OR Spectrum
https://doi.org/10.1007/s00291-024-00795-7

ORIGINAL ARTICLE

Data-based optimisation of intra-hospital patient transport capacity planning

Tobias Kropp[1] · Yuhao Gao[1] · Kunibert Lennerts[1]

Received: 31 August 2023 / Accepted: 22 October 2024

Abstract

Efficient and timely organisational healthcare processes are urgent for patient satisfaction and medical success in hospitals. Despite process analysis and problem identification, there are especially challenges in evaluating and implementing planning alternatives. This is also valid for the planning of resource capacities. There are currently few use cases that offer data-driven, automated solutions and typically significant effort in modeling complex processes and systems is involved. Therefore, we explore the use of a combination of neural networks and metaheuristic algorithms to optimise organisational capacity planning in healthcare. These techniques allow for autonomous learning and optimisation of processes. A Multilayer Perceptron (MLP) is developed in a use case utilising data from approximately 3.5 years of accompanied intra-hospital patient transport in a German hospital in order to be able to make accurate predictions about delayed transports on a day of the week basis. A data preprocessing was performed, aggregating case-wise transportation information into hourly information to serve as input and labelling data for the MLP training. Using a genetic algorithm (GA), hourly input variables such as the number of active transporters, the number of planned transports, or the automation rate of transport dispatching are adapted in order to reduce the model predicted number of delayed transports throughout a day. Through this approach, a theoretical reduction in delayed transports on a day of the week ranging from 27% to 42% could be achieved merely through resource reallocating, without adding additional resources. The performance of both MLP and GA are validated using various measures.

Keywords Organisational healthcare · Patient transportation · Capacity planning optimisation · Artificial neural network · Multilayer perceptron · Genetic algorithm

Extended author information available on the last page of the article

Published online: 22 November 2024

1 Introduction

The growing emphasis on patient-centred care within healthcare organisations underlines the importance of providing efficient and timely healthcare services to patients. Lenz and Reichert (2007) distinguishes between two types of processes that take place in hospitals: on the one hand there are the medical treatment processes, that are directly linked to the patients' health and on the other hand there are the organisational processes. Ensuring that organisational processes function seamlessly is the basis for the success of medical processes (Lenz and Reichert 2007). This paper will aim to optimise the process of intra-hospital patient transportation (IHPT), as it is an important organisational healthcare process. Due to its widespread use, IHPT plays a crucial role to provide efficient and timely medical treatments (Beckmann et al. 2004; Hendrich and Nelson 2005; Ulrich and Zhu 2007). IHPT refers to the internal transfer of patients within a hospital, such as between different wards and functional areas (Nakayama et al. 2012). The effectiveness of this service and its associated processes have a significant impact on clinical outcomes and patient satisfaction (Beckmann et al. 2004). For instance the research findings of a cross-sectional analysis of 191 IHPT incidents showed that 31% of these incidents resulted in adverse outcomes (Beckmann et al. 2004). Another study reported that of 288 transport cases involving brain-injured patients, 36% had at least one significant complication (Picetti 2013). Patient complications can be rooted in increased waiting time due to insufficient service capacity (Meephu et al. 2023). Kropp et al. (2023) highlights inappropriate resource allocation causing waiting times in IHPT as well as the necessity of continuous data-based monitoring of the workflows to improve the IHPT. Due to the heterogeneous nature of hospital organisations, no general IHPT problem can be defined and literature considers different goals and approaches (Klein and Thielen 2024).

Hospital information systems (HIS), but also other software systems, can bear information on logistical processes, like the IHPT (van der Aalst 2016; Martin 2020). Jaroon (2018) has already indicated that the use of a computer-based online patient transfer system could help improve work efficiency and lead to an increase in the overall on-time service delivery rate from around 56% to 66% (and thus by around 18%) in a Southern Thailand hospital case study. The relevant process information is collected in the software systems during the planning and execution of IHPT. Information, that is collected and stored in so called event logs can be very detailed, i.e. specific to single events that took place within a transport ("event attributes"), or somewhat higher-level, i.e. transport case-specific ("case attributes") (van der Aalst 2016). There may even be further general information about the process, such as general responsibilities or guidelines in the process, that are not directly linked to e.g. a transport case or the specific activities during the transport process of this case. Analyzing event logs through process mining techniques can yield valuable insights into the actual execution of healthcare processes and thus help to improve the provided services (van der Aalst 2016; Martin 2020).

We adopt the 4-phase model of Aguirre et al. (2017), that describes different steps of process mining projects, as a basis in order to generally situate our investigations (see Fig. 1). Generally, process mining projects aim to exploit event data in a meaningful way in order to improve processes (van der Aalst 2016). The aim of this paper is to provide a decision support for the last stage in a process mining project, process redesign phase (see Fig. 1). In this paper the event data of the IHPT of around 3.5 years of operation in a German public hospital are being processed for an Artificial Neural Network (ANN) in order to finally enable improved transport capacity planning by solving an optimisation problem through a genetic algorithm (GA) after model development. The implementation of the derived optimisation in real operation will not be investigated in this paper.

de Roock and Martin (2022) analysed 263 papers on process mining in healthcare and emphasises that only about 7.6 per cent of the papers deal with analyses of organisational processes and only another 15.2 per cent deal with organisational processes in part. Most of the analyses deal with medical treatments. Furthermore (de Roock and Martin 2022) concludes, among others, that currently research tends to focus primarily on the analysis stage, but the true value lies in the ability to convert analysis results into concrete actions that drive process improvement within healthcare organisations. The last stage of a process mining project, the process redesign phase, is barely addressed (de Roock and Martin 2022). This highlights the importance of our approach, as our research allows us to propose specific adjustments to the allocation of IHPT resources that are projected to optimise efficiency.

