IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 21 March 2024, accepted 6 April 2024, date of publication 10 April 2024, date of current version 17 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3386887

RESEARCH ARTICLE

Enhancing Warehouse Efficiency With Time Series Clustering: A Hybrid Storage Location Assignment Strategy

HICHAM KALKHA[1], **AZEDDINE KHIAT**[1], **AYOUB BAHNASSE**[2], **AND HASSAN OUAJJI**[1]

[1]Computing, Artificial Intelligence and Cyber Security Laboratory (2IACS), ENSET Mohammedia, University Hassan II of Casablanca, Casablanca 20670, Morocco

[2]Laboratory Engineering of Structures, Processes, Intelligent Systems, and Computer Science (ISPS2I), ENSAM Casablanca, Casablanca 20670, Morocco

Corresponding author: Hicham Kalkha (hicham.kalkha-etu@etu.univh2c.ma)

ABSTRACT In the rapidly evolving domain of e-commerce, effective warehouse management emerges as a critical factor for ensuring timely deliveries. This paper addresses the Storage Location Assignment Problem (SLAP) in e-commerce warehouses, a challenge intensified by varying product volumes and unpredictable demands. We introduce a novel Intelligent Storage Location Assignment (ISLA) method that utilizes advanced time series clustering algorithms specifically, Self-Organizing maps, dynamic time warping-Based k-means, and Agglomerative Hierarchical Clustering (AHC), to optimize order fulfillment and enhance warehouse efficiency. By clustering and positioning items with similar demand patterns, our approach minimizes order preparation time, reduces unnecessary warehouse movements, and improves operational flows. Our empirical evaluation, based on a real-world dataset from Kaggle, demonstrates the superiority of AHC in efficiently grouping high-turnover items, as evidenced by higher silhouette scores. Applying this method in simulations across various picking strategies such as s-shape, mid-point, discrete order picking, zone picking, and batch picking, we achieve significant efficiency improvements. Notably, our ISLA method results in up to 61% and 69% efficiency gains under s-shape and midpoint routing policies, respectively, outperforming traditional random and ABC storage assignments. These results not only highlight the significant potential of Artificial Intelligence (AI) in revolutionizing warehouse operations but also bridge the existing knowledge gap by showcasing a practical and impactful application of AI in SLAP. Our research advances the field of smart logistics, emphasizing the critical role of AI-driven intelligent storage location assignment in optimizing warehouse processes and enhancing the efficiency of the e-commerce supply chain.

INDEX TERMS E-commerce, intelligent storage location assignment, artificial intelligence, order picking, time series clustering, warehouse efficiency, smart logistics.

I. INTRODUCTION

The rapid expansion of e-commerce has brought about an era in which the effective management of warehouses holds a central position in ensuring punctual deliveries. In this dynamic environment, optimizing warehousing operations is not just important; it's crucial to meet the growing demand for swift and precise order fulfillment. Warehousing, within the larger context of supply chain management, carries critical responsibilities that encompass the storage of diverse e-commerce products. To maintain a competitive advantage in a swiftly evolving industry and outperform competitors, warehouses must execute their operations with precision. These tasks range from receiving, put-away, cross-docking, order picking, to shipping. The execution of these operations must prioritize efficiency to guarantee the smooth flow of the supply chain while simultaneously reducing costs. A central objective in warehouse management revolves around the minimization of order preparation cycle time, a process, as illustrated in **FIGURE 1**, commands

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad J. Abdel-Rahman.

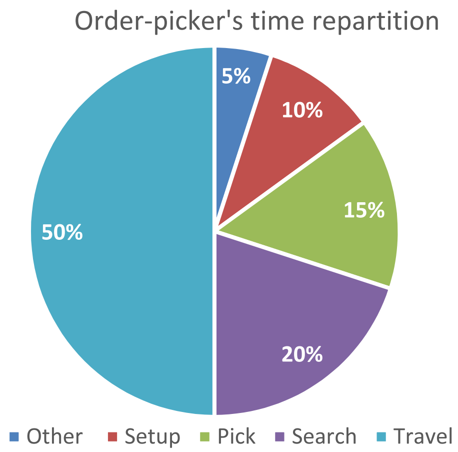

FIGURE 1. Order picker's time repartition.

approximately 60% of the total labor resources, with 50% of that specifically attributed to picking travel. Order preparation consumes nearly 35% of total warehouse operations time [1]. It has become the bottleneck in e-commerce logistics [2]. Consequently, the operational efficiency of order picking, and the management of operating costs hold the key to not only enhancing warehouse performance but also to catalyzing sustainable development across the logistics supply chain in the context of fourth industrial revolution [3] In parallel, a persistent challenge looms large in the form of the Storage Location Assignment Problem (SLAP), which seeks to maximize the efficiency of storage and retrieval operations by minimizing the time and effort required to locate and retrieve items for order fulfillment and various other operational purposes. However, the formidable complexity of this problem, compounded by uncertainties surrounding the future demand for diverse products. Therefore, SLAP has been firmly established as a non-deterministic polynomial-time hard (NP-hard) problem [4], underscoring the need for innovative and data-driven approaches to tackle this complexity. SLAP aims to answer the question: What is the most suitable location for storing an item to minimize picking distance? The input for SLAP is the demand for an item relative to others, and the output is the optimal location for that item. A recent study applied ABC clustering and association rule mining to address the SLAP in cold logistics centers, resulting in strategies that notably improved efficiency optimizing order picking times by up to 8% compared to random placement [5].

In this study, we will emphasize the Intelligent Storage Location Assignment (ISLA) approach as a means of anticipating order picking optimization, specifically during the 'put away' phase. This involves anticipating the items that are typically shipped together or, at the very least, during the same time shift, by grouping similar order shipping time series using time series clustering algorithms such as Self-Organizing Maps (SOM), Dynamic Time Warping-Based K-Means (DTW), and Agglomerative Hierarchical Clustering. The objective is to concentrate picker activities within a single area of the warehouse, thereby reducing logistical inefficiencies (MUDA). It's important to note that this 'put away' phase precedes the optimization of the order picking routes.

The motivation behind this study is to address a gap identified in our recent literature review [6], this systematic literature review and bibliometric analysis over 230 papers in the context of Smart Logistics (SL), analyzes 71 representative papers published between 2005 and 2017, revealed a lack of research interest in the utilization of advanced Information and Communication Technologies (ICT) such as Artificial Intelligence (AI) and Internet of Things (IoT) for tackling SLAP. This is in line with the findings of Juan et al., as corroborated by their literature review on storage location assignment. It calls for further research into management tools integrating machine learning and artificial intelligence and suggests broadening future studies beyond academic papers [7]. Contemporary literature in SL advocates for AI integration in manufacturing, emphasizing its role in optimizing logistics processes for enhanced efficiency in the context of Smart Logistics [8] emphasizing the importance of AI-driven solutions in SL, proposing advanced system that optimizes order dispatch, improves delivery time predictions, and outperforms traditional methods, ultimately enhancing efficiency and reducing costs in warehousing using deep learning [9] However, the proposal to use clustering as a method to tackle SLAP within the context of SL has never been introduced in the literature. It's crucial to emphasize that efficient storage location assignments are the foundation for optimizing order preparation, and this aspect should take precedence over the role of TSP in routing optimization.

Numerous research endeavors have embarked on addressing SLAP by harnessing machine learning algorithms and data mining techniques. These methods encompass the utilization of Association rules to cluster items frequently shipped together in the same order, thus orchestrating a reduction in order preparation cycle time [10]. Furthermore, exploration into deep reinforcement learning techniques has ventured into product assignment, considering historical demand patterns for intelligent routing decisions [11].

Our aim is to contribute to the ongoing debate on storage location assignment by presenting a solution that reduces warehouse movements. We propose a heuristic optimization method that provides practical solutions swiftly and with minimal computing resources. Notably, our contribution entails a hybrid location assignment based on specific product characteristics, remaining unchanged during product receptions and put-away processes over defined time periods.

In the pursuit of advancing our understanding and efficacy in addressing these critical challenges, this study endeavors to answer several pressing research questions:

1) How has the concept of the Storage Location Assignment Problem evolved within the realm of artificial intelligence in literature?

2) What advantages does the practice of item clustering offer in terms of enhancing warehouse performance?
3) Among the myriad storage location assignments, how does our proposed method react with different picking strategies and methods?

To address these questions comprehensively, this paper is meticulously organized as follows:

Section II provides a comprehensive overview of the existing body of work related to data-driven storage location assignment approaches, Setting the stage for our contribution by providing insights into different time series clustering methods. In Section III, we introduce our innovative proposition, which centers around leveraging historical demand clustering as a potent solution to the complex SLAP. In this section, we present a detailed discussion of our proposal. Section IV, offers a critical comparative analysis, pitting our proposal against alternative approaches from the literature. Additionally, we provide a benchmark analysis to ascertain the efficacy of the proposed methodology in practical warehouse scenarios.

II. RELATED WORK

In the realm of warehouse optimization, routing optimization is often treated as a Traveling Salesman Problem (TSP) and as SLAP, each addressing distinct facets of operational efficiency. While TSP focuses on optimizing the path taken through the warehouse to minimize travel distance or time, our method concentrates on the SLAP, specifically targeting the intelligent assignment of items to storage locations. By prioritizing the SLAP, our approach aims to enhance overall warehouse operations by strategically positioning items based on demand patterns and other relevant factors, thereby facilitating more efficient retrieval processes, and improving the foundation for subsequent routing optimizations. In this section we focus on reviewing papers tackling SLAP.

A. STORAGE LOCATION ASSIGNMENT

Efficient warehouse management is essential in the rapidly evolving landscape of e-commerce logistics. At the heart of this efficiency lies the strategic storage location assignment within the warehouse. It is paramount for achieving streamlined operations, reducing order fulfillment times, and minimizing operational costs. This process is crucial for streamlining order fulfillment, minimizing picking times, and ultimately reducing overall operational costs by considering order fulfilment constraint during the put away phase. The goal is to minimize order picking route **FIGURE 2**, by assigning each product to an optimal storage location, taking into consideration various factors such as warehouse layout, product characteristics, demand patterns, and operational constraints.

SLAP is akin to the Quadratic Assignment Problem and is considered NP-hard, indicating its computational complexity.

