Computers & Industrial Engineering 191 (2024) 110175

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A hybrid approach integrating genetic algorithm and machine learning to solve the order picking batch assignment problem considering learning and fatigue of pickers

Matteo Gabellini [a,*], Francesca Calabrese [a], Alberto Regattieri [a], Dominic Loske [b], Matthias Klumpp [b,c]

[a] Department of Industrial Engineering, University of Bologna, viale del risorgimento 4, Bologna 61121, Italy
[b] Institute for Production and Supply Chain Management, Technical University of Darmstadt, Germany
[c] School of Management, Politecnico di Milano, Milan, Italy

ARTICLE INFO

Keywords:
Order picking
Batch assignment
Machine learning
Genetic algorithms

ABSTRACT

Modeling human behaviors has become increasingly relevant to improving the performance of manual order-picking systems. However, although a vast corpus of literature has recently started to consider the human factors in these systems, several gaps remain uncovered. Specifically, mental and physical human factors, like learning and fatigue, and quantitative and spatial features of picking orders have never been considered jointly to estimate the time a human order picker requires to execute a specific picking mission. Furthermore, little attention has been given to assigning and sequencing orders to pickers to minimize the picking time acting on their individual learning and fatigue characteristics. This study thus proposes a novel approach integrating machine learning and genetic algorithms to solve the problem. A non-linear machine learning-based predictive model has been adopted to predict the picking time of batches of orders based on quantitative and spatial features of batches and learning and fatigue indicators of pickers. These predictions have thus been adopted to guide a genetic algorithm to find the best assignment of future planned batches of orders to pickers. One year of picking data collected from the warehouse of a grocery retailer has been adopted to investigate the potential of the proposed approach. Furthermore, multiple comparisons have been performed. First, the advantages of predicting the batch-picking time with the proposed non-linear model have been compared with predictions executed based on linear models. In addition, an ablation analysis has been performed to investigate the advantages of predicting the batch picking time while simultaneously considering the quantitative and spatial features of batches and the learning and fatigue indicators of pickers. Moreover, the advantages of the proposed batch assignment strategy, which considers learning and fatigue indicators, have been compared with an assignment strategy that does not optimize these elements. Lastly, an explainability analysis of the predictive model has been performed to understand how and how much quantitative and spatial features of batches and learning and fatigue indicators of pickers affect the batch picking time.

1. Introduction

In contemporary markets, a diverse range of stock keeping units (SKUs), coupled with elevated labor costs, particularly in Europe and North America, alongside heightened customer demands, are prevalent. Consequently, there is a pressing need to streamline warehouse operations for enhanced productivity and service standards. From an economic perspective, warehousing operations contribute approximately 15 % to logistics costs, with order-picking (OP) activities constituting a significant 55 % of typical warehouse operating expenses (Granillo-Macías, 2020; Giannikas et al., 2017).

The OP process involves retrieving products from storage or buffer areas in response to specific customer requests (Tompkins et al., 2010). OP systems can be categorized into picker-to-part or parts-to-picker configurations (Manzini et al., 2005). In picker-to-part systems, often referred to as manual OP, order pickers traverse the warehouse to gather the requested items from their storage locations. Conversely, in parts-to-picker systems, materials are delivered to the picker. Despite the

* Corresponding author.
E-mail address: matteo.gabellini5@unibo.it (M. Gabellini).

https://doi.org/10.1016/j.cie.2024.110175
Received 18 July 2023; Received in revised form 20 March 2024; Accepted 22 April 2024
Available online 23 April 2024

growing trend of automation in warehouse operations, manual OP remains prevalent in various sectors due to the cognitive skills, flexibility, and adaptability of human workers (Vijayakumar et al., 2022).

In picker-to-part systems, strategic, tactical, and operational decisions aim to optimize OP performance in terms of both traveled distance and picking time (de Koster et al., 2007). Specifically, the total traveled distance is influenced by warehouse layout and item positioning, which are regarded as strategic and tactical decisions. In contrast, picking time is influenced by additional factors, i.e., the set-up time, the search time, and the pick time (Tompkins et al., 2010), which depend on the experience and behavior of who perform the tasks (Grosse & Glock, 2015).

Therefore, intervening at the operational level, specifically through the strategic assignment of orders to pickers, provides advance insight into the time required to complete a certain number of orders each day, thereby improving overall work organization. In addition, it holds the potential for a significant reduction in the picking time, which can lead to a more efficient picking process. The efficiency of such optimization becomes evident when considering multiple pickers, as each may exhibit varying times for specific tasks. An effective assignment strategy must account for the inherent heterogeneity among pickers to derive optimal solutions. Assigning orders to pickers, neglecting the effect of their behavior and heterogeneity in terms of learning and fatigue profiles, can have negative effects on operational efficiency and profitability. In particular, the picking time may be overestimated or underestimated, leading to an over or under dimensioning of resources. Consequently, there would be an increased risk of an underutilized workforce on the one hand, and on the other hand, of failing to complete orders within the required timeframes, resulting in a decrease in customer service level.

However, assigning daily orders to heterogeneous pickers is not trivial since several human factors impact the picking time differently and also influence each other (Asadayoobi et al., 2021), (Winkelhaus et al., 2018), (Vanheusden et al., 2022; Wang et al., 2020). Existing parametric curves used to model the learning process and the fatigue in picking operations often fail to accurately capture the real behaviour of pickers in specific contexts. Similarly, studies leveraging historical data to quantify the learning and fatigue processes, along with other skills and behaviors (Matusiak et al., 2017; Loske & Klummp, 2022), resort to regression curves to model the relationships between pickers' skills and picking time, assuming that these relationships are linear.

Therefore, the primary objective of this paper is thus to propose a methodology for optimizing the assignment of order batches to different pickers, taking into account both batch characteristics and the pickers' heterogeneity. In particular, the simultaneous effect of learning and fatigue on the picking time is investigated and exploited to assign batches to pickers considering the constraints and characteristics of real industrial contexts. To this aim, nonlinear relationships between pickers' learning and fatigue, batch characteristics and picking time are extracted by means of machine learning models, allowing the prediction of the time needed by a picker to complete an order batch. Then, the predicted picking time of batches is integrated into the batch assignment optimization model in order to provide a picker with a list of order batches to complete in a day considering his or her level of experience and fatigue so that the total picking time is minimized.

In summary, the research questions that motivated this study are as follows:

1. How can we devise an integrated approach capable of addressing the batch assignment problem while simultaneously accounting for the impact of both batch characteristics and picker learning and fatigue on picking time?
2. What are the benefits of utilizing a nonlinear machine learning model over a linear one to delineate the relationship between batch characteristics, picker learning, fatigue, and picking time?
3. What advantages does the proposed approach offer compared to an assignment strategy that does not explicitly optimize picker learning and fatigue indicators during the assignment, particularly in terms of the potential reduction in picking time?

The remainder of the paper is organized as follows. First, Section 2 illustrates the research background by reviewing the literature. Then, Section 3 describes the problem under investigation, the proposed approach and the experimental procedure adopted to test and compare the proposed approach. Section 4 shows the results obtained in the case study. In conclusion, Sections 5 and 6 discuss the results and summarize and conclude the work.

2. Literature review

This section explores the existing literature on order picking and human factors in order picking decisions, giving particular attention to the batch assignment problem. The goal is to highlight the research gaps and the main difference between existing studies and this paper. This section first analyses papers on order picking problems aimed at minimizing the total travel distance, then focuses on the batch assignment problem, highlighting how the picking time not only depends on the traveled distance, but also on other times that are picker-dependent and therefore relevant to the assignment problem. Then, studies investigating the relationships between learning, fatigue, skills and the picking time, adopting both linear analytical models and predictive analytical models are presented. These paragraphs aims to highlight the lack of a comprehensive method that considers multiple human factors and the necessity of a transition from analytical to predictive models. Finally, the reviewed literature is discussed, highlighting challenges of existing approaches and opportunities of the proposed approach.

Decisions in order picking management typically evaluate performance through distance-based or time-based metrics (van Gils et al., 2018). Tactical-level decisions, such as storage assignment, predominantly use travel time as a key metric (Lee et al., 2020). At the operational level, distance optimization commonly occurs when considering order batching (OB) and picker routing (PR) (Cano et al., 2020), (Kübler et al., 2020) (Aboelfotoh et al., 2019). These studies are mainly focused on the solving methods used to address challenges posed by large instances in modern business contexts and only consider items' positions and quantity for batch as parameters affecting the traveled distance. In particular, Cano et al., (2020) demonstrated that leveraging the traveled distance in the OB problem leads to saving 5.5 % of total OP costs on average. Kübler et al., (2020) proposed a particle swarm optimization methodology and a local search algorithm combined with a nearest neighbor algorithm to solve OB and PR problems. They validated the procedure on large-scale problems and demonstrated that batches with fewer order lines and numbers of parts per order line led to a higher reduction in travel distance.

At the operational level, once batches and routes are optimized for minimum travel distance, batch assignment becomes crucial. Batch assignment, also known as job assignment, consists of assigning batches to a limited set of pickers (van Gils et al., 2018). In such scenarios, time-related metrics, such as tardiness (Henn, 2015), (Scholz et al., 2017), completion time (Zhang et al., 2017); (Alipour et al., 2020), and makespan (Ardjmand et al., 2020); (Pretolani et al., 2023) are employed. Differently from the traveled distance, these metrics account not only for travel time but also for the time needed to prepare the batch (set-up time), to search for items in the specific aisle and shelf (search time), and to extract items from their locations (pick time), collectively comprising 50 % of the picking time (Tompkins et al., 2010). However, all time components, except for distance, are treated as constant and uniform for all pickers. Therefore, the factors affecting these time metrics are still only related to the items' position and quantity per batch. In particular, Pretolani et al., (2023) propose a batch assignment model only constrained to the capacity of a batch, that does not depend on pickers. Ardjmand et al., (2020), adopt columns generation technique to minimize the makespan, still considering the items' position and quantity per batch as factors determining the optimized metric. Alipour et al., (2020) propose rule-based algorithms to solve the batch assignment problem, considering all pickers as identical, and consequently the search time and the picking velocity are the same for all pickers. Conversely, Zhang et al., (2017) propose a rule-based algorithm considering a constant set-up time for all batches, while the pick time depends on the number of items in an order.

All the presented works consider time components as constant or independent of the specific picker executing the batch. However, human factors, encompassing characteristics, behaviors, and the physical and mental conditions of order pickers, introduce variability into the time required for an operator to complete an OP task (De Vries et al., 2016; De Bruecker et al., 2015). Batt & Gallino, (2019) demonstrated that pick time decreases with worker experience. In addition, workers exhibit heterogeneous learning, and the variance in performance across pickers diminishes with experience. Grosse & Glock, (2015) demonstrated that learning occurs primarily in the search times; that is, the more frequently an item is picked, the better the order picker remembers the positions, and the shorter the search time. Following this stream of research, Zhang et al., (2023) studied the effect of learning on the search time in the problem of batch and zone picking and demonstrated that assigning pickers to smaller picker areas accelerates pickers' familiarization with the locations of items in their zones and accelerates the decline in pickers' search time. Mou (2022) considered the search time as a function of the cumulated items and a specific learning rate.