In primary healthcare processes there are already several applications of ANN to optimise process redesign. For instance, Amato et al. (2013) regards ANN as an invaluable tool to aid doctors in diagnosis and analysis, encompassing tasks such as data processing, reducing the likelihood of overlooking relevant information and shortening diagnosis time, thereby enhancing the reliability of doctors' ultimate diagnostic decisions. Yang et al. (2017a; b) developed a logistic regression model based process recommender system that provides data-driven step-by-step treatment recommendations. The framework introduced in Yang et al. (2017a, b) begins by clustering treatment procedures of trauma resuscitation patients according to context attributes (patient attributes and hospital factors). When a new set of context

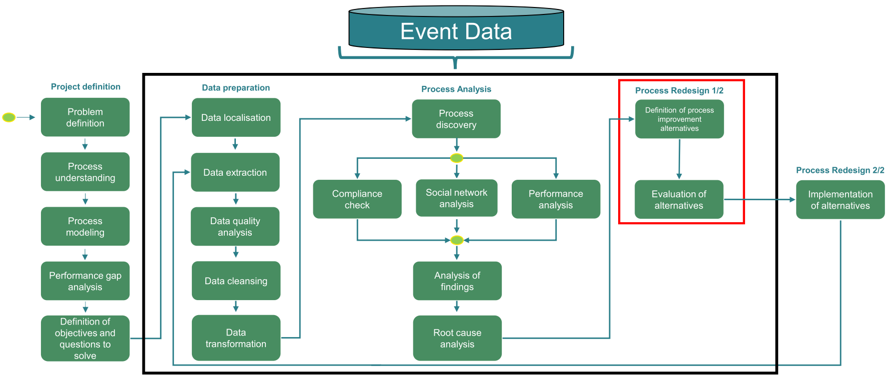

Fig. 1 Process mining project methodology (adapted from Aguirre et al. 2017). The focus of this paper is on process redesign

attributes is fed into the trained regression model, it generates a suggested execution of the patient treatment process (Yang et al. 2017a, b).

In summary, there are few studies providing insights on data analysis of organisational processes in hospitals and furthermore there are no studies for IHPT in which process data were analysed in order to be able to automatically derive process improvements. In Nas and Koyuncu (2019) it is furthermore emphasised that few applications are found in the literature in hourly patient arrival problems. We aim to provide hourly predictions for delayed IHPT depending on several resource conditions in order to optimise overall delayed cases per day of the week by optimising the resource planning.

Our paper is structured as follows: Sect. 2 presents the relevant literature and outlines the contribution of our work, Sect. 3 introduces to fundamentals of the techniques used, Sect. 4 describes our approach in detail and shows our obtained findings, Sect. 5 discusses the results and finally Sect. 6 concludes on our work and indicates future work.

2 Literature review

Subsection 2.1 focuses on literature that generally aims to optimise in a data supported way the process redesign phase of organisational processes in healthcare, especially in the IHPT domain. Subsequently Subsect. 2.2 specifically presents capacity optimisation approaches through the application of ANNs. Based on this, our approach to reorganising the capacity planning of IHPT is outlined in Subsect. 2.3.

2.1 Data supported process redesign in organisational healthcare

Agostinelli et al. (2020) investigated patient care flows in a real case study with process mining. The authors in Agostinelli et al. (2020) were able evaluate e.g. the value of last investments or the temporal distribution of abandonments from emergency room and examinations that had no reservation. Hints for starting points for future improvement ideas to solve some identified problems in the care flow process are given but there were neither specific improvement measures developed nor assessed.

Andrews et al. (2020) analysed transport pathways discovered across the time-critical phase of pre-hospital care for persons involved in road traffic crashes. With the help of domain experts, improvement concepts are proposed that are aimed at improving data quality on the one hand and at future automated decision support through e.g. AI on the other (Andrews et al. 2020). However, specific measures and their effects have not been evaluated.

Badakhshan and Alibabaei (2020) investigated data from an automation system of a pre-hospital emergency room. After process discovery and conformance checks there are some specific improvement ideas given based on previous bottleneck identification (Badakhshan and Alibabaei 2020). No quantitative evaluation of effects of the proposed solution ideas is conducted and furthermore no alternatives to the in Badakhshan and Alibabaei (2020) proposed improvement ideas were investigated.

Canjels et al. (2021) uses process mining techniques to improve the care process for arthrosis patients so that the provided care matches better with the facilities and resources used. The knowledge gained through the process mining analysis of historical data is used to cluster and redistribute potential patient pathways depending on the necessary complexity of the patient care between two hospital sites, so that in the future potentially more patients can be fully or partly treated in the cost-effective outpatient clinic environment (Canjels et al. 2021).

Stefanini et al. (2017) propose a methodology exploiting the benefits of process mining techniques in the healthcare systems to support service reconfiguration and apply it furthermore to historical data of a lung cancer unit. The investigations unveil e.g. the average demand of activities and related resource consumption of an average patient and thus help support managers in taking decisions about the implementation of a new lung cancer unit (Stefanini et al. 2017). Process improvement ideas and what-if analysis are proposed but in the further not conducted or evaluated within Stefanini et al. (2017).

