OPERATIONS AND SUPPLY CHAIN MANAGEMENT
Vol. 17, No. 2, 2024, pp. 236 - 252
ISSN 1979-3561 | EISSN 2579-9363

Optimizing Procurement Strategies for Diverse Product Segments: A Case Study in Pharmaceutical Supply Chain Management

Kaoutar Douaioui
Faculty of Sciences and Techniques,
Laboratoire Ingénierie Laboratory of Industrial Engineering, Management and Innovation (LIMII),
Settat, Morocco
Email: Kaoutar.douaioui@gmail.com (*Corresponding Author*)

Rachid Oucheikh
Physical Geography and Ecosystem Science,
Lund University, Sweden
Email: rachid.oucheikh@nateko.lu.se

Othmane Benmoussa
Euromed University of Fes, UEMF, Morocco
Email: o.benmoussa@ueuromed.org

ABSTRACT

Selecting the most suitable procurement strategy is crucial to the efficient management of supply chain operations and the prevention of stock shortages. Nevertheless, when dealing with a wide variety of products, this task becomes an intricate challenge. While traditional and advanced procurement tools are available, applying them across such diverse product ranges is often impractical. This research is dedicated to determining distinct procurement strategies tailored to each product cluster. These strategies will be designed to accommodate the technical and financial constraints specific to each cluster. To address the optimization challenges associated with clustering algorithms, especially within complex search spaces, metaheuristic algorithms are considered as promising solutions. In this paper, Accelerated Particle Swarm Optimization (APSO) is harnessed for its exploratory capabilities, and Teaching Learning Based Algorithms (TLBO) are leveraged for their high exploitation competence. This innovative approach effectively combines the strengths of both algorithms, ensuring optimal clustering solutions in an efficient manner. The suggested approach outperforms the accuracy of the well-known metaheuristics including Grey Wolf Optimizer and the Whale Optimization Algorithm. This methodology successfully identifies five major clusters and assigns the appropriate procurement strategy to each cluster. The selection of a suitable procurement strategy for each product cluster significantly enhances overall procurement performance. This study introduces a powerful approach to assist managers in adapting procurement strategies for different product clusters. This approach has been implemented within organizations specializing in pharmaceutical freight and holds potential applicability across various product types. This innovation has the capacity to significantly impact and enhance global procurement performance.

Keywords: *clustering algorithms, comparative analysis, data mining, intra-cluster distances, metaheuristic algorithms, optimization, procurement, supply chain management*

1. INTRODUCTION

In the contemporary interconnected global economic landscape (Soledispa-Canarte *et al.*, 2023), procurement in supply chain stands at the center of efficient and effective business operations. This multifaceted domain encompasses the intricate processes and strategies integral to the sourcing, acquiring, and managing the goods, services, and resources essential for seamless functioning of organizations (Gurrala & Hariga, 2022). Amid its pervasive impact across diverse industries, the pharmaceutical sector emerges as a paramount exemplar wherein procurement assumes unparalleled significance (Seddigh *et al.*, 2022). Within the pharmaceutical industry, the procurement of supplies plays a critical role in the context of global healthcare systems, ensuring the constant availability of vital medications, vaccines, and healthcare products. The significance of a well-functioning pharmaceutical procurement cannot be overstated and extends beyond the scope of business and economics.

However, procurement in pharmaceutical supply chain faced several challenges (Sodhi & Tang, 2021) that impact its effectiveness and efficiency. The first challenge is about Global Sourcing for pharmaceuticals products (Ghadge *et al.*, 2023). Pharmaceutical companies often rely on a global network of suppliers for raw materials, active pharmaceutical ingredients, and finished products. Sourcing from multiple countries introduces supply chain vulnerabilities, including geopolitical risks (Fri *et al.*, 2021a), trade conflicts, and disruptions arising from natural disasters or political instability. Ensuring a diversified and resilient procurement strategy becomes a primordial priority for pharmaceutical supply chain. The second challenge is about the Cold Chain Management. Many pharmaceutical products, especially vaccines, biologics, and certain medications, are temperature-sensitive. Maintaining the cold chain throughout procurement and distribution is essential to prevent product degradation or spoilage. The need for specialized cold storage facilities and transportation adds complexity and cost to the supply chain (Fri *et al.*, 2020). The third challenge is about Regulatory Compliance Complexity (Fri *et al.*, 2020). Globally, the pharmaceutical industry is among the most heavily regulated sectors; procurement teams must face the multitude and complex regulations, including Good Manufacturing Practices (GMP), Good Distribution Practices (GDP), and robust quality standards. Making sure that every supplier follows these regulations can be a complicated and time-consuming task, and it can affect the procurement efficiency. The fourth challenge is about Patient-Centric Distribution Models (Bassand *et al.*, 2022). The trend towards patient-centric care requires innovative distribution models, including direct-to-patient deliveries and specialty pharmacy networks. These models aim to meet individual patient needs while ensuring the integrity and security of the supply chain.

The pharmaceutical supply chain must effectively deal with these challenges that are inherent in its procurement operations. Choosing the appropriate tactic relies on numerous factors that include the attributes of the products being considered and the specific circumstances and criteria that govern the purchasing procedure. Consequently, the key question emerges: How can we formulate procurement strategies that enhance the acquisition of healthcare products within the pharmaceutical supply chain, given the complexities of a globalized environment?

Within this context, clustering techniques have emerged as indispensable tools, benefiting not only procurement but various supply chain domains (Liang *et al.*, 2022). The surge in clustering challenges has necessitated the creation of algorithms designed to tackle multidimensional problems. Nevertheless, present-day clustering algorithms face limitations such as sensitivity to initial values, computational complexity, and the risk of stagnating in a local optimum. To overcome these obstacles, metaheuristic algorithms have emerged as powerful contenders capable of discovering global solutions. Amid these challenges, the No-Free Lunch (NFL) theorem underscores the importance of aligning algorithmic capabilities with data and problem structures. As the complexity of clustering within supply chain management unfolds, it emphasizes the need for researchers to leverage domain-specific insights and a nuanced understanding of problems when selecting algorithms.

The structure of this paper is organized as follows: The first section explores the conceptual framework by elucidating the background of clustering techniques that harness metaheuristics. Additionally, a brief overview of Teaching Learning Based Algorithms (TLBO) and APSO is provided to establish the groundwork for the proposed approach. Subsequently, an in-depth presentation of the novel clustering methodology is presented and validated using multiple benchmark datasets with varying complexity and features. This is followed by meticulous details on the experimental setup and comprehensive comparative results emphasizing the relevance of the proposed method to procurement challenges. Finally, the paper concludes with remarks and outlines potential perspectives for future research, particularly in the context of procurement within supply chain management.

2. LITERATURE REVIEW

This section delves into the theoretical foundations and provides a succinct overview of the prior research conducted on to clustering methodologies, particularly those employing metaheuristic approaches. The examination extends from a broad perspective to a more specific focus within the domain of supply chain management.

2.1 Clustering

Clustering is a fundamental technique in data analysis that groups similar data points together based on defined criteria (Darbanian *et al.*, 2024). The dataset undergoes segmentation into clusters with the objective of ensuring that data objects within a given cluster exhibit greater similarity compared to those in different clusters. In other words, the process of grouping data aims to minimize the intra-cluster distance among data objects within the same cluster, enhancing their similarity, while maximizing the inter-cluster distance between different clusters, reflecting dissimilarity. Data clustering allows discovering underlying patterns, relationships, and structures within data, ultimately facilitating a deeper understanding of complex datasets. Across diverse domains, clustering emerges as a critical tool, notably within supply chain management, where it empowers judicious, data-driven decision-making processes (Ezugwu, 2020).

Clustering is a pivotal facet of data mining with widespread applications across diverse domains (Ezugwu, 2020). Clustering plays a crucial role in anomaly detection, assisting in cybersecurity, fraud detection, and quality control by identifying abnormal patterns within datasets (Oucheikh *et al.*, 2020, 2021). In image analysis and computer vision, clustering facilitates image segmentation, allowing for the organization of complex visual data and tasks such as object recognition (Z. Wang *et al.*, 2020). Biomedical research benefits from clustering by categorizing biological data to uncover disease patterns and potential biomarkers (Halner & Bafadhel, s. d.). Document classification, network analysis, spatial data analysis, and pattern recognition are additional areas where clustering contributes to efficient information organization, community detection, spatial pattern identification, and various machine learning applications (Ezugwu *et al.*, 2022). Overall, clustering's ability to reveal hidden patterns and structures in data makes it indispensable for informed decision-making across a broad spectrum of disciplines.

2.2 Clustering in Supply Chain Management

Clustering is imperative within the field of supply chain management, chiefly owing to the dynamic evolution of operational processes and the substantial magnitude of associated datasets. The utilization of clustering methodologies serves as a resilient technological solution, affording adaptability to effectively manage and distill actionable insights from expansive datasets, thereby enabling adept decision-making amidst the nuanced challenges posed by the evolving supply chain landscape.

Literature showcases several applications of clustering techniques in supply chain management. Its significance is evident in customer segmentation, where it helps businesses in tailoring marketing strategies and enhancing customer satisfaction by categorizing customers based on discernible behaviors and preferences (Gichuru & Limiri, 2017). The work (Lahtinen, 2021) aimed to group suppliers based on appropriate features of supplier logistics in order to reduce shipment costs. The outcome of this work can help buyers select cost-effective suppliers for their business requirements. Clustering is also applied in production planning. A volatility clustering algorithm is developed to optimize the production plan of a manufacturing supply chain and reduce uncertainty in the value of a Central Bank Digital Currency (Ding *et al.*, 2022). The proposed solution outperforms GARCH model in volatility clustering. In the distribution phase, a paper proposes a fuzzy C-means clustering algorithm to determine the priority level of emergency material distribution, which classify various emergency materials required in disaster areas and determine the efficiency of each emergency supplies (Hong *et al.*, 2020). Simulation experiments are developed in order to highlight the feasibility of solution in emergency supply materials scheduling. In response to the increasing importance of customer churn management as a competitive advantage for businesses and the need to efficiently handle large-scale industry data, some new techniques based on clustering are proposed. A clustering algorithm named the Semantic-Driven Subtractive Clustering Method (SDSCM) has been introduced to enhance the clustering accuracy of SCM and K-means but also to mitigate the risk of imprecise operations management using axiomatic fuzzy sets (Ghosal *et al.*, 2020).