Its objective function typically seeks to minimize a cost or distance metric associated with the allocation of products to specific storage locations [12]. In a generic form, assuming that each product 'i' is assigned to only one location 'j' within the warehouse, it can be expressed as:

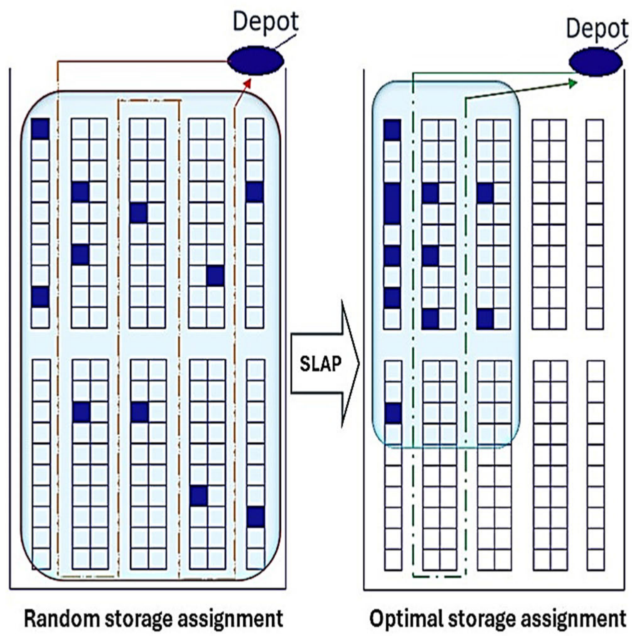

FIGURE 2. The goal of storage location assignment.

$$Min\ Z = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij} \tag{1}$$

- n is the number of products or items
- m is the number of available storage locations
- c_{ij} is the cost or distance associated with assigning product 'i' to storage location 'j'.
- x_{ij} is a binary decision variable representing whether product 'i' is assigned to storage location 'j' (1 if assigned, 0 otherwise).

The specific form of c_{ij} depends on the cost or distance metric chosen, such as travel time, transportation cost, or any other relevant criterion for the storage assignment problem at hand, we consider distance and time for this study.

Historically, warehouses employed conventional methods such as random allocation and ABC classification to assign storage locations [13]. While these methods provided a basic level of the organization relies on the rotation frequency of products without considering interdependencies among them. thus, they often fell short in handling dynamic demand patterns and the increasing complexity of modern e-commerce logistics. Recent advancements in storage location assignment techniques have revolutionized warehousing operations. Machine learning and data-driven approaches, such as clustering algorithms, have gained prominence. AI and optimization techniques have been pivotal in addressing complex storage location assignment challenges. Genetic algorithms, simulated annealing, and deep reinforcement learning have emerged as powerful tools to find optimal storage location assignments while considering constraints and ever-changing demand patterns [14], [15]. These AI-driven methods enable real-time adaptability and responsive warehouse operations. Furthermore, data-driven approaches are gaining traction, harnessing historical data to inform storage location assignments. By analyzing past order data, these methods anticipate future demand patterns, thereby reducing the time and resources needed for storage location adjustments, these dynamic methods, utilizes sophisticated algorithms that, while effective, come with a trade-off in terms of time and computational resource [11]. The impact of efficient storage location assignment goes beyond cost reduction. It directly translates into enhanced warehouse efficiency, faster order fulfillment, and improved customer satisfaction. In the competitive e-commerce landscape, where rapid deliveries and precision are crucial, optimized storage location assignment plays a pivotal role in shaping the logistics ecosystem. In literature, we found two types of storage locations assignment:

Static Storage Location Assignment (SSLA): is simpler, offering a stable and easily manageable warehouse layout with lower initial costs. Yet, it can lead to inefficiencies and poor space utilization due to its inability to adapt to fluctuating demand.

Dynamic Storage Location Assignment (DSLA): offers adaptability and efficiency, optimizing warehouse operations by adjusting to changing demand patterns, which enhances space utilization and reduces picking times. However, it requires sophisticated management systems, incurs higher setup costs, and demands more extensive staff training [16].

Hybrid Storage Location Assignment (HSLA): The approach of using time series clustering for SLAP is dynamic in its analysis and planning stages but can manifest as static in its implementation over fixed periods and positions. The distinction lies in how frequently the analysis results are applied to physical reassignments. If product locations are periodically adjusted based on updated time series clustering analyses, the system embodies a dynamic strategy, albeit with static phases between adjustments. Thus, it's a hybrid model: dynamic in its analytical foundation and potentially static in its execution, depending on the frequency of physical reassignments in the warehouse.

B. OPERATIONAL RESEARCH METHODS

Many methodologies showcase the versatility of Operational Research (OR) techniques in addressing storage location assignment challenges, providing efficient solutions while considering constraints, demand patterns, and real-time adaptability, contributing to enhanced warehouse efficiency and logistics optimization:

1) PARETO-BASED ABC CLASSIFICATIONS

This approach divides inventory into three categories (A, B, and C) based on the Pareto principle, with the aim of reducing the travel distance for forklifts, thereby saving costs, and improving process efficiency. The new warehouse layout design led to significant reductions in forklift travel distances and allowed for a reduction in the number of forklift operators, showcasing tangible improvements in operational efficiency within a real warehouse setting. When compared to time series clustering for the SLAP, several key differences emerge. Time series clustering dynamically adjusts storage locations is focusing on long-term operational efficiency and adaptability to changing demands. It's particularly effective in environments where demand for products fluctuates over time, allowing warehouses to proactively adjust storage strategies to anticipate future needs. On the other hand, the ABC classification method, as applied in the discussed paper, is more static in nature, focusing on optimizing warehouse layout based on the current categorization of items according to their turnover rates. While it does lead to immediate efficiency gains by reducing travel times and operational costs, it may not be as responsive to sudden changes in demand patterns as time series clustering [17]

2) BRANCH AND BOUND ALGORITHMS

These algorithms are widely employed in operational research for storage location assignment due to their ability to systematically explore the solution space while minimizing computational efforts. These algorithms typically start with an initial solution and iteratively branch into subproblems, evaluating lower and upper bounds for each branch. Researchers have improved upon traditional Branch and Bound techniques, as seen in Huang, Liu, and Wang [18] by introducing an advanced Branch and Bound algorithm designed for e-commerce warehousing, emphasizing optimized item placements to reduce order picking times. The algorithm employs sophisticated heuristics and pruning strategies to expedite the search process and enhance overall efficiency. Class-based storage policies have also garnered attention, as they have proven to be efficient in minimizing pick travel distances. A proposed nonlinear integer-programming model considers savings in required storage space due to the random allocation of products within a class. To solve the model, a branch and bound algorithm (BBA) is developed and compared with a benchmark dynamic programming algorithm. Computational experience shows that the class-based policy results in shorter pick-travel distances, with the BBA demonstrating superior computational efficiency [18]. However, these algorithms can be limited by their dependency on initial solutions and heuristics, potentially hindering their adaptability to dynamic warehousing environments. Time series clustering (our approach) addresses these limitations by clustering items based on demand trends, thereby offering a more flexible and foresighted approach to storage location assignment.

3) A-STAR ALGORITHM

Recognized for its pathfinding capabilities, this algorithm has found applications in real-time storage location assignment, particularly in automated warehouses. Operational research endeavors have formulated the assignment problem as a graph-based model, where storage locations represent nodes, and distances between them form the edges. Chen, Wu, and Li exemplify this approach, applying the A-Star algorithm to automate storage location assignment in warehouses [1]. By considering factors like item characteristics, demand patterns, and order picking routes, the algorithm optimizes storage assignments to minimize time and resource wastage in retrieval processes. Real-time adaptability and efficiency are key benefits of this approach. Several studies have been conducted to explore and optimize storage location assignment in warehouses and fulfillment systems using operational research such as branch and bound algorithms, A-star etc., and using artificial intelligence technics such as deep reinforcement learning and genetic algorithm, also data mining such as Apriori algorithm for association rules. One such study focuses on collaborative optimization in robotic mobile fulfillment systems, where storage location assignment and path planning are combined into a single optimization problem. By establishing a sustainable mathematical model, proposes a location assignment strategy that incorporates goods clustering, rack turnover, reservation tables, automated guided vehicle operation rules, and an improved A-star algorithm. Simulation studies confirm the effectiveness of the approach, showcasing significant improvements in order picking efficiency, energy consumption reduction, and overall operating cost reduction [1]. The A-Star algorithm excels in automated warehousing through its real-time adaptability and precise pathfinding, optimizing storage assignments based on various operational factors. However, its efficiency can be contingent upon the accuracy of the initial model and may not dynamically respond to fluctuating demand patterns, potentially leading to suboptimal storage allocations as market trends evolve. Time series clustering addresses these limitations by analyzing historical demand data to forecast future patterns, allowing for storage assignments that adapt to anticipated changes in demand.

4) DYNAMIC PROGRAMMING

It has been harnessed for class-based storage location assignment, catering to specialized warehouse scenarios. Wang and Zhang illustrate this methodology in their case study focused on cold storage warehouses [19]. Dynamic programming optimizes assignments based on classes or categories of items, enabling energy-efficient storage location decisions. The approach considers constraints such as temperature control and item compatibility, ensuring that items are allocated to suitable locations. By utilizing dynamic programming, this research demonstrates significant improvements in energy consumption and storage utilization, highlighting the adaptability of operational research methodologies to specific warehousing needs. This method excels in making calculated decisions that account for complex operational constraints, leading to significant improvements in resource utilization. However, its reliance on predefined item classes may not fully capture the dynamic nature of demand fluctuations over time. In contrast, time series clustering addresses SLAP by analyzing historical demand data, providing a more nuanced understanding of how item demand patterns evolve.

5) NESTED ANNEALING AND HAMMING DISTANCES

This approach utilizes a Markov Chain method for initial candidate assignment sampling, followed by future-forecasted pick-round modifications according to candidate assignments, solving these as TSP. It incorporates advanced techniques like Simulated Annealing and a Hamming-distance location-swap heuristic for optimizing storage assignment. The method is designed to be layout-agnostic and introduces strategies to expedite the search for strong solution candidates, including the use of fast function approximation and algorithm restarts from local minima to enhance computational efficiency [20].

C. DATA MINING METHODS

1) ASSOCIATION RULES

Pang and Chan introduce an algorithm driven by data mining techniques, specifically designed for assigning storage locations to individual items in a randomized picker-to-parts warehouse [10]. The algorithm utilizes the extraction and analysis of association relationships among different products found in customer orders. Its primary objective is to minimize travel distances for both put-away and order-picking tasks, leading to improved efficiency overall, this approach focuses on capturing customer buying behavior by identifying patterns of frequently purchased related products. This approach is particularly beneficial in warehousing scenarios where storing correlated products in proximity can reduce order-picking time and cost. The use of data mining and the Apriori algorithm has been successful in addressing this problem. The output of this algorithm are clusters of products that must be grouped together in the warehouse. However, these algorithms do not consider the temporal aspect and are static in nature, lacking the ability to handle seasonality or non-stationary demand. However, while effective in capturing product associations, this method does not account for the temporal variations in demand, presenting a static solution that might not adapt well to seasonality or changes in demand patterns over time. In contrast, time series clustering for the SLAP offers a dynamic approach by analyzing historical demand data to adjust storage locations.