Antunes et al. (2019) optimised waiting times, queue length and queue occurrences in an Emergency Department through a rescheduling of the weekly and hourly available number of physicians with a mixed-integer programming (MIP) mathematical model. To evaluate the success of the mathematical optimisation, the adapted schedule is tested with Discrete-Event simulation (DES) through a model that was designed and validated with the historical process data and knowledge gained through process mining analysis (Antunes et al. 2019).

van Hulzen et al. (2022), Pourbafrani and van der Aalst (2023), Zhou et al. (2014), Abohamad et al. (2017) developed DES models on the basis of process mining analyses to optimise organisational healthcare processes. In van Hulzen et al. (2022) improvement alternatives of capacity management decisions in the radiology department are evaluated using data-driven Process Simulation through DES. Recommendations regarding the required number of radiology devices, waiting area size, and reception staffing could be derived in van Hulzen et al. (2022). But it is also concluded, that the conformity of a developed simulation model relies highly on the modeler (van Hulzen et al. 2022). Domain knowledge is necessary during the data-driven development and validation of a simulation model (van Hulzen et al. 2022). Pourbafrani and van der Aalst (2023) introduces a reference model for data-driven Simulation, with DES, in Process Mining for production systems. In order to develop the reference model, especially literature on practical approaches for generating DES models of processes is investigated (Pourbafrani and van der Aalst 2023). Also Pourbafrani and van der Aalst (2023) highlights the impact of human factors in creating accurate simulation models of processes. Zhou et al. (2014) uses DES, where the simulation model is based on knowledge obtained from process mining analysis, to evaluate the impact of the number of receptionists, nurses, and doctors to improve the performance of an outpatient clinic. Specific scenarios are quantitatively evaluated to determine the impact of operational changes and sensitivity analyses are furthermore conducted to evaluate when increasing the number of specific personnel resources runs up against an improvement threshold level (Zhou et al. 2014). Abohamad et al. (2017) also uses DES with a simulation model, that is developed through process mining based knowledge, to identify performance bottlenecks and to explore improvement strategies to reduce patients' length of stay in the Emergency Department (ED LOS). Specific scenarios for variation in medical staffing, increasing clinical assessment space, or incorporating a policy, where patients that wait more than a specific time threshold to be admitted to a hospital bed are dismissed, were simulated and quantitatively evaluated (Abohamad et al. 2017).

Akbari et al. (2023), Yazır et al. (2023), Li et al. (2021) optimise home healthcare routing and scheduling with mathematical models and also metaheuristic algorithms. Akbari et al. (2023) use a level-based integer programming (IP) mathematical model, for smaller instances, and a generalised variable neighborhood search-based (GVNS) metaheuristic algorithm, for larger instances. The solutions proposed by Akbari et al. (2023) optimise the planning of multiple home healthcare service provider teams that should visit a given set of patients at their homes according to the locations as well as the severity of the condition or the service urgency of the patients. On the one hand, Yazır et al. (2023) formulate a MIP mathematical model and on the other hand use an adaptive large neighborhood search-based (ALNS) metaheuristic algorithm to optimise the planning of the weekly routes of nurses visiting patients located at a scattered geographic area. This is achieved by minimising the total costs that incorporate e.g. wage costs, charging costs of used vehicles, further transfer costs, and the costs of a patient left unserved (Yazır et al. 2023). In Li et al. (2021), a MIP mathematical model is used to minimise travel costs, waiting times and maximise patients' preference satisfaction under constraints on time windows, workload and skill requirements to optimise the routing and scheduling of home healthcare with consideration of outpatient services. Also, for larger instances Li et al. (2021) develop and propose a hybrid GA for the optimisation.

Molenbruch et al. (2017) uses a Multi-directional local search (MDLS) metaheuristic algorithm for optimising operational costs and service quality with improved driver schedules for a service provider that conducts demand-responsive transportation between patients' homes and healthcare locations.

In Naesens and Gelders (2009) a data analysis of the IHPT process was conducted to propose process improvements. This led to a partially decentralisation of the IHPT organisation (Naesens and Gelders 2009). A quantitative evaluation of the redesigned process was not presented in Naesens and Gelders (2009). Haldar et al. (2019) investigated the reasons of delayed IHPT cases and the effects towards operation theatres' efficiency. In a qualitative approach data was collected by an independent observer and transport cases were labeled delayed, if they arrived later than 35 min (Haldar et al. 2019). Most common reasons for delays included e.g. transporter-associated delays during shift changeovers, unavailable lifts or involvement of the pediatric ward (Haldar et al. 2019). The two main effects of IHPT delays observed in theatres were routine cases being extended beyond the scheduled time (i.e. overrunning of operation theatres) and the cancellation of previously scheduled second cases per day (Haldar et al. 2019). With feasible measures like increasing the summon times (i.e. summoning the patient telephonically from the pick-up location to the operation theatre), sensitisation of transporters and nurses could improve the efficiency in operation theatres' functioning (i.e. more than 6% reduction in delayed arrivals to the operation theatres and a similar reduction in overrunning operation theatres) (Haldar et al. 2019). The Hawthorne effect (the fact that the observation and documentation of a process can lead to marked differences in performance of involved people) is not ruled out in the study (Haldar et al. 2019). Kropp et al. (2024) used process mining techniques to analyse the same IHPT dataset used in the current study, already carrying out capacity evaluations. Ideas for process redesign are developed, but their expected impact is presented qualitatively. The investigations of Kropp et al. (2024) can be seen as preliminary work, which helped to understand the IHPT process and problems in order to develop the approach of the current study.

