Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 9, No. 1, February, 2025

Clustering and Network Analysis of Mobility Patterns as an Analysis Tool for Lean Project

András Rácz-Szabó [1, 2], Tamás Ruppert [1, 3*], János Abonyi [1, 2]

[1] *HUN-REN-PE Complex Systems Monitoring Research Group, University of Pannonia, Egyetem str. 10, POB 158, H-8200 Veszprém, Hungary.*

[2] *Department of Process Engineering, University of Pannonia, Egyetem str. 10, POB 158, H-8200 Veszprém, Hungary.*

[3] *Department of System Engineering, University of Pannonia, Egyetem str. 10, POB 158, H-8200 Veszprém, Hungary.*

Abstract

The study aims to optimize internal logistics processes by applying Lean philosophy and data science tools, with a primary focus on qualifying processes to determine their value-added contribution within the logistics context. Utilizing a novel two-step methodology, the research first employs a modified DBSCAN algorithm to analyze indoor positioning data and categorize activities. This is followed by multi-layer network modeling to understand processes and create a framework that enables the reduction of idle activities through optimization algorithms. A real warehouse case study, using a UWB-based Indoor Positioning System (IPS) to track forklifts, demonstrates the method's effectiveness in identifying non-value-added activities. The results reveal specific opportunities for reducing idle, enhancing resource utilization, and improving operational efficiency. This innovative combination of advanced data analysis techniques and Lean principles provides a comprehensive framework for logistics optimization, significantly enhancing process efficiency through optimized task scheduling and resource allocation.

Keywords:

Indoor Positioning System;
Position Data; Warehouse.
Clustering; Multi-Layer Network.

Article History:

Received:	13	July	2024
Revised:	23	December	2024
Accepted:	05	January	2025
Published:	01	February	2025

1- Introduction

The Lean philosophy and logistics processes are interrelated, as the application of Lean principles helps to optimize logistics activities [1, 2]. Digital technologies such as advanced analytics and IoT (Internet of Things) can support Lean practices in manufacturing and supply chain management by outlining eight waste reduction mechanisms that improve operational efficiency and decision-making, giving companies a clear framework for selecting the most effective technologies to improve their processes [3]. The concept of "Digital Lean" focuses on how digital technologies, such as ITs (Information Technologies) and OTs (Operational Technologies), enhance lean manufacturing by detecting and preventing physical waste through simulations and real-time monitoring while also addressing digital waste that arises from the underuse or overuse of advanced smart manufacturing technologies [4]. Lean 4.0 builds on this by incorporating a broader set of Industry 4.0 technologies like artificial intelligence (AI), robotics, and cloud computing to further optimize processes, enabling more intelligent automation and deeper integration of lean principles into the manufacturing environment [5]. The essence of industrial muda is to identify and reduce waste such as transport, inventory, motion, waiting, over-processing, overproduction, and defects within processes [6]. Idle activities of transportation are activities that do not contribute to value creation; therefore, they can be considered non-value-added activities from the point of view of logistics. Minimizing idle activities is crucial for effective resource utilization and improving process efficiency [7].

* **CONTACT**: ruppert@abonyilab.com

DOI: http://dx.doi.org/10.28991/ESJ-2025-09-01-013

One challenge is to identify these activities. Several methods are used for the solution, for example, the Spaghetti Diagram [8-10], a visualization tool that helps to understand and depict the movement and interactions that occur during a process or activity, which can be classified and characterized. Another solution is Value Stream Mapping (VSM) [11-13], which involves reviewing intralogistics processes to identify idle times and eliminate unnecessary steps, improving material flow and reducing downtime. Gemba Walks allow leaders and workers to directly observe work areas, also providing opportunities to make intralogistics processes more efficient and reduce downtime [14]. Another approach is to use manufacturing execution system (MES) and indoor positioning system (IPS) data, combined with data analytics and data science techniques, to identify non-value-added activities. MES, a software system that monitors and controls manufacturing operations on the shop floor, provides detailed, real-time production data, while IPS provides precise location information for tools, materials, and personnel. By integrating these data sources, advanced analytics can be used to identify inefficiencies such as idle time, unnecessary movements, or bottlenecks, enabling a more detailed understanding and optimization of intralogistics processes.

IPS provides accurate, real-time location data, significantly improving logistics efficiency by enabling accurate tracking of material flows, goods, and machinery within warehouse environments [15, 16]. These systems address critical logistics challenges such as improving the accuracy of warehouse positioning and ensuring reliable monitoring of goods in transit within large facilities [17]. In addition, IPS facilitates real-time VSM, helping to identify bottlenecks and optimize processes through accurate state tracking [18]. In manufacturing environments, IPS can also monitor tool locations and calculate utilization, despite challenges related to measurement uncertainty [19]. Ultra-wideband-based (UWB-based) IPS also utilize precise position data to predict human movement in shared workspaces, enabling safer and more efficient navigation for automated guided vehicles (AGVs) in manufacturing and warehouse environments [20]. Clustering techniques, such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise), can be applied to position data provided by the IPS to identify patterns in resource movements, helping to distinguish between value-added and non-value-added activities, thus optimizing logistics processes and improving overall efficiency.

Clustering [21, 22] is a method that can assist in identifying areas or periods where the movement or activity of resources significantly deviates from the norm or in identifying potential areas for optimization. Time-series classification (TSC) is widely utilized in manufacturing systems, including applications such as supply chain optimization. The importance of TSC has grown significantly in smart manufacturing systems, driven by the integration of Machine Learning (ML) and Deep Learning (DL) algorithms, which are essential for processing the vast amounts of time-series data generated by these systems [23]. Additionally, ML-based methods [24] facilitate the creation and training of models to recognize different states, analyzing data to detect anomalies, such as periods of idleness, through various algorithms. Movement Trajectory Analysis [25] examines resource movement patterns to identify areas or routes where resources are less active or exhibit significantly different trajectories. Algorithms like K-Means and DBSCAN, which can also be applied in logistics, enable the automatic detection of the types of work performed by AGVs and the identification of anomalies, thus further improving the optimization of manufacturing processes and supporting predictive maintenance efforts [26]. The DBSCAN algorithm is particularly robust for identifying dense regions within datasets, making it suitable for applications requiring spatial cluster delineation. In logistics, it can be used to determine customer clusters and optimize truck planning based on shipping demands and capacity constraints [27]. Hierarchical clustering methods can also be applied to detect throughput bottlenecks in production systems [28]. In the context of internal logistics, DBSCAN is effectively utilized to classify forklift movements, distinguishing between value-added and non-value-added activities to enhance process efficiency and resource allocation [29].

A multilayer network framework enables the analysis of complex systems by capturing multiple types of interactions between entities across different layers, providing a comprehensive understanding of interconnected subsystems [30]. These networks are also utilized for inventory optimization in e-commerce platforms, integrating material flow, inventory management, and pricing strategies to improve supply chain efficiency [31]. Furthermore, a multilayered temporal network-based model distinguishes value-added from nonvalue-added resource flows, enabling a comprehensive view of the flow of resources in the system [32].

Based on the literature, it is evident that clustering is an important research topic in prioritization tasks related to logistics and production. However, in the context of Lean philosophy, multi-layer network-based clustering—specifically for identifying value-added and non-value-added processes—can greatly assist in identifying and understanding the underlying issues. By accurately distinguishing between value-added and non-value-added activities, organizations can more effectively optimize their processes, reduce waste, and improve overall efficiency. Our work addresses this specific challenge by introducing a novel approach that uses clustering to identify these types of processes within a multi-layer network framework, a method that has not been extensively explored in previous research. This approach provides valuable insights into process optimization by providing a clearer understanding of where inefficiencies exist and how they can be minimized.