Clustering is also used for supply chain risk evaluation and mitigation (M. Wang *et al.*, 2023). In the articles (Rana & Cheok, s. d.) and (D. Wang *et al.*, 2022) the authors proposed a risk evaluation and prediction model employing K-means clustering and LightGBM classifier. Using a dataset from an equipment group, the model categorizes supply chain risks into four types, primarily emphasizing market environment issues and resource constraints, demonstrating the efficacy of the approach. In a related research topic, K-Means, FCM, and SUB models are used to cluster bank customers and analyze their behavioral patterns for supply chain finance, revealing that creditworthiness, education, job, collateral value, collateral type, loan term, and age are the most influential factors (Nazari *et al.*, 2019). Ultimately, K-Means is identified as the most suitable clustering technique among the evaluated methods. Moreover, certain papers explore hybrid approaches by combining clustering algorithms with other methods, as demonstrated in (Gálvez *et al.*, 2020), which introduces a novel framework integrating clustering through K-means and fuzzy-AHP. The suggested solution aims to enhance the performance of the supply chain, particularly in industries with numerous suppliers necessitating development programs. Another study employs a K-means clustering model on environmental data from mainland Chinese provinces to determine the optimal location for a supply chain center of natural resources (D. Chen *et al.*, 2022). The workflow of the approach involves clustering provinces, examining global environmental protection trends, and seeking similarities with the Sihanba ecological site. Generally applying clustering can provide a multitude benefit for supply chain management such as the reduction of different costs, choosing the appropriate strategy or classify product, suppliers or costumers. Overall, applying clustering in supply chain management can assist the manager in decision making and contribute to the efficiency and effectiveness of industries. The application of clustering in supply chain management offers numerous benefits, including cost reduction, sustainability, adaptation to changing dynamics, and the classification of products, suppliers, or customers. Overall, the integration of clustering in SCM supports managers in decision-making processes, contributing to the overall efficiency and effectiveness of industries (Oucheikh *et al.*, 2022).

2.3 Categories of Clustering Methods

In data mining, clustering encompasses various techniques for grouping akin data points together. One common approach is hierarchical clustering, which arranges data into a tree-like structure, either by merging clusters in an agglomerative manner or by recursively dividing them in a divisive fashion. Another widely used method is partitioning clustering, such as K-means algorithm, which divides data into non-overlapping subsets. They can be either hard with unique membership for each datapoints or fuzzy by allowing data points to belong to multiple clusters with varying degrees of membership (Douaioui *et al.*, 2021, 2022). Density-based clustering, typified by DBSCAN, identifies clusters based on areas of higher data point density, while grid-based clustering streamlines the process through the imposition of a structured grid organization on the data. Model-based clustering involves statistical models to describe cluster distributions. Another categorization is centroid-based clustering that assigns data points to the nearest centroids, and subspace clustering that identifies clusters in different subspaces of the data.

The selection of a clustering algorithm relies on various factors such as the inherent attributes of the dataset and the specific objectives of the analysis. Hierarchical clustering is particularly suitable for learning hierarchical relationships, while density-based methods are more effective in identification of clusters with heterogeneous shapes and sizes (Srinivasan & Balaji, 2020). Model-based clustering is advantageous in scenarios where underlying statistical models are presumed to govern cluster formations. But, in general, each clustering paradigm offers distinctive insights into patterns within the data, affording analysts the flexibility to choose the most appropriate method contingent on the nuanced characteristics of the dataset and the overarching goals of the analysis (Azadeh *et al.*, 2012). A recent paper reviews the full taxonomy and explains various techniques under each category (Ezugwu *et al.*, 2022).

2.4 Clustering Validity Indices: Insights and Equations

Clustering validity indices establish a rigorous framework for assessing the quality of clustering solutions, offering essential insights into the optimization of clustering algorithms. These mathematical functions serve a dual purpose: determining the optimal number of clusters and evaluating the partition quality of data. Through iterative executions of clustering algorithms with varying cluster counts, the partitioning that yields the highest validity measure is deemed the optimal solution, reinforcing the robustness and reliability of obtained results (Smith et Johnson, 2020).

Clustering validity indices can be categorized into three distinct types: internal, external, and relative indices (Ullmann *et al.*, 2022). Internal indices evaluate clustering quality solely by analyzing inherent data attributes. On the other hand, external indices incorporate external information for a more comprehensive evaluation, bringing in additional context for validation. Lastly, relative indices play a crucial role in comparing the merits of different clustering solutions, assisting in the identification of superior alternatives. These categories represent diverse approaches, each offering valuable perspectives to enhance our understanding and assessment of clustering outcomes. These cutting-edge indices intelligently balance the precision of partitioning evaluation with a deep understanding of the underlying data structures, thereby transcending traditional boundaries and opening avenues for more adaptive and insightful clustering analyses.

Clustering validity indices play a pivotal role in discerning the ideal cluster count for a given dataset and appraising the efficacy of clustering algorithms across diverse scenarios. The Silhouette coefficient (Bagirov *et al.*, 2023), Dunn index (Gagolewski *et al.*, 2021), and Calinski-Harabasz index (Gagolewski *et al.*, 2021) stand out as widely employed measures. These indices measure the similarity or dissimilarity between data points within and across clusters, providing a valuable framework for evaluating clustering results within supply chain management and beyond (Arbelaitz *et al.*, 2013).

The Silhouette coefficient (Dinh *et al.*, 2019) can be calculated using the following formula:

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{1}$$

Where $a(i)$ represents the average distance between the data point i and other data points in the same cluster, and $b(i)$ represents the smallest average distance between the data point i and data points in different clusters.

The Dunn index (Gagolewski *et al.*, 2021) is formulated as Eq (2):

$$D = \frac{\underset{1 \le i \le k, i \ne j}{min} \delta(C_i, C_j)}{\underset{1 \le l \le k}{max} \operatorname{diam}(C_l)} \tag{2}$$

Where C_i represents cluster k,i is the number of clusters, $\delta(C_i,C_j)$ is the distance between clusters C_i and C_j, and diam (C_l) is the diameter of cluster C_l. The Calinski-Harabasz index, also known as the Variance Ratio Criterion (Gagolewski *et al.*, 2021) , is defined as Eq (3):

$$CH = \frac{B_k}{W_k} \times \frac{N-k}{k-1} \tag{3}$$

Where B_k represents the between-cluster dispersion, W_k represents the within-cluster dispersion, N is the total number of data points, and k is the number of clusters. These mathematical formulations encapsulate the essence of clustering validity indices, allowing the quantitative evaluation of clustering solutions, and play a pivotal role in gauging the efficacy of various algorithms within supply chain management and beyond.

In fact, the clustering problem can be formulated as an optimization problem that lends itself to solution by both multi-objective and single-objective metaheuristics. Most notably, these approaches leverage internal validity indices as objective functions for optimization. The aim is to guide the formation of clusters with low intragroup dissimilarity and high intergroup similarity, without requiring external information (Gagolewski *et al.*, 2021). empirically evaluated a range of internal clustering validation indices using diverse datasets, comparing the behavior of different indices. Their findings affirm that there is no universally “perfect” index; instead, the choice of index should be tailored to the dataset’s characteristics. A multitude of internal validity indices exists in the literature, each expressing the clustering objective from different angles, such as centroid distance, Silhouette coefficient, intra and inter-cluster distances, variance ratio criterion, and more (Gagolewski *et al.*, 2021).

Additionally, there is a category of external indices that relies on ground-truth data for evaluation. These indices encompass measures such as Rand index, Jaccard coefficient, Mallows index, and Fowlkes index (X. Wang & Xu, 2019). They involve cluster-level assessments and the counting of paired data points’ co-location within or across clusters, offering a holistic perspective on the performance of clustering solutions.

In our paper, we consider partitioning-based clustering formulated as an optimization problem. This problem is classified as an NP-hard problem due to its inherent computational complexity that arises from the challenge of determining an optimal partition of a dataset into a predefined number of clusters, with the objective of optimizing a specified criterion (Ariyaratne & Fernando, 2023). The exponential growth in potential partitions, influenced by dataset size and cluster number, makes exhaustive search impracticality. By reducing the problem to NP-hard problems like the Traveling Salesman Problem, it is proven that achieving an optimal solution for partitioning-based clustering is at least as challenging as solving the reduced NP-hard problem. Therefore, heuristic and approximation algorithms are often employed to find satisfactory solutions within reasonable computational time for large datasets. In addition, the algorithm is executed multiple times with different initial partitions and the optimal solution is the clustering output that achieves the best performance across all runs.

Upon framing the issue as an optimization problem, a spectrum of techniques becomes available for resolution, among which nature-inspired methods take precedence in the context of this paper. These methods draw inspiration from principles observed in natural systems and phenomena to develop algorithms for data clustering. They mimic the adaptive and self-organizing behaviors found in nature to solve complex optimization problems. For example, Ant Colony Optimization algorithm is inspired by ant foraging behavior, employ artificial ants guided by pheromone trails to iteratively explore and refine paths in the solution space based on accumulated knowledge of successful solutions. On the other hand, Genetic Algorithm (GA) is inspired from natural selection entailing the evolution of a population of potential solutions, where chromosomes representing solutions undergo crossover and mutation operations, and the fittest solutions are selectively passed on to the next generation. A recent paper includes a comprehensive list of the techniques along with their descriptions (Too & Abdullah, 2021).

The technique may also be hybrid, capitalizing on the strengths of two or more techniques. An instance of this hybrid approach involves leveraging the eagle strategy to synergistically integrate the exploratory capabilities facilitated by Lévy Flight with the exploitation prowess inherent in the Slime Mould Algorithm (Oucheikh *et al.*, 2021, 2022).

3. MATERIAL AND METHODS

This section introduces the methodologies employed to address the clustering challenges within the context of supply chain management. The proposed approach harnesses the combined capabilities of Teaching Learning Based Algorithms (TLBO) and Accelerated Particle Swarm Optimization with Negatively Correlated Search (APSO-NCS) to optimize data clustering more efficiently and effectively. The strategy involves a two-step process: first, TLBO is executed for a specific number of generations, and then APSO is applied directly to the solutions generated by TLBO without undergoing fitness evaluation. This sequential approach leverages the exploratory strengths of TLBO and the inherent exploitation abilities of APSO. However, to prevent premature convergence or suboptimal outcomes, NCS is conditionally used after APSO by leveraging the information about the diversity within the population. If the diversity measure falls below a predefined threshold, we will implement the NCS. Otherwise, if the diversity measure surpasses this threshold, we will directly calculate the fitness function, assess the quality of the solution, and move on to the subsequent round. This approach ensures the discovery of superior and diverse solutions. In the next paragraphs, we elaborate on the three methods employed in our approach and elucidate their role in enhancing the optimization of clustering metrics.