2) ASSOCIATION RULES-BASED ABC CLASSIFICATION

This approach employs ABC clustering and association rule mining from historical orders, demonstrating substantial improvements in picking and waiting times compared to random placements [5]. Comparatively, time series clustering for SLAP dynamically adjusts storage locations based on historical demand patterns, aiming for long-term operational efficiency by anticipating future storage needs. While both methods seek to optimize warehouse operations, the former emphasizes immediate efficiency gains in cold storage environments through strategic placement, whereas time series clustering focuses on adapting to changing demand patterns for general warehousing scenarios.

D. ARTIFICIAL INTELLIGENCE METHODS

1) DEEP REINFORCEMENT LEARNING (DRL)

DRL has gained significant attention in optimizing storage location assignments. Its agents are trained using historical data of storage and retrieval operations. These agents learn optimal storage location assignment strategies by interacting with the environment. Recent advancements in DRL have demonstrated remarkable efficiency improvements, especially in minimizing transportation costs and enhancing overall warehouse performance [14]. One research paper investigates the use of constrained clustering methods integrated with principal component analysis. By incorporating item characteristics and practical storage constraints, the proposed method aims to cluster stored items effectively. Additionally, the application of genetic algorithms in warehouse management systems is examined to solve the storage location assignment problem. By utilizing data analytics and a genetic algorithm, a smart logistics solution is proposed to reduce picking time. Another research article addresses the dynamic Storage Location Assignment Problem (DSLAP) by training a DRL agent on historical data of storage and retrieval operations, the study derives a suitable storage location assignment strategy. The evaluation of new data demonstrates a notable decrease in transportation costs compared to manual ABC classifications. This highlights the competitiveness of DRL as an alternative solution for DSLAP and related problems in the warehousing industry [14], [15]. DRL is highly adaptable to complex, dynamic environments. It can continuously learn and improve from new data, making it suitable for environments where conditions frequently change, or new patterns emerge. However, this approach requires significant computational resources and expertise to implement. The learning process can be time-consuming and finding the right balance.

2) GENETIC ALGORITHMS (GA)

GA are employed to tackle the storage location assignment problem by optimizing storage placements based on genetic operators. These algorithms utilize data analytics to propose intelligent logistics solutions, resulting in reduced order picking times and improved warehouse efficiency. GA-based approaches are effective when dealing with large and complex storage facilities [21].

3) MACHINE LEARNING (ML)-BASED ABC CLASSIFICATION

The study uses machine learning models, including ordinary least squares regression, regression tree, random forest, and multilayer perceptron, to predict optimal zone sizes based on factors like warehouse layout, demand characteristics, and storage and routing policies. In relation to SLAP, this paper's contribution is significant. It offers a sophisticated approach to determining the best zone sizes for ABC classified items in a warehouse, which directly impacts the efficiency of order picking processes, comparing this approach to time series clustering for SLAP, which adjusts storage locations based on historical demand patterns, presents a contrast between static and dynamic methods of optimizing warehouse operations. Time series clustering focuses on long-term efficiency and adaptability, making it suitable for environments with fluctuating demand. On the other hand, the ABC classification method discussed in this paper, while being more static, offers a direct way to achieve immediate efficiency by optimizing warehouse layout based on the current categorization of items according to their turnover rates.

4) LSTM (LONG-SHORT-TERM-MEMORY)

Dynamic storage location assignment algorithms continuously monitor order data and adapt storage locations to ensure that frequently picked items remain easily accessible. This real-time adaptability is essential for efficient and responsive warehouse operations, especially in fast-paced environments, Niu and Wang introduces a model-based deep reinforcement learning approach for a simplified storage assignment problem, leveraging an LSTM network order predictor and approximate value iteration. The algorithm effectively addresses the tradeoff between travel-time efficiency and reposition costs, outperforming random assignment and heuristics in various simulated environments [11]. In summary, the landscape of storage location assignment has undergone a significant transformation, shifting from conventional methodologies to contemporary, data-centric approaches. These advancements not only enhance warehouse operations but also play a vital role in streamlining e-commerce logistics, a crucial aspect in the continuously expanding realm of online retail. While many studies have addressed this challenge through operations research, our literature review underscores the scarcity of AI-centric investigations. Additionally, we highlight the practicality of static storage assignment using heuristics as a viable solution for determining optimal storage locations.

5) TIME SERIES CLUSTERING (OUR PROPOSAL)

Our research leverages time series clustering algorithms to analyze e-commerce order datasets, aiming to identify patterns of products frequently purchased together. By developing specialized clustering algorithms, we intend to group products based on high demand correlation, subsequently optimizing their storage location within the warehouse for efficiency in a hybrid way. To evaluate the effectiveness of our approach, we plan to simulate order preparation in 6 picking scenarios using the clustering method that achieves the highest silhouette score, focusing on key performance metrics such as total route distance and order collection travel time. These metrics will be contrasted with those derived from random and ABC storage location assignment methods, serving as foundational references for our storage location

TABLE 1. Literature review summary.

Research	Type of assignments	Leveraging TSP?	Method	Type of method	Proposal
[18]	Static	Yes	Branch and bound	Optimal	Use of Branch and Bound algorithms to enhance e-commerce warehousing by optimizing item placements, employing sophisticated heuristics, and pruning strategies.
[17]	Static	No	Pareto-based ABC	Heuristic	Pareto-based ABC classification optimizes SLAP by categorizing items, reducing forklift travel and costs.
[13]	Static	No	ML-based ABC	Predictive modeling	SLAP optimization through ML prediction of efficient ABC zone sizes.
[5]	Static	No	Association rules-based ABC	Heuristic	ABC classification using association rules for SLAP: optimizes immediate efficiency, contrasts dynamic demand adaptation.
[1]	Static	Yes	A-Star	Optimal	The use of A-Star to automate warehouse storage, focusing on path efficiency.
[19]	Static	yes	Dynamic programming	Dynamic programming	Dynamic programming optimizes cold storage assignments for energy efficiency.
[10]	Static		Association rules	heuristic	Use of data mining to optimizes storage by product association.
[14]	Dynamic	Yes	DRL	Meta-heuristic	Use of deep reinforcement learning to optimize warehouse storage, reducing costs, enhancing efficiency through learning.
[21]	Static	Yes	Genetic algorithm	Meta-heuristic	Use of evolution principles offered by genetic algorithm for dynamic warehouse storage optimization solutions.
[11]	Dynamic	Yes	DRL and LSTM	Predictive modeling	Dynamic storage assignment uses DRL and LSTM for adaptability, cutting costs by optimizing item placement in real-time.
[20]	Static	Yes	Nested Annealing	Heuristic	Iteratively improve storage assignments, focusing on minimizing total travel distances for order picking in warehouses.
This research	Hybrid	No	TS Clustering	Heuristic	Our study employs data-driven clustering to optimize warehouse storage by analyzing e-commerce patterns,

assignment theory. The study will encompass various order picking strategies and methods, with the goal of statistically validating our findings through ANOVA and Tukey's tests, thus providing a comprehensive analysis of the potential benefits of our proposed clustering-based storage location assignment model.

TABLE 1. provides a summary of the contributions considered in our literature review.

III. PROPOSED METHODOLOGY

A. INTRODUCTION TO INTELLIGENT STORAGE LOCATION ASSIGNMENT TECHNIC

Our research utilizes data-driven techniques to analyze e-commerce orders dataset, identifying patterns of co-occurring products in the same order. We will develop clustering algorithms specifically designed to group products with high demand correlation. Subsequently, these clusters will be optimally assigned to storage locations within the warehouse. To gauge the effectiveness of this approach, we will simulate the order preparation of items using the clustering method that yields the highest silhouette score. Metrics such as total route distance to collect orders will be employed. This routing metrics will be compared with random location and ABC location assignment methods, As the foundational references for storage location assignment theory. Various scenarios related to picking strategies and methods will be considered. To statistically validate our hypothesis, we will use ANOVA and Tukey's tests.

B. Assumptions

- **Data Assumption:** The data used for analysis consists of univariate time series, representing the variation in demand for each product over time.
- **Storage Constraints:** The model assumes no constraints related to storage, weight limitations, or compatibility between stored items.
- **Product Assignment:** Each product is assigned to only one position within the warehouse. There is a one-to-one mapping between products and storage positions.
- **Single Depot:** The model considers a single depot or storage grouping for all products.
- **The number of batches (Batch picking strategy):** we assume that the picker can collect items for 6 orders simultaneously.
- **The number of zones (Zone picking strategy):** we assume that the number of zones is equivalent to the number of clusters identified by our clustering method.
- **Metric:** route distance in meter (m)
- **Distance Calculation:** The distance calculation considers only the proximity between individual items within the warehouse, and not between the depot and items.
- **Sorting time (Batch and zone picking strategies):** our proposal doesn't cover the sorting time impact.

C. ORDER TIME SERIES CLUSTERING

The selection of the clustering algorithm is contingent upon the nature of the data. We opt for algorithms that are suitable for univariate time series analysis such as: K-mean, Self-Organizing Map (SOM), and Hierarchical Clustering (HC). The silhouette scores of these algorithms are compared, and the one demonstrating the highest compatibility will be retained for subsequent phases of our study.

1) K-MEAN TS CLUSTERING

K-means algorithm minimizes the sum of squared distances between data points and their assigned cluster centroids, by minimizing the following objective function:

$$Min\ Z = \sum\nolimits_{i=1}^{k} \sum\nolimits_{j=1}^{n_i} |x_{ij} - c_i|^2 \tag{2}$$

With:

- x_{ij}: Data points
- c_i: Centroid of the cluster
- $|x_{ij} - c_i|$: Distance metric, Represents the dissimilarity between a data point 'x_{ij}' and the centroid c_i of its assigned cluster. We use Euclidean distance or Dynamic Time Warping
- k: Number of Clusters, we identify the optimal number of clusters using the Elbow Method

In our context, each data point represents an order time series. The algorithm iteratively assigns orders to clusters based on the similarity of their time series patterns. We aim to identify clusters that exhibit cohesive temporal behaviors, allowing us to discern patterns in the order data. The centroid of each cluster serves as a representative profile, capturing the common characteristics of orders within that cluster, we use for our case study both Euclidean and DTW metric.