Kallrath (2005), Séguin et al. (2019), Kuchera and Rohleder (2011), Gopal (2016), Maka et al. (2022), Bouabdallah et al. (2013), Turan et al. (2011), Elmbach et al. (2015) optimise IHPT by solving mathematical models. Kallrath (2005) solved MIP mathematical models, as well as used a branch-and-bound approach, a column enumeration approach and (meta-)heuristic algorithms (recommended for larger problem instances) to optimise IHPT routing and scheduling. Kallrath (2005) proposed a general, theoretical framework incorporating the different solution approaches mentioned, and also conducted comparative experiments using real-life instances from a German hospital to reduce e.g. patient waiting times, transport delays and uneven occupancy of transport vehicles. Séguin et al. (2019); Kuchera and Rohleder (2011); Gopal (2016) developed a MIP mathematical model to optimise the hourly staff planning per day of the week and to reduce completion times of transports or patient waiting times. Séguin et al. (2019) reaches a 16% reduction in daily staff capacity. At the same time possible delay minutes of transports per specific hour could be mostly decreased between 27% and 71% (in contrast, less active scheduled transporters partly led to an increase of up to 58% in possible delay minutes per hour) (Séguin et al. 2019). Kuchera and Rohleder (2011) validate their optimisation proposal with positive observations of the patient service quality in real operation and overall two FTE (Full-Time Equivalent) could be saved through the approach. The relative improvement in FTE is not reported in Kuchera and Rohleder (2011). Gopal (2016) uses DES to validate that the proposed solution does not adversely affect the quality of patient service by evaluating the resulting average time from pending to completion of the transports. Depending on the scenario, the approach of Gopal (2016) reduced process throughput times by up to 13% (and by up to 25% in sub-processes depending on the sub-process). At the same time, between around 1% and 8% FTE could be saved, depending on the scenario (Gopal 2016). Maka et al. (2022) also propose a MIP mathematical model for an optimised planning of IHPT to minimise the total cost of operation. The model helps first select a minimum number with the best locations as depots from a choice of locations within a hospital and then allocate different resources (e.g. wheelchairs, stretchers, oxygen tanks, staff) to each depot accordingly (Maka et al. 2022). Bouabdallah et al. (2013) developed a MIP mathematical model that minimises the sum of the empty stretcher moves between missions in IHPT. Turan et al. (2011) provides an optimised IHPT planning for patient routing with fixed randomly generated appointments through solving a weighted sum mathematical model. Thus, Turan et al. (2011) aim to minimise patient transporters' travel time as well as the patients' waiting time. Also, aspects like the number of different transporters per patient and empty runs of transporters between transportation assignments are considered within the mathematical model (Turan et al. 2011). Compared to firstly optimised schedules regarding minimised transporters' travel time as well as the patients' waiting time, the patients' inconvenience of having to deal with different transporters could further be improved by reducing the number of transporters per patient by 27%, but at the cost of around 12% increased transporter travel times (Turan et al. 2011). Furthermore, Turan et al. (2011) indicates, that the developed model has computational limitations and that optimised planning of more than 40 transport requests per hour will require the use and development of (meta-)heuristic approaches. Elmbach et al. (2015) formulate a mathematical model for the scheduling of IHPT with respect to transporters' ergonomic stress and investigate optimisation solutions through dynamic programming (for small instances) and beam search-based heuristic algorithms (for small and large instances). The ergonomic liability at the case study hospital could for large (real world) instances theoretically be reduced by an average of around 36% (and a maximum of around 79%) using the beam search-based heuristic algorithm compared to simple decision rules of a human decision maker (Elmbach et al. 2015).

Beaudry et al. (2010), Kergosien et al. (2011), Fröhlich Von Elmbach et al. (2019), Schmid and Doerner (2014), Fiegl and Pontow (2009), Xiao et al. (2022), Bärmann et al. (2024), Vancroonenburg et al. (2016), Hanne et al. (2009) use (meta-) heuristic algorithms to optimise IHPT. Vancroonenburg et al. (2016); Hanne et al. (2009) furthermore incorporate DES in their approach. Beaudry et al. (2010); Kergosien et al. (2011); Fröhlich Von Elmbach et al. (2019) use tabu search to optimise IHPT routing and scheduling. Waiting times for patients were reduced while using fewer vehicles in Beaudry et al. (2010). Through the approach of Kergosien et al. (2011) the hospital under investigation was theoretically able to handle 10% more requested transports independently and required fewer subcontracted transports to conduct all transports, which were furthermore able to meet the suggested time windows. Kergosien et al. (2011) also tested their approach against integer linear programming (ILP). Fröhlich Von Elmbach et al. (2019) reach average staff savings of about 8% compared to simple decision rules of a human decision maker while improving ergonomic liability of transporters. Fröhlich Von Elmbach et al. (2019) compare the performance of their tabu search approach furthermore with a MIP solver. Schmid and Doerner (2014) developed a hybrid large neighbourhood seach-based algorithm to optimise IHPT routing and scheduling with respect to resource- and client-centered perspectives. For smaller instances exact MIP solutions and for larger (real world) instances ten hour runtime MIP solutions are used to compare the algorithm's performance (Schmid and Doerner 2014). Fiegl and Pontow (2009) developed a heuristic algorithm based on scheduling and graph theory to maximise the possible task throughput in IHPT scheduling. Like this around 17% reduction in task flow time could be achieved compared to usual scheduling in a hospital case study (Fiegl and Pontow 2009). Xiao et al. (2022) develop and compare a greedy and a column generation-based heuristic algorithm to optimise IHPT routing and scheduling. The latter integrates furthermore either MIP or GA-based metaheuristics for solving subproblems (Xiao et al. 2022). Bärmann et al. (2024) developed a lexicographic branch-and-bound column generation-based approach for optimising