3.1 Teaching Learning-Based Optimization

Teaching Learning-Based Optimization (TLBO) is a population-based optimization algorithm that draws inspiration from the pedagogical process observed in educational settings (Ma *et al.*, 2024). In TLBO, optimization is conceptualized as a classroom scenario, where the exchange of knowledge occurs akin to a teacher instructing a group of students see **Figure 1**. It operates through the following phases:

- Teaching Phase
  Analogous to a knowledgeable instructor, a teacher solution is identified, representing the best solution found so far. During the teaching phase, the teacher imparts wisdom to the entire class of potential solutions in the population. This knowledge sharing is directed toward guiding the class toward improved solutions.
- Learning Phase
  In the learning phase, individual solutions within the population, referred to as students, learn from both the teacher and their peers. They adjust their own solutions based on the information gleaned from the teacher and interactions with fellow students.
- Iterative Refinement
  TLBO iteratively refines the solutions within the population. The teacher and students collaborate to enhance the overall performance of the class. Over successive iterations, the population strives to converge towards an optimal or near-optimal solution to the optimization problem at hand.

The steps of the algorithm TLBO

Step 1: Initialize Population
Create an initial population of candidate solutions randomly.
Step 2: Evaluate Fitness
Evaluate the fitness of each candidate solution using an objective function.
Step 3: Calculate Mean
Calculate the mean value of the population.
Step 4: Teacher Phase
Identify the best solution in the population and update the other solutions by moving them towards this best solution.
Step 5: Learner Phase
Identify the worst solution in the population and update it by moving it towards the mean value of the population.
Step 6: Update Population
Replace the worst solution in the population with the updated solution from the learner phase.
Step 7: Repeat
Repeat steps 3 to 6 until a stopping criterion is met, such as a maximum number of iterations or reaching a satisfactory solution.

Figure 1 The steps of the algorithm TLBO

3.2 Accelerated Particle Swarm Optimization

Teaching Learning-Based Optimization (TLBO) is a population-based optimization algorithm that draws inspiration from the pedagogical process observed in educational settings (Ma *et al.*, 2024). In TLBO, optimization is conceptualized as a classroom scenario, where the exchange of knowledge occurs akin to a teacher instructing a group of students see **Figure 1**. It operates through the following phases:

Particle Swarm Optimization (PSO) is a computational optimization algorithm inspired by swarm behavior. It uses a population of particles to explore a multidimensional search space and find optimal or near-optimal solutions. Particles adjust their positions based on their experiences and information from neighbors. Each particle adjusts its position in the search space based on its own experience, the experience of its neighbors, and the best-known solution found by the entire swarm.

The swarm is defined as a set of n particles $P = \{p_1, p_2, \dots, p_n\}$, where each particle is considered as candidate solution with D dimensions. At each iteration t, each particle has a position vector x_i^t and a velocity vector v_i^t.

Velocity updates in PSO are determined by a combination of factors, including their prior velocities, individual learning, and information shared by neighboring particles. The use of the gained memory and interplay between personal and collective learning helps particles navigate the search space effectively. For this reason, we record at each iteration the best position of each particle i, noted the p_{best} and the best position found by the whole swarm, noted g_{best} , see Eq (4):

$$v_i^{t+1} = wv_i^t + c_1 r_1 (p_{best}^t - x_i^t) + c_2 r_2 (g_{best}^t - x_i^t) \quad (4)$$

Where w is the weight of inertia, r_1 and r_2 are random values derived from N(0,1) which are used to maintain the algorithm diversity; c1 and c2 are acceleration constants, which are used to push the swarm towards the local and global best position and usually tuned through experience.

In 2009, Yang developed an accelerated particle swarm optimization (APSO) (Yang *et al.*, 2009) based on the traditional PSO. It takes a different approach by utilizing only the global best position (g_{best}) for velocity updates. This can be expressed using the following formula Eq (5):

$$v_i^{t+1} = v_i^t + \alpha r + \beta\left(g_{best} - x_i^t\right) \quad (5)$$

Where r is a random number in the interval [0,1], and the parameter is chosen from the interval [0.1D, 0.5D], and β is derived from the interval [0.2, 0.7]. The equation essentially states that a particle's new velocity v_i^{t+1} is determined by combining three factors: The particle's previous velocity, which allows the particle to retain some of its previous momentum. The weighted cognitive information αq. The weighted social information $\beta(g_{best} - x_i^t)$, which guides the particle toward the best solution discovered by any particle in the entire swarm. So, the update of the particle position in a subsequent iteration *t+1* can be written as Eq (6):

$$x_i^{t+1} = (1 - \beta)x_i^t + \beta\ g_{best} + \alpha r \quad (6)$$

This version gives a higher order of convergence than in PSO. It is worth pointing out that the velocity does not appear in Eq. (6), and thus there is no need to deal with the initialization of velocity vectors. For that, the APSO is very easy to implement. The random term αr provides the ability to avoid any local optima if a is chosen to be consistent with the scales of the system, while r can be provided from a probability distribution such as Gaussian.

Comparing with many PSO variants, the mechanism of the APSO is simple to understand and uses only two parameters (α and β), This means that a monotonically decreasing function can be used. In our implementation, so we use $\alpha = \delta^t$.

Where:

- $0 < \delta < 1$ is an annealing-like parameter whose value can be taken as 0.7 to 0.9 for most cases.
- $t \in [0, t_{max}]$ such that t_{max} is the maximum of iterations.

Finally, convergence analysis is an important aspect showing the quality of any search algorithm.

APSO is proven to have great ability to converge rapidly to optimal or near-optimal solutions. Hence, we strategically deploy it immediately follows TLBO to extensively leverage its acquired search results. So, it serves as a key component for search exploitation in the approach. However, judicious use is essential to avoid local optimum traps. Hence, we combine it with NCS, to maintain diversity in the search space. This synergy ensures faster convergence to high-quality solutions, safeguarding against premature convergence to suboptimal solutions.

3.3 *Negatively Correlated Search*

Negatively Correlated Search (NCS) is used in our method to maintain an optimal level of diversity among the population. In fact, maintaining diversity among candidate solutions is essential to avoid premature convergence to suboptimal results. As solutions evolve to achieve higher fitness levels, a concomitant reduction in their diversity is observed, underscoring the intricate trade-off between quality and variability inherent in optimization processes. So, the key idea to avoid premature convergence is to ensure a negative correlation between fitness and diversity.

This study leverages the NCS algorithm to explicitly encourage diverse search behaviors by accentuating differences among probability distributions (Tang *et al.*, 2016). The NCS algorithm introduces a novel model for promoting cooperation among individuals within a population, drawing inspiration from cooperative human behaviors (H. Chen *et al.*, 2018). When facing complex tasks, teams often collaborate by addressing different aspects and communicating to prevent redundancy. Analogously, NCS consists of multiple search processes running cooperatively, aiming to discover improved candidate solutions. The cooperation mechanism of NCS can be perceived as a preservation strategy that directly fosters diversity at the level of search behaviors (Liu & Yao, 1999). Empirical investigations underscore NCS's competitiveness against established search methods, demonstrating superior performance across 20 multimodal continuous optimization problems (Tang *et al.*, 2016). Notably, NCS explicitly favors solutions that contribute to diversity while holding the potential to yield favorable outcomes (Bali & Bebok, 2015).

To quantify the diversity index of the population, we use the following formula Eq (7):

$$Diversity\ (D) = \frac{1}{n\,D} \sum_{d=1}^{D} \sqrt{\sum_{i=1}^{n} (x_i^d - X^d)^2} \quad (7)$$

Where n is the population size, D is the dimension of the search space, x_i^d represents the ith particle value in the dth dimension and X^j denotes the average of all particle values within that specific dimension. If the computed diversity index is less than a predefined threshold, then the update rule that should be applied is the following Eq (8):

$$\theta_i(t+1) = \theta_i(t) + \alpha \cdot \beta \cdot sign(Z) \cdot sign\left(\theta_j(t) - \theta_i(t)\right) \quad (8)$$

Where $\theta_i(t)$ and $\theta_j(t)$ represent the solutions of individuals i and j at iteration t, Z denotes a standard normal random variable, α and β are scaling factors, and sign() extracts the sign (positive or negative) of a given number. This equation contributes to the selection of individuals with lower fitness and higher diversity by introducing two key elements: randomness expressed by Z and a diversity-preserving mechanism expressed by the relative positioning of individuals i and j in the search space, i.e., repulsion.

By using the new update rule, NCS dynamically adjusts particle behavior, enabling the swarm to navigate complex solution spaces more effectively. Unlike schemes reliant on intricate coordination among diverse search operators, NCS achieves diversity preservation intrinsically.

Combining APSO with the NCS technique offers several advantages over using APSO alone see **Figure 2**. The combination enhances exploration and exploitation balance, maintains diversity, and improves robustness. It helps avoid premature convergence to local optima, making it more effective for complex optimization problems. While it may introduce some additional computational complexity, the potential benefits in optimization quality make it a valuable choice for challenging scenarios.

The steps of APSO-NCS

Step 1: Initialization
Initialize population
Assign random or predefined initial positions and velocities to individuals
Step 2: Fitness Computation
Evaluate fitness for each individual
Repeat until termination condition met:
Step 3: Update Position
Update position based on current position, velocity, and best-known solutions
Update pheromone levels on paths based on solution quality
Step 4: Diversity Index (DI) Computation
Monitor and promote diversity in the swarm
Step 5: if DI is lower than a threshold
Apply NCS
Step 6: Repeat iteratively the previous steps
Check if termination condition is met
Common conditions include max iterations, target fitness, or time limit
Step 7: Termination
Return best solution found

Figure 2 The steps of APSO-NCS

3.4 *Proposed Method for Procurement Data Clustering in Scm*

Formulating the updating equation for individuals is a crucial step in the development of swarm-based nature-inspired algorithms, as it determines how individuals explore and construct new solutions. The updating equation varies between different algorithms, and it is a defining characteristic of each swarm-based algorithm. The updating equation typically involves the selection of the best candidate solutions, such as in the bat algorithm (Yang, 2010) or the whale optimization algorithm (Mirjalili *et al.*, 2014). In some algorithms, historical best positions may also play a role, as in TLBO algorithm. Additionally, in larger swarms, multiple best candidates may be introduced to guide the individuals during the iterations. By carefully selecting and refining the updating equation, swarm-based nature-inspired algorithms can achieve more efficient and effective solutions across a range of applications.

In this study, a novel approach will be developed in the framework of procurement to segment products using a combination of the SMA and TLBO algorithms. The proposed approach aims to improve the performance of the existing clustering techniques by avoiding local minima traps and searching for the best cluster centroids. The segmentation will be based on several features, including Product Group, Country, Shipment Mode, Freight Cost (USD), and Weight (Kilograms). The approach, TLBO iteratively directs solutions toward improved fitness values and promising regions. Then, APSO exploits these regions extensively while preserving some level of diversity in the solution through NCS mechanism by introducing negative correlations between particles' positions and velocities.