2) SELF-ORGANIZING MAP (SOM)

We use SOM algorithm to cluster order time series by minimizing the following objective function:

$$Min\ Z = \lambda\ x\ Topological\ Error + \left(1 - \lambda\right)\ Quantization\ Error \tag{3}$$

With:

- 'λ' Is the is a weight parameter between 0 and 1 that determines the balance between preserving topology and minimizing quantization error.

The topological error component of the objective function relates to how well the SOM preserves the temporal relationships between different orders, ensuring that orders with similar temporal patterns are grouped together on the SOM map. Simultaneously, the quantization error measures how accurately the SOM represents clusters of orders with similar time series patterns. Our goal is to uncover and visualize meaningful patterns in the order time series data, emphasizing the preservation of temporal relationships and the accurate representation of order clusters. During the training process, the SOM dynamically adjusts its nodes to minimize the combined objective function, facilitating the identification of clusters of orders with similar temporal characteristics [22].

The grid size we used for SOM is the square of dataset size. It has a direct impact on the number of clusters which we can't parameter directly (the case of hierarchical clustering and K-mean TS clustering)

3) AGGLOMERATIVE HIERARCHICAL CLUSTERING (AHC)

The objective is to group orders with similar temporal patterns, providing insights for warehouse optimization. The dissimilarity matrix $Dissim\,(x, y)$ plays a pivotal role within the HC, capturing pairwise dissimilarities between each order's time series. For distance measurement, we opt for the *Euclidean metric* due to its efficiency, avoiding resource-intensive alternatives like Dynamic Time Warping. The mathematical model involves initializing singleton clusters, computing the dissimilarity matrix, and iteratively merging clusters based on the complete linkage method. The objective function for linkage is defined as:

$$CompleteLinkage = max\{Dissim\,(x, y) \tag{4}$$

With:

- $x \in C_i$: x datapoint from time series 'C_i'
- $y \in C_j$: y datapoint from time series 'C_j'

reflecting the maximum dissimilarity between individual time series in two clusters. This process continues until all orders are part of a single cluster. Visualization is achieved through a dendrogram, providing a hierarchical representation of cluster merging at different dissimilarity levels.

4) MODEL SELECTION AND VALIDATION

In our comprehensive study, we compared various clustering algorithms. Since the findings from the benchmark study on time series clustering revealed no single algorithm consistently excels across all datasets, underscoring the importance of dataset-specific algorithm selection [23]. The Adjusted Rand Index highlighted significant variability in the performance of partitional, hierarchical, and density-based methods. Conversely, another study demonstrated how silhouette scores help the time series model selection and validation, by evaluating cluster compactness and separation [24]. Concretely, we compared the silhouette scores of the aforementioned models. This score provided a quantifiable measure of how well each algorithm captured the inherent structure of our dataset, emphasizing the compactness and separation of clusters. The algorithm yielding the highest silhouette score was selected for further analysis and simulation, as it demonstrated the highest compatibility with our dataset's nature.

D. SIMULATION ENVIRONMENT & SCENARIOS

Order picking is a logistical process where items are selected from a warehouse or storage location to fulfill customer orders. Our simulation environment meticulously incorporates two distinct picking policies; S-shape and Midpoint, the S-shape is highlighted as the most efficient beside the VRP in the case study made by Shetty et al. [25] and the Midpoint as the worst one among 4 picking policies, in order to diversify the scenarios of our proposal. We also consider three diverse picking strategies: Discrete Order Picking, Batch picking, and Zone picking. A comprehensive approach allows us to thoroughly assess the impact of our proposal across a spectrum of scenarios, ensuring adaptability and versatility. In addition to its accessibility and straightforward implementation, our simulation environment underscores the notion that the potency of artificial intelligence isn't solely derived from its

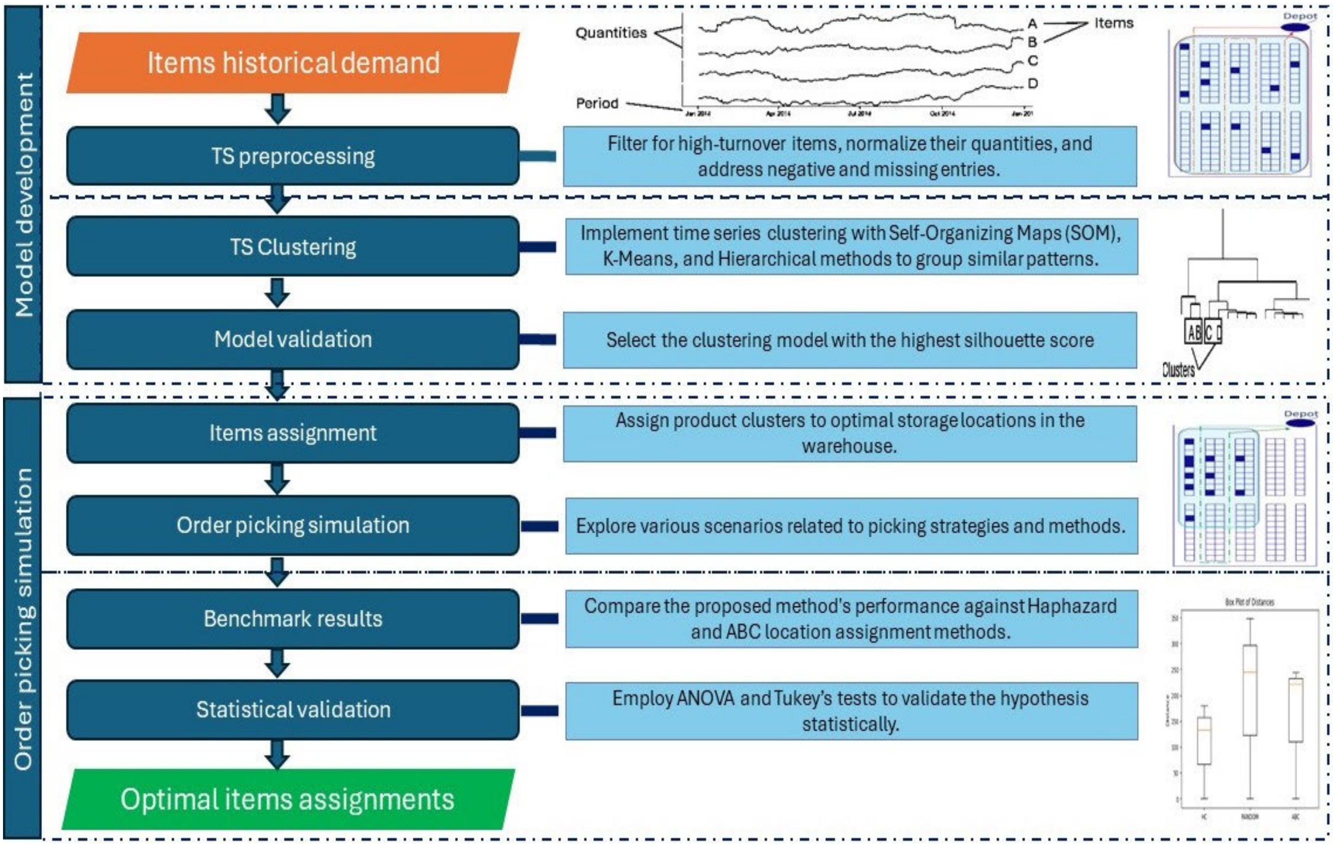

FIGURE 3. Intelligent storage location assignment model.

complexity. Instead, it shines in its ability to address various scenarios effectively, demonstrating that the true power of artificial intelligence lies in its practical versatility. Studying static storage location assignment involves items maintaining a fixed position over time without changing with each reception or put-away. We've simulated a warehouse with 788 positions, 19 aisles, and bays of 20 positions each, adopting a mono-block setup (no cross-aisle).

1) S-SHAPE PICKING POLICY

In the S-shape policy, the route taken by the order pickers forms the shape of an S. This means that any aisle containing at least one pick is traversed entirely by the order picker, as depicted in **FIGURE** *4* Aisles devoid of picks are bypassed, and the order picker returns to the drop-off location (depot) starting from the last visited aisle.

2) MID-POINT PICKING POLICY

In the mid-point policy, the warehouse is divided into two halves. Picks from the bottom half are retrieved from the bottom cross aisle, while picks from the top half are retrieved from the top cross aisle, which is illustrated in **FIGURE** *4*. If the number of picks per aisle is small, this policy provides better results than the S-shape policy.

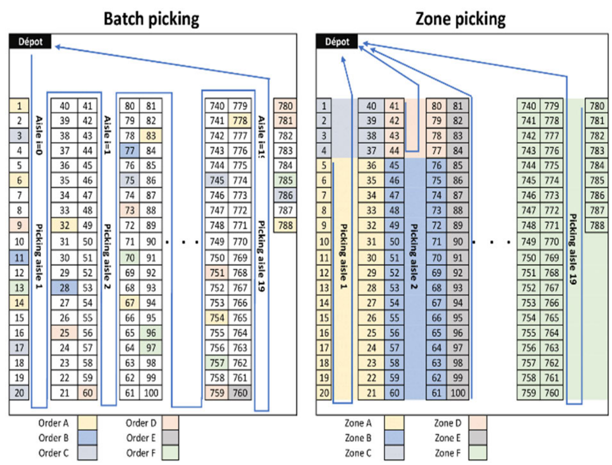

FIGURE 4. *S*-Shape and Mid-Point picking policies.

3) DISCRETE ORDER PICKING STRATEGY

This represents the most prevalent form of order picking due to its fundamental and straightforward nature. In a discrete order picking approach, each order-picker handles a single order, addressing one line at a time. The benefits of employing this order picking method include its simplicity, suitability for paper-based picking systems, swift response times for order fulfillment, and the ease of tracking order picker accuracy, and doesn't need sorting.

4) BATCH PICKING STRATEGY

Batch picking is a strategy where a single picker handles a group or batch of orders simultaneously [26] as depicted in **FIGURE 5**. This approach proves beneficial when multiple orders share the same item, allowing the order picker to make a single trip to the pick location for that specific item, fulfilling multiple orders efficiently. The primary advantage of choosing batch picking lies in reduced travel time, contributing to increased overall productivity.

5) ZONE PICKING STRATEGY

In the zone picking strategy, each order picker is assigned a distinct and physically defined zone within the pick area. The assigned picker is responsible for retrieving all items located in their designated zone for each order as depicted in **FIGURE 5**. If an order includes items from multiple zones, the order is sequentially processed through each zone, a method commonly known as ''pick and pass''. The goal of our proposal is to reduce sorting time by reducing the number of zones engaged to fulfill the order and the decrease the total lead time to fulfill order.