Besides learning, fatigue is also considered as human factor in the previous literature. In particular, the efficiency of a worker decreases because of the fatiguing effect (Battini et al., 2016). While several studies explicitly consider the fatiguing effect in tactical problems, only the work of Feng & Hu, (2021) takes into account different fatigue curves for different pickers in the batch assignment problem. In particular, the authors used a generalized logistic function to characterize the work fatiguing effect and estimate the processing time according to the order size and working efficiency, obtaining a more balanced assignment.

Finally, skills, intended as the speed at which a worker can perform a task, are others human factors affecting OP performance (De Lombaert et al., 2023). In particular, Jamal et al., (2022) defined four indicators, namely, the items' weight, volume, and storage height and the distance to be traveled to collect the object, to classify pickers according to their skills. They then proposed a batch assignment model to assign batches according to different skill levels. Similarly, Matusiak et al., (2017) demonstrate that different batch execution times reflect the different skill levels in a set of heterogeneous pickers. In particular, the authors addressed the batch assignment problem by considering the past performance of each picker, extracted from data in the WMS, such as the number of pick lines in a batch and the total batch travel distance These skills were then used to forecast the batch execution time through multiple linear regression models.

Table 1 provides a summary of studies aimed at addressing operational challenges in order picking systems. Each article is characterized by the specific problem it tackles, the approach employed for resolution, the type of optimized metric, the methodology for metric estimation, and factors influencing the metric.

Notably, only Kübler et al., (2020), Jamal et al., (2022) and Aboelfotoh et al., (2019) propose linear analytical models to optimize distance-based metrics in operational problems. All these papers incorporate both quantitative and spatial features, with only one considering picker skills in order batching and routing. When examining papers focusing on the batch assignment problem, predominant use of time-based metrics is observed. Most studies adopt linear analytical models to correlate these metrics with quantitative and spatial features of batches. Learning aspects are considered in two papers, while fatigue is addressed in just one. Learning is typically associated with search and pick time, and fatigue is expressed in terms of order size, item position, or follows a specific curve. Only one paper introduces a linear predictive model for the batch assignment problem, incorporating picker skills but neglecting the impact of learning and fatigue on picking time and their mutual interaction.

In comparing these studies, the integration of various factors emerge as significant aspect in optimizing time-based metrics for the batch assignment problem. However, concurrently considering all these factors poses a considerable challenge. Linear analytical models constrain the development of metrics dependent on batch features, picker learning, fatigue, and skills, resulting in marginal improvements in picking performance. These models treat both picker-independent and picker-dependent factors as fixed parameters, whereas a more effective approach would involve taking the time a picker with specific characteristics requires to complete a batch as an input parameter. This complexity necessitates a more advanced metric estimation methodology, such as non linear predictive models. Given the intricacy of human behavior and the interrelation of human factors, nonlinear predictive models may offer a more accurate representation. Moreover, the dynamic nature of behavior adds an additional challenge, emphasizing the

Table 1
Comparison of existing literature on operational problems in order picking.

Article	Problem			Solving approach	Optimized metric typology	Metric Estimation methodology	Factors affecting metric				
	B	A	R				Q	S	L	F	PS
(Cano et al., 2020)	✓	–	–	GA	Dist. based	Assumed	✓	✓			
(Kübler et al., 2020)	✓	–	✓	PSO LS + NN	Dist. based	Linear analytical model	✓	✓			
(Jamal et al., 2022)	✓	–	✓	B&C	Dist. based	Linear analytical model	✓	✓			✓
(Aboelfotoh et al., 2019)	✓	✓	–	CH	Dist. based	Linear analytical model	✓	✓			
(Alipour et al., 2020)	✓	–	–	RBH	Time based	Linear analytical model	✓	✓			
(Henn, 2015)	✓	✓	–	VNS	Time based	Linear analytical model	✓	✓			
(Zhang et al., 2017)	✓	✓	–	RBH	Time based	Linear analytical model	✓	✓			
(Scholz et al., 2017)	✓	✓	✓	VND	Time based	Linear analytical model	✓	✓			
(Ardjmand et al., 2020)	✓	✓	✓	CG	Time based	Linear analytical model	✓	✓			
(Pretolani et al., 2023)	✓	✓	✓	B&C	Time based	Linear analytical model	✓	✓			
(Mou, 2022)	✓	✓	–	GA VND	Time based	Linear analytical model	✓	✓	✓		✓
(Feng & Hu, 2021)	✓	✓	–	GA	Time based	Linear analytical model	✓	✓		✓	
(Zhang et al., 2023)	–	✓		GA	Time based	Linear analytical model	✓	✓	✓		
(Matusiak et al., 2017)	✓	✓	✓	ALNS	Time based	Linear predictive model	✓	✓			✓
This paper	–	✓	–	**GA**	**Time based**	**Non-linear predictive model**	✓	✓	✓	✓	

Q: Quantitative features of batches, S: Spatial features of batches, L: Learning of pickers, F: Fatigue of pickers, PS: Skills of pickers, B: Batching, R: Routing, A: Assignment, LS: Local Search, NN: Nearest Neighborhood, CH: Custom Heuristic, PSO: Particle Swarm Optimization, VNS: Variable Neighbourhood Search, VND: Variable Neighbourhood Descendent, RBH: Rule Based Heuristic, CG: Column Generation, B&C: Branch & Cut, ALNS: Adaptive Large Neighbourhood Search, GA: Genetic Algorithm.

need for a solving approach that considers past assignments.

3. Materials and methods

In this section, an overview of the investigated problem and its mathematical formulation is provided. Then, the framework of the approach proposed to solve the problem is presented, and the modules composing the proposed approach are detailed. Lastly, the experimental procedure followed to test and compare the effectiveness of the proposed approach is described.

3.1. Problem description

Manual OP systems process hundreds of customers' orders every day. Orders are usually composed of multiple picking lines, each referring to a specific quantity of an article to pick at a particular warehouse location. The need to optimize the efficiency of these systems has led practitioners and researchers to investigate different strategies to reduce the picking time required to complete all the picking missions. In particular, reducing the traveling distance has been a crucial research area in reducing picking time. As a result, a commonly adopted strategy involves grouping orders to create batches. Thereafter, a routing policy is applied to minimize the distance traveled to pick all the required products in a batch. Finally, these optimized batches are assigned to different pickers, who perform the picking missions related to each batch.

However, manual OP systems rely heavily on the human workforce and, as has emerged from the related literature, are strongly affected by human behavior. Therefore, acting solely on the traveling distance may lead to suboptimal solutions. An optimized assignment of batches to pickers could, in fact, positively affect system performance. Human pickers are subject to fatigue and learning mechanisms, which can affect the time required to complete all the picking missions.

In light of this evidence, the problem under investigation in this paper is how to assign future batches of orders to pickers to minimize the overall time required to complete all the picking missions. In particular, the assignment needs to optimize the learning and fatigue mechanism of pickers while considering the quantitative and spatial features of batches. The formulated problem thus relies on the following assumptions:

- The future orders to process are known in advance;
- Orders are already grouped into batches, and routing policies within batches are already optimized;
- The batch picking time (i.e., the time required by a specific picker to execute all the picking missions within a batch of orders) is considered a non-linear function of variables related to the pickers' learning and fatigue indicators and to the quantitative and spatial features of batches;
- Different assignment decisions lead to different values of the pickers' learning and fatigue indicators;
- Different assignment decisions do not lead to different values of quantitative and spatial features of batches;
- Learning indicators are expressed by values cumulated over the picker's entire working life, while fatigue indicators are represented by values cumulated within the picker's working day.
- No task variability is considered over time. Therefore, different assignment strategies can lead to different values of learning and fatigue indicators but not to different learning or fatigue profiles of pickers (Glock et al., 2019, Grosse & Glock, 2015, Grosse et al., 2015).

3.2. Problem formulation

In this section, a mathematical formulation of the problem is proposed. First, the sets, indexes, parameters, and variables are introduced. Thereafter, the model formulation is expressed.

3.2.1. Indexes and sets

$i \in I$: index of pickers
$j \in J$: index of batches
$c \in C$: index of the quantitative features of a batch
$p \in P$: index of the spatial features of a batch
$s \in S$: index of the positions (in sequence) in which a batch can be executed

3.2.2. Parameters

C_i: Maximum daily working time for picker i.

F_{ijsc}: Fatigue indicator obtained by cumulating all the values of the quantitative feature c, for all the batches executed within a specific day, by picker i before executing batch j in position s.

F_{ijsp}: Fatigue indicator obtained by cumulating all the different values of spatial features p, for all the batches executed within a specific day, by picker i before executing batch j in position s.

L_{ijsc}: Learning indicator obtained by cumulating all the values of the quantitative feature c, for all the batches executed by picker i before executing batch j in position s.

L_{ijsp}: Learning indicator obtained by cumulating all the different values of the spatial features p, for all the batches executed by picker i before executing batch j in position s.

$mean_{cj}$: Mean value of the quantitative feature c of all the articles grouped in batch j.

PL_{ic}: Learning indicator related to the quantitative feature c cumulated by picker i on previous working days.

PL_{ip}: Learning indicator related to the spatial features p cumulated by picker i on previous working days.

sum_{cj}: Sum of the values of the quantitative feature c of all the articles grouped on batch j.

$unique_{pj}$: Number of different values assumed by the spatial features p while picking all the articles in batch j.

3.2.3. Decision variables

x_{ijs}: Binary decision variable, equal to 1 if batch j has been assigned to picker i in position s, 0 otherwise.

3.2.4. Objective function

Following the notation, the objective function can thus be expressed as follows:

$$Min \sum_{i\in I}\sum_{j\in J}\sum_{s\in S} x_{ijs}T_{ijs} \tag{1}$$

T_{ijs} represents the picking time required by picker i to execute all the picking missions related to batch j when the batch is executed in position s.

3.2.5. Constraints

$$\sum_{i\in I}\sum_{s\in S} x_{ijs} = 1 \ \forall j \tag{2}$$

$$\sum_{i\in I}\sum_{j\in J} x_{ijs} = 1 \ \forall s \tag{3}$$

$$\sum_{j\in J}\sum_{s\in S} T_{ijs}x_{ijs} \leq C_i \ \forall i \tag{4}$$

$$\sum_{k\in J}\sum_{q=1,\cdots,s} sum_{ck}x_{ikq} = F_{ijsc} \ \forall i, \ \forall j, \ \forall s, \ \forall c \tag{5}$$

$$\sum_{k\in J}\sum_{q=1,\cdots,s} unique_{pk}x_{ikq} = F_{ijsp} \ \forall i, \forall j, \forall s, \forall p \tag{6}$$

$$PL_{ic} + F_{ijsc} = L_{ijsc} \ \forall i, \forall j, \forall s, \forall c \tag{7}$$

$$PL_{ip} + F_{ijsp} = L_{ijsp} \ \forall i, \forall j, \forall s, \forall p \tag{8}$$

$$x_{ijs} \in \{0, 1\} \ \forall i, \forall j, \forall s \tag{9}$$

Eqs. (2) and (3) represent the assignment constraints. The former expresses that each batch j can be assigned to only picker i in only one position s. The latter expresses that for each position s, there can be only one picker i in charge of a specific batch j. Eq. (4) represents the capacity constraint. In particular, for each picker i, the sum of all the picking times related to batches assigned to picker i in the different positions s cannot be higher than the maximum working time of picker i. Eqs. (5) and (6) allow the fatigue indicators to be computed. In particular, the fatigue of picker i when executing batch j in position s related to the quantitative features c is equivalent to the sum of all the sum_{ck} values related to all the batches k executed by picker i before batch j. Similarly, the fatigue of picker i, when performing batch j in position s related to spatial features p, is equivalent to the sum of all the $unique_{pk}$ values related to all the batches k executed by picker i before batch j. Eqs. (7) and (8) allow the learning indicators to be computed. They express that the learning reached by picker i when executing batch j in position s for a specific quantitative feature c (Eq. (7)) and a particular spatial feature p (Eq. (8)) is equal to the sum of the learning cumulated by that picker in previous working days and the fatigue indicator estimated for that picker. Eq. (9) expresses the integrality constraints.