The combined approach of TLBO and APSO-NCS empowers the optimization of data clustering within supply chain management. The procedural details of the new algorithm that is devloped in the context of this work is schowing in **Figure 3**, and the new proposed algorithem is named Improved Teaching Learning Based Optimization via Accelerated Particle Swarm Optimization (ITLBO).

The approach will be evaluated based on the speed of convergence, and statistical tests such as Wilcoxon will be used to compare the results with eight well-known metaheuristics, including the Grey Wolf Optimizer (GWO), Slime Mould Algorithm (SMA), Whale Optimization Algorithm (WOA), Harris Hawks Optimization (HHO), Sine Cosine Algorithm (SCA), Multi-Verse Optimizer (MVO), Moth-Flame Optimization (MFO), and Teaching-Learning-Based Optimization (TLBO). The performance of the proposed approach will be tested using ten different datasets, including shapes, UCI repositories, and SCM datasets. The proposed method expects to contribute to the development of more effective product segmentation techniques in supply chain management.

The steps of the proposed method:

Step 1: Initialize Population:
Create an initial population of candidate solutions randomly.
Step 2: Evaluate Fitness:
Evaluate the fitness of each candidate solution using an objective function.
Step 3: Calculate Mean:
Calculate the mean value of the population.
Step 4: Teacher Phase:
Identify the best solution in the population and update the other solutions by moving them towards this best solution.
Step 5: Learner Phase:
Identify the worst solution in the population and update it by moving it towards the mean value of the population.
Step 6: Apply **APSO-NCS** as shown in the previous flowchart
Step 7: Repeat:
Repeat steps 3 to 6 until a stopping criterion is met, such as a maximum number of iterations or reaching a satisfactory solution.

Figure 3 The steps of the proposed method

3.5 *Validation of the Method on Uci Datasets*

Before applying the developed method on SCM data, we validate it on UCI dataset from machine learning repository (Blake *et al.*, 1998). This data set has been widely used in benchmarking the performance of various learning systems. In fact, by validating clustering algorithms on a diverse set of datasets, researchers can assess their algorithms' ability to generalize across different data types and domains.

UCI datasets encompass a diverse set of domains, including but not limited to healthcare, finance, social sciences, and natural sciences. Each of these datasets offered unique insights and challenges within their respective domains. The iris dataset, for instance, is a classic example used for species classification, while the glass dataset posed questions about glass type classification based on its chemical composition. In addition, the datasets offer varying levels of complexity and dimensions, ranging from relatively small datasets like the iris dataset with 150 samples and 4 features (dimensions) to more complex datasets such as the wine quality dataset with 11 features, providing a diverse set of data sizes and attributes essential for exploring clustering algorithms and patterns. This diversity allows us to evaluate the robustness algorithm, meaning less sensitive to variations in data, noise, or outliers. This resilience to various data conditions is a highly valuable characteristic in real-world applications, specifically in supply chain. It also validates the generalization property as if an algorithm performs well across a variety of datasets, it is more likely to be applicable to real-world problems.

In pursuit of this goal, we assessed the performance of each algorithm by employing various metrics, namely standard deviation, best, worst, and average values of the used objective measure, which is the sum of intra-cluster distances, across all the clusters. We also examined the convergence behavior of these algorithms. The specific metric values for the UCI datasets are detailed in **Table 3**.

Evidently, **Table 3** highlights that the novel approach introduced within this framework consistently ensures the minimum of the best results of the sum of intra-cluster distances on the majority of the datasets, demonstrating its superior performance. This demonstrates that ITLBO serves as a highly efficient alternative for clustering, adept at identifying high-quality clusters that effectively encapsulate cohesion while minimizing dispersion in the representation space, outperforming other algorithms in most datasets. It is worth noting, however, that in the case of Seeds, Balance, and Yeast datasets, ITLBO's performance in terms of the worst and average sum of intra-cluster distances was surpassed by some other techniques.

In addition to these results, **Figure 16** provide two-dimensional visualizations of the clustering results for SCM datasets, employing distinct colors to distinguish individual clusters. In addition, the centroids of the clusters are illustrated by circles. Notably, our approach consistently surpasses the performance of all the algorithms under consideration. Furthermore, the small standard deviation (SD) values associated with ITLBO demonstrates the method's stability. In addition, the centroids of the clusters are illustrated by circles. Interestingly, the graphical analysis has shown clearly distinguished groups in all datasets. To make the comparison clearer, we used two more measures based on the best obtained values. We first calculate the average of the best values over all the datasets as reported in **Table 3**. The average of the best values obtained using our method is the smallest in all datasets except in followed by TLBO and MVO. To enhance the clarity of our comparison, we employed two additional metrics derived from the obtained optimal results. Initially, we computed the average of the best values across all the datasets, as detailed in **Table 3**. The average of the best values obtained using our method is the smallest, followed by the one obtained using TLBO and MVO. Subsequently, we ranked the algorithms based on their respective best values and calculated the average rankings. Our approach secured the top average ranking, followed by MVO and TLBO.

3.6 Experimental Data Sets and Settings

The primary dataset used in this analysis was collected from Kaggle website and includes Antiretroviral (ARV) and HIV lab shipments to supported countries. The dataset includes information on commodity pricing and supply chain expenses associated with moving the commodities to their destination countries for use. This information is similar to that found in the Global Fund's Price, Quality, and Reporting (PQR) data. PEPFAR and the

Global Fund are the two largest procurers of HIV health commodities. This study used the SCM dataset, which contains 10,325 entries and 33 columns of information related to the supply chain. However, we only focused on a subset of this information for each entry, including Product Group, origin Country, Shipment Mode, Freight Cost (USD), and Weight (Kilograms).

However, before integrating it into our algorithms, the data underwent preprocessing to address missing values and nominal features. This involved the removal of instances with missing values and the transformation of nominal values into numerical ones across all data sets, ensuring their compatibility with our analysis. It is noteworthy that this data set encompasses a range of attributes, encompassing demographic, clinical, and behavioral factors, all associated with diabetes progression. In total, the data set comprises X instances and Y features, and it has gained substantial traction in training and testing prediction models for diabetes diagnosis and prognosis.

4. RESULTS AND DISCUSSION

It is worth reminding that the evaluation of our approach has been validated through comparison with well-known algorithms over UCI benchmark dataset in the previous section. It was shown that the novel approach proposed ensures the minimum of the best and average of the sum of intra-cluster distances in the majority of the datasets. This underscores the efficiency of ITLBO as a viable alternative for clustering tasks. It demonstrates ITLBO's ability to discover high-quality clusters that effectively capture cohesion characteristics and exhibit minimal dispersion in the representation space when compared to other algorithms.

This section presents the comparison of our proposed method with the same established algorithms and analyzes the experimental results obtained by them on the SCM dataset. This evaluation will focus on measuring the performance of each algorithm by utilizing standard deviation, the best, worst, and average values of a key objective metric – specifically, the sum of intra-cluster distances. Additionally, we will examine the convergence behavior of these algorithms.

Table 1 Best, worst, average, and standard deviation of WCSS obtained by the considered techniques on the SCM datasets

Dataset name	Criteria	Algorithms								
		GWO	SCA	WOA	HHO	MVO	MFO	SMA	TLBO	ITLBO
SCM	Best	1,52E+05	2,14E+05	1,62E+05	2,03E+05	1,90E+05	2,61E+05	1,89E+05	1,26E+05	**2,12E+04**
	Worst	1,80E+05	2,27E+05	1,67E+05	2,21E+05	2,51E+05	2,66E+05	2,36E+05	1,69E+05	**2,52E+04**
	Average	1,64E+05	2,21E+05	1,64E+05	2,13E+05	2,31E+05	2,64E+05	2,16E+05	1,42E+05	**2.38E+04**
	Std	1,45E+04	6,92E+03	2,54E+03	9,12E+03	3,50E+04	2,48E+03	2,47E+04	2,32E+04	**1,27E+04**

The average values of the Within-Cluster Sum of Squares (WCSS) in all the experiments are reported in **Table 3** and **Table 1**. The WCSS aim to measure the compactness or cohesion of clusters in a clustering algorithm. So, according to tables, it's clear that our approach (ITLBO) outperforms, again, clearly all the considered algorithms. The obtained value for standard deviation (std) is small and shows the stability of ITLBO. It indicates that its solutions are consistently close to the mean or average solution and that it is resilient to changes in initial conditions or problem scenarios. In addition to these results, **Figure 13** visualize in two dimensions the obtained clustering for supply chain management datasets. Each cluster is uniquely identified by a distinct color, and the centroids of these clusters are depicted using circles. Interestingly, the graphical analysis shows clearly distinguished groups when using our method.

4.1 Statistical Analysis

In order to perform an accurate evaluation process, a non-parametric statistical test is used on the experimental results. The Wilcoxon test is conducted on a commonly used significance level of 0.05 (5%). This test determines whether the results are statistically significant or not and obtaining a smaller p-value (i.e., $p-$ value < 0.05) corresponds to more robust evidence against the null hypothesis. In other words, p-values less than 0.05 suggest compelling evidence to reject the null hypothesis and imply a statistically significant distinction between ITLBO and the respective algorithm. **Table 2** contains the obtained $p-$ value of ITLBO compared to the chosen meta-heuristic on the SCM dataset. It can be confirmed that ITLBO outperforms all the other algorithms. Using this statistical test, only MVO can be considered as a competitor of our approach since the p-value obtained is slightly higher than the significance level.

Table 2 Resulting values of Wilcoxon test for statistically significance level at α=0.05 (SCM dataset)

ITLBO vs	GWO	SCA	WOA	HHO	SMA	MVO	MFO	TLBO
SCM	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,04E-01	1,83E-04	1,83E-04

4.2 Convergence Analysis

Figure 4 to **Figure 15** show convergence behavior of the algorithms graphically for the SCM dataset. The vertical axis represents the fitness values obtained in each iteration lying in the horizontal axis while the horizontal axis refers to the number of iterations. ITLBO has shown the best behavior and the curve demonstrates again a rapid switch from exploration to exploitation confirming the results obtained on the benchmark datasets. Thus, the ITLBO proves an efficient performance compared to the well-known algorithm on the supply chain data.

In accordance with this study, it is recommended to use ITLBO metaheuristic as an efficient optimization algorithm, particularly for hybrid applications such as data clustering. Most importantly, due to its high exploratory capabilities that prevent getting stuck in a local optimum when optimizing the distance objectives. Additionally, its high exploitative property explains why the ITLBO based clustering can converge efficiently towards the global optimum in complex search space.

A low standard deviation obtained in the previous results (**Table 2**) also indicate the high-quality of the convergence. It indicates that the method has reached a good state of convergence and has effectively explored the solution space and is producing solutions that are very similar, if not identical, to the optimal or near-optimal solutions.