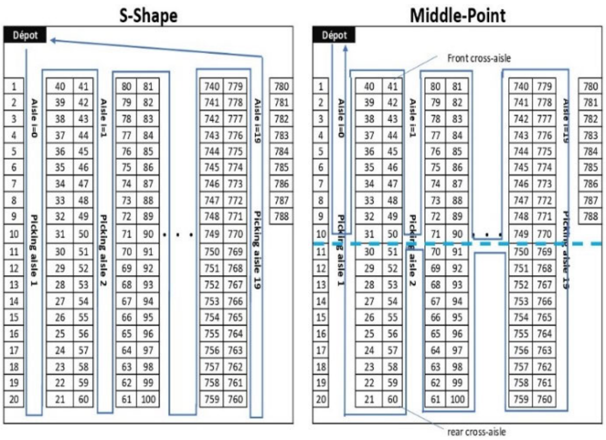

FIGURE 5. Batch picking and Zone picking strategies.

IV. EXPERIMENT RESULTS AND ANALYSIS

This section outlines our computational experiments, validating the proposed model from the preceding section. For our SLAP study, we utilized Google Colab's Nvidia RTX 3060 GPU and Intel Xeon(R) Platinum 8259CL CPU @ 2.50GHz with 16 cores. This setup enabled efficient execution of SOM and K-Means DTW algorithms leveraging GPU [27]. Utilizing a real-world dataset from Kaggle. Initially, we employ our developed clustering algorithms to group items typically shipped together in the same order. Subsequently, we simulate order picking under various scenarios, comparing mean time for order collection. The section concludes with statistical validations. With a clean and focused dataset, our implementation of developed clustering algorithms on Google Colab efficiently grouped items that are typically shipped together, leveraging the computational power and simplicity of the integrated Jupyter notebook, with the assistance of ChatGPT in debugging the code [28]. The simulation of various order picking scenarios, informed by these clusters, provided valuable insights into optimizing order collection times. Statistical validations underscored the robustness of our findings, offering a compelling case for the applied methodologies. After implementing each of the four models, we meticulously evaluated their performance, ultimately selecting the model that yielded the highest silhouette score. This score served as a quantitative measure of how well each model clustered the dataset, with a higher silhouette score indicating a better fit of the model to the data's inherent structure. The chosen model was identified as the most compatible with our dataset's nature, effectively capturing the nuances of item groupings based on their co-occurrence in orders, to showcase our approach's significance in picking optimization through simulations across six varied scenarios.

A. ITEM CLUSRERING

1) DATASET DESCRIPTION AND PREPROCESSING

Our study utilizes a comprehensive e-commerce dataset sourced from Kaggle [29], featuring 541,909 order lines spanning over a business period from January 10, 2011, to September 9, 2011. The dataset encompasses a wide array of products, totaling 4,070 unique items across 25,900 distinct orders. The data attributes include InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. These attributes offer a multifaceted view of the retail operations, from transaction details to item specifics and customer geography.

a: DATASET PREPROCESSING

The preprocessing phase was critical in refining the dataset for our clustering and simulation analyses. Our approach was methodical, aimed at ensuring data integrity and relevance for high-impact insights. The steps undertaken include:

Data cleansing: We began by removing entries with negative quantities to exclude returns or inventory adjustments, ensuring our analysis focused on actual sales data. Additionally, records with missing values, particularly in key fields such as CustomerID and Description, were omitted to maintain data consistency and reliability.

Focusing on high-runner Items: To align our analysis with the most commercially impactful products, we applied the Pareto principle, isolating the top 788 items by turnover, known as the ''Pareto head.'' This concentration allowed for a more targeted examination of products driving the majority of sales. Temporal Transformation: Invoice dates were converted into minutes of the year to achieve a uniform temporal scale, facilitating more nuanced temporal analyses and clustering based on time of purchase.

b: DATA STANDARDIZATION

We structured monovariate dataframes for each of the Pareto items, focusing on essential attributes for clustering, such as quantity and invoice data. To ensure consistency across

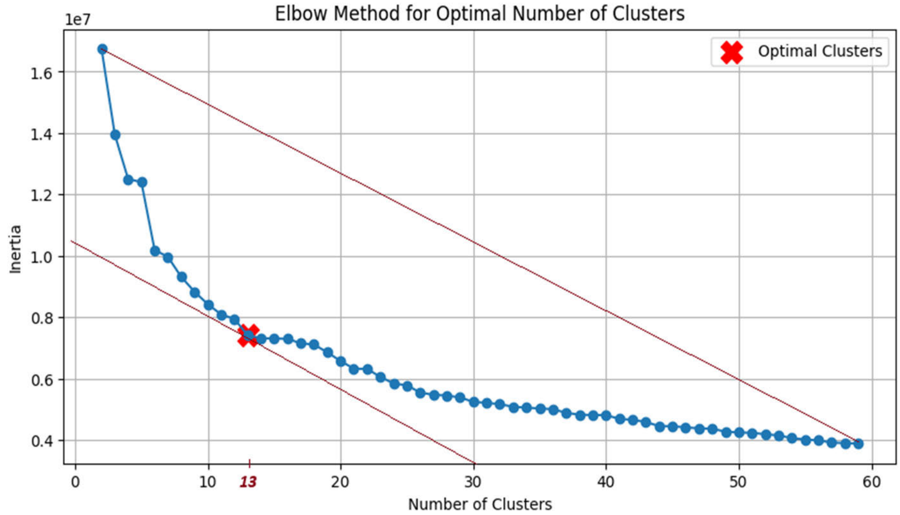

FIGURE 6. **Elbow method.**

these dataframes, we applied padding where necessary to standardize their lengths.

c: DATA NORMALIZATION

Recognizing the variance in quantity scales across items, we employed Min-Max scaling to normalize these values. This normalization was crucial for our clustering algorithms, ensuring that differences in order magnitude did not bias the analysis.

The preprocessing steps significantly enhanced the dataset's suitability for our objectives. By focusing on high-turnover items and ensuring data quality and uniformity, we laid a solid foundation for extracting meaningful patterns and insights. This refined dataset not only facilitated effective item clustering but also allowed for the simulation of order picking scenarios with greater accuracy and relevance to real-world operations.

2) K-MEAN CLUSTERING

We identified the optimal number of clusters through the elbow method '**FIGURE *6***', As explained by Delgado et al., the elbow is identified by the tangent to the curve and the parallel to the straight line connecting the first and the last point on the curve [30] revealed that 13 clusters are optimal for our analysis. This information guided the parameterization of both K-means and AHC [31].

In the analysis, we employed both Euclidean and DTW metrics for K-means, finding that the Euclidean metric outperformed DTW in terms of computational efficiency but the negative silhouette score suggests that the clusters may be poorly defined or overlapping. DTW, while effective, was notably slower compared to other clustering algorithms utilized in our study, and the low positive silhouette score indicates slightly better-defined clusters compared to Euclidean distance, but still not strongly distinct.

3) AGGLOMERATIVE HIERARCHICAL CLUSTERING (AHC)

In our analysis, we applied AHC using the complete linkage criterion to dissect our dataset. This approach determines cluster dissimilarity by identifying the greatest distance between points in separate clusters. Guided by the elbow method's insights, we partitioned the data into 13 distinct clusters, achieving an optimal silhouette score. The truncated dendrogram, depicted in **FIGURE 7**, meticulously illustrates the grouping of items based on their mutual characteristics, revealing the data's hierarchical organization. This division into 13 clusters is a deliberate choice, mirroring the methodology outlined in the case study by Xu and Beard [31]. The silhouette score obtained, which quantifies the clustering effectiveness, is marked as 0.22, underscoring the method's robustness in identifying coherent product groupings.

4) SELF-ORGANIZING MAP (SOM)

We conducted training for 100 epochs with the implementation of an early stopping mechanism to prevent overfitting. This precaution is crucial, particularly when dealing with datasets that share similar characteristics to avoid the risk of grouping all items into a single cluster. Through experimentation, we identified the optimal hyperparameters for the SOM model, including a best sigma value of 0.1 and a learning rate of 1.0. Additionally, the grid size was determined as the square root of the total number of items, resulting in the creation of 9 distinct clusters **FIGURE 8**. The positive silhouette score suggests relatively well-separated clusters with instances belonging clearly to their assigned clusters.

5) MODEL SELECTION AND VALIDATION

The clustering outcomes are concisely presented in the TABLE 2. The AHC method that exhibits superior clustering performance, as evidenced by the silhouette score, is carefully selected for further simulations.

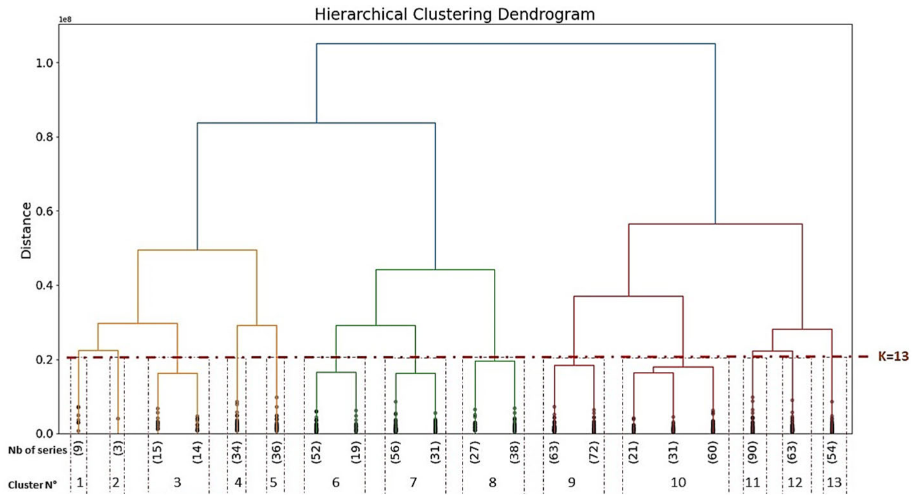

FIGURE 7. Hierarchical Clustering Dendrogram.

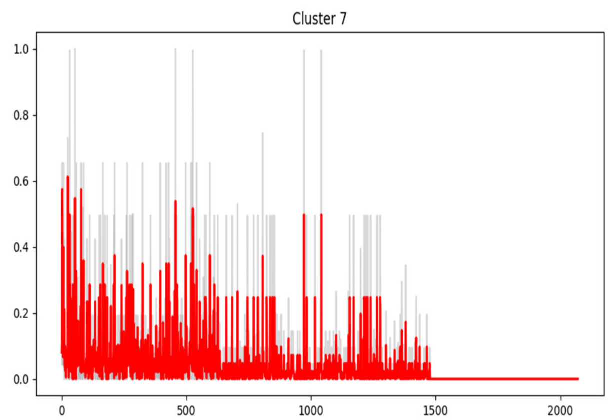

FIGURE 8. Clustered demand time series using SOM.