The purpose of this paper is to present a methodology for identifying non-value-added activities in logistics processes. The methodology uses the DBSCAN algorithm to distinguish between value-added and non-value-added networks. The effectiveness of this approach is demonstrated through a case study that analyzes forklift movements based on indoor positioning data within a warehouse environment. The paper is structured as follows: the next section outlines the proposed methodology, followed by a detailed case study, and concludes with a discussion of the results and implications for future research.

2- Methodology of Clustering the IPS-Based Data

In this section, we provide a detailed methodological overview based on the steps illustrated in Figure 1.

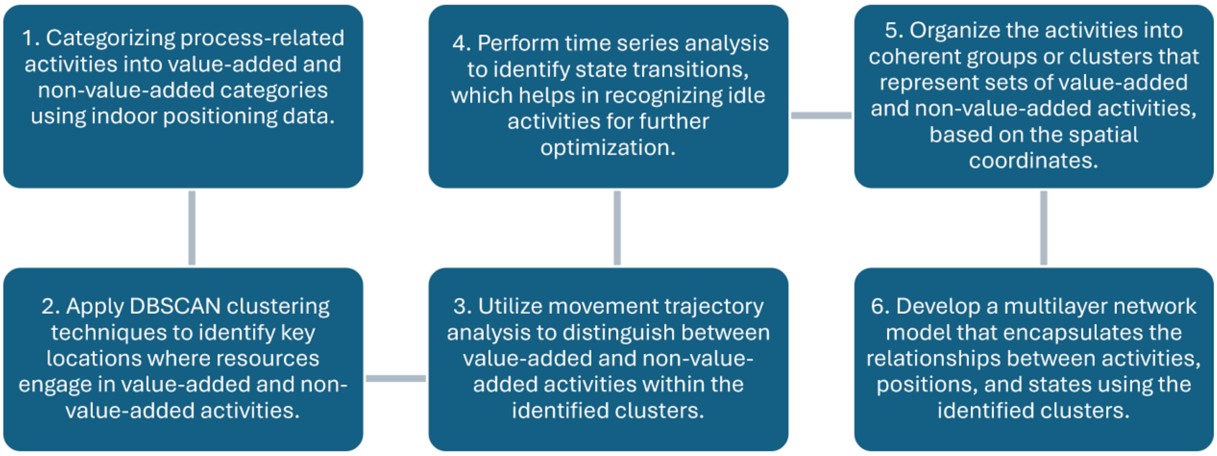

Figure 1. Classification and optimization of activities based on IPS

The aim of the methodology is to classify process-related activities into value-added and non-value-added categories based on indoor positioning data. Our method combines clustering, time series analysis, and movement trajectory analysis. By clustering, we identify the places where the resource performs value-added and non-value-added activities, for which we use movement trajectory analysis. After that, we can identify the state transitions with a time series analysis, thereby identifying the idle activities for optimization. Activities are discerned through the analysis of positional data, wherein the spatial coordinates serve as crucial indicators. The contextual background of the data plays a pivotal role in the identification of various states associated with these activities. To enhance the precision of this process, positional data undergoes clustering through the application of DBSCAN [33], which is meticulously tuned to delve into the typical positions and places where activities unfold. In this study, we utilize a modified version of DBSCAN, which focuses on identifying whether a resource is engaged in value-added or non-value-added activities based on positional data. Without this modification, we would not be able to construct the multilayer-based clustering, as the standard DBSCAN algorithm would not sufficiently differentiate between value-added and non-value-added activities based on positional data. This approach creates separate cluster groups depending on the activity category, and these clusters are represented as multilayers. This multilayer representation assists in uncovering inefficiencies and optimizing processes. Additionally, the modified DBSCAN is specifically tailored to support the processing of positional data, enhancing the precision of activity classification. This clustering results in the formation of coherent groups or clusters, each representing the set of value-added and non-value-added activities. Subsequently, a multilayer network is constructed, drawing upon the tuples that encapsulate the relationships between activities, positions, and states. The two-step methodology, based on the provided text, includes the following steps:

1. Step - Identification and Grouping of Activities:

- Activities are discerned through the analysis of positional data using spatial coordinates.
- The contextual background of the data plays a pivotal role in identifying various states associated with these activities.
- Positional data undergoes clustering using DBSCAN (Density-Based Spatial Clustering of Applications with Noise), meticulously tuned to explore typical positions and places.

2. Step - Network Modelling and Optimization:

- The groups or clusters formed in the previous step represent activities, including both value-added and non-value-added activities.
- A multilayer network is constructed using tuples encapsulating relationships between activities, positions, and states.

The available $\boldsymbol{T_k}$ database is structured as follows:

$$T_k = (t_k, d_k, r_l, l_k) \tag{1}$$

where $k = 1 \dots N$ represents the index of the recorded position data, t_k is the time of the k-th recorded data, $\boldsymbol{d_k} = [d_x\ d_y\ d_z] \in \boldsymbol{D}$ is the k-th position data vector, r_l is the unique identification number of the $r_l \in R$ resources and $l_k = \{\alpha, \beta\}$ is the activity qualifying flag, what identify the value-added ($\alpha = 0$) and non-value-added ($\beta = 1$) activity of the r_l resource at the k-th position.

The next step is to define value-added and non-value-added activities based on position data with DBSCAN algorithm. Figure 2 shows the steps of applying the algorithm:

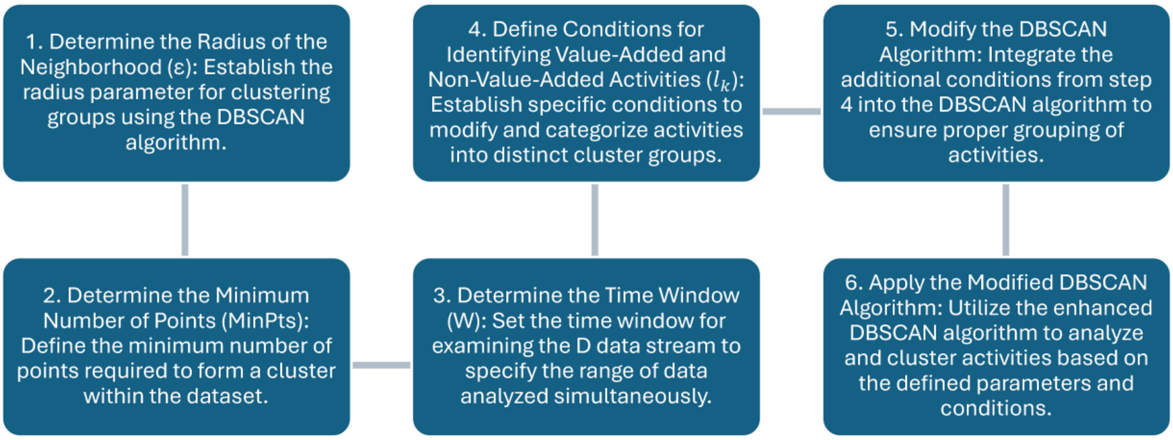

Figure 2. Steps of applying Modified DBSCAN algorithm

When using the algorithm, we define a **W** vector with a window size for the time-dependent $\boldsymbol{D}$ data stream for clustering. This is necessary because we are dealing with time series data, so it is important to see the sequence among the clusters created. It is possible that several clusters are defined for the same position of the layout, but we interact with them at different times in chronological order.