3.3. Framework of the proposed approach

The proposed approach adopted to solve the problem is composed of three different modules: a features engineering module, a predictive module, and a prescriptive module. The framework of the proposed approach is shown in Fig. 1.

For a specific future day t + 1, the features engineering module models Eqs. (5) to (8). It receives as input the orders that need to be processed in day t + 1, their respective quantitative and spatial aggregate information (sum_{cj}, $mean_{cj}$, $unique_{pj}$), a random initial assignment of these orders to pickers, and the previously computed value of the parameters PL_{ic} and PL_{ip} containing the values of the learning variables related to all the pickers up to day t.

As a result, it generates the values of the parameters $F_{ijsp}, L_{ijsp}, F_{ijsc}, L_{ijsc}$ for the day t + 1. These values are adopted in the predictive module, together with the quantitative and spatial features of batches, to predict the batch picking time T_{ijs} for all the batches that will be executed on day t + 1. To capture the complexity and heterogeneity of manual OP systems, T_{ijs} is estimated in the predictive module through an empirical function H: $R^N \rightarrow R$, whose equation is directly learned from historical data through supervised ML techniques. The value of T_{ijs} can thus be expressed as:

$$T_{ijs} = H\left(i, sum_{cj}, mean_{cj}, unique_{pj}, F_{ijsp}, L_{ijsp}, F_{ijsc}, L_{ijsc}\right) \quad \forall i, \forall j, \forall s \tag{10}$$

where $\left(i, sum_{cj}, mean_{cj}, unique_{pj} F_{ijsc}, L_{ijsc}, F_{ijsp}, L_{ijsp}\right)$ express the concatenation of vectors such that Eq. (10) is equivalent to:

$$H\left(i, \left(sum_{1j}, \cdots, sum_{|C|j}\right), \left(mean_{1j}, \cdots, mean_{|C|j}\right), \cdots, \left(L_{ijs1}, \cdots, L_{ijs|C|}\right)\right) \forall i, \forall j, \forall s \tag{11}$$

$$H\left(i, sum_{1j}, \cdots, sum_{|C|j}, mean_{1j}, \cdots, mean_{|C|j}, \cdots, L_{ijs1}, \cdots L_{ijs|C|}\right) \ \forall i, \ \forall j, \ \forall s \tag{12}$$

The predictive module, which models function H, has thus been previously trained with historical data of the variables sum_{cj}, $mean_{cj}$, $unique_{pj}$, F_{ijsp}, L_{ijsp}, F_{ijsc}, L_{ijsc}, and T_{ijs} up to day t to learn the historical relationship between the pickers' learning and fatigue, the quantitative and spatial features of batches, and the batch picking time. Once trained, this module is able to provide predictions about the values T_{ijs} on day t + 1 resulting from a specific assignment of batches to picker. Lastly, the prescriptive module aims to find the best assignment of batches to pickers that minimizes the objective function described by Eq. (1) and satisfies the constraints reported by Eqs. (2) to (9). It receives as input

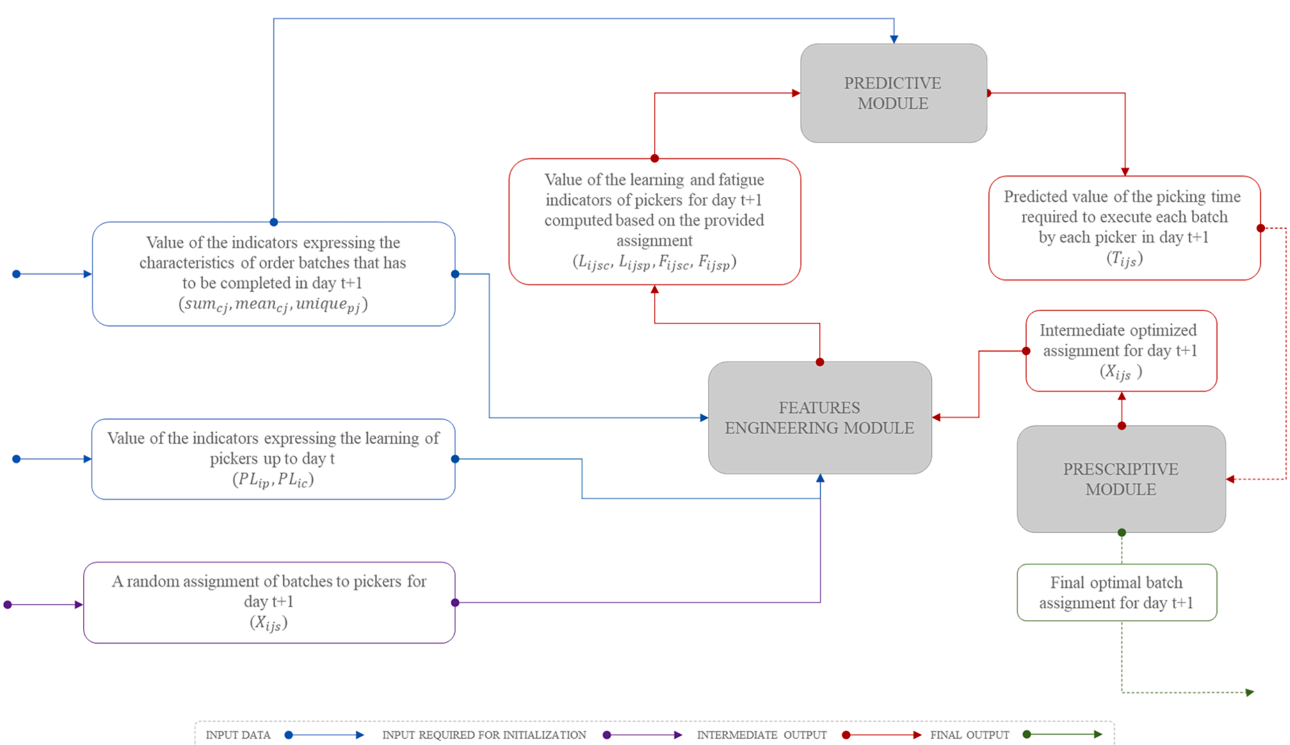

Fig. 1. Framework of the proposed approach.

the predicted time T_{ijs} coming from the predictive modules and progressively builds better assignments of orders to pickers.

3.3.1. Predictive module

As shown in Fig. 1, the predictive module adopted in the proposed approach communicates with the features engineering and prescriptive modules. In particular, it receives the value of the quantitative and spatial features of the batch (modeled by the parameters sum_{cj}, $mean_{cj}$, $unique_{pj}$) and the value computed in the feature engineering module of the variables related to the pickers' fatigue (modeled by the parameters F_{ijsc}, F_{ijsp}) and variables related to their learning mechanism (modeled by the parameters L_{ijsc}, L_{ijsp}). Although different options could have been adopted to build the predictive module, a ML model called Catboost was proposed for the predictive module.

The decision to utilize CatBoost, a non-linear gradient boosting model, for the predictive module was made for several reasons. Firstly, while linear models are typically quicker to train (Hastie & Pregibon, 2017), non-linear machine learning models such as CatBoost excel at capturing complex relationships and patterns, particularly in large datasets (Mahmoudi, 2018; Strang et al., 2018). Secondly, compared to other non-linear black box models like neural networks or support vector machines, gradient boosting models offer higher explainability (Burkart & Huber, 2021). Lastly, among gradient boosting algorithms, CatBoost, despite requiring longer training times than LightGBM, has demonstrated superior predictive performance when applied to tabular data (Dorogush et al., 2018). Given the complexity of the relationships in the dataset and the availability of ample tabular training data, CatBoost was deemed the most suitable choice for the predictive module.

3.3.2. Prescriptive module

The prescriptive module adopted to solve the problem formulated in Section 3.2 was obtained by adapting the GA proposed by Chu and Beasley (1997) to solve the general assignment problem.

Multiple alternative solutions have been evaluated to solve the problem described in Sections 3.1 and 3.2 based on their applicability and potential performance. In particular, linear programming, GAs, and other metaheuristic approaches like simulated annealing, particle swarm optimization, and ant colony optimization were all considered as possible solutions. However, linear programming was ruled out due to its limitation to problems with linear objective functions and constraints. Since the mathematical formulation of the problem under investigation featured a non-linear objective function (as expressed by Eq. (10), linear programming was deemed impractical.

Conversely, GAs, simulated annealing, particle swarm optimization, and ant colony optimization were all deemed suitable for handling non-linear functions. Considering that the efficacy of different metaheuristic techniques often depends on the specific problem at hand, a review of the literature was conducted to identify the most suitable algorithm. The literature survey revealed a prevalent use of GAs in solving operational optimization problems in OP, as evidenced by numerous studies (Cano et al., 2021, Cano et al., 2022, Cano et al., 2023; Pan et al., 2015; Schrotenboer et al., 2017). Consequently, a GA was chosen to be implemented as the solving technique in the prescriptive module.

A GA was thus implemented in the prescriptive module to communicate with the predictive one and progressively build the best assignment of batches to pickers. In the following, the encoding format of the solution and the initial population generation mechanism, the fitness and unfitness function, the population update procedure, and the termination criteria of the GA are detailed.

3.3.2.1. Solution encoding format and initial population. In the proposed GA, a generic solution (i.e., an individual) is an ordered vector of integer values. The length of the vector is equal to the number of batches to execute in a working day, while the integer numbers reported in each vector position (i.e., genes) indicate the picker in charge of that batch. This representation ensures that Eqs. (2) and (3) of the problem formulated in Section 3.2 are automatically satisfied since exactly one picker is assigned to each batch in a specific position. However, this representation does not guarantee that the capacity constraints expressed by Eq. (4) are satisfied. According to this encoding format, the initial population of the GA is thus generated by N individuals, constructed by randomly assigning a picker to each batch. Fig. 2 illustrates each individual's encoding format and how these individuals constitute the population.

3.3.2.2. Fitness and unfitness functions. For each individual of the population, the value of two functions is computed to guide the GA in generating progressively better solutions. These two values are obtained by integrating the predictive module with the prescriptive and features engineering ones. The fitness function f_k, of the solution k is equal to:

$$f_k = \sum_{j\in J} t^{pred}_{s_{kj},j} \tag{13}$$

where s_{kj} represents the picker assigned to batch j in solution k, and $t^{pred}_{s_{kj},j}$ is the time the picker requires to complete batch j in solution k obtained from the predictive module once the learning and fatigue features of the new solution are computed by the feature engineering module.

In the same way, the unfitness value of a solution k, u_k, which is a measure of its infeasibility, is computed as follows:

$$u_k = \sum_{i\in I} \max\left[0, \left(\sum_{j\in J, s_{kj}=i} t^{pred}_{s_{kj},j}\right) - C_i\right] \tag{14}$$

For a specific solution, Eq. (14) thus compares the predicted picking time required by the picker to complete all the missions assigned to them and their respective maximum available working time. Therefore, a feasible solution to the problem is obtained only when $u_k = 0$, that is, the predicted time for each picker to complete all the missions assigned, does not exceed their maximum available daily working time.

