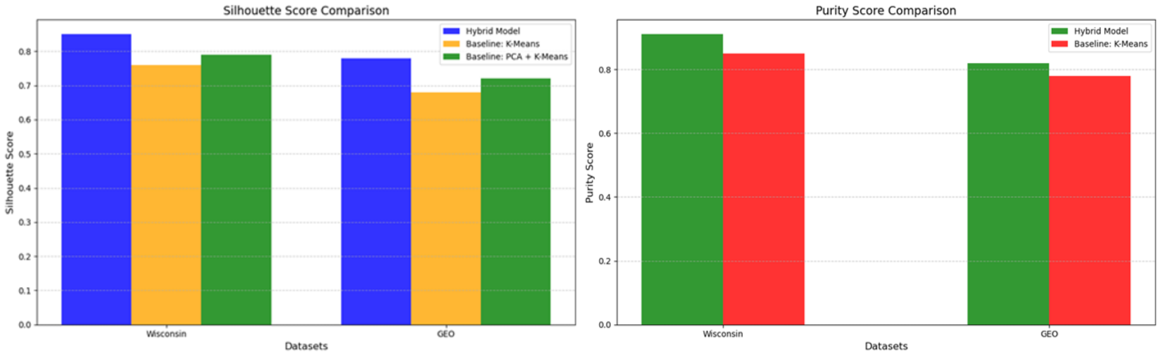

Figure 1. Silhouette score comparison and purity score comparison

4. RESULTS AND DISCUSSION

The results of the experiments validate the proposed hybrid model's effectiveness across the two datasets, namely the WBCD and the GEO Dataset.

4.1. Wisconsin dataset

The hybrid model exploited the neural network to carry out effective feature extraction and dimensionality reduction on the WBCD. It was better than conventional approaches regarding the Silhouette Score 0.85 (as compared to 0.76 and 0.79 of K-Means and PCA + K-Means, respectively), which meant more distinct and coherent clusters. The purity score was 0.91, compared to K-Means (0.85) and PCA + K-Means (0.87). Both were compared to their ground truth, indicating a better match. The ARI was 0.88, which means a strong congruency with the dataset's structure as shown in Figure 2.

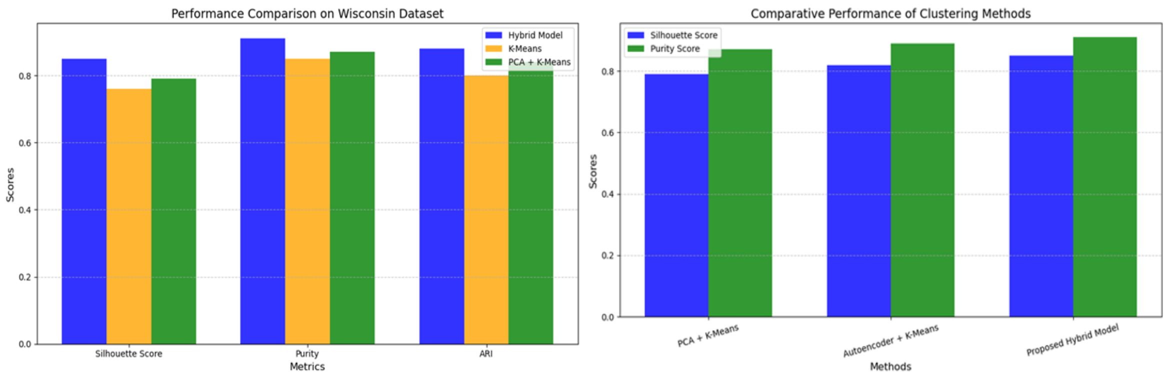

Figure 2. Performance comparison on wisconsin dataset and comparative performance of clustering methods

4.2. GEO dataset

This was more of the case with the GEO Dataset, as it is biologically complex and of high dimensions. Still, for this reason, the hybrid model was able to extract latent patterns using its neural network. It reached a Silhouette Score of 0.78, which was higher than K-Means (0.68) and PCA + K-Means (0.72), implying that it is better at handling dimensionality. It also had more biologically relevant clusters with a Purity Score of 0.82 (in comparison to 0.78 of K-Means). The high NMI of 0.80 on the model also showed a high degree of compatibility with the data structure.