IHPT routing and scheduling with respect to transport delays, walked distances of transporters between transportation assignments and equal transporter utilisation. Performance evaluations in comparison with a standard branch-and-bound column search approach, classical column generation-based methods and MIP are furthermore conducted using large instances from two European hospitals (Bärmann et al. 2024). The approach of Bärmann et al. (2024) is also tested against the commercial routing and scheduling software used by the hospitals achieving approximately 20% improvement in both transport delays and empty runs by transporters. Finally (Bärmann et al. 2024) deployed and evaluated their approach in a German hospital for 9 months, resulting in promising performance outcomes. Vancroonenburg et al. (2016) use a cheapest insertion and a local search heuristic algorithm, as well as DES, to optimise the scheduling and assignment of IHPT to transporters. Different scenarios were designed and compared to a baseline scenario, using the evenly weighted sum of delayed transport time, transporters' windows of availability without carrying out transports, and the total transporters' travel time carrying out assignments (Vancroonenburg et al. 2016). Through the approach of Vancroonenburg et al. (2016) the sum of these three factors could be improved by up to 31% depending on parameters like transport request arrival rates and the number of transporters. The process could be even more improved in scenarios where transporters are allowed to combine transports (multiple pick-ups in sequence, before performing deliveries) (Vancroonenburg et al. 2016). Hanne et al. (2009) improved patient waiting times in IHPT through different metaheuristic algorithms (GA, etc.) and DES that aim to optimise the scheduling and assignment of transports to vehicles or transport teams. The approach of Hanne et al. (2009) could theoretically reduce e.g. average patient waiting times by around 20% to 26% and average patient travel times by around 10%. Furthermore, the approach was deployed and evaluated in practice in a German hospital leading to reduced transportation costs of around 20% and at the same time contributing to improved patient satisfaction (Hanne et al. 2009).

Meephu et al. (2023) used DES to investigate 20 scenarios with different improvement strategies of IHPT to find the best scenario among these. In the best scenario mean patient waiting times could be reduced by around 22% (Meephu et al. 2023).

A hybrid Analytical Hierarchy Process and Artificial Neural Network (AHP-ANN) model was utilised in Fashoto et al. (2016) to model the vendor selection process for university health centers. The study determined the priority sequence of five criteria (service, delivery, cost, risk, quality) and provided data-supported recommendations for decision-making (Fashoto et al. 2016).

As ANN approaches are effective and efficient in providing a high level of capability in modeling complex problems (Abiodun et al. 2018), in the following subsection explicitly ANN approaches for optimising capacity planning in healthcare are presented.

2.2 Artifical neural networks to optimise capacity planning in healthcare

In Rajakumari and Madhunisha (2020) there is no case study conducted but a four staged framework to create and use an intelligent and Convolutional Neural Network

(CNN) to predict and evaluate improvement options for a Smart Hospital and Patient Scheduling System is portrayed on schematic and theoretic level. After the first stage ("data preparation") and the second stage ("process mining analysis"), in the third stage ("simulation modeling and evaluation"), a CNN is to be built and reality conformity of the model predictions are to be evaluated (Rajakumari and Madhunisha 2020). In the last stage ("Experiments and decision support") improvement options are to be explored to support decisions with the best option (Rajakumari and Madhunisha 2020).

In Mesabbah et al. (2019) a framework is proposed to offer an accurate and process mining-based auto-generated DES model using healthcare event logs for improved resources handling to support complex decision-making processes around hospital staff planning. Therein activity durations and next activities within the DES model are to be predicted by a Machine Learning (ML) model given different patient features and process information (Mesabbah et al. 2019). The ML model is not specified and not yet implemented so that no specific case study is investigated with the holistic proposed framework in Mesabbah et al. (2019).

Nas and Koyuncu (2019) use a Recurrent Neural Network (RNN) and DES approach to optimise the emergency department (ED) capacity planning. With the help of RNN predictions of patients' arrival rates were conducted and used as an input parameter for the DES model (Nas and Koyuncu 2019). For further simulation model input, the route of the patients in the ED was extracted from the analysis of the hospital's data (Nas and Koyuncu 2019). Treatment and service times were determined from observations with the help of the hospital experts (Nas and Koyuncu 2019). After conducting simulations, the number of beds could be identified as one of the process bottlenecks impacting the waiting time of patients at the ED (Nas and Koyuncu 2019). Further exemplary simulations helped to find the optimal number of beds and it is highlighted that the simulation part in the study could also be replaced using ML methods (Nas and Koyuncu 2019).

To address the high patient demand faced by the Emergency Department (ED) during peak hours, Gul and Guneri (2015) utilised an ANN considering different sets of variables with the aim of modeling and forecasting the patient ED LOS. However in Gul and Guneri (2015) it is mentioned, that there is still room for improvement in the accuracy of the prediction results and so far no predictions are presented after the modeling phase.

2.3 Outline of the work

Table 1 summarises the literature on process redesign of organisational healthcare processes and Table 2 summarises the literature specifically on process redesign of IHPT processes. Some of the found literature give only qualitative optimisation ideas. In Andrews et al. (2020), it is already emphasised that automating the derivation of improvement opportunities in a data-driven way is a future necessity. It can be concluded that if there was an optimisation quantitatively approached in the literature, mostly DES models (Nas and Koyuncu 2019, Antunes et al. 2019, van Hulzen et al. 2022, Pourbafrani and van der Aalst 2023, Zhou et al. 2014, Abohamad et al.

Table 1 Literature overview - redesign phase of organisational healthcare processes

Literature	Type	Presented evaluation of process improvement alternatives	Used main approach to create and evaluate process improvement alternatives	Comments on process redesign and process improvement alternatives
Agostinelli et al. (2020)	Case study	Qualitative	Knowledge from PM analysis	Only ideas, no evaluation of alternatives.
Andrews et al. (2020)	Case study	Qualitative	Domain experts, knowledge from PM analysis	Only ideas, no evaluation of alternatives.
Badakhshan and Alibabaei (2020)	Case study	Qualitative	Knowledge from PM analysis	Only ideas, no evaluation of alternatives.
Canjels et al. (2021)	Case study	Quantitative	Domain experts, knowledge from PM analysis	Specific optimisation proposed, but only one alternative is evaluated.
Stefanini et al. (2017)	Framework/ Case study	Qualitative	Domain experts, knowledge from PM analysis	Only ideas, no evaluation of alternatives.
Antunes et al. (2019)	Case study	Quantitative	MIP mathematical model and DES	Alternatives are evaluated to find best solution.
van Hulzen et al. (2022)	Case study	Quantitative	DES	Alternatives are evaluated to find best solution.
Pourbafrani and van der Aalst (2023)	Framework	Quantitative	DES	Framework for data-based model developing, no evaluation of alternatives.
Zhou et al. (2014)	Case study	Quantitative	DES	Specific alternatives are evaluated to find the best solution among them and sensitivity analysis towards personnel resources is conducted.
Abohamad et al. (2017)	Framework/ Case study	Quantitative	DES	Specific alternatives are evaluated to find the best solution among them and sensitivity analysis of different resources is conducted.