For DBSCAN algorithm the following parameters should be set:

- Data stream contains all of $\boldsymbol{d_k}$ positions in chronological order based on t_k: $\boldsymbol{D}$
- Radius of the neighborhood: ε
- Vector used to collect data that meets the determined conditions: $\boldsymbol{idx}$
- Minimum number of points required to from a $\boldsymbol{C}$ cluster: $MinPts$
- $\boldsymbol{W}$ vector with window size of the time-dependent $\boldsymbol{D}$ data stream for clustering: $i{:}w$
- $\boldsymbol{D}$ data stream within $i{:}w$ time window: $\boldsymbol{D_{i:w}}$
- Set of clusters: $\boldsymbol{C} = \ \boldsymbol{C_1}, \boldsymbol{C_2}, \dots, \boldsymbol{C_k}$, where k is the number of the created cluster and $(\boldsymbol{d}^{C_k}, \boldsymbol{t}^{C_k}) \in \boldsymbol{C_k}$
- $\boldsymbol{C^j}$ tuple of the $\boldsymbol{C}$ clusters based on $l_k = \{\alpha, \beta\}$ activity qualifying flag

The method is the following:

Algorithm 1. The pseudo code of the DBSCAN algorithm

```
for j = 0:1
Initialize C as an empty list and i = 1
Initialize vector, Window W with size w forall each new data point d ϵ D
Mark all d ϵ D as not visited
if |D| < w
Add all d ∈ D_{i:w} to W
        d_1 = W_1
        if all elements of W are not visited
for d ∈ W
if |d − d_i| < ε AND (l_k = j) then add d to idx
                end for
                if |idx| ≥ MinPts
                        Let C = idx be a new cluster
                        Add all idx related t time of the recorded data to C
                        Mark all d ∈ C in D as visited (no overlap)
                        Add C to C^j
else
Mark d_i point as outlier
                end if
                Clear the elements of idx
            end if
        Clear the elements of W
            w = w + 1
            i = i + 1
end if
end for
```

From the $\boldsymbol{C^j}$ qualified clusters - where the value-added activities are from $j = 0$ clusters and the non-value-added activities are from $j = 1$ and by matching the position data-based cluster centers to the layout, the $\boldsymbol{V}$ state vector could be determined. The elements of the state vector were based on the positions of layout-based clusters to identify where activity occurred within the examined production or logistics environment under study. The number of the unique states are the number of areas defined based on the layout $|\boldsymbol{V}| = N_V$. Since the cluster itself identifies a value-added activity, for each state there is a state transition that points back to itself, which shows the activity in the position according to the given layout.

Based on the previous paragraph, we already know the state of the resources, which indicates their layout-based position and the type of activity they perform (value-added or non-value-added). However, it is important to study the state transitions, because by classifying these transitions we can also characterize the transport activities and minimize the idle. This requires a multi-layered network analysis.

Based on the t_k time data, the t_s^C start and t_e^C end times in the given state can also be determined, therefore the Time Series Analysis with focus on the state can created the $\boldsymbol{E} = \boldsymbol{V} \times \boldsymbol{V}$ state transition matrix. We can create based on this information a $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{F})$ multi-layer network, where $(r, l) \in \boldsymbol{F}$ set of dimensions (labels of resources and utilizations). If we focus on one resource, the multi-layer network based on α dimension is the following:

$$G^{\alpha} = (V^{\alpha}, E^{\alpha}) \tag{2}$$

where $\boldsymbol{V^{\alpha}}$ are the nodes of the multi-layers and $\boldsymbol{E^{\alpha}}$ is the set of intra-layer connections are represented by the elements of the $\boldsymbol{A^{\alpha}}$ adjacency matrices defined as:

$$\boldsymbol{A^{\alpha}} = \left[a_{p,q}^{\alpha}\right]_{N_V x N_V} \tag{3}$$

$$a_{p,q}^{\alpha} = \begin{cases} > 0, & if\ (v_p^{\alpha}, v_q^{\alpha}) \in \boldsymbol{E^{\alpha}} \\ 0, & otherwise \end{cases} \tag{4}$$

where $a_{p,q}^{\alpha}$ represent the weight of the state transitions from the p state to q according to the α type of activity.

The $a_{p,q}^{\alpha,\beta}$ interlayer connections between the nodes of the layers are defined as:

$$a_{p,q}^{\alpha,\beta} = \begin{cases} \geq 1, & if\ (v_p^{\alpha}, v_q^{\alpha}) \in \boldsymbol{E^{\alpha,\beta}} \\ 0, & otherwise \end{cases} \tag{5}$$

State transitions can be:

$$a_{p,q} = \begin{cases} a_{p,q}^{\alpha,\alpha} = a_{p,q}^{\alpha}\text{: } value - added\ activites \\ a_{p,q}^{\beta,\beta} = a_{p,q}^{\beta}\text{: } non - value - added\ activites \\ a_{p,q}^{\alpha,\beta}\text{: } value - added\ activites \\ a_{p,q}^{\beta,\alpha}\text{: } non - value - added\ activites \end{cases} \tag{6}$$

Figure 3 illustrates how connections define the value-added and non-value-added transitions. With the help of the method, we can qualify the processes, which is indirectly useful for process optimization. Resource utilization can also be determined, and non-value-added processes (as idle route) can be reduced to perform value-added tasks by reducing them with an effective optimization procedure.

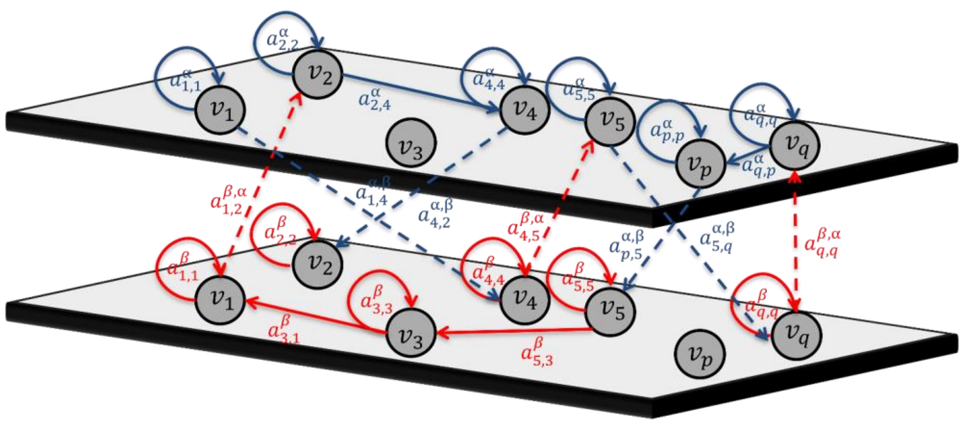

Figure 3. Multilayer network representation of the value-added (upper) and non-value-added (lower) activities of one resource

3- Results of the Applied IPS Based DBSCAN Clustering

3-1- Description of the Environment and the Used Resources in the Case Study

Our use case originates from a real logistics environment, where forklifts are tracked by a UWB-based IPS during the storage process. The forklifts are equipped with one tag (as seen in Figure 4) moves with the fork and five forklifts are monitored.

Figure 4. **Sensor placed on the forklift**

These tags continuously transmit real-time position data at specified intervals during movement. If the forklift is not moving, the data recording time intervals are longer. The Figure 5 represents the layout of warehouse environment where material flow is managed using the aforementioned forklifts and blue markings are the movement of one forklift based on the integrated IPS sensor tag data. This includes storage operations as from the delivery zone to the high warehouse, as well as the reverse process as picking out.