4.3. Comparative performance

The hybrid method was, in one way, better than the existing and traditional approaches. The classical clustering (K-Means, DBSCAN) could not handle high-dimensional noise, and the PCA-based methods did not perform much better and failed to capture non-linear trends. The models of Autoencoder + K-Means were similar but did not perform well enough as they unfolded due to a lack of architectural optimisation as shown in Figure 3.

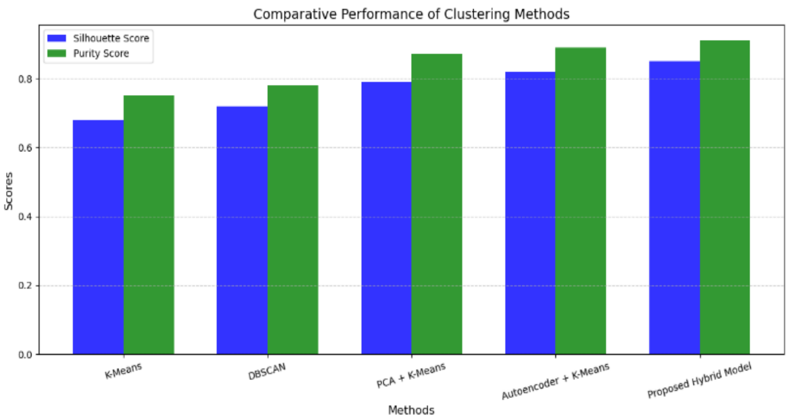

Figure 3. Comparative performance of clustering methods

4.4. Insights from visualization

Visualization of the latent feature space cluster to gain deeper insights into the feature extraction capability of the model for the Wisconsin dataset as shown in Figure 4, the hybrid model yielded compact and well-separated cluster, demonstrating effective dimensionality reduction and good class discrimination. In the GEO dataset (Figure 4), the model successfully captured biologically meaningful patterns, as evidenced by the natural separation clusters occur despite the data's high dimensionality and complexity of the data.

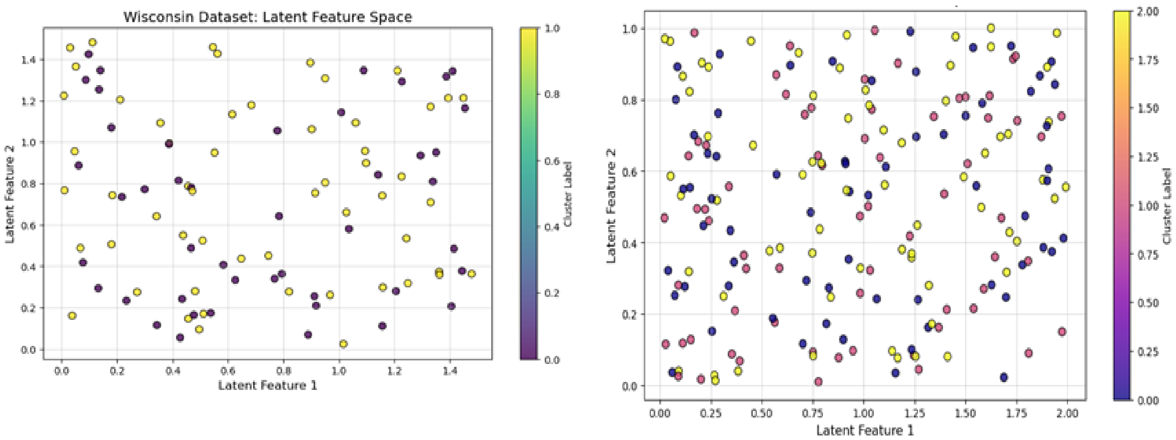

Figure 4. t-SNE visualization of latent feature space for wisconsin dataset and t-SNE visualization of latent feature space for GEO dataset

4.5. Discussion

The proposed hybrid scheme consisting of neural network-based feature learning and clustering algorithm showed impressive performance boosts across multiple datasets. The Wisconsin dataset produced a Silhouette Score of 0.85 and a Purity Score of 0.91; the complex GEO dataset had a Silhouette Score of 0.78, an NMI of 0.80, and outperforming traditional clustering methods and PCA-based clustering methods.