The results of this work show that although evolutionary algorithms have high exploration and can achieve high accuracy in determining the centroid positions, as shown by the use of ground truth data, the problem of data clustering requires intelligently avoiding local optima throughout the optimization process. The achieved results show that the ITLBO is very efficient in this regard. It is worth mentioning that ITLBO is strongly recommended when the considered problem and dataset are complex and multi-dimensional with a large number of features. Otherwise, it is more convenient to use gradient-based training algorithms where the dataset is small and includes very few features as it will be faster and less computationally expensive.

4.3 Product Clusters and Proposed Procurement Strategies

ITLBO (Improved Teaching Learning Based Optimization via Accelerated Particle Swarm Optimization), was applied to address the complex challenges of product segmentation within procurement using the SCM dataset. The utilization of ITLBO yielded valuable insights into the organization of products based on diverse features (product group, supplier's origin, chosen shipment mode, freight cost (in USD), late delivery, and item weight (in kilograms)), enabling the development of tailored procurement strategies.

The analysis resulted in the identification of five distinct clusters, each characterized by specific features and unique management behaviors.

- Cluster 1

  Healthcare products destined for various African countries emerged as a prominent category. These products, primarily consisting of Antiretroviral (ARV) and HIV lab supplies, exhibited high freight costs (USD) and substantial weight (Kilograms), with the predominant shipment mode being air. To optimize the procurement of these essential healthcare items, we propose the implementation of Regional Procurement Centers (RPCs) strategically (Reichman, 2006) positioned across Africa. RPCs can facilitate consolidated orders, streamline customs processes, and ensure efficient distribution to specific regions, leveraging local knowledge and infrastructure to enhance supply chain efficiency. Additionally, for this cluster, exploring long-term contracts with airfreight carriers can help negotiate more favorable rates, thereby reducing overall transportation costs.

- Cluster 2

  A noteworthy cluster centered on late delivery of orders to varied destination countries. This cluster displayed lower to moderate freight costs (Fri et al., 2021b), moderate weight, and primarily utilized truck transport with extended delivery times. Given the critical nature of healthcare supplies, particularly in the context of HIV/AIDS management, addressing late deliveries is imperative. To mitigate this issue, collaboration with third-party logistics (Zhou et al., 2023) providers is recommended. Selecting reliable suppliers is equally crucial to ensure the success of such partnerships and the timely delivery of healthcare commodities. Furthermore, optimizing inventory management through the adoption of advanced inventory control systems can help minimize lead times (Ouchiekh et al., 2021), reducing the likelihood of late deliveries (Lamii et al., 2022).

- Clusters 3&4

  These clusters highlighted neighboring countries that purchase similar healthcare products, primarily ARVs, with high freight costs, significant weight, and air as the primary shipment mode. A strategy of group procurement (Evans, 1987)can be employed to maximize efficiency and cost-effectiveness in these cases, as neighboring countries often share similar healthcare needs. Collaborative procurement efforts can lead to bulk purchasing and potentially reduced costs. Furthermore, exploring the possibility of joint contracts with suppliers can provide leverage for negotiating lower prices due to the increased volume of orders from multiple countries.

- Clusters 3&4

  This cluster comprised healthcare products, including both ARV and HRDT, which lacked specific proprietary characteristics. These items carried high freight costs, substantial weight, and were commonly shipped via ocean transport. To enhance procurement for this cluster, the implementation of the Just-in-Time (JIT) strategy is recommended. JIT focuses on acquiring goods precisely when needed, reducing storage costs and the risk of product expiration. This approach relies on accurate demand forecasting, close collaboration with suppliers, and the use of frequent, small orders based on real-time demand. Furthermore, optimizing ocean freight logistics through partnerships with reliable carriers and container-sharing agreements can provide additional cost savings in transportation.

These clusters underscore the importance of tailoring procurement strategies to the unique attributes and requirements of each product category within the healthcare supply chain. By aligning supply chain practices with the specific needs of each cluster, organizations can achieve greater efficiency, cost-effectiveness, and ultimately, improved procurement in the healthcare domain. This approach not only optimizes the allocation of resources but also ensures the timely delivery of vital healthcare products to those who depend on them.

5. CONCLUSION

The pharmaceutical supply chain faces multifaceted challenges that require effective procurement strategies. The research highlights the critical role of data-driven decision-making and the integration of clustering techniques in optimizing pharmaceutical supply chains procurement. through hybridization of three algorithms, APSO, TLBO, and NCS, a novel algorithm named (ITLBO) was developed. This innovative algorithm not only demonstrates his efficiency against well knowing metaheuristic, but also achieves high accuracy in clustering problem. Using this new algorithm, ITLBO, this study successfully clusters pharmaceutical products into five distinct clusters within the supply chain. Each cluster is paired with an appropriate procurement strategy. This work will help the decision-makers to manage products procurement efficiently to increase the resiliency of the pharmaceutical supply chains procurement. Furthermore, the designed clustering method and Dataset used in this study focuses on pharmaceutical context and the clustering method can be applied in other context for clustering product to identify the products clusters and the best procurement strategy for each cluster.

It is essential to acknowledge several limitations in this research. Firstly, while the clustering methodology showcased effectiveness within the pharmaceutical supply chain, its applicability might vary in different supply chain contexts or industries due to contextual factors and variations in supply chain structures. Additionally, the performance of clustering algorithms can be influenced by many factors such as the choice of distance metrics and initial parameters which may require further tuning. Furthermore, this study focused on specific aspects of procurement within the pharmaceutical supply chain, and its findings may not cover all potential challenges or strategies. Future research should explore broader applications and adaptability while considering these limitations.

Table 3 Best, worst, average, and standard deviation of WCSS obtained by the considered techniques on the UCI datasets

UCI Dataset	**Criteria**	**Algorithms**								
		GWO	**SCA**	**WOA**	**HHO**	**MVO**	**MFO**	**SMA**	**TLBO**	**ITLBO**
Iris	Best	96,68	133,55	97,32	96,74	96,66	96,66	96,66	**96,66**	96,66
	Worst	97,83	161,67	128,58	133,78	127,67	123,63	96,67	97,26	**96,66**
	Average	97,15	147,70	105,51	117,03	99,89	109,33	96,66	96,88	**96,66**
	Std	0,46	8,04	9,58	13,72	9,76	11,43	0,00	0,29	**0,00**
Seeds	Best	312,34	416,59	356,39	319,48	312,24	311,98	311,84	**311,80**	311,80
	Worst	313,67	486,37	444,01	419,16	313,43	382,86	312,04	**311,80**	311,90
	Average	312,88	453,88	385,29	376,62	312,59	330,19	311,90	**311,80**	311,81
	Std	0,43	19,48	31,80	32,60	0,34	21,52	0,06	**0,00**	0,03
Glass	Best	302,53	395,42	355,51	357,72	298,33	256,76	237,12	213,37	**211,33**
	Worst	348,12	478,12	433,00	464,27	391,78	347,22	260,12	257,98	**246,88**
	Average	323,53	433,61	384,59	403,31	350,53	286,25	247,44	238,69	**234,22**
	Std	15,29	24,46	29,42	36,73	33,57	24,58	**8,79**	15,17	12,10
Cancer	Best	2964,41	3300,59	2971,22	3203,22	2964,49	2966,32	2964,40	**2964,39**	2964,39
	Worst	2964,49	3499,06	3003,58	3751,13	2965,11	3312,54	2964,42	3433,52	**2964,39**
	Average	2964,43	3415,15	2982,74	3515,05	2964,78	3082,76	2964,41	3012,10	**2964,39**
	Std	0,02	59,16	9,83	218,87	0,24	118,07	0,01	148,09	**0,00**
Balance	Best	1423,83	1442,85	1429,73	1429,36	1423,82	**1423,82**	1423,82	1424,17	1423,82
	Worst	1425,55	1457,37	1436,68	1438,17	**1424,05**	1432,59	1425,73	1425,90	1425,87
	Average	1424,00	1450,51	1433,57	1433,22	**1423,85**	1425,48	1424,39	1425,42	1424,78
	Std	0,54	4,95	2,34	2,64	**0,07**	2,66	0,85	0,64	0,92
Haberman	Best	2569,43	2662,95	2570,89	2567,63	2566,99	**2566,99**	2566,99	**2566,99**	**2566,99**
	Worst	2660,70	2701,84	2624,40	2602,01	2567,87	2623,03	2567,83	2569,15	**2567,82**
	Average	2588,89	2682,2	2596,85	2580,33	2567,26	2578,28	2567,16	2567,62	**2567,07**
	Std	27,20	11,88	22,34	12,09	0,41	19,41	0,35	0,68	**0,26**
Wine	Best	16307,73	16510,43	16332,64	16395,53	16330,87	16302,35	16293,36	16293,39	**16292,69**
	Worst	16355,48	17146,63	16508,25	16657,83	16450,76	16412,02	16295,11	16300,64	**16294,35**
	Average	16330,46	16904,20	16422,41	16488,73	16391,82	16326,16	16294,35	16296,50	**16293,47**
	Std	14,97	217,21	58,46	99,14	38,94	33,60	**0,52**	2,46	0,78
Cmc	Best	5806,13	6608,33	5972,31	5903,15	5746,45	5716,61	5694,03	**5693,73**	5693,73
	Worst	6081,79	7409,58	7388,83	7333,89	5887,50	6304,09	5695,31	5723,26	**5693,90**
	Average	5863,83	7008,74	6385,59	6249,22	5816,24	5902,14	5694,47	5697,07	**5693,81**
	Std	82,00	249,68	435,67	410,17	53,17	163,75	0,40	9,25	**0,06**

Table 4 Best, worst, average, and standard deviation of WCSS obtained by the considered techniques on the UCI datasets (Con't)

UCI Dataset	**Criteria**	**Algorithms**								
		GWO	**SCA**	**WOA**	**HHO**	**MVO**	**MFO**	**SMA**	**TLBO**	**ITLBO**
Heart	Best	10638,90	11462,14	10714,83	10966,80	10640,25	10630,09	10623,71	**10622,98**	10622,98
	Worst	10681,04	12767,14	11203,06	13292,02	10741,73	11678,05	10628,35	10626,48	**10622,99**
	Average	10652,08	12187,18	10920,19	11914,02	10667,81	10769,73	10625,36	10623,42	**10622,99**
	Std	16,68	488,94	176,62	748,46	30,71	322,72	1,49	1,08	**0,00**
Yeast	Best	316,35	402,82	352,28	302,30	302,66	326,13	278,66	**257,27**	262,01
	Worst	375,23	584,23	379,65	351,08	379,05	378,05	296,11	**288,65**	321,58
	Average	344,51	485,66	368,91	327,22	358,37	362,57	284,65	**272,34**	273,51
	Std	22,38	72,95	9,14	16,31	27,40	15,01	**6,70**	9,94	18,14
Art	Best	514,31	581,36	514,74	514,24	513,90	**513,90**	513,90	**513,90**	**513,90**
	Worst	516,15	709,58	523,49	908,89	513,91	789,06	513,91	513,90	**513,90**
	Average	515,23	622,20	517,44	740,50	513,91	541,42	513,91	513,90	**513,90**
	Std	0,58	35,86	3,28	193,91	0,00	87,01	0,00	0,00	**0,00**