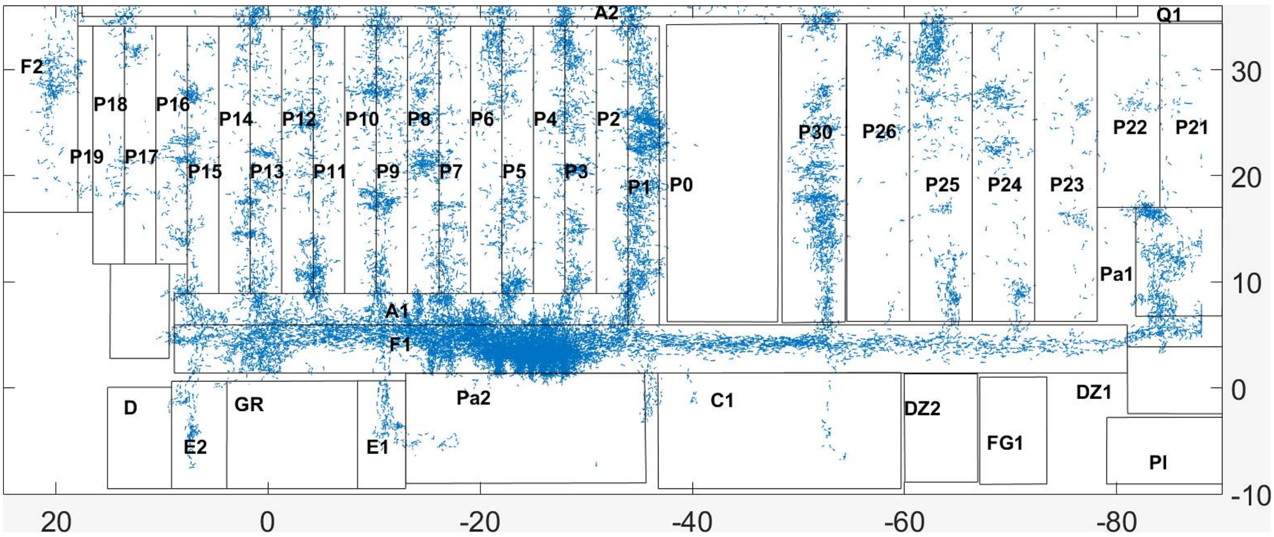

Figure 5. **Indoor positioning data-based movement of forklift in a warehouse environment. Sufficiently accurate location data is important for identifying the logistics area**

3-2- Definition of the Used States for DBSCAN Clustering

Based on $\boldsymbol{d_k}$ position data we should determine two states: when the forklift is active (moving) and not active (waiting). These can be identified from the δ_k data recording frequency:

$$\delta_k = t_k - t_{k-1} = \begin{cases} \leq d_t \rightarrow \boldsymbol{d_k} \in \boldsymbol{D_a} \\ > d_t \rightarrow \boldsymbol{d_k} \in \boldsymbol{D_n} \end{cases} \tag{7}$$

$$\boldsymbol{D_a} \cap \boldsymbol{D_n} = \boldsymbol{0} \tag{8}$$

where d_t is the limit of the active recording frequency, $\boldsymbol{D_a}$ is the vector of active position data and $\boldsymbol{D_n}$ is the vector of non-active position data. In our case d_t is three seconds. The working of the forklift can be defined by the fact that the member attached to the fork moves in the z-coordinate direction:

$$\boldsymbol{d_k}(d_z) = \begin{cases} > m \rightarrow working \\ \leq m \rightarrow not\ working \end{cases} \tag{9}$$

where m is the movement limit in z direction. The Table 1 summarizes, what states should be defined based on the position data, and the Table 2 represents the classification of the activities based on given state transitions. The methods of defining the states are presented in detail in the remaining parts of the chapter.

Table 1. Defined states based on position data driven clustering

	Forklift is active ($\boldsymbol{d_t} \leq 3$)	**Forklift is not active ($\boldsymbol{d_t} > 3$)**
Forklift is working ($d_k(d_z) > m$)	Cluster 1: value-added activity	Cluster 3: not applicable
Forklift is not working ($d_k(d_z) \leq m$)	Cluster 2: non-value-added activity	Cluster 4: waiting

Table 2. State transitions based classification of the activities

State transition	**Classification of the activities**
1 - 1	Value-added (on itself)
1 - 2	Value-added
1 - 3	Going to waiting
2 - 1	Non-value-added
2 - 2	Non-value-added (on itself)
2 - 3	Going to waiting
3 - 1	Going to do value-added activity
3 - 2	Going to do non-value-added activity
3 - 3	Waiting (on itself)

Second step we have to determine the clusters of the Table 3 with DBSCAN algorithm. For DBSCAN algorithm the following parameters should be set:

- Data stream contains all of active positions in chronological order based on t_k: $\boldsymbol{D^a}$
- Data stream contains all of non-active positions in chronological order based on t_k: $\boldsymbol{D^n}$
- Radius of the neighborhood: ε
- Movement limit in z direction: m
- Vector used to collect data that meets the determined conditions: $\boldsymbol{idx}$
- Minimum number of points required to from a $\boldsymbol{C}$ cluster: $MinPts$
- **W** vector with window size of the time-dependent $\boldsymbol{D^a}$ or $\boldsymbol{D^n}$ data stream for clustering: $i{:}w$
- $\boldsymbol{I}$ matrix contains the $\boldsymbol{D^a}$ and $\boldsymbol{D^n}$ data stream separated by m condition
- $\boldsymbol{D^a}$ or $\boldsymbol{D^n}$ data stream within $i{:}w$ time window: $\boldsymbol{D^a_{i:w}}$, $\boldsymbol{D^n_{i:w}}$
- Set of clusters: $\boldsymbol{C} = \boldsymbol{C_1}, \boldsymbol{C_2}, \ldots, \boldsymbol{C_k}$, where k is the number of the created cluster
- $\boldsymbol{C_{f,b}}$ tuple of the $\boldsymbol{C}$ clusters based on
- Variables to handle the number of for loop iterations: *f*, *b*

The method of the cluster determination is the following:

Algorithm 2. The pseudo code of the modified DBSCAN algorithm

```
Initialize C as an empty list
D = [D^a D^n]
A = D(d_z) > m
B = D(d_z) ≤ m
I = [A; B] = [D^a(d_z) > m   D^a(d_z) ≤ m; D^n(d_z) > m   D^n(d_z) ≤ m]
for f = 1:2
for b = 1:2
Initialize vector, Window W with size w forall each new data point d ϵ I_{b,f}
Mark all d ϵ I_{b,f}  as not visited
if |I_{b,f}| < w
        Add all d ϵ I_{b,f_{i:w}} to W
        d_1 = W_1
        if all elements of W are not visited
                for  d ϵ W
                        if |d − d_1| < ε then add d to idx
end for
                if |idx| ≥ MinPts
                        Let C = idx be a new cluster
                        Mark all d ϵ C in I_{b,f} as visited (no overlap)
                        Add C to C_{f,b}
else
                        Mark d_1 point as outlier
                end if
                Clear the elements of W
                w = w + 1
                i = i + 1
        end if
end for
end for
```

Table 3. Defined clusters based on the states

	$D_1 = D_a$: Forklift is active	$D_2 = D_n$: Forklift is not active
$I_1 = A = idx(d_z) > m$: Forklift is working	$C_{1,1}$: Clusters of value-added activities	$C_{2,1}$: Empy clusters (not applicable the not active worklift and working state together)
$I_2 = B = idx(d_z) \leq m$: Forklift is not working	$C_{1,2}$: Clusters of non-value-added activities	$C_{2,2}$: Clusters of waiting

4- Result of the DBSCAN clustering

The cluster groups created based on DBSCAN are shown in Figures 6 and 7. The first layout in Figure 6. shows the $C_{1,1}$ clusters of value-added activities, the second the $C_{1,2}$clusters of non-value-added activities and the third the $C_{2,2}$ clusters of waiting.