Table 1 (continued)

Literature	Type	Presented evaluation of process improvement alternatives	Used main approach to create and evaluate process improvement alternatives	Comments on process redesign and process improvement alternatives
Akbari et al. (2023)	Case study	Quantitative	IP mathematical model and metaheuristic algorithm (GVNS)	Alternatives are evaluated to find best solution, both approaches are compared.
Yazır et al. (2023)	Case study	Quantitative	MIP mathematical model and metaheuristic algorithm (ALNS)	Alternatives are evaluated to find best solution, both approaches are compared.
Li et al. (2021)	Case study	Quantitative	MIP mathematical model and metaheuristic algorithm (hybrid GA)	Alternatives are evaluated to find best solution, both approaches are compared.
Molenbruch et al. (2017)	Case study	Quantitative	Metaheuristic algorithm (MDLS)	Alternatives are evaluated to find best solution.
Fashoto et al. (2016)	Case study	Quantitative	AHP-ANN	Specific alternatives are evaluated to find the best solution among them.
Rajakumari and Madhunisha (2020)	Framework	Quantitative	CNN	Coarse framework without providing details on implementation, no alternatives are evaluated.
Mesabbah et al. (2019)	Framework	Quantitative	DES and ML approach	Coarse Framework without providing details on implementation, no alternatives are evaluated.
Nas and Koyuncu (2019)	Case study	Quantitative	ML approaches (RNN etc.) and DES	Specific alternatives are evaluated by a simulation model that uses ML results as input parameter to find best solution among them.
Gul and Guneri (2015)	Case study	Quantitative	ANN	Focus is on development of ANN that allows for evaluation of alternatives, no alternatives are evaluated.

Table 2 Literature overview—redesign phase of IHPT processes

Literature	Type	Presented evaluation of process improvement alternatives	Used main approach to create and evaluate process improvement alternatives	Comments on process redesign and process improvement alternatives
Naesens and Gelders (2009)	Case study	Qualitative	Process analysis	Only ideas, no evaluation of alternatives.
Haldar et al. (2019)	Case study	Qualitative	Process analysis	Specific optimisation implemented, but only one alternative is evaluated.
Kropp et al. (2024)	Case study	Qualitative	Knowledge from PM analysis	Only ideas, no evaluation of alternatives.
Kallrath (2005)	Framework/ Case study	Quantitative	MIP mathematical model, branch-and-bound, column enumeration, (meta-)heuristic algorithms	Alternatives are evaluated to find best solution.
Séguin et al. (2019)	Case study	Quantitative	MIP mathematical model	Alternatives are evaluated to find best solution.
Kuchera and Rohleder (2011)	Case study	Quantitative	MIP mathematical model	Alternatives are evaluated to find best solution.
Gopal (2016)	Case study	Quantitative	MIP mathematical model and DES	Alternatives are evaluated with MIP model to find best solution. DES is used to validate the found solution.
Maka et al. (2022)	Case study	Quantitative	MIP mathematical model	Alternatives are evaluated to find best solution.
Bouabdallah et al. (2013)	Case study	Quantitative	MIP mathematical model	Alternatives are evaluated to find best solution.
Turan et al. (2011)	Case study	Quantitative	Weighted sum mathematical model	Alternatives are evaluated to find best solution.
Elmbach et al. (2015)	Case study	Quantitative	Heuristic algorithm (beam search-based) and dynamic programming	Alternatives are evaluated to find best solution.
Beaudry et al. (2010)	Case study	Quantitative	Metaheuristic algorithm (tabu search)	Alternatives are evaluated to find best solution.

Table 2 (continued)

Literature	Type	Presented evaluation of process improvement alternatives	Used main approach to create and evaluate process improvement alternatives	Comments on process redesign and process improvement alternatives
Kergosien et al. (2011)	Case study	Quantitative	Metaheuristic algorithm (tabu search) and ILP	Alternatives are evaluated to find best solution.
Fröhlich Von Elmbach et al. (2019)	Case study	Quantitative	Metaheuristic algorithm (tabu search) and MIP	Alternatives are evaluated to find best solution.
Schmid and Doerner (2014)	Case study	Quantitative	Metaheuristic algorithms (LNS-based hybrid algorithm) and MIP	Alternatives are evaluated to find best solution.
Fiegl and Pontow (2009)	Case study	Quantitative	Heuristic algorithms (based on scheduling and graph theory)	Alternatives are evaluated to find best solution.
Xiao et al. (2022)	Case study	Quantitative	Heuristic greedy or heuristic column generation-based algorithm (with integrated metaheuristic GA or integrated MIP)	Alternatives are evaluated to find best solution.
Bärmann et al. (2024)	Case study	Quantitative	Heuristic (lexicographic) branch-and-bound column generation-based algorithms, classical column generation-based algorithms and MIP	Alternatives are evaluated to find best solution.
Vancroonenburg et al. (2016)	Case study	Quantitative	Heuristic algorithms (cheapest insertion and local search) and DES	Alternatives are evaluated to find best solution.
Hanne et al. (2009)	Case study	Quantitative	Metaheuristic algorithms (GA, etc.) and DES	Alternatives are evaluated to find best solution.
Meephu et al. (2023)	Case study	Quantitative	DES	Specific alternatives are evaluated to find the best solution among them.