TABLE 2. Summary of clustering.

Algorithm	Number of clusters	Execution time	Silhouette score
K-mean Euclid.	13	19s	-0.01
K-mean DTW	13	189s	0.01
SOM	9	40s	0.17
AHC	**13**	**34s**	**0.22**

B. ORDER FULFILMENT SIMULATIONS AND ANALYSIS

In this simulation phase, we are comparing the Clustering-based storage location assignment, employing the AHC model that we selected for its superior silhouette score, against two alternative approaches: Random storage assignment and the traditional ABC storage assignment. The evaluation metric employed is the distance in meters, and we assess the Mean Time to Collect Orders. The simulation specifically includes orders containing the top Pareto items selected during data preprocessing. The order picking simulation encompasses 788 picking positions, 19 aisles, and follows a mono-block setup (without cross aisles). Various picking policies and strategies are considered, resulting in a total of six scenarios for our study.

In the ABC clustering scenario, items are classified into classes A, B, and C based on turnover. Class A includes high-priority items assigned to positions 1-190, Class B comprises items with moderate turnover in positions 191-570, and Class C involves items with lower turnover beyond position 570. For random assignment, products are distributed randomly throughout the warehouse. We simulate the order picking from the initial dataset 320150 order lines that contains these 788 Items. To evaluate and compare the effectiveness of our approach using statistical methods, we employ a one-way ANOVA to test the equality of means across multiple groups. Subsequently, we apply Tukey's multiple comparison test to identify which storage assignment demonstrates superior performance in terms of average travel distance when compared pairwise. In the context of ANOVA, the null hypothesis posits that all travel time means are equal, while the alternative hypothesis suggests that at least one mean differs. Symbolically, $H0 = u1 = u2 = u3$, and Ha indicates that at least one pair of means differs from each other. Here, H0 and Ha represent the null and alternate hypotheses, respectively, and u1, u2, and u3 denote the population means of travel distances for three location assignment approaches. A significance level of 5% is chosen, and the analysis assumes equal variance.

1) S-SHAPE: DISCRETE ORDER PICKING STRATEGY

Mean distance to collect orders for the first scenario:

- Random: 404.86 m
- ABC: 387.50 m
- AHC: 358.91m

Notably, AHC outperforms Random by 11.38% and demonstrates a 7.5% improvement compared to the ABC strategy

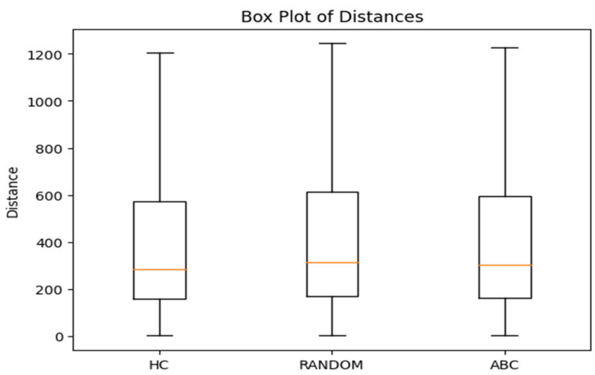

FIGURE 9. Travel distance Boxplot for first scenario.

FIGURE 9. These results highlight the effectiveness of our approach in optimizing order collection distances within the given warehouse context.

The ANOVA one-way rejected the null hypothesis, means that there is a significant difference in mean distance to collect orders in between various assignment methods, and Tukey's test confirm the effectiveness of our approach TABLE 3.

TABLE 3. Tukey simultaneous tests for first scenario.

Scenario 1	Mean-diff	P-adj	95% CI	Reject
ABC-AHC	-28.59	0	(-35.13, -22.05)	True
ABC-Random	17.35	0	(10.80, 23.89)	True
AHC-Random	45.94	0	(39.40, 52.49)	True

2) S-SHAPE: BATCH PICKING STRATEGY

Mean distance to collect orders for the second scenario:

- Random: 157.17m
- ABC: 69.15m
- AHC: 66.25m

In comparison to ABC, the AHC assignment exhibits superior performance by 4.34%, while surpassing Random by a significant margin of 57%. Notably, the maximum distance in the Random approach reaches 229, whereas both AHC and ABC limit this to 106 and 109, respectively. However, it's worth noting that there are some aberrant distances exceeding 200 meters in all three strategies.

Interestingly, ABC and AHC showcase relatively similar performance levels, especially in the context of batch picking within the given parameters, and this is confirmed by the fact that the null hypothesis for AHC and ABC cannot be rejected based on the results of Tukey's TABLE 4.

TABLE 4. Tukey simultaneous tests for second scenario.

Scenario 2	Mean-diff	P-adj	95% CI	Reject
ABC – AHC	-2.9	0.4	(-8.18, 2.38)	False
ABC-Random	88.01	0	(82.72, 93.29)	True
AHC-Random	90.91	0	(85.62, 96.19)	True

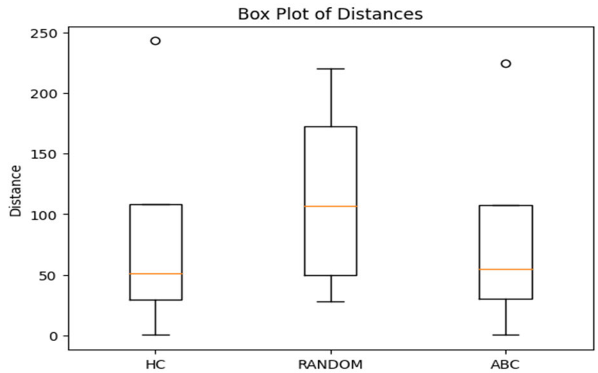

FIGURE 10. Travel distance Boxplot for scenario 2.

Our approach outperforms Random by 61% and ABC by 43%. The clusters corresponding to zones in our approach exhibit better results than random storage and ABC assignment.

3) S-SHAPE: ZONE PICKING STRATEGY

In this scenario, our approach demonstrates superior results:

- Random: 348.47m
- ABC: 244.75m
- AHC: 133.82m

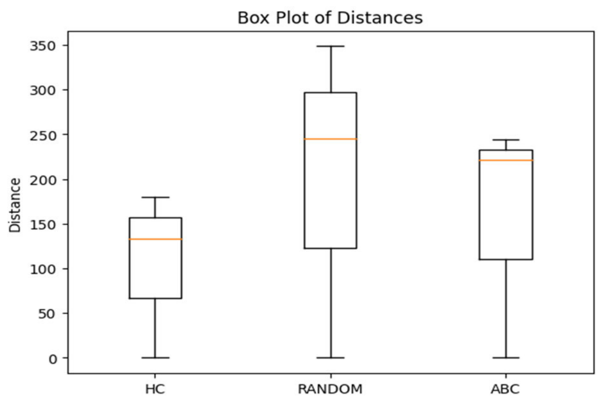

FIGURE 11. Travel distance Boxplot for scenario 3.

The interval of distances is lower in AHC compared to the others FIGURE 11 contributing to better flow management and aiding in balancing the load capacity within the warehouse. This characteristic of AHC enhances its ability to optimize the distribution of distances, leading to more efficient operations and improved overall capacity utilization in the warehouse.

This superiority is further validated by Tukey's and ANOVA tests TABLE 5, affirming the effectiveness of our storage assignment approach in optimizing travel distances.

4) MID-POINT: DISCRETE ORDER PICKING STRATEGY

Mean distance to collect orders for the fourth scenario:

TABLE 5. Tukey simultaneous tests for third scenario.

Scenario 3	Mean-diff	P-adj	95% CI	Reject
ABC - AHC	-108.96	0	(-115.59, -102.34)	True
ABC-Random	-12.39	0	(-19.01, -5.77)	True
AHC-Random	96.57	0	(89.95,103.19)	True

- Random: 385.79m
- ABC: 423.45m
- AHC: 313.58m

In comparison to Random, AHC outperforms by 18%, and against ABC, it shows a superior performance by 26%. These percentages highlight the efficiency gains our approach, particularly AHC, achieves over both Random and ABC in scenario 4. And the results are confirmed by Tukey's and ANOVA tests ***TABLE 6***, Also The interval of distances is lower in AHC compared to the others.

TABLE 6. Tukey simultaneous tests for 4th scenario.

Scenario 4	Mean-diff	P-adj	95% CI	Reject
ABC - AHC	-109.87	0	(-116.70, -103.04)	True
ABC-Random	-37.66	0	(-44.49, -30.83)	True
AHC-Random	72.21	0	(65.38, 79.04)	True

5) MID-POINT: BATCH PICKING STRATEGY

Mean distance to collect orders for the fifth scenario:

- Random: 193.56m
- ABC: 213.00m
- AHC: 66.50m

AHC outperforms ABC by 69%, and against Random, it demonstrates a superior performance by 65%. These percentages indicate the efficiency gains of HC over both ABC and Random in terms of mean distances in this scenario.

The statistical analysis indicates that there is a significant difference between AHC and ABC TABLE 7, as well as between AHC and Random in terms of mean distances, but there is no significant difference is observed between ABC and Random.

TABLE 7. Tukey simultaneous tests for 5th scenario.

Scenario 5	Mean-diff	P-adj	95% CI	Reject
ABC -AHC	-146.5	0	(-228.90, -64.09)	True
ABC-Random	-19.44	0.8	(-101.84, 62.96)	False
AHC-Random	127.05	0	(44.65, 209.46)	True

6) MID-POINT: ZONE PICKING STRATEGY

In the final scenario:

- Random: 341.03m
- ABC: 1183.46m
- AHC: 723.32m

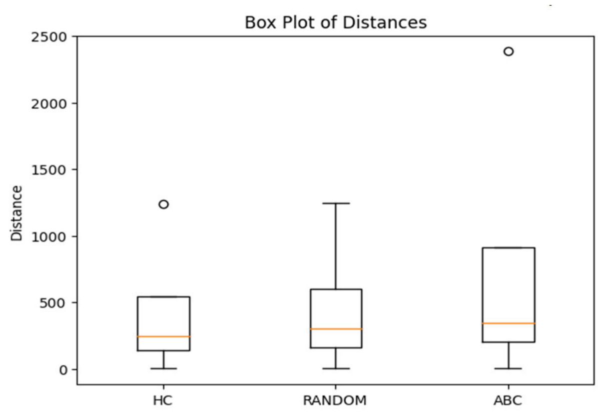

FIGURE 12. Travel distance Boxplot for scenario 4.