3.3.2.3. Population update mechanism. Once the fitness and unfitness functions have been computed, two individuals (i.e., parents) are selected for reproduction. A binary tournament selection method is adopted, as in Chu and Beasley (1997), to generate each parent. In each binary tournament, two individuals are randomly selected, and the one with the lowest fitness function is chosen to create a parent. Thereafter, a one-point crossover operator is applied to generate a new individual (i.e., a child). Here, a crossover point p is randomly selected, and the resulting child is composed of the first p genes from their first parent and the remaining ones from the second parent. Furthermore, a mutation procedure is also applied to the generated child at each iteration. In the mutation procedure, two randomly selected genes are exchanged (i.e., batches are exchanged between two pickers). The crossover and mutation procedures are executed at each iteration of the GA (i.e., with a crossover and mutation probability equal to 1).

Lastly, a local improvement is performed to improve the feasibility of the generated child. For each picker i, if their working capacity is exceeded, a batch executed by picker i is randomly selected and assigned to the first picker in set I with enough free capacity. The generated child is thus adopted to replace the individual in the population with the highest unfitness value (i.e., the most infeasible solution). If the population consists of all feasible solutions ($u_k = 0, \forall k$), the individual with the lowest fitness is replaced.

3.3.2.4. Termination criteria. Starting from the randomly initially generated population, the fitness and unfitness computation step, the population updating mechanism, and the computation of the fitness and unfitness values for the newly generated child are repeated until M non-duplicate children have been generated without improving the best solution found so far. In addition, according to the daily nature of the

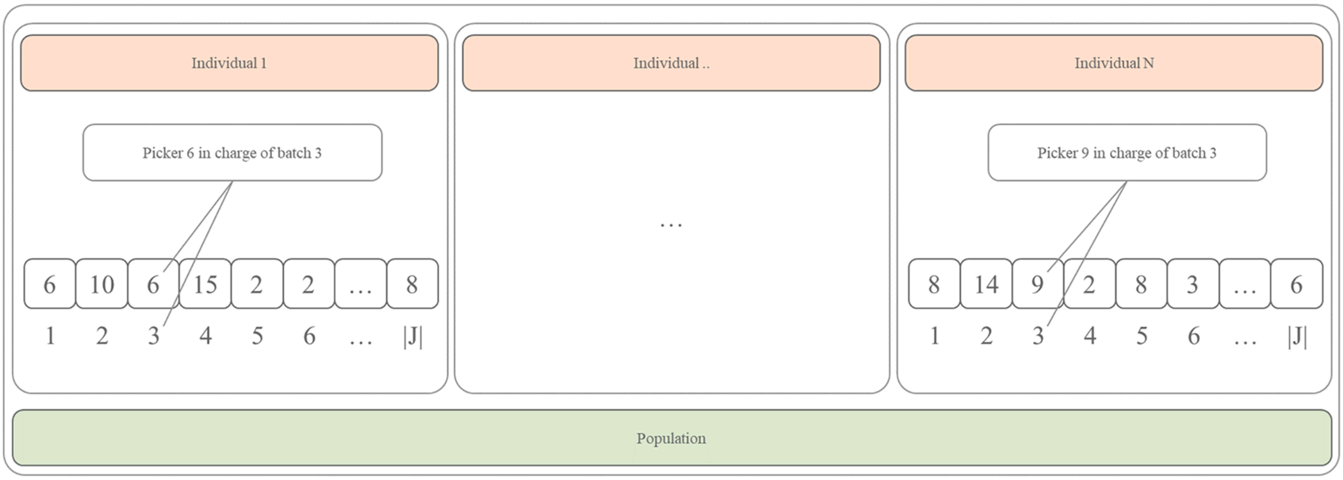

Fig. 2. GA population and individuals encoding format.

problem, a time limit of 8 h is adopted as an additional termination criterion.

3.4. Experimental design

In this section, the data collected to test and compare the proposed approach against different benchmarks are illustrated. Thereafter, the different comparisons and analyses made are detailed together with the metrics adopted to evaluate each comparison. Lastly, the settings adopted to conduct the experimental comparisons and analysis are presented.

3.4.1. Data collection

A real case study of a retail grocery store was used to evaluate the potential of the proposed approach. The investigated warehouse consists of 27 areas, 50 different aisles, 13 picking levels, and 985 different picking locations. In the experiments, it was assumed that the batch creation and routing policies are not modifiable, as set out in Section 3.1. Furthermore, managers are supposed to know the picking missions that must be executed a day in advance due to planning activities. Data were taken from the historical record provided by the WMS over one year, from 2021 to 2022. The period selected to test the approach was identified based on interviews with domain experts, who described a representative number of days to capture the dynamics occurring in the warehouse. The level of picker turnover identified in the warehouse did indeed suggest that a longer period would not lead to additional information. Furthermore, more than 300 pickers were found to be working over the selected period, which was assumed to be a sufficient number to investigate learning and fatigue differences. In addition, no task variability has been observed over this period. A statistical summary of the main daily operational indicators related to the investigated case study is reported in Table 2.

Table 2
Statistical summary of warehouse operational indicators.

Indicators	Mean	Standard deviation	1° Quartile	2° Quartile	3° Quartile
Number of daily pickers	133,4	26,8	125	141	150
Daily picking time [h]	668,2	195,0	590,0	693,2	765,1
Number of daily batches	1689,2	429,9	1501,5	1787,0	1953,0
Number of different areas visited by a picker in a day	1,7	1,1	1	1	2
Number of different aisle visited by a picker in a day	8,6	5,5	6	8	11
Number of different picklocation visited by a picker in a day	179,4	64,4	148,0	192,0	223,0
Number of different levels visited by a picker in a day	4,2	2,0	3,0	3,0	6,0

3.4.2. Experimental comparisons and analyses

Three different comparisons were performed to investigate the effectiveness of the three different modules composing the approach described in Section 3.3. In addition, a specific analysis was performed to explain how the quantitative and spatial features of batches and pickers' fatigue and learning indicators affect the picking time.

3.4.2.1. Predictive accuracy comparison. In the first comparison, the effectiveness of the proposed predictive modules based on a non-linear CatBoost model was tested by first comparing it against a linear predictor and, subsequent, against other non-linear models. A linear regression model (LINEAR) was used as the linear predictor benchmark, while a multi-layer perceptron (MLP) and an XGBoost (XGBOOST) model were adopted as non-linear benchmarks. All these models were implemented in Python, recurring to the Scikit-Learn library (Pedregosa et al., 2011). The mean absolute error (MAE), mean squared error (MSE), and coefficient of correlation (R^2) were used to compare the performance of the proposed CatBoost model against those of the benchmarks. Furthermore, the training time required by each model to learn the relationship between the picking time and the adopted features is reported in seconds. The MAE, MSE, and R^2 metrics are defined as follows:

$$MAE = \frac{1}{N}\sum_{j=1}^{N}\left|Y_j - \widehat{Y}_j\right| \quad (15)$$

$$MSE = \frac{1}{N}\sum_{j=1}^{N}\left(Y_j - \widehat{Y}_j\right)^2 \quad (16)$$

$$R^2 = 1 - \sum_{t=1}^{T}\frac{\left(Y_j - \widehat{Y}_j\right)^2}{\left(Y_j - \overline{Y}\right)^2} \quad (17)$$

where Y_j is the true value of the batch picking time for batch j, $\overline{Y}$ is the mean value of the batch picking time over the N observations, and $\widehat{Y}_j$ is the predicted picking time for batch j.

3.4.2.2. Features engineering comparison. In the second comparison, the effectiveness of the features engineering module was evaluated through an ablation analysis. In an ablation analysis, the performance of systems is investigated by removing components to understand the contribution of each component to the system. Therefore, the predictive capability of the proposed model when using the overall set of features (ORDER & FATIGUE & LEARNING) was compared with the predictive capability of the proposed model when a progressively reduced group of features was adopted. First, only the features related to quantitative and spatial features of batches and learning (ORDER & LEARNING) were adopted as benchmarks. Thereafter, only the features related to the quantitative and spatial features of batches and fatigue (ORDER & FATIGUE) were selected. Lastly, only the features related to the quantitative and spatial features of batches (ORDER) were provided to the model. The same metrics adopted in Section 3.4.2.1 were chosen for this analysis.

3.4.2.3. Batch assignment strategy comparison. In the third comparison, the advantages of the proposed approach were tested against an assignment strategy that did not explicitly consider learning and fatigue during the assignment. The procedure adopted for the comparison is illustrated in Fig. 3.

In the proposed strategy, the prescriptive module (GA) is guided by a predictive module (H_{BLF}) that predicts the future batch picking time (T_{ijs}) considering from the beginning the quantitative and spatial features of batches (B) and the learning (L) and fatigue indicators (F) of pickers resulting from a given assignment (X). On the other hand, to simulate the benchmark strategy, the prescriptive module was guided in this case by a predictive module (H_B) that predicts the batch picking time only based on the quantitative and spatial features of batches (B).

Once the final optimal assignments were obtained respectively from the benchmark strategy (X_B^*) and the proposed strategy (X_{BLF}^*), an ex-post computation of pickers' learning and fatigue indicators related to the optimal solution obtained from the proposed approach and from the benchmark strategy was performed. Thereafter, the learning and fatigue indicators computed ex post, together with the quantitative and spatial features of the batches, were used to recompute the picking time based on a unique common predictive module (H_{BLF}). Lastly, relying on the new predicted picking time (T_{ijs}^*) the fitness ($F_{BLF}\left(T_{ijs}^*\right)$) and unfitness ($U_{BLF}\left(T_{ijs}^*\right)$) value of the optimal solution identified by the proposed approach and the fitness ($F_B\left(T_{ijs}^*\right)$) and unfitness $U_B\left(T_{ijs}^*\right)$ value of the optimal solution identified by the benchmark strategy were compared according to Eqs. (13) and (14) and adopted as metrics for the comparison.

The adopted procedure was chosen to provide a fair simulated comparison between the two approaches. Indeed, it was necessary to use a common unique predictive module (H_{BLF}) to avoid differences in the results related to the different levels of accuracy of the two predictive modules. Furthermore, the decision to adopt the predictive module (H_{BLF}) for the comparison was based on the assumption that the picking time predicted using as input the quantitative and spatial features of batches and the learning and fatigue indicators of pickers represents the closest estimation to the reality of the picking time a picker would have required to complete a batch.

A comparison between an assignment strategy that assigns batches to pickers without considering their learning and fatigue and the proposed approach, which explicitly considers these elements during the assignment, thus aims to understand if the latter approach can lead to more feasible solutions (lower unfitness) or a reduction of the overall picking time (lower fitness). On the one hand, an assignment of batches to pickers based on a prediction that does not consider their learning and fatigue mechanism from the beginning could result, in reality, in an amount of work that exceeds the daily working time of pickers. On the other hand, assuming that a feasible solution is obtained in both cases, not optimizing learning and fatigue could, in reality, lead to a higher picking time due to the fact that this aspect was not taken into consideration.