2017, Gopal 2016, Vancroonenburg et al. 2016, Hanne et al. 2009, Meephu et al. 2023), mathematical models (Antunes et al. 2019, Akbari et al. 2023, Yazır et al. 2023, Li et al. 2021, Molenbruch et al. 2017, Kallrath 2005, Séguin et al. 2019, Kuchera and Rohleder 2011, Gopal 2016, Maka et al. 2022, Bouabdallah et al. 2013, Turan et al. 2011, Elmbach et al. 2015, Beaudry et al. 2010, Kergosien et al. 2011, Fröhlich Von Elmbach et al. 2019, Schmid and Doerner 2014, Fiegl and Pontow 2009, Xiao et al. 2022, Bärmann et al. 2024) and (meta-)heuristic algorithms (Akbari et al. 2023, Yazır et al. 2023; Li et al. 2021, Molenbruch et al. 2017, Kallrath 2005, Elmbach et al. 2015, Beaudry et al. 2010, Kergosien et al. 2011, Fröhlich Von Elmbach et al. 2019, Schmid and Doerner 2014, Fiegl and Pontow 2009, Xiao et al. 2022, Bärmann et al. 2024, Vancroonenburg et al. 2016, Hanne et al. 2009) were developed and used to simulate and evaluate specific scenarios. In Subsect. 2.1 it was already mentioned, that the conformity of simulation models are highly dependent on modelers and the provided domain knowledge. The same is valid for mathematical models and metaheuristic algorithms. Exactly here, ANNs can provide a remedy. Furthermore, Camargo et al. (2021) derived, that ANN models outperform automatically from process data derived simulation models when trained with large logs by comparing the relative accuracy for generating activity durations and process control flows. ANNs may be able to learn dependencies that cannot be captured by the process discovery algorithms that are the basis of data-driven process simulation approaches (Camargo et al. 2021). Literature has shown, that ANNs are able to model complex behaviour without having to assume certain function forms and degrees of non-linearities in advance (Mitrea et al. 2009; Gardner and Dorling 1998).

Like concluded in Dumitru and Maria (2013), West et al. (1997), Salami et al. (2016), Izadifar and Abdolahi (2006), Nikzad et al. (2012), Al-Waeli et al. (2019), Neto and Fiorelli (2008), ANNs offer superior predictive capabilities over traditional statistical or mathematical methods, especially where (non-linear) relationships are difficult to capture and describe. If sufficient data involving a wide range of all variables is available, ANNs are able to model complicated and multi-variable dependent processes (Izadifar and Abdolahi 2006).

Nas and Koyuncu (2019), Fashoto et al. (2016), Rajakumari and Madhunisha (2020), Mesabbah et al. (2019), Gul and Guneri (2015) already provide investigations on the usage of ANNs, or more generally of ML approaches, for automatic process redesign in healthcare organisation. ANNs can automatically recognise interrelationships between different process-relevant entities in the system (Abiodun 2019; Sharma and Kaur 2013).

The research question is whether the combination of an ANN and a metaheuristic algorithm is suitable to use process data to optimise the capacity planning in the IHPT through more efficient allocation of resources involved in the process. The objective is to investigate if and how the combination of both methods can lead to a significant, reliable and comprehensible improvement of the IHPT process. Therefore, in our approach, we develop a Multilayer Perceptron (MLP), i.e. an ANN, to automatically model dependencies between selected attributes within the IHPT based on historical data in order to accurately predict delayed cases. Subsequently, through the application of a GA, that utilises the MLP to predict effects of resource adaptions, optimises the resource allocation of the involved capacities (e.g. transporters). The validation of the proposed optimisation, that is derived by the GA (see Subsect. 4.4), will be given by the achieved performance of the MLP model (see Subsect. 4.3). The effectiveness of the GA is also assessed in Subsect. 4.4.3 on the basis of a problem with reduced complexity. Thus, the contribution of our paper is to show, based on the IHPT process, that optimisations in capacity planning are possible by combining the two data-based techniques mentioned. In this context, we also address how the raw data for the two techniques must be prepared and made available. We will use real data to improve and evaluate the predictive performance of our MLP models. This goes beyond IHPT literature, that mostly not use real data or only use broader statistics of real data to feed instances into simulation or mathematical models. So far, only the approaches of Kuchera and Rohleder (2011); Bärmann et al. (2024); Hanne et al. (2009) were transferred into practice and tested. Predictive performance is mostly not evaluated in detail in IHPT literature, and if it is, it is almost exclusively qualitative. Only (Meephu et al. 2023) investigated differences in mean ratings between historical and simulation data by applying the two-sample T-test and confidence interval. We will present relevant predictive performance metrics for our developed MLP model (see Subsect. 4.3).

Our redesign results will further serve as an input at a tactical or strategical level to the IHPT literature addressing the routing and scheduling planning in detail at the operational level (see Subsect. 2.1 and also Table 2) with the appropriate instances in place. We want to communicate the benefits of redesigning IHPT capacity planning by combining an ANN and a metaheuristic algorithm in IHPT with exemplary techniques recommended in the literature.

3 Materials and methods

In order to introduce the techniques used in our use case (see Sect. 4), this chapter briefly presents general information about MLPs in Subsect. 3.1 and about GA in Subsect. 3.2.

3.1 Multilayer perceptron

The MLP is one type of ANN architectures (feed-forward network) (Gardner and Dorling 1998). It is the most common ANN (Guneri and Gumus 2008) and the focus of the majority of ANN research (Ncibi et al. 2017). Generally, an ANN is considered to be a massive parallel combination of simple processing units, also known as neurons or nodes, which can acquire knowledge from environment through a learning process and store the knowledge in its connections (Haykin 1999). It is widely utilised in cognitive tasks, such as learning and optimisation (Lawrence and Luedeking 1993).