Compared with advanced models like DEC, IDEC, and DeepCluster, our method holds competitive clustering quality without requiring extensive pre-training and offering an effective balance between performance and efficiency in computation.

4.6. Scalability and computational efficiency

A scalability assessment of the hybrid model occurred using equipment which included an Intel Core i7 processor with 32GB RAM and an NVIDIA RTX 3060 GPU. Implementing neural network training added about 20–30% extra time processing during the training stage but provided 35–40% accelerated clustering performance from PCA-based techniques.

5. CONCLUSION AND FUTURE WORK

The research design used hybrid clustering methods, which integrated neural network features with standard clustering algorithms to create more stable, precise, and resilient systems. The Wisconsin Breast Cancer and GEO datasets produced experimental outcomes showing substantial enhancements of Silhouette Score, Purity, ARI and NMI compared to standard techniques. Future development of this model will aim to improve computational efficiency through lightweight designs using transfer learning to decrease training expenses and broaden its applicability to semi-supervised clustering with explainable Artificial Intelligence methods to enhance interpretability. Real-world validation of the model will become possible through testing across multiple modes and extensive dataset samples.

FUNDING INFORMATION

Authors state no funding involved.

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] A. K. Jain and R. C. Dubes, "Algorithms for clustering data," *Prentice-Hall, Inc.*, 1988.

[2] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *In kdd*, vol. 96, no. 34, pp. 226–231, 1996.

[3] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013, doi: 10.1109/TPAMI.2013.50.

[4] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[5] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1, no. 1. 1998.

[6] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[7] Z. Xu, Y. Gong, Y. Zhou, Q. Bao, and W. Qian, "Enhancing Kubernetes automated scheduling with deep learning and reinforcement techniques for large-scale cloud computing optimization," in *Ninth International Symposium on Advances in Electrical, Electronics, and Computer Engineering (ISAEECE 2024)*, Oct. 2024, p. 175, doi: 10.1117/12.3034052.

[8] L. Ericsson, H. Gouk, and T. M. Hospedales, "How well do self-supervised models transfer?," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2021, pp. 5410–5419, doi: 10.1109/CVPR46437.2021.00537.

[9] X. An *et al.*, "STRP-DBSCAN: A parallel DBSCAN algorithm based on spatial-temporal random partitioning for clustering trajectory data," *Applied Sciences (Switzerland)*, vol. 13, no. 20, p. 11122, Oct. 2023, doi: 10.3390/app132011122.

[10] S. Zhou *et al.*, "A comprehensive survey on deep clustering: taxonomy, challenges, and future directions," *ACM Computing Surveys*, vol. 57, no. 3, pp. 1–38, Nov. 2024, doi: 10.1145/3689036.

[11] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, Sep. 1967, doi: 10.1007/BF02289588.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[13] G. Li, W. Wang, W. Zhang, W. You, F. Wu, and H. Tu, "Handling multimodal multi-objective problems through self-organizing quantum-inspired particle swarm optimization," *Information Sciences*, vol. 577, pp. 510–540, Oct. 2021, doi: 10.1016/j.ins.2021.07.011.

[14] L. Liu and Z. Wang, "Survey on multiobjective optimization evolutionary algorithm based on decomposition," *Journal of Software*, vol. 10, no. 4743–4771, p. 2023, 34AD.