Table 4 Resulting values of Wilcoxon test for statistically signifcance level at α=0.05 (UCI dataset)

p-value (UCI dataset)								
ITLBO vs	**GWO**	**SCA**	**WOA**	**HHO**	**SMA**	**MVO**	**MFO**	**TLBO**
Iris	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,04E-01	1,83E-04	1,21E-01
Seeds	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	7,69E-04	1,83E-04
Glass	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	2,11E-02	3,45E-01
Cancer	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,40E-01
Balance	2,12E-01	1,83E-04	1,83E-04	1,83E-04	1,62E-01	9,70E-01	6,78E-01	6,40E-02
Haberman	1,49E-04	1,49E-04	1,49E-04	2,03E-04	1,13E-03	4,43E-01	1,48E-03	1,56E-01
Wine	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,73E-02	2,20E-03
Cmc	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	5,71E-01
Heart	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	1,83E-04	6,78E-01
Yeast	2,46E-04	1,83E-04	1,83E-04	5,83E-04	2,46E-04	1,83E-04	3,61E-03	7,34E-01
Art	1,69E-04	1,69E-04	1,69E-04	1,69E-04	1,69E-04	1,87E-02	1,69E-04	5,11E-02

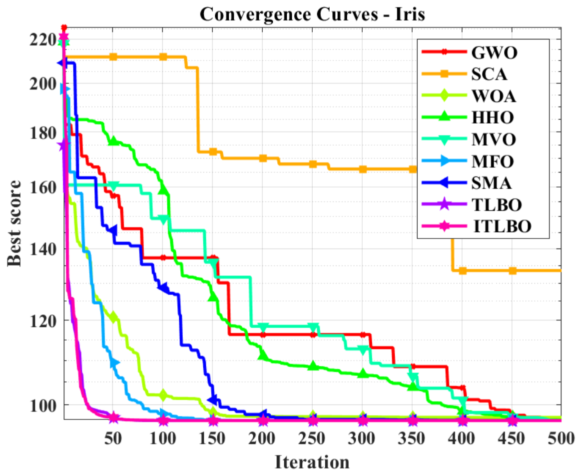

Figure 4 Convergence graph of the metaheuristics for Iris dataset

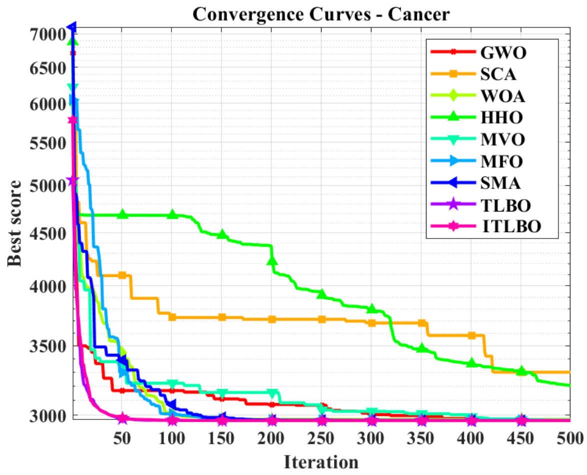

Figure 7 Convergence graph of the metaheuristics for Cancer dataset

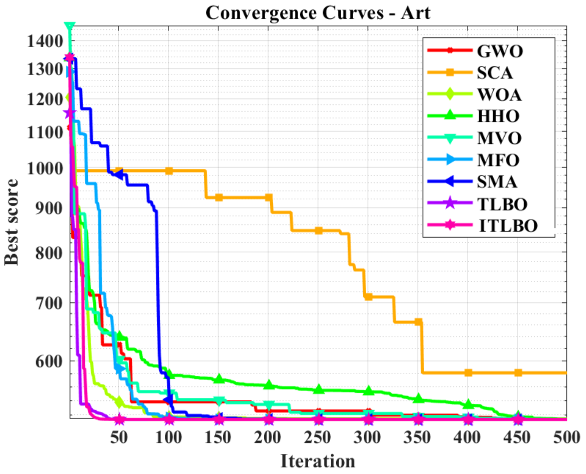

Figure 5 Convergence graph of the metaheuristics for Art dataset

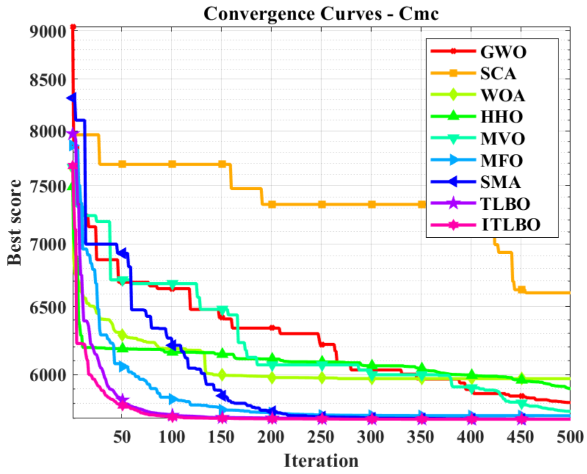

Figure 8 Convergence graph of the metaheuristics for Cmc dataset

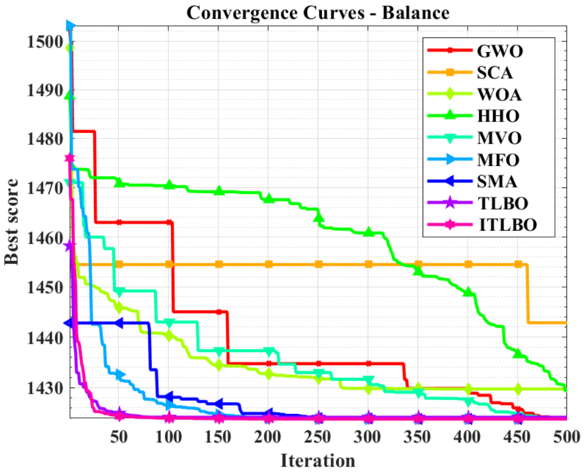

Figure 6 Convergence graph of the metaheuristics for Balance dataset

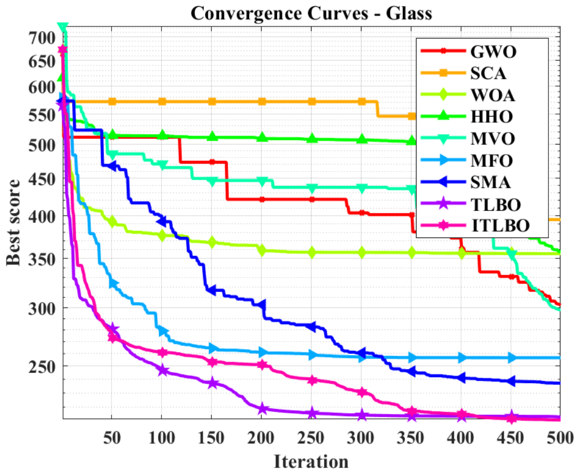

Figure 9 Convergence graph of the metaheuristics for Glass dataset

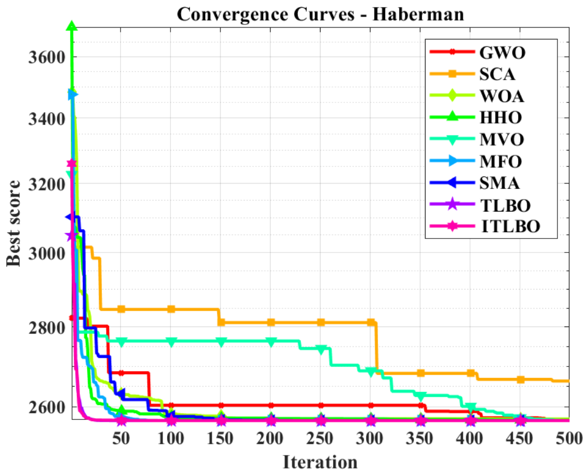

Figure 10 Convergence graph of the metaheuristics for Haberman dataset

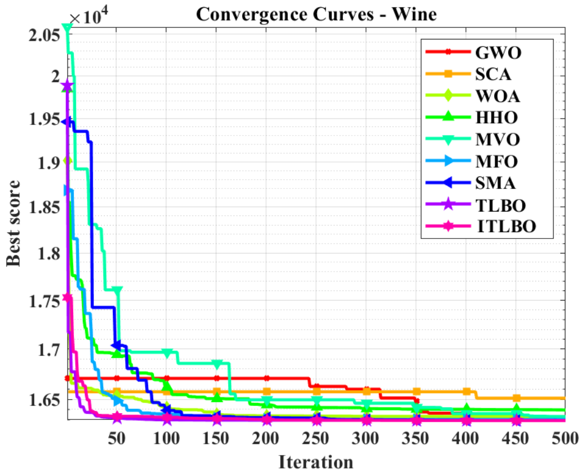

Figure 13 Convergence graph of the metaheuristics for Wine dataset

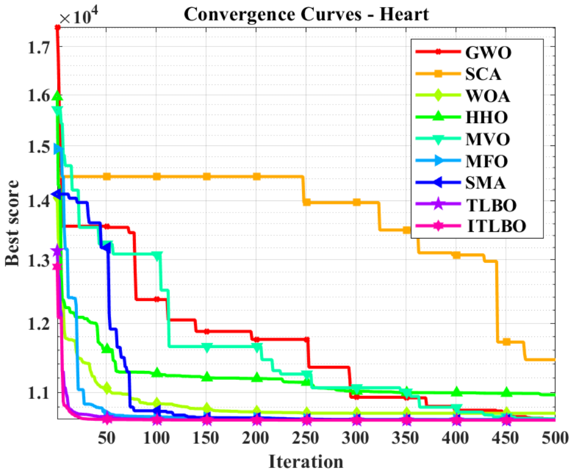

Figure 11 Convergence graph of the metaheuristics for Heart dataset

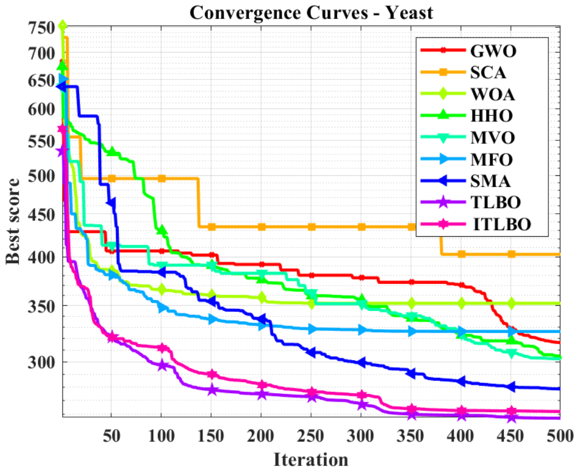

Figure 14 Convergence graph of the metaheuristics for Yeast dataset

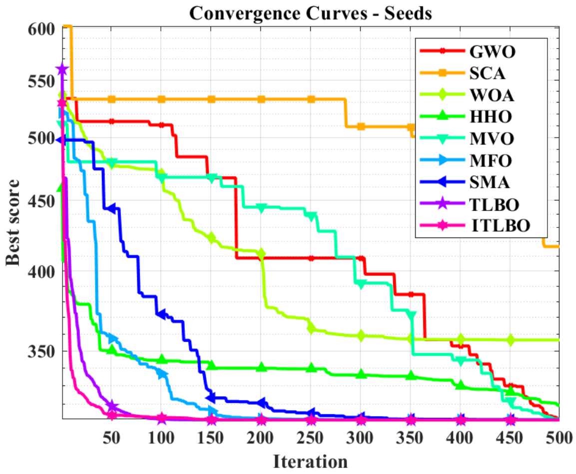

Figure 12 Convergence graph of the metaheuristics for Seeds dataset

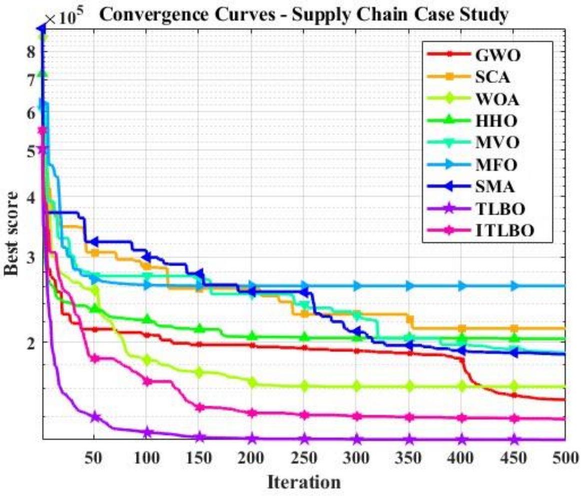

Figure 15 Convergence graph of the metaheuristics for SCM dataset

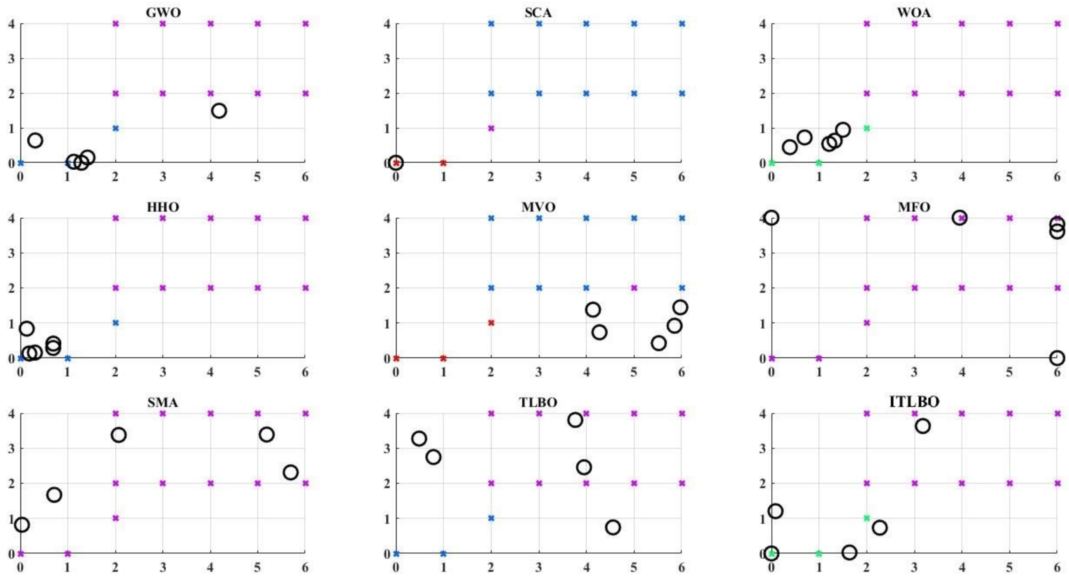

Figure 16 Clustering results for SCM dataset

REFERENCES

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An Extensive Comparative Study of Cluster Validity Indices. *Pattern Recognition*, *46(1)*, pp. 243-256. https://doi.org/10.1016/j.patcog.2012.07.021

Ariyaratne, M., & Fernando, T. (2023). A Comprehensive Review of the Firefly Algorithms for Data Clustering. In A. Biswas, C. B. Kalayci, & S. Mirjalili (Éds.), *Advances in Swarm Intelligence*, 1054, pp. 217-239. Springer International Publishing. https://doi.org/10.1007/978-3-031-09835-2_12

Azadeh, A., Osanloo, M., Khoshalhan, F., & Amiri, M. H. (2012). A Novel Hierarchical Clustering Method Using the Hybrid of K-Means and Gravitational Search Algorithms. *Expert Systems with Applications*, *39(17)*, pp. 13057-13066.

Bagirov, A. M., Aliguliyev, R. M., & Sultanova, N. (2023). Finding Compact and Well-Separated Clusters: Clustering Using Silhouette Coefficients. *Pattern Recognition*, *135*, pp. 109144. https://doi.org/10.1016/j.patcog.2022.109144

Bali, V., & Bebok, Z. (2015). Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *The International Journal of Biochemistry & Cell Biology*, *64*, pp. 58-74. https://doi.org/10.1016/j.biocel.2015.03.011

Bassand, C., Villois, A., Gianola, L., Laue, G., Ramazani, F., Riebesehl, B., Sanchez-Felix, M., Sedo, K., Ullrich, T., & Duvnjak Romic, M. (2022). Smart Design of Patient-Centric Long-Acting Products: From Preclinical to Marketed Pipeline Trends and Opportunities. *Expert Opinion on Drug Delivery*, *19(10)*, pp. 1265-1283. https://doi.org/10.1080/17425247.2022.2106213