3.4.2.4. Model explainability analysis. In conclusion, to investigate the relevance of the overall set of adopted features and understand how these features affect the picking time, SHAP values (Lundberg & Lee, 2017) were adopted to explain the predictive module. For each batch of picking orders considered in the case study, SHAP values provided information about how each feature contributes to the final predictions of the picking time of that batch. In particular, the SHAP value of each feature reports the positive or negative contribution of the feature with

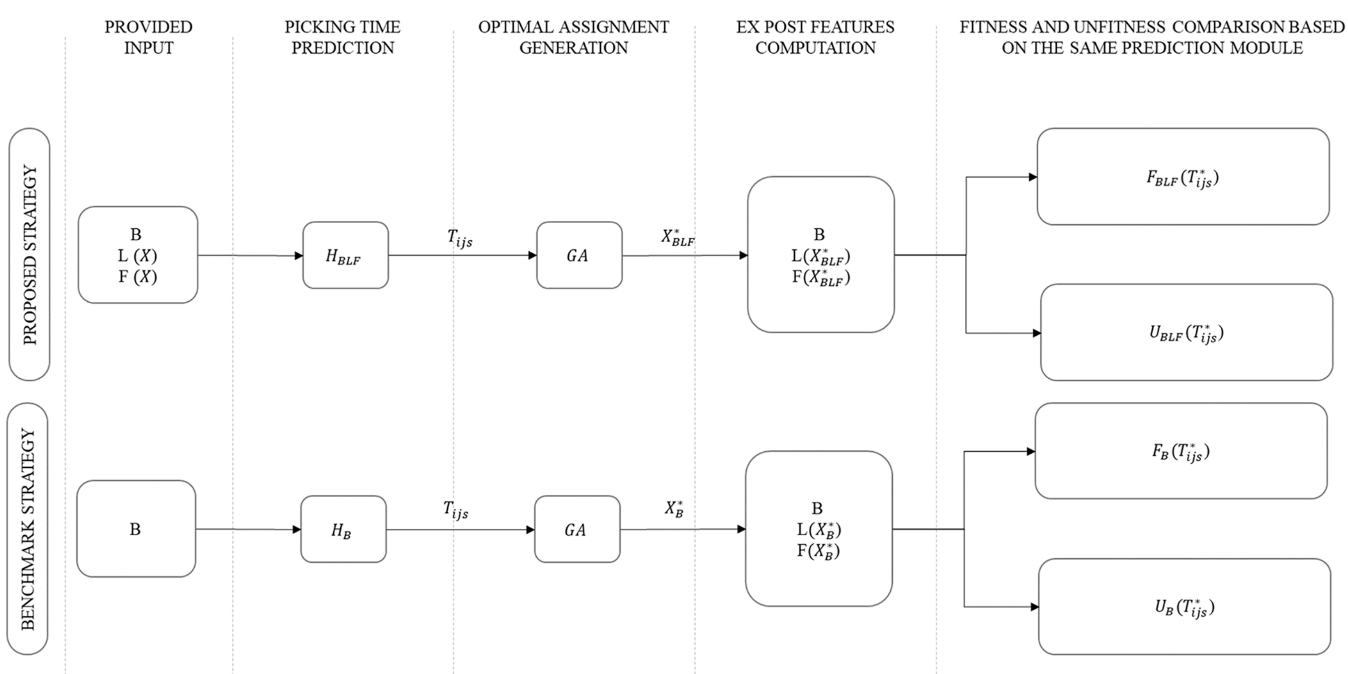

Fig. 3. Batch Assignment Strategy Comparison.

respect to the mean picking time computed over all the considered batches. A SHAP value of 10 for a specific feature when considering a specific batch thus indicates that the value reported for that feature seems to increase the picking time of that batch of 10 units with respect to the mean picking time computed over all the batches.

3.4.3. Experimental set-up

The collected historical data described in Section 3.4.1 were split into two consecutive temporal subsets to perform the comparisons and analyses described in the previous sections. The first 219 days of the dataset, corresponding to 90 % of the total observations, were used as the training set, while the remaining 21 days were used as the test set. The training set was used to allow the predictive module to learn the historical relationship between the batch picking time, the pickers' learning and fatigue indicators, and the quantitative and spatial features of batches. The test set, which was not used to train the predictive module, was used to perform the comparisons described in Section 3.4.2.

A description of the cardinality and values of sets I, J, C, and P defined in Section 3.2.1 for the 21 days constituting the test set is provided in Table 3. Furthermore, a description of the 28 different features used in the prediction is reported in Tables 4, 5, and 6. In particular, Table 4 reports the 12 features adopted to model the quantitative and spatial features of batches (representing the parameters sum_{cj}, $mean_{cj}$, and $unique_{pj}$). Table 5 reports the remaining eight features describing pickers' learning mechanisms (representing the parameters L_{ijsc} and L_{ijsp}). Table 6 reports the eight features adopted to express pickers' fatigue (representing the parameters F_{ijsc} and F_{ijsp}). The hyperparameters adopted for the predictive models are reported in Table 7. These hyperparameters were selected by applying a random search strategy over the training set. Lastly, details about the GA parameters adopted in the comparison described in Section 3.4.2.3 are reported in Table 8. The number of solutions per population and the number of accepted attempts made without improvement were chosen as reported in the original algorithm proposed by Chu and Beasley (1997).

Table 3
Cardinality and values of the parameters adopted in the prescriptive modules related to the days of the test set.

Day	|I|	|J|	|S|	Set C	Set P
1	151	1785	1785	weight, volume, quantity,distance	aisle, pick level, pick location
2	153	1913	1913	weight, volume, quantity,distance	aisle,pick level, pick location
3	147	1966	1966	weight, volume, quantity,distance	aisle,pick level, pick location
4	162	1995	1995	weight, volume, quantity,distance	aisle,pick level, pick location
5	151	2054	2054	weight, volume, quantity,distance	aisle,pick level, pick location
6	147	1989	1989	weight, volume, quantity,distance	aisle,pick level, pick location
7	146	2107	2107	weight, volume, quantity,distance	aisle,pick level, pick location
8	152	1864	1864	weight, volume, quantity,distance	aisle,pick level, pick location
9	155	2036	2036	weight, volume, quantity,distance	aisle,pick level, pick location
10	144	1766	1766	weight, volume, quantity,distance	aisle,pick level, pick location
11	149	1877	1877	weight, volume, quantity,distance	aisle,pick level, pick location
12	145	1589	1589	weight, volume, quantity,distance	aisle,pick level, pick location
13	142	1329	1329	weight, volume, quantity,distance	aisle,pick level, pick location
14	140	1518	1518	weight, volume, quantity,distance	aisle,pick level, pick location
15	160	1549	1549	weight, volume, quantity,distance	aisle,pick level, pick location
16	155	1542	1542	weight, volume, quantity,distance	aisle,pick level, pick location
17	149	1820	1820	weight, volume, quantity,distance	aisle,pick level, pick location
18	153	1637	1637	weight, volume, quantity,distance	aisle,pick level, pick location
19	155	1630	1630	weight, volume, quantity,distance	aisle,pick level, pick location
20	149	1551	1551	weight, volume, quantity,distance	aisle,pick level, pick location
21	155	1673	1673	weight, volume, quantity,distance	aisle,pick level, pick location

Table 4
Quantitative and spatial features adopted in the predictive module.

ID	Category	Description
SUM_PICKED_WEIGHT_[KG]	sum_{cj}	Sum of the weight of all the articles to pick in batch j
SUM_PICKED_VOLUME_[L]	sum_{cj}	Sum of the volume of all the articles to pick in batch j
SUM_PICKED_QTY	sum_{cj}	Sum of the quantity of all the articles to pick in batch j
SUM_DISTANCE	sum_{cj}	Sum of the distance to travel to pick all the articles in batch j
MEAN_PICKED_WEIGHT_[KG]	$mean_{cj}$	Mean of the weight of all the articles to pick in batch j
MEAN_PICKED_VOLUME_[L]	$mean_{cj}$	Mean of the volume of all the articles to pick in batch j
MEAN_PICKED_QTY	$mean_{cj}$	Mean of the quantity of all the articles to pick in batch j
MEAN_DISTANCE	$mean_{cj}$	Mean distance to travel to pick all the articles in batch j
UNIQUE_AREAS	$unique_{pj}$	Number of different areas to visit to pick all the articles in batch j
UNIQUE_AISLE	$unique_{pj}$	Number of different aisles to visit to pick all the articles in batch j
UNIQUE_PICKLOCATION	$unique_{pj}$	Number of different pick locations to visit to pick all the articles in batch j
UNIQUE_LEVELS	$unique_{pj}$	Number of different levels to visit to pick all the articles in batch j

Table 5
Learning features adopted in the predictive module.

ID	Category	Description
LEA_SUM_PICKED_WEIGHT_[KG]	L_{ijsc}	Historical cumulated sum of picked weight up to batch j
LEA_SUM_PICKED_VOLUME_[L]	L_{ijsc}	Historical cumulated sum of picked volume up to batch j
LEA_SUM_PICKED_QTY	L_{ijsc}	Historical cumulated sum of picked quantity up to batch j
LEA_SUM_DISTANCE	L_{ijsc}	Historical sum of traveled distance up to batch j
LEA_UNIQUE_AREAS	L_{ijsp}	Historical cumulated number of different visited areas of the warehouse up to batch j
LEA_UNIQUE_AISLE	L_{ijsp}	Historical cumulated number of different visited aisles of the warehouse up to batch j
LEA_UNIQUE_PICKLOCATION	L_{ijsp}	Historical cumulated number of different visited pick locations of the warehouse up to batch j
LEA_UNIQUE_LEVELS	L_{ijsp}	Historical cumulated number of different visited levels of the warehouse up to batch j

4. Results

In this section, the results of the comparisons and analyses described in Section 3.4.2 are reported.

Table 6
Fatigue features adopted in the predictive module.

ID	Category	Description
FAT_SUM_PICKED_WEIGHT_ [KG]	F_{ijsc}	Daily cumulated sum of picked weight up to batch j
FAT_SUM_PICKED_VOLUME_ [L]	F_{ijsc}	Daily cumulated sum of picked volume up to batch j
FAT_SUM_PICKED_QTY	F_{ijsc}	Daily cumulated sum of picked quantity up to batch j
FAT_SUM_DISTANCE	F_{ijsc}	Daily sum of traveled distance up to batch j
FAT_UNIQUE_AREAS	F_{ijsp}	Daily cumulated number of different visited areas of the warehouse up to batch j
FAT_UNIQUE_AISLE	F_{ijsp}	Daily cumulated number of different visited aisles of the warehouse up to batch j
FAT_UNIQUE_PICKLOCATION	F_{ijsp}	Daily cumulated number of different visited pick locations of the warehouse up to batch j
FAT_UNIQUE_LEVELS	F_{ijsp}	Daily cumulated number of different visited levels of the warehouse up to batch j

Table 7
Hyperparameter values of predictive models.

Model	Hyperparameters	Values
MLP	Loss function	MSE
MLP	Hidden_layer_sizes	100
MLP	Learning_rate	0.001
MLP	Iterations	200
XGBOOST	Loss function	MSE
XGBOOST	Learning rate	0.3
XGBOOST	Iterations	100
CATBOOST	Loss function	MSE
CATBOOST	Learning rate	0.1
CATBOOST	Iterations	1000

Table 8
Parameters of GA.