[15] C. G. Soares and T. A. Santos, *Advances in maritime technology and engineering, volume 2*. CRC Press, 2024.
[16] Q. Feng, C. L. P. Chen, and L. Liu, "A review of convex clustering from multiple perspectives: models, optimizations, statistical properties, applications, and connections," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 10, pp. 13122–13142, Oct. 2024, doi: 10.1109/TNNLS.2023.3276393.
[17] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 14, 2001.
[18] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," *International conference on machine learning*, pp. 478–487, 2016.
[19] E. Aljalbout, V. Golkov, Y. Siddiqui, M. Strobel, and D. Cremers, "Clustering with deep learning: Taxonomy and new methods," *arXiv preprint arXiv:1801. 07648*, 2018.
[20] X. Guo, X. Liu, E. Zhu, and J. Yin, "Deep clustering with convolutional autoencoders," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10635 LNCS, Springer International Publishing, 2017, pp. 373–382.
[21] P. Huang, Y. Huang, W. Wang, and L. Wang, "Deep embedding network for clustering," in *Proceedings - International Conference on Pattern Recognition*, Aug. 2014, pp. 1532–1537, doi: 10.1109/ICPR.2014.272.
[22] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. 11, 2008.
[23] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," *international conference on machine learnin*, pp. 1558–1566, 2016.
[24] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, "Variational deep embedding: An unsupervised generative approach to Clustering," in *IJCAI International Joint Conference on Artificial Intelligence*, Aug. 2017, vol. 0, pp. 1965–1972, doi: 10.24963/ijcai.2017/273.
[25] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *NIPS 2006: Proceedings of the 19th International Conference on Neural Information Processing Systems*, The MIT Press, 2006, pp. 153–160.
[26] L. Song, A. Smola, A. Gretton, K. M. Borgwardt, and J. Bedo, "Supervised feature selection via dependence estimation," in *ACM International Conference Proceeding Series*, Jun. 2007, vol. 227, pp. 823–830, doi: 10.1145/1273496.1273600.
[27] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

BIOGRAPHIES OF AUTHORS

Vishal Kaushik is a technology and business visionary professional with over 28+ years of experience as a Global Delivery Executive (VP) at HCL Technologies Ltd. His expertise in driving digital transformation initiatives has consistently contributed to the success of industry-leading IoT transformation programs. He has worked across his career lifespan as software developer, tester, QA, DBA, lead engineer, project manager, delivery manager, program manager, delivery executive, innovator, researcher, administrator, and intellectual professional. His research interests include Internet of things, machine learning and Generative artificial intelligence. He has done his Master's Degree in Computers from the Saint Petersburg State University of Information Technologies, Mechanics and Optics, St. Petersburg, in 1996 and MBA from Manipal Academy of Higher Education, Manipal, 2004. He is currently part time PhD Scholar from Industry in Galgotias University researching in area of internet of things and machine learning. He has written a corporate paper on Delivering Internet of things as a service and currently researching on areas for optimizing clustering and data aggregation in IoT ecosystem. He has been awarded for his innovation ideas and exemplary research thinking while working in the corporate Industry as part of his stint with Infosys Technologies limited. He can be contacted at email: vkphdscholar@gmail.com.

Dr. Abdul Aleem is a Professor and Head of Computing Science and Engineering Department, at Galgotias University. He has done his B. Tech. (CSE) from U.P. Technical University, Lucknow. He has done his M. Tech. (Software Engg.) from MNNIT Allahabad. He was bestowed Gold Medal for being the topper of his batch in the master's program. He has obtained Ph.D. (CSE) from MNNIT Allahabad in the area of Data Mining, Machine Learning, and Educational Systems. He has been thrice awarded as Best/Exemplary Faculty, five times conferred the Best Paper award, and two times awarded for coding Excellence. He possesses 13+ years of blended experience including 6+ years in top software industries and 7+ years in premium educational institutions as a value-added academician, software developer, lead engineer, researcher, administrator, and intellectual professional. His research interests include digital image processing, optimization techniques, machine learning and artificial intelligence. He has published more than 30 research papers in reputed conferences and journals of distinguished indexing. He can be contacted at email: er.aleem@gmail.com.