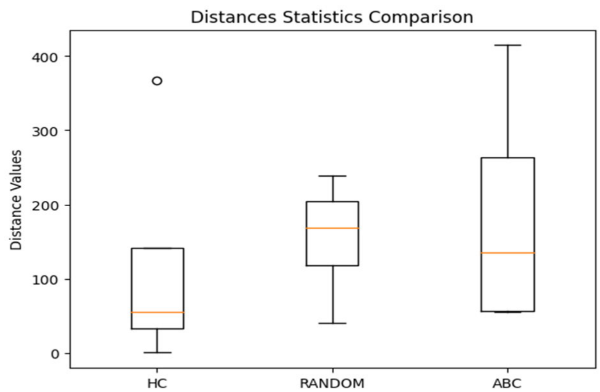

FIGURE 13. Travel distance Boxplot for scenario 5.

TABLE 8. Tukey simultaneous tests for 6th scenario.

Scenario 6	Mean-diff	P-adj	95% CI	Reject
ABC - AHC	-108.96	0	(-115.59, -102.34)	True
ABC-Random	-12.39	0	(-19.01, -5.77)	True
AHC-Random	96.57	0	(89.95,103.19)	True

The performance of our model is not satisfactory in this scenario, and similarly, ABC does not align well with the midpoint zone picking strategy **FIGURE *14***, the results are summarized in TABLE 8. This can be attributed to the fact that zones are separated by the midpoint policy, forcing pickers to start in the front aisle and then move to the back aisle within the same zone. This practice increases the overall travel distance. To enhance the performance of these approaches, it is essential to adapt zones for the front aisle and separate zones for the back aisles. This adjustment can potentially optimize the picking strategy and reduce travel distances of our proposal in such scenarios to fit the result we have got in the S-Shape policy.

C. SUMMARY OF THE RESULTS

The results could be summarized for S-shape in TABLE 9:

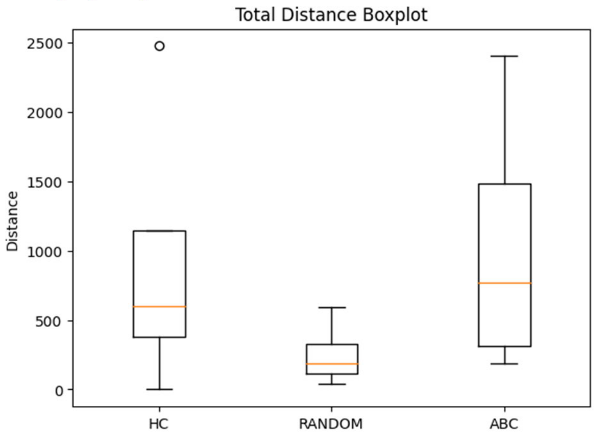

FIGURE 14. Travel distance Boxplot for scenario 6.

TABLE 9. Summary of results S-shape.

S-Shape	Discrete Picking	Batch Picking	Zone Picking
Random	404.86	157.17	348.47
ABC	387.50 (-4%)	69.15 (-56%)	244.75 (-30%)
AHC	358.91 (-11%)	66.25 (-58%)	133.82 (-62%)

And results of Mid-point policies in TABLE 10:

TABLE 10. Summary of results Mid-point.

Mid-Point	Discrete Picking	Batch Picking	Zone Picking
Random	385.79	193.56	341.03
ABC	423.45 (+10%)	213.2 (+10%)	1183 (+247%)
AHC	313.58(-18%)	66.5 (-66%)	723.3 (+113%)

V. DISCUSSION

In this work, we've explored various methodologies for optimizing storage location assignments, drawing on both traditional operational research methods and advanced artificial intelligence techniques. We delved into the potential of Pareto-based ABC classifications, Branch and Bound Algorithms, and the A-Star Algorithm, each offering unique advantages for warehouse efficiency. Additionally, we examined data mining methods like association rules and AI approaches, including DRL and GA for their adaptability and efficiency in dynamic environments. Our proposal stands out by focusing on a data-driven clustering approach, validated through comprehensive computational experiments. By employing clustering algorithms and simulating order preparation using methods yielding the highest silhouette scores, we've showcased substantial improvements in order picking performance. Our model's success, particularly in AHC scenarios, demonstrates its superiority over random and ABC methods in various picking strategies, confirmed by rigorous statistical analysis. This section aims to highlight the innovative aspects of our research, particularly our clustering-based model's ability to enhance efficiency and adapt to complex warehouse operations. In doing so, we acknowledge the contributions and limitations of existing works, setting a foundation for our study's significance in advancing storage location assignment strategies. Our work not only contributes to the theoretical landscape but also offers practical insights for implementing AI-driven logistics solutions in real-world warehousing scenarios. There is potential for further optimization of our clustering algorithm by hyperparameter tuning, including adjusting the number of epochs, finding the optimal grid size for SOM, and selecting the best linkage model for AHC, could enhance overall performance. Additionally, it's essential to acknowledge that the nature of our dataset, with closely similar demand patterns among items, poses a challenge, explained by the low silhouette score. Despite this, our clustering-based storage location assignment method remains a powerful tool for optimization. Continuous refinement, both in hyperparameter tuning and adapting to specific dataset characteristics, is key to unlocking even better results. Our time series clustering approach for storage location assignment represents a novel hybrid model within the spectrum of solutions discussed in the literature. Unlike purely static methods like Branch and Bound, Pareto-based ABC, ML-based ABC, Association rules-based ABC, A-Star, and Dynamic Programming, our method dynamically analyzes demand patterns through time series clustering, enabling more responsive storage decisions. This approach allows for adaptation to demand fluctuations without the constant need for physical reassignments, addressing a key limitation of static methods. Furthermore, compared to purely dynamic methods that require frequent and resource-intensive changes to storage positions, our hybrid model offers a balanced solution by incorporating dynamic analysis with less frequent implementation phases, thereby conserving operational resources.

Additionally, our method distinguishes itself by adopting a heuristic approach, aiming to provide an optimal solution with less computational resource consumption. This contrasts with more complex or exhaustive techniques like genetic algorithms or deep reinforcement learning, which, while powerful, may demand significant computational power and expertise. By leveraging the strengths of both static and dynamic methodologies and utilizing a heuristic for efficient computation, our time series clustering approach offers a promising alternative for storage location assignment, particularly in environments where demand patterns exhibit variability, but operational flexibility is constrained. Our study stands out from the existing literature by comprehensively covering six distinct scenarios related to various picking methods and strategies. This extensive coverage ensures a holistic understanding of how our time series clustering approach can be optimized across different operational contexts. Unlike other studies that may focus on a singular aspect of warehouse operations or apply a one-size-fits-all approach, our research delves into the nuances of multiple picking scenarios. This includes examining the efficiency of batch picking, zone picking, and different strategies within these frameworks, such as S-Shape and Mid-Point picking. This diversity in scenario analysis highlights our study's unique contribution to the field, offering insights into the adaptability and effectiveness of our method across a range of warehouse activities, thereby providing a richer, more nuanced understanding of its potential benefits and applicability.

The results of simulations consistently highlight the prowess of our ISLA approach in optimizing order picking performance across various scenarios. In S-Shape policies, AHC consistently outperforms both random and ABC, showcasing efficiency gains ranging from 4.34% to 61%. These improvements are substantiated by statistical tests, affirming the significance of our proposal. In Mid-Point scenarios, AHC demonstrates remarkable efficiency gains, surpassing random by 18% to 69% and ABC by 26% to 65%. Tukey's and ANOVA tests further validate the effectiveness of AHC in these contexts. However, in the 6th scenario involving Zone picking with a Mid-Point constraint, our model faces challenges due to the chevauchee between implemented zones and the routing policy. Without addressing this constraint, our proposed method might not yield optimal results, as demonstrated in the superior performance observed in S-Shape Scenario 3. To address this constraint, we recommend adapting zones for front and back aisles, potentially aligning the results with the success observed in S-Shape Scenario 3.

VI. CONCLUSION

In conclusion, our study represents a significant advancement in addressing SLAP within the realm of e-commerce warehouse management. By introducing and validating the Intelligent Storage Location Assignment model, we have not only demonstrated its robustness in optimizing warehouse operations but also highlighted its adaptability to diverse logistical scenarios. Our approach stands out for its innovative integration of time series clustering and spatial analysis, tailored specifically for the dynamic requirements of e-commerce warehouses. The strategic value of our research lies in its potential to revolutionize e-commerce logistics by offering a scalable, flexible solution to storage location assignment that can significantly enhance order fulfillment processes. With its proven effectiveness, particularly in scenarios employing S-Shape policies, our method paves the way for more efficient, cost-effective warehouse management practices. Looking ahead, we are poised to explore further optimizations, with a keen interest in refining Mid-point policies and Zone picking strategies. Our future work will also delve into the integration of collaborative filtering and the inclusion of more granular warehouse and product characteristics to fine-tune our model. This ongoing research endeavor aims not only to refine the practical applicability of our approach but also to contribute to the broader academic and operational discourse on e-commerce logistics management. This study, therefore, does not merely present a novel approach to SLAP but also lays the groundwork for subsequent innovations in warehouse management. It underscores the critical role of advanced, data-driven models in enhancing the efficiency and effectiveness of e-commerce warehouse operations, setting a new benchmark for future research in this field.

ACKNOWLEDGMENT

To their dedicated project team, sincere appreciation for their collaborative efforts. Special thanks to Kaggle for providing the dataset they used to evaluate the performance of their approach and Google Colab for offering the best coding environment. Additionally, they acknowledge the invaluable assistance of ChatGPT in code generation for clustering and simulation. Grateful for their indispensable contributions.

REFERENCES

[1] J. Cai, X. Li, Y. Liang, and S. Ouyang, "Collaborative optimization of storage location assignment and path planning in robotic mobile fulfillment systems," *Sustainability*, vol. 13, no. 10, p. 5644, May 2021, doi: 10.3390/su13105644.

[2] B. J. Ma, Y. Jiang, Y.-H. Kuo, and G. Q. Huang, "RubikCell: Towards robotic cellular warehousing system for e-commerce logistics," *SSRN Electron. J.*, p. 3, Jan. 2024, doi: 10.2139/ssrn.4136564.

[3] N. Chbaik, A. Khiat, A. Bahnasse, and H. Ouajji, "The application of smart supply chain technologies in the Moroccan logistics," *Proc. Comput. Sci.*, vol. 198, pp. 578–583, 2022, doi: 10.1016/j.procs.2021.12.289.

[4] Y. Zhang, "Correlated storage assignment strategy to reduce travel distance in order picking," *IFAC-PapersOnLine*, vol. 49, no. 2, pp. 30–35, 2016, doi: 10.1016/j.ifacol.2016.03.006.