Parameters	Values
Solution per population	100
Number of accepted attempts without improvement	500′000
Parent selection method	Binary tournament
Child generation method	One point crossover
Mutation method	One point exchange

4.1. Results of the predictive accuracy comparison

Fig. 4 reports the accuracy reached in different metrics by the adopted non-linear CatBoost model against that reached by the benchmarks when the entire set of 28 features modeling the spatial and the quantitative features of batches and pickers' learning and fatigue mechanisms are adopted. According to the chart, the proposed CatBoost model outperforms both the linear and non-linear models in every accuracy metric. The highest performance difference is reported between the proposed CatBoost model and the linear regression model. In particular, a reduction of the prediction error of 43.0 % in terms of MAE, 83.0 % in terms of MSE, and an increase of 5,3% in terms of R^2 is reported. This improvement comes, however, at the price of an increase in the training time of the model of 98.4 %. On the other hand, minor improvements can be found against the non-linear benchmarks. In particular, considering the XGBoost model, which represents the best model in the group of non-linear benchmarks, a reduction of the prediction error of 2.8 % in terms of MAE, 10.2 % in terms of MSE, and an increase of 0.6 % in terms of R^2 is reported. In addition, the XGboost model was found to be 91.5 % faster than the proposed one.

4.2. Results of the features engineering comparison

Fig. 5 reports the accuracy reached in different metrics by the proposed non-linear CatBoost model when different sets of features are provided as input to the model. Compared to the set of features that only considers the quantitative and spatial features of a batch (ORDER), the addition of features related to pickers' fatigue (ORDER & FATIGUE) or to pickers' learning (ORDER & LEARNING) can, respectively, lead to a reduction of the prediction error of 0.6 % and 20.3 % in terms of MAE. Furthermore, considering the entire set of features (ORDER & FATIGUE & LEARNING) enables an overall reduction of 21.5 % in terms of MAE compared to the initial set of features (ORDER). The reduction of the prediction error comes at the cost of an increase in training time of the model of 6.1 % when considering the ORDER & FATIGUE set of features, 42.7 % when considering the ORDER & LEARNING set of features, and 46.1 % when considering the entire set of features (ORDER & FATIGUE & LEARNING).

4.3. Results of the batch assignment strategy comparison

Table 9 reports the comparison between the predicted fitness and unfitness values computed for the optimal solutions obtained from the proposed approach and from the benchmark strategy according to the procedure described in Section 3.4.2.4.

The results reported in Table 9 were analyzed to investigate whether there was a significant difference between the fitness values resulting from the proposed approach and those resulting from adopting the benchmark strategy. The normal distribution of data was checked and confirmed by conducting a Shapiro–Wilk test (Shapiro & Wilk, 1965) and a *t*-test was adopted for the investigation. A p-value of 0.9983, thus greater than the adopted significance level of 0.005, was found, suggesting that there is no significant difference between the mean of the fitness values reported from the proposed approach and that reported from the benchmark strategy.

The same procedure was used to investigate whether there was a significant difference between the unfitness values resulting from the proposed approach and the benchmark. In this case, however, the normal distribution of data was not confirmed by the Shapiro–Wilk test. A Wilcoxon signed-ranked test (Wilcoxon, 1945) was thus performed in this case. Here, a p-value of 8.85 x 10^{-5}, thus smaller than the adopted significance level of 0.005, was found, suggesting that there is a significant difference between the mean of the unfitness values reported from the proposed approach and that reported when using the benchmark strategy.

Indeed, an assignment strategy that assigns batches to pickers based on a batch picking time predicted without considering pickers' learning and fatigue indicators results, on average, in 95 % of unfeasible solutions compared to the 24 % obtained from the proposed strategy.

4.4. Results of the model explainability analysis

Fig. 6 reports the distribution of the SHAP values related to the first 20 most relevant features of the 28 considered. The features are ranked from the highest to lowest according to their relevance. Furthermore, how the value of each feature affects its SHAP value is reported. According to the chart, the most relevant features are the quantitative and spatial features of batches, such as the number of different pick locations to visit and the mean volume to pick in a batch. The higher their value, the higher the picking time required to complete the batch. Learning indicators follow: the higher their value, the lower the picking time to complete the batch. This is true for all the learning indicators, with the exceptions of those related to the cumulated sum of traveled distance and the cumulated number of different visited areas of the warehouse. Lastly, fatigue indicators report a lower relevance. Furthermore, the higher they are, the lower the picking time they generate.

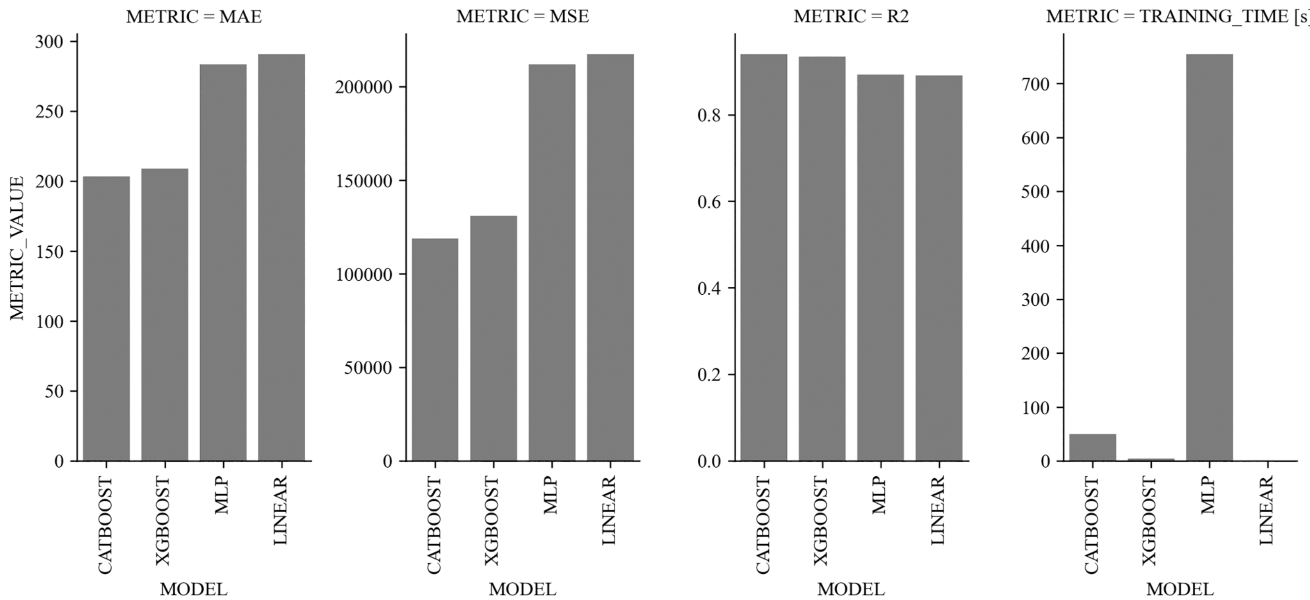

Fig. 4. Model accuracy comparison when quantitative and spatial features of batches and pickers' learning and fatigue indicators are adopted as input features for the predictive models.

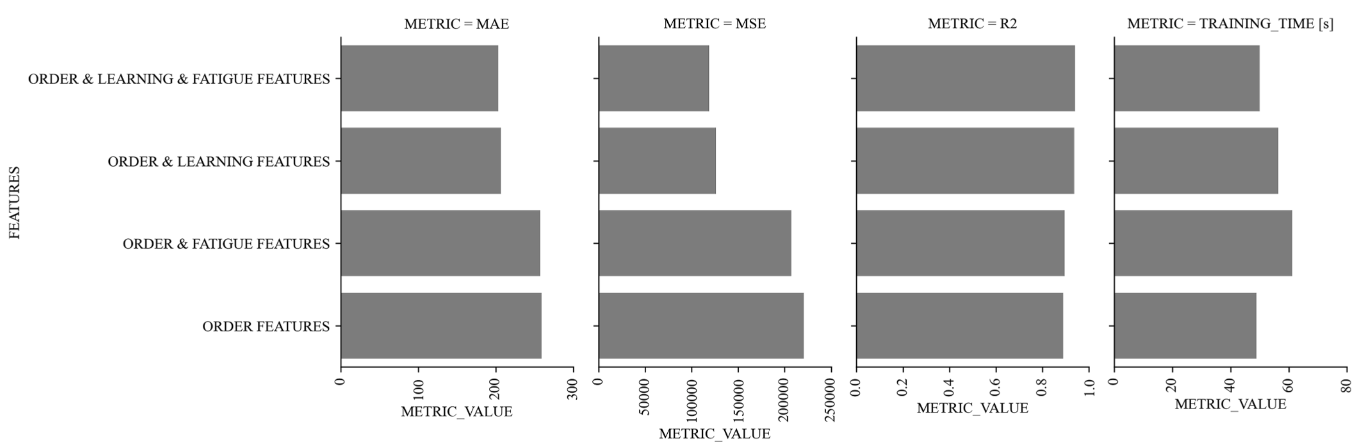

Fig. 5. Comparison of the accuracy reached by the proposed CatBoost model when different sets of features are provided as input to the predictive model.

5. Discussion

This paper investigates the problem of simultaneously considering mental (learning) and physical (fatigue) human factors with quantitative and spatial features of batches when planning the future assignment of batches to pickers in manual order-picking systems. In particular, a new approach is proposed to solve the problem. In the proposed approach, a non-linear machine learning model has been adopted to predict the batch picking time based on quantitative and spatial features of batches and on fatigue and learning indicators of pickers. The predicted picking time is then integrated within a customized GA to produce optimized assignment of batches to pickers considering their learning and fatigue mechanisms.

Multiple experiments have been performed based on a real dataset reporting 1 year of data from a manual warehouse of grocery retail to investigate the advantages of the modules composing the proposed integrated approach. First, the predictive accuracy of the proposed non linear model has been compared with those resulting from other linear and non linear benchmark. Afterward, an ablation analysis has been performed to investigate the relevance of considering different features when predicting the batch picking time. Moreover, a comparison between an assignment strategy that assigns batches to pickers without considering their learning and fatigue indicators from the beginning is performed to test the advantages of the proposed assignment strategy. Lastly, a model explainability analysis has been performed to investigate how and how much quantitative and spatial features of batches and learning and fatigue indicators of pickers affect the batch picking time.

Empirical results suggested significative advantages in terms of reduction of prediction error when predictions of the batch picking time are performed based on the proposed non linear model rather than a linear one. This evidence is supported by a wide corpus of literature reporting the advantages of ML when there is the need to handle high dimensional data to describe complex environments. Moreover, although machine learning has been rarely adopted to predict the batch picking time, this evidence is in line with the study conducted in Falkenberg & Spinler, (2023), where machine learning models have been found to work well to predict operators' productivity in manual order picking systems.

Moreover, the improved predictive accuracy reported when considering simultaneously quantitative and spatial features of batches and learning and fatigue indicators of pickers support the necessity also reported in other studies to consider human factors in order picking

Table 9
Comparison between the predicted fitness and unfitness values of solutions obtained from the proposed assignment strategy against a benchmark strategy that does not consider fatigue and learning aspects during the assignment.