Blake, C., Keogh, E., & Merz, C. J. (1998). UCI Repository of Machine Learning Databases. Department of Information and Computer Science, University of California, Irvine, CA. http://www.ics.uci.edu/mlearn/MLRepository.htm

Chen, D., Zhao, G., & Yang, Y. (2022). Research on Location of Supply Chain Center of Natural Resources based on K-means Clustering Model. 2022 4th International Conference on Communications, *Information System and Computer Engineering (CISCE)*, pp. 674-677. https://ieeexplore.ieee.org/abstract/document/9850975/

Chen, H., Peng, Q., Li, X., Todo, Y., & Gao, S. (2018). An Efficient Negative Correlation Gravitational Search Algorithm. *2018 IEEE International Conference on Progress in Informatics and Computing (PIC)*, pp. 73-79.

Darbanian, F., Brandtner, P., Falatouri, T., & Nasseri, M. (2024). Data Analytics in Supply Chain Management: A State-of-the-Art Literature Review. *Operations and Supply Chain Management: An International Journal*, pp. 1-31. https://doi.org/10.31387/oscm0560411

Ding, S., Cui, T., Wu, X., & Du, M. (2022). Supply chain Management Based on Volatility Clustering: The Effect of CBDC Volatility. *Research in International Business and Finance*, *62*, 101690.

Dinh, D.-T., Fujinami, T., & Huynh, V.-N. (2019). Estimating the optimal number of clusters in categorical data clustering by silhouette coefficient. Knowledge and Systems Sciences: 20th International Symposium, KSS 2019, Da Nang, Vietnam, November 29–December 1, 2019, *Proceedings 20*, pp. 1-17.

Douaioui, K., Fri, M., Mabrouki, C., & Semma, E. A. (2021). Improved Meta-Heuristic for Multi-Objective Integrated Procurement, Production and Distribution Problem of Supply Chain Network Under Fuzziness Uncertainties. *Journal of Computer Science*, *17(12)*, pp. 1196-1209. https://doi.org/10.3844/jcssp.2021.1196.1209

Douaioui, K., Fri, M., Mabrouki, C., & Semma, E. A. (2022). A Multiobjective Integrated Procurement, Production, and Distribution Problem of Supply Chain Network. In J. Kacprzyk, V. E. Balas, & M. Ezziyyani (Éds.), *Advanced Intelligent Systems for Sustainable Development (AI2SD'2020)*, pp. 1019-1035. Springer International Publishing.

Evans, R. G. (1987). Public health insurance: The Collective Purchase of Individual Care. *Health Policy*, *7(2)*, pp. 115-134. https://doi.org/10.1016/0168-8510(87)90026-1

Ezugwu, A. E. (2020). Nature-Inspired Metaheuristic Techniques for Automatic Clustering: A Survey and Performance Study. *SN Applied Sciences*, *2(2)*, pp. 273. https://doi.org/10.1007/s42452-020-2073-0

Ezugwu, A. E., Agushaka, J. O., Abualigah, L., Mirjalili, S., & Gandomi, A. H. (2022). Prairie Dog Optimization Algorithm. *Neural Computing and Applications*, *34(22)*, pp. 20017-20065. https://doi.org/10.1007/s00521-022-07530-9

Fri, M., Douaioui, K., Lamii, N., Mabrouki, C., & Semma, E. A.

(2020). A Hybrid Framework for Evaluating the Performance of Port Container Terminal Operations: Moroccan Case Study. *Pomorstvo*, *34(2)*, pp. 261-269. https://doi.org/10.31217/p.34.2.7

Fri, M., Douaioui, K., Mabrouki, C., & El Alami, S. (2021a). Reducing Inconsistency in Performance Analysis for Container Terminals. *International Journal of Supply and Operations Management*, *8(3)*, pp. 328-346. https://doi.org/10.22034/ijsom.2021.3.6

Fri, M., Douaioui, K., Mabrouki, C., & El Alami, S. (2021b). Reducing Inconsistency in Performance Analysis for Container Terminals. *International Journal of Supply And Operations Management*, *8(3)*, pp. 328-346. https://doi.org/10.22034/ijsom.2021.3.6

Gagolewski, M., Bartoszuk, M., & Cena, A. (2021). Are Cluster Validity Measures (in) Valid? *Information Sciences*, *581*, pp. 620-636.