[5] M. F. Benaglia, M.-H. Chen, S.-H. Lu, K.-M. Tsai, and S.-H. Hung, "Improving the picking efficiency of a cold warehouse to avoid temperature abuse," *Int. J. Logistics Manag.*, p. 1, 2024, doi: 10.1108/IJLM-01-2023-0044.

[6] H. Kalkha, A. Khiat, A. Bahnasse, and H. Ouajji, "The rising trends of smart e-commerce logistics," *IEEE Access*, vol. 11, pp. 33839–33857, 2023, doi: 10.1109/ACCESS.2023.3252566.

[7] J. J. R. Reyes, E. L. Solano-Charris, and J. R. Montoya-Torres, "The storage location assignment problem: A literature review," *Int. J. Ind. Eng. Computations*, vol. 10, no. 2, pp. 199–224, 2019, doi: 10.5267/j.ijiec.2018.8.001.

[8] K. Haricha, A. Khiat, Y. Issaoui, A. Bahnasse, and H. Ouajji, "Recent technological progress to empower smart manufacturing: Review and potential guidelines," *IEEE Access*, vol. 11, pp. 77929–77951, 2023, doi: 10.1109/ACCESS.2023.3246029.

[9] Y. Issaoui, A. Khiat, K. Haricha, A. Bahnasse, and H. Ouajji, "An advanced system to enhance and optimize delivery operations in a smart logistics environment," *IEEE Access*, vol. 10, pp. 6175–6193, 2022, doi: 10.1109/ACCESS.2022.3141311.

[10] K.-W. Pang and H.-L. Chan, "Data mining-based algorithm for storage location assignment in a randomised warehouse," *Int. J. Prod. Res.*, vol. 55, no. 14, pp. 4035–4052, Jul. 2017, doi: 10.1080/00207543.2016.1244615.

[11] F. Niu and Z. Wang, "model based reinforcement learning for simplified storage assignment optimization," pp. 1–12, 2019. [Online]. Available: https://github.com/znwang25/SmartStorage

[12] J. B. Zuñiga, J. A. S. Martínez, T. E. S. Fierro, and J. A. M. Saucedo, "Optimization of the storage location assignment and the picker-routing problem by using mathematical programming," *Appl. Sci.*, vol. 10, no. 2, p. 534, Jan. 2020, doi: 10.3390/app10020534.

[13] A. Silva, K. J. Roodbergen, L. C. Coelho, and M. Darvish, "Estimating optimal ABC zone sizes in manual warehouses," *Int. J. Prod. Econ.*, vol. 252, Oct. 2022, Art. no. 108579, doi: 10.1016/j.ijpe.2022.108579.

[14] C. Li and M. Qiu, *Reinforcement Learning for Cyber-Physical Systems: With Cybersecurity Case Studies*, 1st ed. Boca Raton, FL, USA: CRC Press, 2019, pp. 125–154, doi: 10.1201/9781351006620.

[15] P. Courtin, A. Grimault, M. Lhommeau, and J.-B. Fasquel, ''Dynamic storage location assignment problem at knapp company: Toward reinforcement learning?'' in *Problematic and State of the Art. Angers, FR: Artificial Intelligence Platform*. 2020, pp. 1–8.

[16] C. Waubert de Puiseau, D. T. Nanfack, H. Tercan, J. Löbbert-Plattfaut, and T. Meisen, ''Dynamic storage location assignment in warehouses using deep reinforcement learning,'' *Technologies*, vol. 10, no. 6, p. 129, Dec. 2022, doi: 10.3390/technologies10060129.

[17] M. Jemelka, B. Chramcov, and P. Kříž, ''Design of the storage location based on the ABC analyses,'' in *Proc. AIP Conf.*, vol. 1738, 2016, p. 1, doi: 10.1063/1.4951909.

[18] V. R. Muppani (Muppant) and G. K. Adil, ''A branch and bound algorithm for class based storage location assignment,'' *Eur. J. Oper. Res.*, vol. 189, no. 2, pp. 492–507, Sep. 2008, doi: 10.1016/j.ejor.2007.05.050.

[19] M. Wang and R.-Q. Zhang, ''A dynamic programming approach for storage location assignment planning problem,'' *Proc. CIRP*, vol. 83, pp. 513–516, 2019, doi: 10.1016/j.procir.2019.04.113.

[20] J. Oxenstierna, L. van Rensburg, P. Stuckey, and V. Krueger, ''Storage assignment using nested annealing and Hamming distances,'' in *Proc. 12th Int. Conf. Operations Res. Enterprise Syst.*, 2023, pp. 94–105, doi: 10.5220/0011785100003396.

[21] H. Saleet, ''Smart solution for enhancing storage location assignments in WMS using genetic algorithm,'' *Int. J. Eng. Res. Technol.*, vol. 13, no. 11, p. 3456, Nov. 2020, doi: 10.37624/ijert/13.11.2020.3456-3463.

[22] A. Javed, D. M. Rizzo, B. S. Lee, and R. Gramling, ''Somtimes: Self organizing maps for time series clustering and its application to serious illness conversations,'' *Data Mining Knowl. Discovery*, vol. 38, no. 2, pp. 1–36, 2023, doi: 10.1007/s10618-023-00979-9.

[23] A. Javed, B. S. Lee, and D. M. Rizzo, ''A benchmark study on time series clustering,'' *Mach. Learn. Appl.*, vol. 1, Sep. 2020, Art. no. 100001, doi: 10.1016/j.mlwa.2020.100001.

[24] K. R. Shahapure and C. Nicholas, ''Cluster quality analysis using silhouette score,'' in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Analytics (DSAA)*, Oct. 2020, pp. 747–748, doi: 10.1109/DSAA49011.2020.00096.

[25] N. Shetty, B. Sah, and S. H. Chung, ''Route optimization for warehouse order picking operations via vehicle routing and simulation,'' *Social Netw. Appl. Sci.*, vol. 2, no. 2, pp. 1–18, Feb. 2020, doi: 10.1007/s42452-020-2076-x.

[26] T. van Gils, K. Ramaekers, K. Braekers, B. Depaire, and A. Caris, ''Increasing order picking efficiency by integrating storage, batching, zone picking, and routing policy decisions,'' *Int. J. Prod. Econ.*, vol. 197, pp. 243–261, Mar. 2018, doi: 10.1016/j.ijpe.2017.11.021.

[27] *Google Colab*. Accessed: Aug. 22, 2023. [Online]. Available: https://colab.research.google.com/

[28] *ChatGPT*. Accessed: Aug. 22, 2023. [Online]. Available: https://chat.openai.com/

[29] *Kaggle*. Accessed: Aug. 20, 2023. [Online]. Available: https://www.kaggle.com/code/surekharamireddy/e-commerce-data-set/input

[30] H. Delgado, X. Anguera, C. Fredouille, and J. Serrano, ''Novel clustering selection criterion for fast binary key speaker diarization,'' in *Proc. Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Sep. 2015, pp. 3091–3095, doi: 10.21437/interspeech.2015-108.

[31] F. Xu and K. Beard, ''A comparison of prospective space-time scan statistics and spatiotemporal event sequence based clustering for COVID-19 surveillance,'' *PLoS ONE*, vol. 16, no. 6, Jun. 2021, Art. no. e0252990, doi: 10.1371/journal.pone.0252990.

HICHAM KALKHA received the degree in engineering from ESITH Casablanca, with versatile training in industrial management and logistics. He is currently pursuing the Ph.D. degree with the Computer Science, Artificial Intelligence, and Cyber Security Laboratory (L2IACS), University Hassan II of Casablanca. He is a Logistics Engineer. His research interests include information and communications technology (ICT) and its applications in e-commerce logistics.

AZEDDINE KHIAT received the H.D.R. and Ph.D. degrees in computer science, networks, and telecommunications from ENSET Mohammedia, University Hassan II of Casablanca, Morocco. He is currently a Research Professor with the Department Mathematics and Computer Science. He is a Research Member with the Computing, Artificial Intelligence and Cyber Security Laboratory (2IACS), Artificial Intelligence, Cybersecurity and Digital Trust Team (IACSCN), ENSET Mohammedia, University Hassan II of Casablanca. He has published tens of refereed journal and conference papers. His research interests include wireless networks, QoS on mobile networks, handover on wireless networks, networks, and telecommunications, SDN, new generation networks, security, cybersecurity, mobile learning, the IoT, smart city, smart grids, and industry 4.0. He is certified on several products and solutions among them: the Fortinet NSE4, the Microsoft Azure Administrator Associate, the Juniper Associate, the Juniper Cloud Associate, the Huawei RS Associate, the Oracle Associate, and the Red Hat Associate. He is a organizing committee member and a technical program committee member of tens of international conferences. He is also a outstanding reviewer of various indexed journals.

AYOUB BAHNASSE received the Ph.D. degree in networks and telecommunication from Chouaib Doukkali University, Morocco. He is currently a Professor and a Researcher with ENSAM Casablanca, University Hassan II of Casablanca, Morocco. He is also a Research Associate with the Information Treatment Laboratory, Faculty of Sciences Ben M'sik, University Hassan II of Casablanca. He has published tens of refereed journals and conference papers. He is a member of the ''Big Data and Internet of Things Project for Urban Service Project, University Hassan II of Casablanca. His research interests include new-generation networks, security, mobile learning, wireless networks, quality of service, multimedia systems, the IoT, smart city, and MPLS. He is certified in several products and solutions, among them are Fortinet NSE4, the Microsoft Azure Administrator Associate, the Juniper DevOps Associate, the Juniper Cloud Associate, the Juniper Design Associate, and the Huawei RS Associate. He is an organizing committee member and a technical program committee member of tens of international conferences in around 20 countries. He is an Outstanding Reviewer of various indexed journals, such as *Computer Networks* (Elsevier), *Computers and Electrical Engineering* (Elsevier), *Computers and Security* (Elsevier), IEEE Transactions on Communications, *Telecommunication Systems* (Springer), *Wireless Personal Communications* (Springer), *Mobile Networks and Applications* (Springer), and tens of indexed journals.

HASSAN OUAJJI received the degree (higher education) from the Department of Electronic and Electrical Engineering, ENSET Mohammedia. He is currently a Full Professor with the Department of Electronic and Electrical Engineering, ENSET Mohammedia. He is also a Research Member with the SSDIA Laboratory and the Head of the Network and Telecommunications Team, University Hassan II of Casablanca, Morocco. He has published tens of refereed journals and conference papers. He is a member of several scientific projects with the University Hassan II of Casablanca. His research interests include acoustic electrical engineering, signal processing, electronic, networks, and telecommunications. He is an organizing committee member and a technical program committee member of tens of international conferences. He is an outstanding reviewer of various indexed journals.