Day	Fitness value		Unfitness value		Number of operators	
	Proposed	Benchmark	Proposed	Benchmark	Proposed	Benchmark
1	2.385.471	2.376.083	0	26.354	151	151
2	2.452.008	2.438.553	0	6.534	153	153
3	2.337.147	2.308.838	0	21.481	147	147
4	2.554.113	2.561.188	0	21.176	162	162
5	2.719.931	2.731.041	4.202	41.402	151	151
6	2.603.606	2.585.442	2.171	47.745	148	148
7	2.940.025	2.944.637	20.508	84.296	146	146
8	2.388.639	2.384.664	0	36.743	152	152
9	2.632.972	2.629.194	0	29.759	155	155
10	2.502.851	2.507.411	651	40.882	144	144
11	2.458.728	2.457.352	0	24.898	149	149
12	2.352.343	2.343.415	0	36.886	146	146
13	1.738.270	1.769.415	0	7.842	142	142
14	2.332.028	2.348.254	0	9.167	140	140
15	2.521.202	2.509.703	0	39.035	160	160
16	2.596.117	2.616.591	596	60.503	155	155
17	2.469.213	2.465.966	0	49.673	150	150
18	2.241.687	2.224.987	0	0	153	153
19	2.069.211	2.086.261	0	6.995	156	156
20	2.129.173	2.138.582	0	7.143	149	149
21	2.356.388	2.350.036	0	7.534	155	155

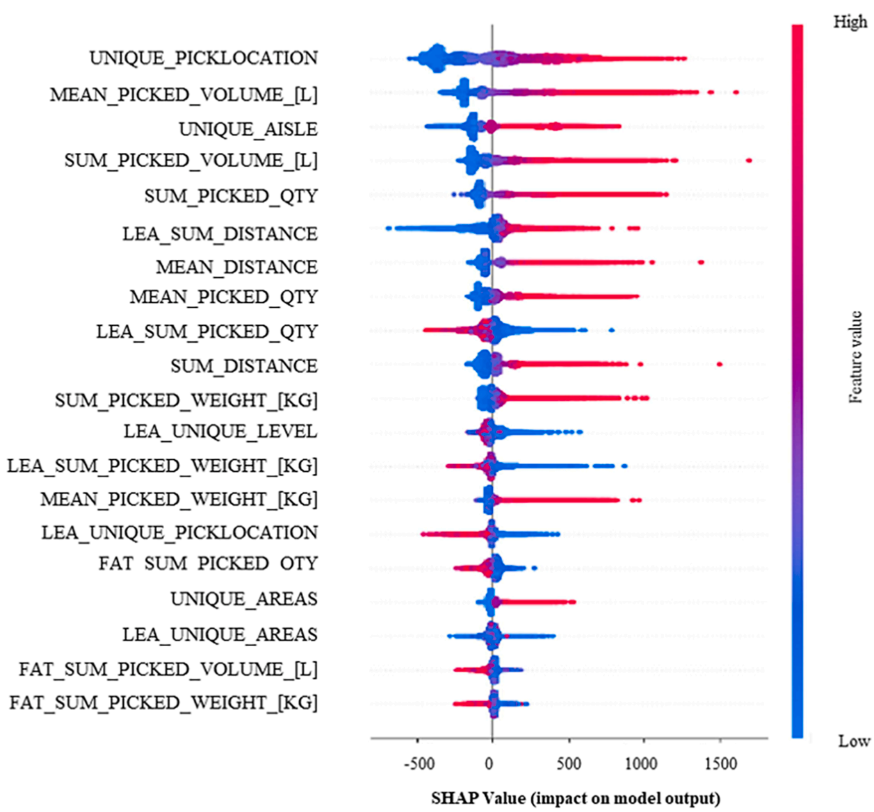

Fig. 6. Predictive model explanation obtained by computing the SHAP values of the model.

problems (Grosse et al., 2017). This necessity is made even more evident from the results of the comparison performed between the proposed assignment strategy and an assignment strategy that assigns batches to pickers based on a batch picking time predicted without considering learning and fatigue indicators. The latter assignment has indeed shown to lead to unfeasible solutions for the assignment problem. The lack of consideration of learning and fatigue mechanism can indeed lead to inaccurate estimations of the batch picking time leading to assignments that exceed the working time of pickers when adopted in reality.

Lastly, some unexpected results were found when the non-linear predictive model was explained, recurring with the use of SHAP values. While learning indicators were expected to decrease the picking time and fatigue ones to increase it, this expectation has not always been met. Learning indicators related to the cumulated sum of traveled distance and the cumulated number of different visited areas of the warehouse have been shown to lead to an increase in picking time. On the contrary, fatigue indicators that were expected to increase the picking time have been found to reduce it. Explanations for these results could be due to the procedure adopted to compute this indicator. Learning indicators have been indeed assumed to be computed by cumulating a value over the entire working history of a picker, while fatigue indicators have been assumed to be computed over each working day of a picker. According to these explanations, learning could also happen daily, while fatigue can also be viewed across a whole week, month or even a picker's entire working life. Similar results have been found in Winkelhaus et al., (2018), where a small amount of fatigue has been found to support learning effects. Furthermore, in Daria et al., (2017) an overuse of operators has been identified as a possible source of lower efficiency in the long-term period.

6. Conclusion and future research

6.1. Practical implications

The results of this study lead to three main practical implications. First, results highlight the capability of the presented approach to better predict the batch picking time in manual order picking systems. Predictions of task performance times bear significant relevance for companies, as they enable optimized resource planning and allocation as well as enhanced operational efficiency. Precise time predictions are instrumental and required for streamlining workflow, reducing bottlenecks, and ensuring a more balanced workload distribution. Furthermore, the accuracy in forecasting performance times contributes to improved customer satisfaction by ensuring timely order announcement and fulfillment, a critical aspect in e-commerce operations.

Second, the presented approach allows to generate more human-centric assignment routines for allocating batches towards pickers. This aspect is particularly relevant as it takes human differences into account on an individual level when allocating tasks. Managers can apply the proposed methodology to mitigate the detrimental effects of fatigue while simultaneously amplifying the beneficial impacts of accumulated experience. Recognizing and accounting for these human factors is not merely about adapting to individual capabilities, it also involves quantifying and using the individual strengths of human workers and tailoring task assignments accordingly. These alignments are especially relevant in the evolving landscape of Industry 5.0, where the synergy between human-centricity and technological advancements aims at promoting worker well-being and job satisfaction while maintaining or increasing high productivity levels.

Third, the results stress the general necessity for managers to consider human factors in manual order-picking systems. Prior research on work-rate variability has identified the task itself, the work environment, and the human worker executing the task as major sources of variance in average process performance. Human-related performance variance mainly originates from physical and mental conditions, social aspects, or work experience. In the context of Operations Management, specifically in the case of order picking, Matusiak (2017) quantifies significant differences of up to 10 % in average task performance between individual workers. Therefore, our study underlines the necessity of integrating human factors in optimization models for order picking systems because they are an essential element. If not considered, they can indeed lead to infeasible assignment planning.

6.2. Methodological implications

From a methodological perspective, the findings suggest advanced machine learning models as a prerequisite to comprehensively model complex industrial systems involving humans. Prior research has mostly applied methods from the domain of multivariate statistics, e.g. ordinary least square regressions, multi-level models, or accelerated failure time models. While these parametric methods hold strong assumptions for the distribution of dependent or independent variables, they might over or underestimate experience and fatigue effects. This study addresses these shortcomings and proposes the application of an alternative methodology. This has a very high transfer potential to other areas and application fields in operations and supply chain management and beyond.

Furthermore, our study provides the first-time proof regarding the possibility of integrating predictive and prescriptive tools to optimize order-picking systems. The convergence of predictive analytics with prescriptive methodologies embodies a transformative step in data-driven decision-making, allowing researchers and scholars to proactively adapt and optimize operations in real-time based on data insights.

6.3. Limitations and further research directions

Our study has limitations, which mainly arise from the use of secondary data. First, we are limited to the data made available by the partnering firm. Even though we can identify order pickers by unique user identification codes and measure their performance on a granular storage location level, we cannot access detailed video tracking data for more comprehensive operationalization of our variables. We are still confident that our main results are robust, as the secondary data variables are determined objectively.

Another limitation is that we do not have detailed individual data for more comprehensive control variables (i.e., demographic characteristics or prior experience). However, we quantity experience by accumulated repetitions which is a common approach in literature on learning. Therefore, we are convinced that any statistical issues related to omitting a variable that indicates individual-level data are effectively mitigated.

Although the behavior of more than 300 pickers has been examined and the case study under investigation involves a typical manual order picking system, extending the analysis to more case studies could be an interesting feature research. Lastly, there are limitations to the generalizability of our findings. We study a context in which order pickers complete standardized order picking tasks, i.e., traveling to storage locations and retrieving products and no task variability is observed. Our findings should be generalizable to other similar standardized order picking tasks. However, some of our results could be dependent on the type of task that order pickers are completing and may not replicate in settings involving additional tasks, i.e., replenishing racks or storage locations or settings with high task variability.

Our research lays the groundwork for future studies aimed at developing comprehensive approaches to effectively address the complexities of operational challenges in human order picking systems. One potential direction for future research involves incorporating a broader range of personal factors into predictive models. By exploring the advantages, costs, and ethical implications of collecting additional human factors indicators, researchers can test more complex predictive models based on this data. This could include investigating various classifications of human factors and considering factors such as mental aspects, motivation, stress, and how they interact with environmental indicators like light levels, noise, or temperature. Additionally, future research could expand beyond predicting system efficiency to forecasting and optimizing other important aspects such as picking accuracy, picker turnover, and injuries.

Another promising avenue for future research is extending prescriptive modules to integrate batch assignment with other typical operational order picking tasks. While our proposed approach assumes fixed batch compositions and routing policies, an integrated design of batches, routing policies, and assignment strategies could maximize overall system efficiency. Moreover, considering additional task-related constraints commonly encountered in order picking contexts, such as order precedence or time windows, can enhance the usability and practical applicability of these models among practitioners.

CRediT authorship contribution statement

Matteo Gabellini: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Francesca Calabrese:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Alberto Regattieri:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Dominic Loske:** Writing – review & editing, Supervision, Project administration, Data curation, Conceptualization. **Matthias Klumpp:** Writing – review & editing, Conceptualization, Project administration, Supervision.

Data availability

The corporate data that has been used is confidential.

References

Aboelfotoh, A., Singh, M., & Suer, G. (2019). Order batching optimization for warehouses with cluster-picking. *Procedia Manufacturing, 39*(2019), 1464–1473. https://doi.org/10.1016/j.promfg.2020.01.302

Alipour, M., Zare Mehrjedrdi, Y., & Mostafaeipour, A. (2020). A rule-based heuristic algorithm for on-line order batching and scheduling in an order picking warehouse with multiple pickers. *RAIRO - Operations Research, 54*(1), 101–117. https://doi.org/10.1051/ro/2018069

Ardjmand, E., Ghalehkhondabi, I., Young, W. A., Sadeghi, A., Weckman, G. R., & Shakeri, H. (2020). A hybrid artificial neural network, genetic algorithm and column generation heuristic for minimizing makespan in manual order picking operations. *Expert Systems with Applications, 159*, Article 113566. https://doi.org/10.1016/j.eswa.2020.113566

Asadayoobi, N., Jaber, M. Y., & Taghipour, S. (2021). A new learning curve with fatigue-dependent learning rate. *Applied Mathematical Modelling, 93*, 644–656.

Batt, R. J., & Gallino, S. (2019). Finding a needle in a haystack: The effects of searching and learning on pick-worker performance. *Management Science, 65*(6), 2624–2645. https://doi.org/10.1287/mnsc.2018.3059

Battini, D., Glock, C. H., Grosse, E. H., Persona, A., & Sgarbossa, F. (2016). Human energy expenditure in order picking storage assignment: A bi-objective method. *Computers & Industrial Engineering, 94*, 147–157. https://doi.org/10.1016/j.cie.2016.01.020

Burkart, N., & Huber, M. F. (2021). A Survey on the Explainability of Supervised Machine Learning. In *Journal of Artificial Intelligence Research, 70*.