Gálvez, J., Cuevas, E., Becerra, H., & Avalos, O. (2020). A Hybrid Optimization Approach Based on Clustering and Chaotic Sequences. *International Journal of Machine Learning and Cybernetics*, *11(2)*, pp. 359-401. https://doi.org/10.1007/s13042-019-00979-6

Ghadge, A., Bourlakis, M., Kamble, S., & Seuring, S. (2023). Blockchain Implementation in Pharmaceutical Supply Chains: A Review and Conceptual Framework. *International Journal of Production Research*, *61(19)*, pp. 6633-6651. https://doi.org/10.1080/00207543.2022.2125595

Ghosal, A., Nandy, A., Das, A. K., Goswami, S., & Panday, M. (2020). A Short Review on Different Clustering Techniques and Their Applications. In J. K. Mandal & D. Bhattacharya (Éds.), *Emerging Technology in Modelling and Graphics*, *937*, pp. 69-83. Springer Singapore. https://doi.org/10.1007/978-981-13-7403-6_9

Gichuru, M. J., & Limiri, E. K. (2017). Market Segmentation as A Strategy for Customer Satisfaction and Retention. *International Journal of Economics, Commerce and Management*, *5(12)*, pp. 544-553.

Gurrala, K., & Hariga, M. (2022). Key Food Supply Chain Challenges: A Review of the Literature and Research Gaps. *Operations and Supply Chain Management*: An International Journal, pp. 441-460. https://doi.org/10.31387/oscm0510358

Halner, A., & Bafadhel, M. (s. d.). Usefulness of Biological Clustering Patterns in Chronic Obstructive Pulmonary Disease. Consulté 2 juin 2024, à l'adresse https://pdfs.semanticscholar.org/d206/5a8fc00038e7e37c9667378170dd3c9f5b84.pdf

Hong, X., Yuan, W., Zhifeng, L., Ruidi, X., & Zhuoran, X. (2020). Research on Emergency Material Optimization Scheduling and Importance Decision in Cluster Supply Chain. *2020 Chinese Control And Decision Conference (CCDC)*, pp. 2816-2821. https://ieeexplore.ieee.org/abstract/document/9164284/

Lahtinen, J. (2021). Clustering and Classification of Material Suppliers Using Machine Learning Algorithms. https://lutpub.lut.fi/handle/10024/162545

Lamii, N., Fri, M., Mabrouki, C., & Semma, E. A. (2022). Using Artificial Neural Network Model for Berth Congestion Risk Prediction. *IFAC-PapersOnLine*, *55(12)*, pp. 592-597. https://doi.org/10.1016/j.ifacol.2022.07.376

Liang, F., Qian, C., Yu, W., Griffith, D., & Golmie, N. (2022). Survey of graph neural networks and applications. *Wireless Communications and Mobile Computing*, *2022(1)*.

Liu, Y., & Yao, X. (1999). Ensemble Learning Via Negative Correlation. *Neural networks*, *12(10)*, pp. 1399-1404.

Ma, Y., Li, Y., & Yong, L. (2024). Teaching–Learning-Based Optimization Algorithm with Stochastic Crossover Self-Learning and Blended Learning Model and Its Application. *Mathematics*, *12(10)*, pp. 1596.

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Let A Biogeography-Based Optimizer Train Your Multi-Layer Perceptron. *Information Sciences*, *269*, pp. 188-209.

Nazari, A., Mehregan, M., & Tehrani, R. (2019). Credit Scoring of Bank Depositor with Clustering Techniques for Supply Chain Finance. *International Journal of Supply Chain Management*, *8*, pp. 374-383.

Oucheikh, R., Fri, M., Fedouaki, F., & Hain, M. (2020). Deep Real-Time Anomaly Detection for Connected Autonomous Vehicles. *Procedia Computer Science*, *177*, pp. 456-461.

Oucheikh, R., Fri, M., Fedouaki, F., & Hain, M. (2021). Deep Anomaly Detector Based on Spatio-Temporal Clustering for Connected Autonomous Vehicles. In L. Foschini & M. El Kamili (Éds.), *Ad Hoc Networks*, *345*, pp. 201-212. Springer International Publishing. https://doi.org/10.1007/978-3-030-67369-7_15

Oucheikh, R., Touil, A., & Fri, M. (2022). Data Clustering using Two-Stage Eagle Strategy Based on Slime Mould Algorithm. *Journal of Computer Science*, *18(11)*, pp. 1062-1084. https://doi.org/10.3844/jcssp.2022.1062.1084

Ouchiekh, R., Fri, M., Touil, A., & Echchatbi, A. (2021). Total Weighted Tardiness in the Permutation Flow Shop under Uncertainty. *IFAC-PapersOnLine*, *54(1)*, pp. 1174-1180. https://doi.org/10.1016/j.ifacol.2021.08.138

Rana, S. U. A., & Cheok, A. D. (s. d.). Application of Data Mining Combined with K-means Clustering Algorithm in Enterprises' Risk Audit. Consulté 2 juin 2024, à l'adresse https://www.qeios.com/work-supplementary-data/E03BE8/g9g0s3.pdf

Reichman, J. H. (2006). Procuring Essential Medicines Under the Amended TRIPS Provisions: The Prospects for Regional Pharmaceutical Supply Centers. Intellectual Property, Trade and Development Conference, Chicago-Kent College of Law.

Seddigh, M., Shokoohyar, S., Ghanadpour, S., & Moradi, H. (2022). Pharmaceutical Supply Chain Sustainability under the Torchlight of Social Media. *Operations and Supply Chain Management: An International Journal*, pp. 486-504. https://doi.org/10.31387/oscm0510361

Smith, J., & Johnson, K. (2020). Determination of Optimum Number of Classes in Clustering. *Journal of Data Mining*, *10(3)*, pp. 123-136. https://doi.org/10.1234/abcd1234

Sodhi, M. S., & Tang, C. S. (2021). Supply Chain Management for Extreme Conditions: Research Opportunities. *Journal of Supply Chain Management*, *57(1)*, pp. 7-16. https://doi.org/10.1111/jscm.12255

Soledispa-Canarte, B., Pibaque-Pionce, M., Merchan-Ponce, N., Mex Alvarez, D., Tovar-Quintero, J., Escobar-Molina, D., Cedeno-Ramirez, J., & Rincon-Guio, C. (2023). The Role of Logistics 4.0 in Agribusiness Sustainability and Competitiveness, A Bibliometric and Systematic Literature Review. *Operations and Supply Chain Management: An International Journal*, pp. 109-120. https://doi.org/10.31387/oscm0520376

Srinivasan, D., & Balaji, B. (2020). Hierarchical Clustering Approach for Optimizing Inventory Management in Multi-Outlet Food Chain. *Journal of Cleaner Production, 278,* 123947.

Tang, K., Yang, P., & Yao, X. (2016). Negatively Correlated Search. IEEE Journal on Selected Areas in Communications, *34(3)*, pp. 542-550.

Too, J., & Abdullah, A. R. (2021). A New and Fast Rival Genetic Algorithm for Feature Selection. *The Journal of Supercomputing*, *77(3)*, pp. 2844-2874. https://doi.org/10.1007/s11227-020-03378-9

Ullmann, T., Hennig, C., & Boulesteix, A. (2022). Validation of Cluster Analysis Results on Validation Data: A Systematic Framework. *WIREs Data Mining and Knowledge Discovery*, *12(3)*, e1444. https://doi.org/10.1002/widm.1444

Wang, D., Li, L., & Zhao, D. (2022). Corporate Finance Risk Prediction Based on LightGBM. *Information Sciences*, *602*, pp. 259-268.

Wang, M., Radics, R., Islam, S., & Hwang, K.-S. (2023). Towards Forest Supply Chain Risks. *Operations and Supply Chain Management: An International Journal*, pp. 97-108. https://doi.org/10.31387/oscm0520375

Wang, X., & Xu, Y. (2019). An Improved Index for Clustering Validation Based on Silhouette Index and Calinski-Harabasz Index. *IOP Conference Series: Materials Science and Engineering*, *569(5)*, 052024. https://doi.org/10.1088/1757-899X/569/5/052024

Wang, Z., Wang, E., & Zhu, Y. (2020). Image Segmentation Evaluation: A Survey of Methods. *Artificial Intelligence Review*, *53(8)*, pp. 5637-5674. https://doi.org/10.1007/s10462-020-09830-9

Yang, X.-S. (2010). Nature-Inspired Metaheuristic Algorithms.

Zhou, H., Wang, Q., Wang, L., Zhao, X., & Feng, G. (2023). Digitalization and Third-Party Logistics Performance: Exploring the Roles of Customer Collaboration and Government Support. *International Journal of Physical Distribution & Logistics Management*, *53(4)*, pp. 467-488. https://doi.org/10.1108/IJPDLM-12-2021-0532

Kaoutar Douaioui is a researcher with a Ph.D. in Industry and Logistics from Hassan 1st University in Settat, Morocco. She also holds an engineer's degree in Industrial Management from the same university, earned in 2015. Kaoutar is a member of the Laboratory of Engineering, Industrial Management, and Innovation. Her main research areas include Industry 4.0, Smart Logistics, Smart Ports, Supply chain management, and machine learning.

Rachid Oucheikh received his Engineering degree in computer sciences and networking systems from the National Institute of Posts and Telecommunications in Rabat, Morocco, in 2012. He then pursued a Ph.D. in Computer Sciences and Modeling from the Faculty of Science at Sidi Mohamed Ben Abdellah University, Morocco. Currently, he is a researcher at Lund University, Sweden. His current research interests include artificial intelligence (AI) applications in multiple domains such as autonomous systems, real-time systems, geospatial analysis, and intelligent transportation systems. He has actively participated in several national projects, receiving funding from organizations such as the Swedish Research Council and the Knowledge Foundation.

Othmane Benmoussa joined the Euromed University of Fes in January 2019. He is currently Dean of the Euromed Polytechnic School (EPS). In addition to his experience in the academic area in Switzerland and Morocco, Pr. Benmoussa has over 15 years' industry work experience as a senior executive director in strategy, airport's management, logistics, transportation, real estate, and purchasing for many Moroccan and international public and private organizations. With training as a civil engineer, Pr. Benmoussa has also earned in 2003 a Ph.D in management of technology, entrepreneurship, and breakthrough project management both from the EPFL (Swiss Federal Institute of Technology - Lausanne), as well as a postgraduate diploma in transportation economics and modelling from the MIT (Massachusetts Institute of Technology) and an advanced certificate for executives in Management, Innovation, and Technology from the MIT Sloan School of Management.