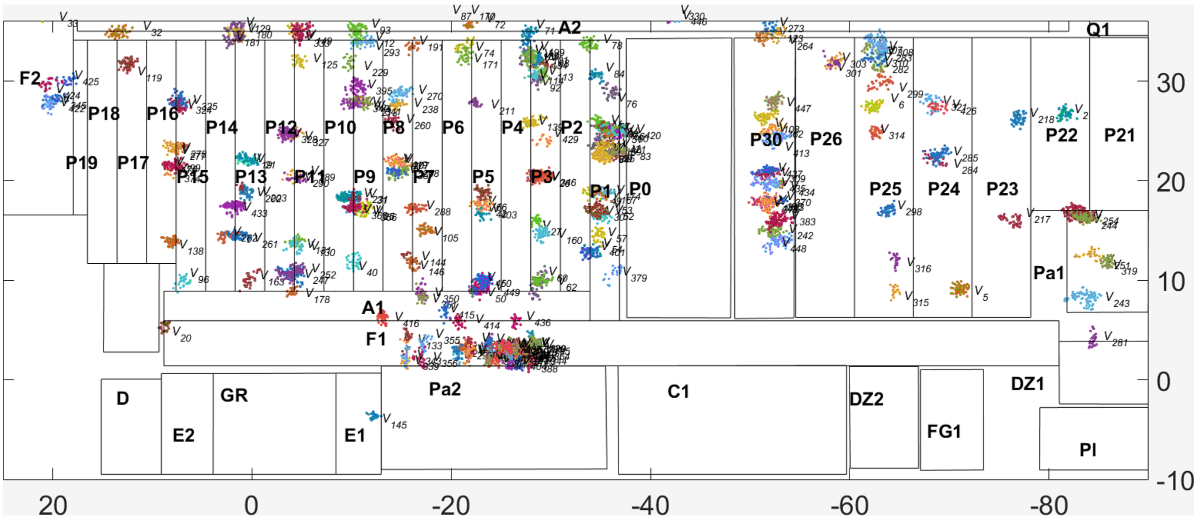

Figure 6. Indoor positioning data-based clustering in a warehouse environment

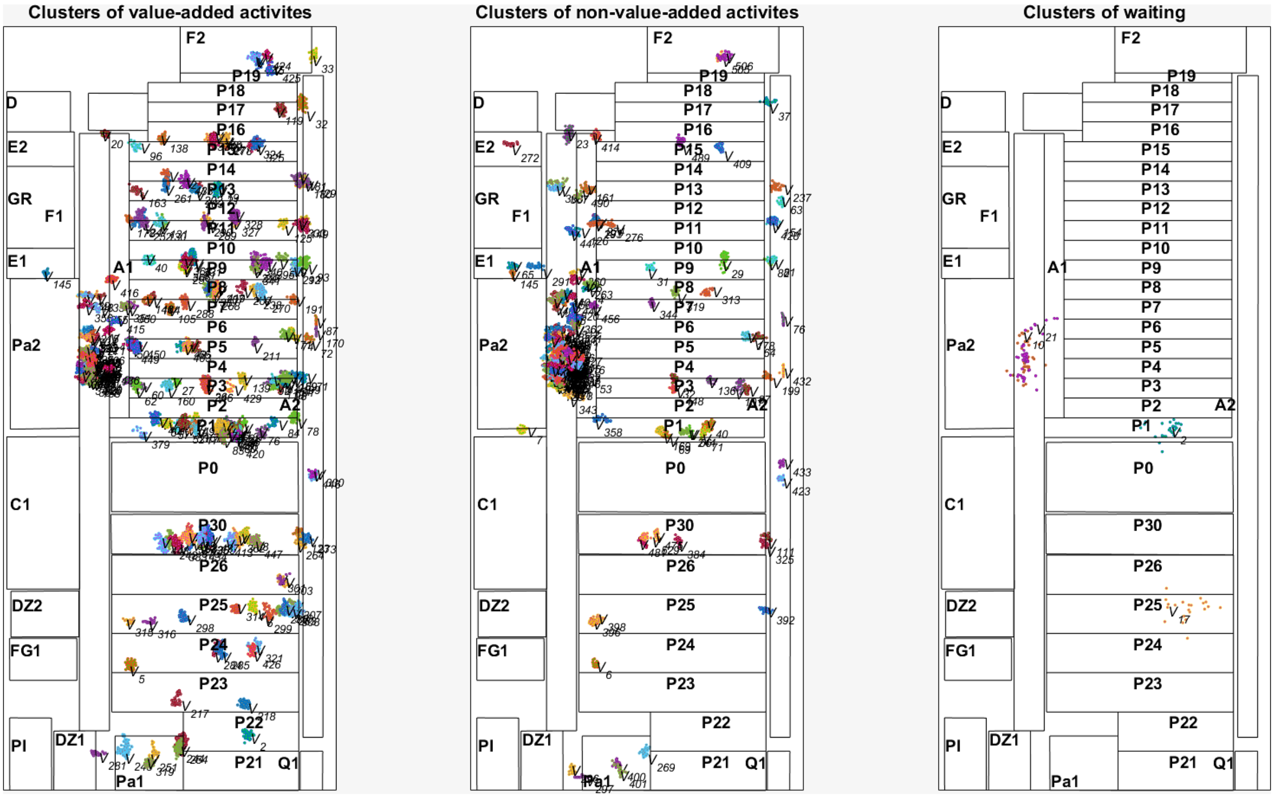

Figure 7. **Created clusters based on Table 3 activities**

With the center of the cluster groups, they can determine which area of the warehouse environment the given activity applies to (Figure 8):