Cano, J. A., Correa-Espinal, A., & Gómez-Montoya, R. (2020). Using genetic algorithms for order batching in multi-parallel-aisle picker-to-parts systems. *International Journal of Applied Decision Sciences, 13*(4), 435–447. https://doi.org/10.1504/IJADS.2020.110606

Cano, J. A., Cortés, P., Campo, E. A., & Correa-Espinal, A. A. (2021). Solving the order batching and sequencing problem with multiple pickers: A grouped genetic algorithm. *International Journal of Electrical and Computer Engineering, 11*(3), 2516–2524. https://doi.org/10.11591/ijece.v11i3.pp2516-2524.

Cano, J. A., Cortés, P., Campo, E. A., & Correa-Espinal, A. A. (2022). Multi-objective grouping genetic algorithm for the joint order batching, batch assignment, and sequencing problem. *International Journal of Management Science and Engineering Management, 17*(3), 188–204. https://doi.org/10.1080/17509653.2021.1991852

Cano, J. A., Cortés, P., Muñuzuri, J., & Correa-Espinal, A. (2023). Solving the picker routing problem in multi-block high-level storage systems using metaheuristics. *Flexible Services and Manufacturing Journal, 35*(2), 376–415. https://doi.org/10.1007/s10696-022-09445-y

Chu, P. C., & Beasley, J. E. (1997). A genetic algorithm for the generalised assignment problem. *Journal of the Operational Research Society, 24*(1), 17–23. https://doi.org/10.1057/palgrave.jors.2600431

Daria, B., Martina, C., Alessandro, P., Fabio, S., & Valentina, V. (2017). Fatigue and recovery: Research opportunities in order picking systems. *IFAC-PapersOnLine, 50*(1), 6882–6887. https://doi.org/10.1016/j.ifacol.2017.08.1211

De Bruecker, P., Van Den Bergh, J., Beliën, J., & Demeulemeester, E. (2015). Workforce planning incorporating skills: State of the art. *European Journal of Operational Research, 243*(1), 1–16. https://doi.org/10.1016/j.ejor.2014.10.038

De Lombaert, T., Braekers, K., De Koster, R., & Ramaekers, K. (2023). In pursuit of humanised order picking planning: Methodological review, literature classification and input from practice. *International Journal of Production Research, 61*(10), 3300–3330.

De Koster, R., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research, 182*(2), 481–501. https://doi.org/10.1016/j.ejor.2006.07.009

De Vries, J., De Koster, R., & Stam, D. (2016). Exploring the role of picker personality in predicting picking performance with pick by voice, pick to light and RF-terminal picking. *International Journal of Production Research, 54*(8), 2260–2274. https://doi.org/10.1080/00207543.2015.1064184

A.V. Dorogush V. Ershov A. Gulin CatBoost: Gradient boosting with categorical features support 2018 http://arxiv.org/abs/1810.11363.

Falkenberg, S. F., & Spinler, S. (2023). Integrating operational and human factors to predict daily productivity of warehouse employees using extreme gradient boosting. *International Journal of Production Research, 61*(24), 8654–8673. https://doi.org/10.1080/00207543.2022.2159563

Feng, X., & Hu, X. (2021). A Heuristic Solution Approach to Order Batching and Sequencing for Manual Picking and Packing Lines considering Fatiguing Effect. *Scientific Programming, 2021*. https://doi.org/10.1155/2021/8863391

Giannikas, V., Lu, W., Robertson, B., & McFarlane, D. (2017). An interventionist strategy for warehouse order picking: Evidence from two case studies. *International Journal of Production Economics, 189*, 63–76. https://doi.org/10.1016/j.ijpe.2017.04.002

Granillo-Macías, R. (2020). Inventory management and logistics optimization: A data mining practical approach. *Logforum, 16*(4), 535–547. https://doi.org/10.17270/J.LOG.2020.512.

Grosse, E. H., & Glock, C. H. (2015). The effect of worker learning on manual order picking processes. *International Journal of Production Economics, 170*, 882–890. https://doi.org/10.1016/j.ijpe.2014.12.018

Glock, C. H., Grosse, E. H., Jaber, M. Y., & Smunt, T. L. (2019). Applications of learning curves in production and operations management: A systematic literature review. *Computers & Industrial Engineering, 131*, 422–441. https://doi.org/10.1016/j.cie.2018.10.030

Grosse, E. H., Glock, C. H., & Müller, S. (2015). Production economics and the learning curve: A meta-analysis, *International Journal of Production Economics. Part B, 170*, 401–412. https://doi.org/10.1016/j.ijpe.2015.06.021

Grosse, E. H., Glock, C. H., & Neumann, W. P. (2017). Human factors in order picking: A content analysis of the literature. *International Journal of Production Research, 55*(5), 1260–1276. https://doi.org/10.1080/00207543.2016.1186296

Hastie, T. J., & Pregibon, D. (2017). In *Generalized linear models* (pp. 195–247). Routledge.

Henn, S. (2015). Order batching and sequencing for the minimization of the total tardiness in picker-to-part warehouses. *Flexible Services and Manufacturing Journal, 27*(1), 86–114. https://doi.org/10.1007/s10696-012-9164-1

Jamal, J., Loske, D., Klumpp, M., Chou, X., Renzo, D. F. D. A., Dell'amico, M., & Montemanni, R. (2022). Skill-Based Joint Order Batching and Picker Routing Problem. *ACM International Conference Proceeding Series, 64–69*. https://doi.org/10.1145/3523132.3523143

Kübler, P., Glock, C. H., & Bauernhansl, T. (2020). A new iterative method for solving the joint dynamic storage location assignment, order batching and picker routing problem in manual picker-to-parts warehouses. *Computers and Industrial Engineering, 147*(August 2019), Article 106645. https://doi.org/10.1016/j.cie.2020.106645

Lee, I. G., Chung, S. H., & Yoon, S. W. (2020). Two-stage storage assignment to minimize travel time and congestion for warehouse order picking operations. *Computers and Industrial Engineering, 139*(August 2019), Article 106129. https://doi.org/10.1016/j.cie.2019.106129

Loske, D., & Klumpp, M. (2022). Quantifying heterogeneity in human behavior : An empirical analysis of forklift operations through multilevel modeling. *Logistics Research, 15*(1), 1–17. https://doi.org/DOI_10.23773/2022_1.

S. Lundberg S.-I. Lee A Unified Approach to Interpreting Model Predictions 2017 http://arxiv.org/abs/1705.07874.

Mahmoudi, M. R. (2019). On comparing two dependent linear and nonlinear regression models. *Journal of Testing and Evaluation, 47*(1), 449–458.

Manzini, R., Gamberi, M., & Regattieri, A. (2005). Design and control of a flexible order-picking system (FOPS). *Journal of Manufacturing Technology Management, 16*(1), 18–35. https://doi.org/10.1108/17410380510574068

Matusiak, M., de Koster, R., & Saarinen, J. (2017). Utilizing individual picker skills to improve order batching in a warehouse. *European Journal of Operational Research, 263*(3), 888–899. https://doi.org/10.1016/j.ejor.2017.05.002

Mou, S. (2022). Integrated Order Picking and Multi-Skilled Picker Scheduling in Omni-Channel Retail Stores. *Mathematics, 10*(9). https://doi.org/10.3390/math10091484

Pan, J. C. H., Shih, P. H., Wu, M. H., & Lin, J. H. (2015). A storage assignment heuristic method based on genetic algorithm for a pick-and-pass warehousing system. *Computers and Industrial Engineering, 81*, 1–13. https://doi.org/10.1016/j.cie.2014.12.010

Pretolani, D., Chou, X., Loske, D., & Klumpp, M. (2023). *Mathematical Models for Order Batching and Assignment in a Deep-Frozen Warehouse*. 158–168. https://doi.org/10.3233/ATDE230020.

Scholz, A., Schubert, D., & Wäscher, G. (2017). Order picking with multiple pickers and due dates – Simultaneous solution of Order Batching, Batch Assignment and Sequencing, and Picker Routing Problems. *European Journal of Operational Research, 263*(2), 461–478. https://doi.org/10.1016/j.ejor.2017.04.038

Schrotenboer, A. H., Wruck, S., Roodbergen, K. J., Veenstra, M., & Dijkstra, A. S. (2017). Order picker routing with product returns and interaction delays. *International Journal of Production Research, 55*(21), 6394–6406. https://doi.org/10.1080/00207543.2016.1206982

Shapiro, S. S., & Wilk, M. B. (1965). *An Analysis of Variance Test for Normality (Complete Samples)* (Vol. 52, Issue 3). https://www.jstor.org/stable/2333709.

Strang, B., Putten, P. V. D., Rijn, J. N. V., & Hutter, F. (2018). Don't rule out simple models prematurely: A large scale benchmark comparing linear and non-linear classifiers in OpenML. In Advances in Intelligent Data Analysis XVII: 17th International Symposium, IDA 2018,'s-Hertogenbosch, The Netherlands, October 24–26, 2018, Proceedings 17 (pp. 303-315). Springer International Publishing.

Tompkins, J. A. (2010). *Facilities planning* (4th ed.). Hoboken, NJ: Wiley http://www.loc.gov/catdir/enhancements/fy1011/2010455781-d.html.

Van Gils, T., Ramaekers, K., Caris, A., & de Koster, R. B. M. (2018). Designing efficient order picking systems by combining planning problems: State-of-the-art classification and review. *European Journal of Operational Research, 267*(1), 1–15. https://doi.org/10.1016/j.ejor.2017.09.002

Vanheusden, S., van Gils, T., Ramaekers, K., Cornelissens, T., & Caris, A. (2022). Practical factors in order picking planning: State-of-the-art classification and review. *International Journal of Production Research, 1–25*. https://doi.org/10.1080/00207543.2022.2053223

Vijayakumar, V., Sgarbossa, F., Neumann, W. P., & Sobhani, A. (2022). Framework for incorporating human factors into production and logistics systems. *International*

Journal of Production Research, 60(2), 402–419. https://doi.org/10.1080/00207543.2021.1983225

Wang, M., Zhang, R. Q., & Fan, K. (2020). Improving order-picking operation through efficient storage location assignment: A new approach. *Computers and Industrial Engineering, 139*(November 2019), Article 106186. https://doi.org/10.1016/j.cie.2019.106186

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. In *Bulletin* (Vol. 1, Issue 6). https://www.jstor.org/stable/3001968.

Winkelhaus, S., Sgarbossa, F., Calzavara, M., & Grosse, E. H. (2018). The effects of human fatigue on learning in order picking: An explorative experimental investigation. *IFAC-PapersOnLine, 51*(11), 832–837. https://doi.org/10.1016/j.ifacol.2018.08.442

Zhang, J., Wang, X., Chan, F. T. S., & Ruan, J. (2017). On-line order batching and sequencing problem with multiple pickers: A hybrid rule-based algorithm. *Applied Mathematical Modelling, 45*, 271–284. https://doi.org/10.1016/j.apm.2016.12.012

Zhang, J., Zhang, X., & Zhang, N. (2023). Considering pickers' learning effects in selecting between batch picking and batch-synchronized zone picking for online-to-offline groceries. *Applied Mathematical Modelling, 113*, 358–375. https://doi.org/10.1016/j.apm.2022.09.009