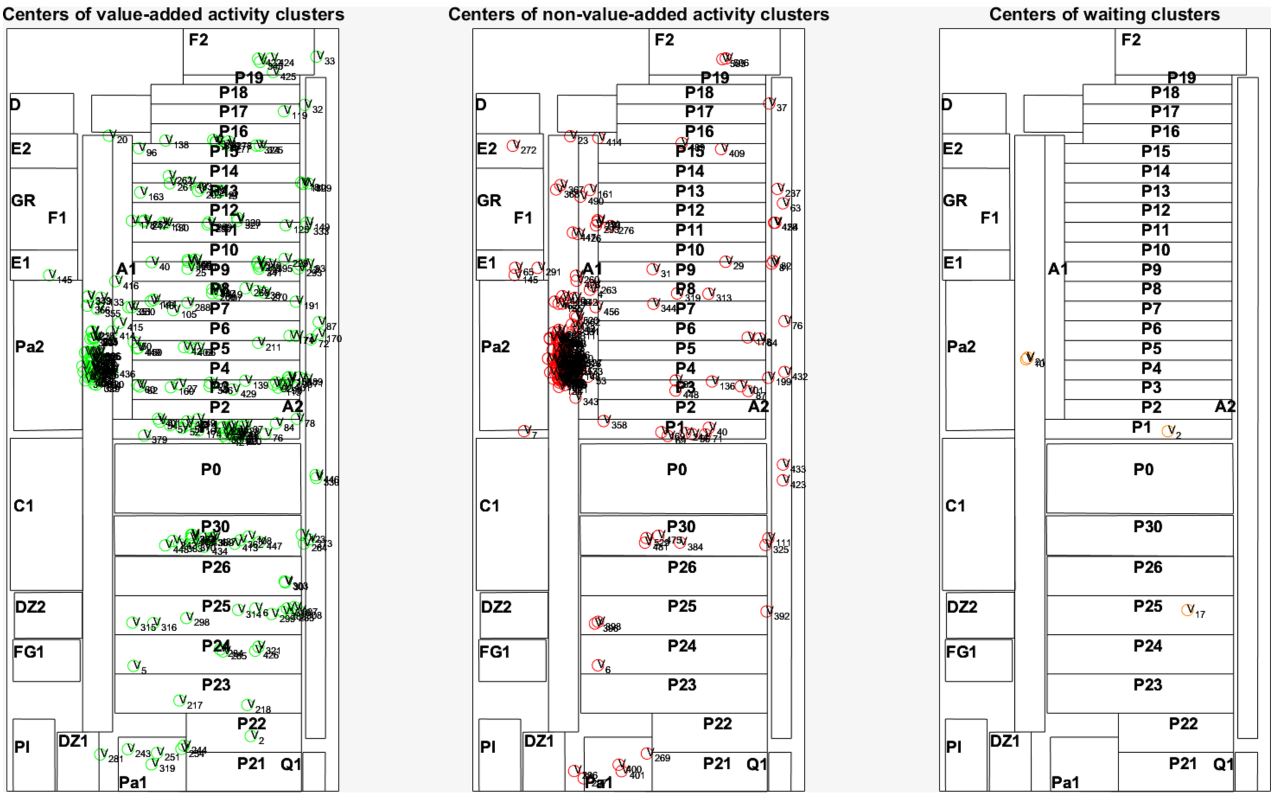

Figure 8. **Centre of clusters for identifying the layout-based states**

With the help of the cluster groups, we can classify individual states of the forklift, and by defining the state transitions defined in Table 2, the classifications of the movement of the forklift can also be determined. The multi-layer network of the non-value-added (red arrows) and value-added (green arrows) activities is seeable in Figure 9. The value 45 marks the locations forklifts arrive from outside the warehouse. Item for storage is temporarily stored in Aisle I (A1). As can be seen in the left-hand diagram, non-value-added routes typically lead from Aisles I (A1) and II (A2) to the high-bay area marked "P". This is understandable, as this is where forklifts depart to picking tasks. Conversely, there are movements in the opposite direction, albeit to a lesser extent, indicating instances where forklifts embark on storage tasks, also idle. Typical empty trips occur between Packing (Pa) and Aisle I (A1), indicating the idle before the packed products transport or idle after transport of awaiting packing products. There are also movements towards the Delivery zone (DZ), explaining forklifts travelling to pick up finished products. Frequent value-added movements can be seen in the right-hand diagram, typically go from the high-bay area, marked "P", to Aisle I (A1), which makes sense as forklifts are carrying out storage operations. Of course, there are also movements in the opposite direction, representing picking operations. Value-added activities from the delivery zone are also storage activities.

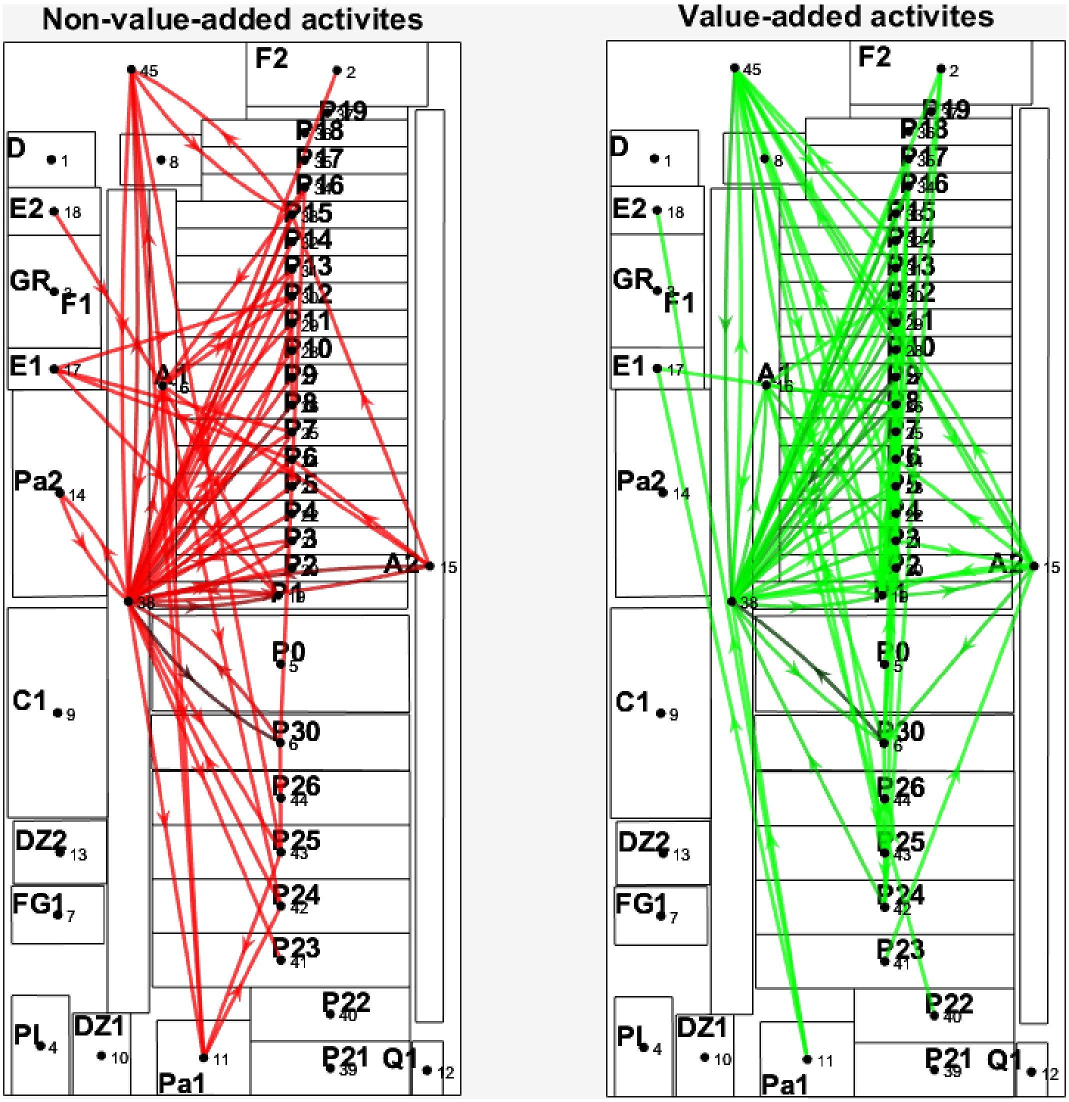

Figure 9. **Representation the multi-layer network of the forklift activities**

By adopting a multi-layered network approach, we open possibilities for optimizing both value-added and non-value-added processes. Through techniques such as resource reallocation, we can effectively minimize idle and reduce the number of required resources and improve utilization. This comprehensive strategy allows ultimately leading to improved productivity and cost-effectiveness.

Without the modified DBSCAN algorithm, we would only be able to identify clusters. However, by monitoring the movement of the forklift forks, we can distinguish these clusters more effectively, as demonstrated in Figure 10. Naturally, the way these activities should be defined and incorporated into the algorithm as additional conditions depends on the specific case being examined. If you have any suggestions or adjustments in mind, they can be tailored to better fit the context of the study.

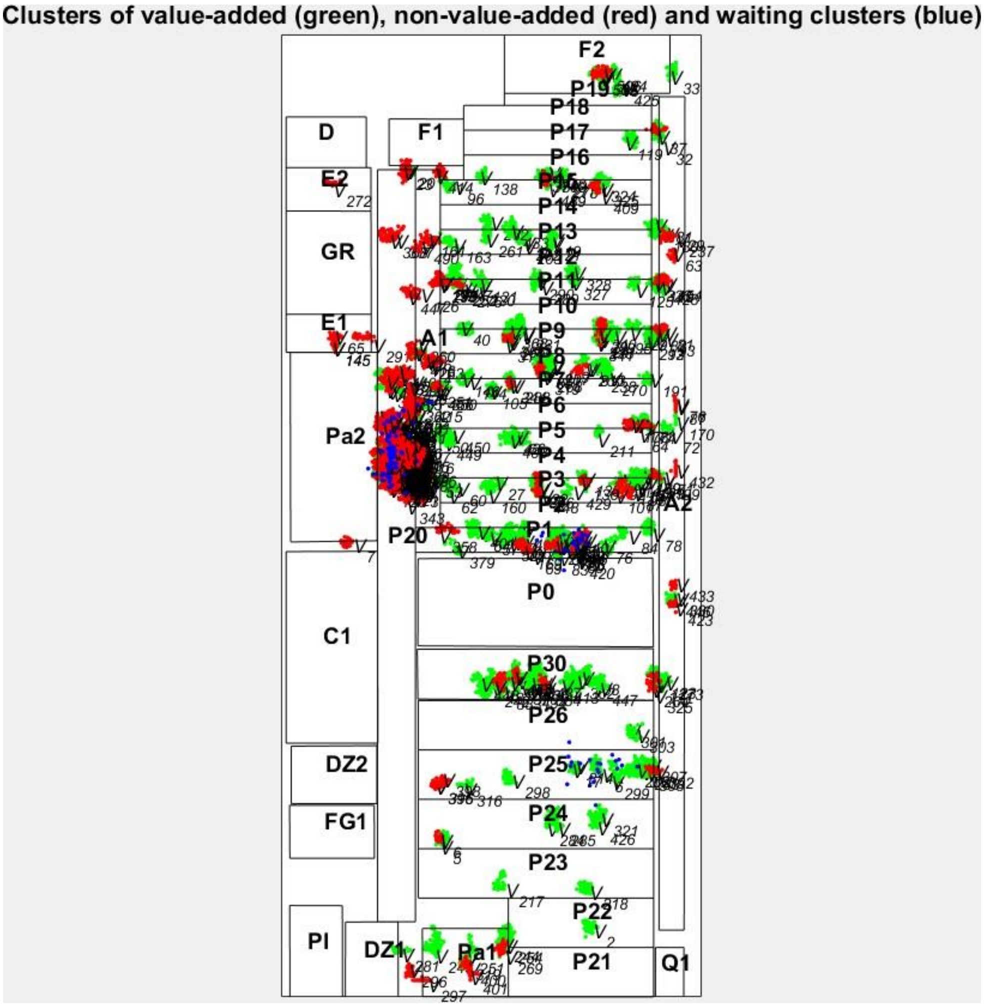

Figure 10. With the modified DBSCAN algorithm, the position data-based clusters can be classified and differentiated based on forklift fork movements, enabling the identification of distinct value-added, non-value-added, and waiting activities

The algorithm processes position data from five forklifts, representing more than 800,000 data points collected over three days, in 27.877153 seconds in a MATLAB environment, including the generation of map-based visualizations. While the algorithm can be further optimized to reduce processing time, the current limit on the number of resources that can be analyzed is primarily determined by the computational power and time required to process the data.

5- Conclusion

This article proposed a method to classify resource activities as value-added and non-value-added based on indoor positioning system data with the DBSCAN algorithm. The presented multilayer network-based solution can be used to provide information for the analysis of supply chains. By classifying the position data and with the help of layout-based clustering, we can determine the status of the activity of the resource belonging to the given cluster group (performing value-added activity, performing non-value-added activity, waiting). The state transitions also characterize transport activities as value-added or non-value-added, and these can be represented on a multi-layer network. The presentation methodology is illustrated by processing an indoor position database from a logistics environment, where we monitor five forklifts. The sensors providing the position data are placed on their forks, enabling us to determine and classify the states of the activities based on the movement and the different data recording frequencies. Using this information, we can identify state transitions, which helps classify transport activities. A multi-layer network-based representation of these transport activities helps us understand our processes and can reveal potential optimization opportunities, which may be the focus of future research.

In addition, this method has the potential for broader applicability in various industries where tracking and optimizing resource activities is critical. Future research could explore the implementation of this approach in manufacturing or warehousing environments to improve operational efficiency. Compared to existing methods, the proposed approach provides a more granular and dynamic classification of activities by leveraging the strengths of the multilayer network representation and the DBSCAN algorithm. This allows for a more detailed analysis of state transitions and activity patterns, providing insights that static or less sophisticated models may not capture. The practical implications of this study are significant for logistics and supply chain management. By accurately classifying activities and identifying value-added processes, companies can streamline operations, reduce waste, and improve overall efficiency. The multi-layered network visualization also helps to better understand complex processes, making it easier to identify bottlenecks or inefficiencies. However, it is important to note some limitations of this study. The accuracy of the classification is highly dependent on the quality of the positioning data and the specific characteristics of the logistics environment. Future research could address these limitations by integrating more advanced sensor technologies or by applying the method in different contexts to assess its robustness and generalizability.

One of the main challenges in implementing this methodology in the warehouse environment was the accuracy of the position data. The accuracy of the data was inhomogeneous due to environmental factors such as noise and physical obstructions. To address this issue, we developed a data reconciliation method to improve the accuracy of the position data. This method involved the use of advanced filtering techniques and algorithms to reconcile discrepancies and improve data accuracy [34]. Five forklifts were equipped with IPS tags, which transmitted positional data. The sample case demonstrates the applicability of the method; however, it would fundamentally require the involvement of more equipment to effectively implement resource reallocation and optimization.

Another significant challenge was clustering, especially when cluster groups were in close proximity. This situation occasionally resulted in "jumping" between clusters, where data points inaccurately moved from one cluster to another. This issue was mitigated by refining the clustering algorithms and further improving the data accuracy through our reconciliation process. These combined efforts helped stabilize the cluster assignments and provided a more reliable representation of the position data, ultimately improving the overall effectiveness of the methodology.

6- Declarations

6-1- Author Contributions

Conceptualization, J.A. and A.R.-Sz.; methodology, J.A. and A.R.-Sz.; data curation, A.R-Sz.; writing—original draft preparation, A.R-Sz.; writing—review and editing, T.R and J.A.; visualization, A.R-Sz. and T.R.; supervision, J.A. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available in the article.

6-3- Funding

This work has been implemented by the OTKA 143482 (Monitoring Complex Systems by goal-oriented clustering algorithms) project with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund. The work of András Rácz-Szabó was supported by the project no. KDP-11-3/PALY-2022 financed under the KDP-2021 funding scheme.

6-4- Institutional Review Board Statement

Not applicable.

6-5- Informed Consent Statement

Not applicable.

6-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Sopadang, A., Wichaisri, S., & Sekhari, A. (2014, August). The conceptual framework of lean sustainable logistics. Materiały z konferencji, International Conference on Transportation and Logistics. 27-29 August, 2017, Kuala Lumpur - Malaysia.

[2] Campos, V. M. K., Cotrim, S. L., Galdamez, E. V. C., & Leal, G. C. L. (2016). Introduction of Lean Manufacturing Philosophy by Kaizen Event: Case Study on a Metalmechanical Industry. Independent Journal of Management & Production, 7(1), 151–167. doi:10.14807/ijmp.v7i1.388.

[3] Cifone, F. D., Hoberg, K., Holweg, M., & Staudacher, A. P. (2021). 'Lean 4.0': How can digital technologies support lean practices? International Journal of Production Economics, 241, 108258. doi:10.1016/j.ijpe.2021.108258.

[4] Romero, D., Gaiardelli, P., Powell, D., Wuest, T., Thürer, M. (2018). Digital Lean Cyber-Physical Production Systems: The Emergence of Digital Lean Manufacturing and the Significance of Digital Waste. Advances in Production Management Systems. Production Management for Data-Driven, Intelligent, Collaborative, and Sustainable Manufacturing. APMS 2018. IFIP Advances in Information and Communication Technology, Springer, Cham, Switzerland. doi:10.1007/978-3-319-99704-9_2.

[5] Javaid, M., Haleem, A., Singh, R. P., Rab, S., Suman, R., & Khan, S. (2021). Exploring relationships between Lean 4.0 and manufacturing industry. Industrial Robot: The International Journal of Robotics Research and Application, 49(3), 402–414. doi:10.1108/ir-08-2021-0184.

[6] Arunagiri, P., & Gnanavelbabu, A. (2014). Identification of major lean production waste in automobile industries using weighted average method. Procedia Engineering, 97, 2167–2175. doi:10.1016/j.proeng.2014.12.460.

[7] Aminzadegan, S., Tamannaei, M., & Rasti-Barzoki, M. (2019). Multi-agent supply chain scheduling problem by considering resource allocation and transportation. Computers & Industrial Engineering, 137, 106003. doi:10.1016/j.cie.2019.106003.

[8] Szabó, P., Janík, S., MĺKva, M., & Czifra, G. (2024). Waste Elimination in the Assembly Process using Lean Tools. Acta Polytechnica Hungarica, 21(4), 165–184. doi:10.12700/APH.21.4.2024.4.9.

[9] Gebeyehu, S. G., Abebe, M., & Gochel, A. (2022). Production lead time improvement through lean manufacturing. Cogent Engineering, 9(1), 2034255. doi:10.1080/23311916.2022.2034255.

[10] Correia, N., Teixeira, L., & Ramos, A. L. (2020). Implementing an AGV system to transport finished goods to the warehouse. Advances in Science, Technology and Engineering Systems, 5(2), 241–247. doi:10.25046/aj050231.

[11] Lian, Y.-H., & Landeghem, H. Van. (2002). An Application of Simulation and Value Stream Mapping in Lean Manufacturing. Proceedings 14th European Simulation Symposium, 23-26 October, 2002, Dresden, Germany.

[12] Schoeman, Y., Oberholster, P., & Somerset, V. (2020). Value Stream Mapping as a Supporting Management Tool to Identify the Flow of Industrial Waste: A Case Study. Sustainability, 13(1), 91. doi:10.3390/su13010091.

[13] Venkataraman, K., Ramnath, B. V., Kumar, V. M., & Elanchezhian, C. (2014). Application of Value Stream Mapping for Reduction of Cycle Time in a Machining Process. Procedia Materials Science, 6, 1187–1196. doi:10.1016/j.mspro.2014.07.192.

[14] Romero, D., Gaiardelli, P., Wuest, T., Powell, D., Thürer, M. (2020). New Forms of Gemba Walks and Their Digital Tools in the Digital Lean Manufacturing World. Advances in Production Management Systems. Towards Smart and Digital Manufacturing. APMS 2020. IFIP Advances in Information and Communication Technology, Springer, Cham, Switzerland. doi:10.1007/978-3-030-57997-5_50.

[15] Rácz-Szabó, A., Ruppert, T., Bántay, L., Löcklin, A., Jakab, L., & Abonyi, J. (2020). Real-Time Locating System in Production Management. Sensors, 20(23), 6766. doi:10.3390/s20236766.

[16] Elsanhoury, M., Makela, P., Koljonen, J., Valisuo, P., Shamsuzzoha, A., Mantere, T., Elmusrati, M., & Kuusniemi, H. (2022). Precision Positioning for Smart Logistics Using Ultra-Wideband Technology-Based Indoor Navigation: A Review. IEEE Access, 10, 44413–44445. doi:10.1109/ACCESS.2022.3169267.

[17] Pan, C., & Liu, M. (2021). Optimization of intelligent logistics supply chain management system based on wireless sensor network and RFID technology. Journal of Sensors, 2021(1), 8111909. doi:10.1155/2021/8111909.

[18] Rácz-Szabó, A, Ruppert, T, & Abonyi, J. (2023). Improving micro-logistics processes by indoor positioning system and the tools of data science PCS Science: Eclipse Solutions Ltd, 10-16.

[19] Darányi, A., Dörgő, G., Ruppert, T., & Abonyi, J. (2022). Processing indoor positioning data by goal-oriented supervised fuzzy clustering for tool management. Journal of Manufacturing Systems, 63, 15–22. doi:10.1016/j.jmsy.2022.02.010.

[20] Locklin, A., Ruppert, T., Jakab, L., Libert, R., Jazdi, N., & Weyrich, M. (2020). Trajectory Prediction of Humans in Factories and Warehouses with Real-Time Locating Systems. 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 1317–1320. doi:10.1109/etfa46521.2020.9211913.

[21] Khaswala, Z. N., & Irani, S. A. (2001). Value network mapping (VNM): visualization and analysis of multiple flows in value stream maps. Proceedings of the Lean Management Solutions Conference, 10-11 September, 2001, St Louis, United States.

[22] Kolich, D., Yao, Y. L., Fafandjel, N., & Hadjina, M. (2014). Value Stream Mapping Micropanel Assembly with Clustering to improve Flow in a Shipyard. International Conference on Innovative Technologies, 28-29 December, 2014, Bangkok, Thailand.

[23] Farahani, M. A., McCormick, M. R., Harik, R., & Wuest, T. (2025). Time-series classification in smart manufacturing systems: An experimental evaluation of state-of-the-art machine learning algorithms. Robotics and Computer-Integrated Manufacturing, 91, 102839. doi:10.1016/j.rcim.2024.102839.

[24] Muehlbauer, K., Wuennenberg, M., Meissner, S., & Fottner, J. (2022). Data driven logistics-oriented value stream mapping 4.0: A guideline for practitioners. IFAC-PapersOnLine, 55(16), 364–369. doi:10.1016/j.ifacol.2022.09.051.

[25] Löcklin, A., Dettinger, F., Artelt, M., Jazdi, N., & Weyrich, M. (2022). Trajectory Prediction of Workers to Improve AGV and AMR Operation based on the Manufacturing Schedule. Procedia CIRP, 107, 283–288. doi:10.1016/j.procir.2022.04.046.

[26] Steclik, T., Cupek, R., & Drewniak, M. (2022). Automatic grouping of production data in Industry 4.0: The use case of internal logistics systems based on Automated Guided Vehicles. Journal of Computational Science, 62, 101693. doi:10.1016/j.jocs.2022.101693.

[27] Lyu, Z., Pons, D., Palliparampil, G., & Zhang, Y. (2023). Optimising Urban Freight Logistics Using Discrete-Event Simulation and Cluster Analysis: A Stochastic Two-Tier Hub-and-Spoke Architecture Approach. Smart Cities, 6(5), 2347–2366. doi:10.3390/smartcities6050107.

[28] Subramaniyan, M., Skoogh, A., Muhammad, A. S., Bokrantz, J., Johansson, B., & Roser, C. (2020). A generic hierarchical clustering approach for detecting bottlenecks in manufacturing. Journal of Manufacturing Systems, 55, 143–158. doi:10.1016/j.jmsy.2020.02.011.

[29] Pilati, F., Sbaragli, A., Ruppert, T., & Abonyi, J. (2024). Goal-oriented clustering algorithm to monitor the efficiency of logistic processes through real-time locating systems. International Journal of Computer Integrated Manufacturing, 1–17. doi:10.1080/0951192X.2024.2372272.

[30] Racz-Szabo, A., Ruppert, T., & Abonyi, J. (2024). Multilayer Network-Based Evaluation of the Efficiency and Resilience of Network Flows. Complexity, 2024(1), 6940097. doi:10.1155/2024/6940097.

[31] Lepuschitz, W., Jirkovsky, V., Kadera, P., & Vrba, P. (2012). A Multi-Layer Approach for Failure Detection in a Manufacturing System Based on Automation Agents. 2012 Ninth International Conference on Information Technology - New Generations, 1–6. doi:10.1109/itng.2012.8.

[32] Zhou, J. (2023). Logistics inventory optimization method for agricultural e-commerce platforms based on a multilayer feedforward neural network. Pakistan Journal of Agricultural Sciences, 60, 487-496. doi:10.21162/PAKJAS/23.131.

[33] Rácz-Szabó, A., Ruppert, T., & Abonyi, J. (2024). Data reconciliation of indoor positioning data: Improve position data accuracy in warehouse environment. MethodsX, 13, 102838. doi:10.1016/j.mex.2024.102838.

[34] Cheng, D., Yue, G., Pei, T., & Wu, M. (2021). Clustering indoor positioning data using E-DBSCAN. ISPRS International Journal of Geo-Information, 10(10), 669. doi:10.3390/ijgi10100669.