An MLP specifically consists of an input layer, multiple hidden layers, and an output layer (Gardner and Dorling 1998). Figure 2 shows schematically a MLP with exemplary inputs and outputs. The input can be numbers, characters, audios,

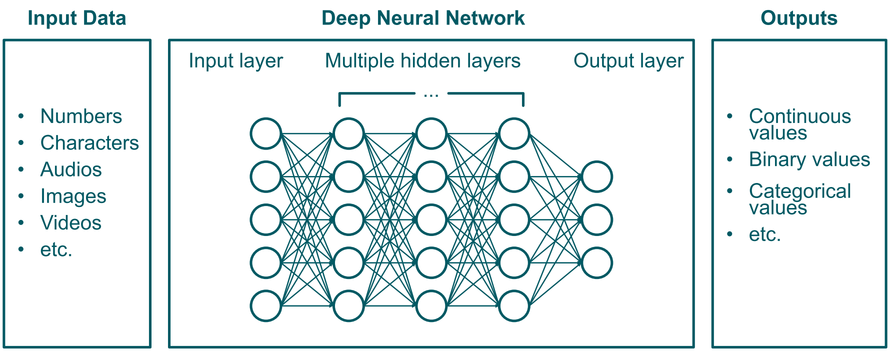

Fig. 2 Diagram of a MLP-ANN (adopted from Zhu et al. 2023)

images, etc. that are decomposed into binary data that can be processed by a computer (Zhu et al. 2023). The output can be e.g. of continuous, binary or categorical values, depending on the specific task (Zhu et al. 2023). Each layer consists of neurons or nodes (Gardner and Dorling 1998). With the exception of the input nodes, each node within the MLP network represents a neuron that leverages a non-linear activation function (Gardner and Dorling 1998). Figure 3 shows schematically the perception of a neuron (in the hidden or output layer). A neuron first multiplies each of its inputs x_i by an associated weight w_i, then sums these weighted inputs and adds a preset number b called the bias (Zhu et al. 2023; Russell and Norvig 2010). The result of this computation is then adjusted by an activation function *g(x)* (Zhu et al. 2023; Russell and Norvig 2010). In Fig. 3, the exemplary activation function, and one of the most popular, is called Rectified Linear Unit (ReLU) (Zhu et al. 2023; Russell and Norvig 2021). There are also other types of non-linear activation functions, e.g. the popular Sigmoid, Tanh or Softplus functions (see Russell and Norvig 2021 for further information), that can be used to activate the neurons (Zhu et al. 2023; Russell and Norvig 2021). A general challenge in developing an ANN is that there is no good theory for achieving an optimal ANN structure (Russell and Norvig 2010). There is always a degree of fiddling to achieve the appropriate network structure for the underlying problem and data (Russell and Norvig 2010).

An MLP does not rely on any prior assumptions about the distribution of the utilised data (Gardner and Dorling 1998). It possesses the capability to model complex

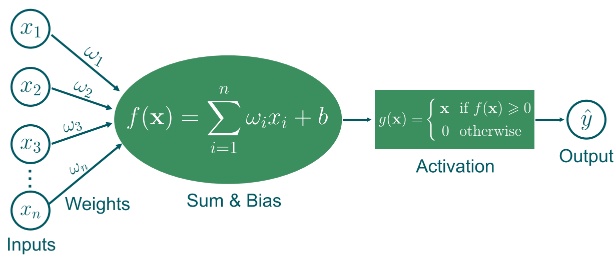

Fig. 3 Diagram of a neuron's perception (adapted from Zhu et al. 2023)

non-linear functions and can be trained to generalise accurately when confronted with previously unseen data (Gardner and Dorling 1998). These intrinsic properties of the MLP make it a compelling substitute against traditional mathematical models and serve as a viable choice when selecting among statistical methodologies (Gardner and Dorling 1998). Other ANN architectures, such as RNN or CNN, are also being successfully used more recently in healthcare prediction tasks on time series data (Morid et al. 2021). Comparative investigations of these architectures are beyond the scope of this paper, but are recommended for further studies to be conducted in the future (see Sect. 6).

3.2 Genetic algorithm

Combinatorial optimisation problems can generally be computed through exact solving methods or approximate methods (Blum and Roli 2003). In IHPT problems, that are usually characterised through complex models, excessively long computation times for optimisation need to be overcome in order to increase the practical implementation and relevance (Klein and Thielen 2024). As exact solving methods for complex problems might consume exponential computation time, that is too high for practical purposes, approximate methods received more attention in the recent years (Blum and Roli 2003). Among these approximate methods, metaheuristic algorithms have emerged (Blum and Roli 2003).

GA are search algorithms taking natural genetics and natural selection as role models (Goldberg 1989). They belong to the family of metaheuristic algorithms used to solve real-life complex problems (e.g. multi-objective problems) and are population based (Katoch et al. 2021; Goldberg 1989). This means that they search from a population of solutions and use reproduction, crossover and mutation of successful solutions (evaluated on the basis of an objective function) to iteratively improve performance towards an optimal point or points (Goldberg 1989). Like this GA maintain the diversity in population and avoid the solutions getting stuck in local optima (Katoch et al. 2021).

The GA outperforms other population-based algorithms, like e.g. more recently evolved swarm-intelligence based ones, when the computational budget is high (Piotrowski et al. 2017). Swarm-intelligence algorithms are inspired by the collective behaviour of species such as ants, bees, wasps, etc., and are therefore called behaviourally inspired algorithms (Talbi 2009; Janga Reddy and Nagesh Kumar 2020; Bonabeau et al. 1999). A comparative analysis of different combinatorial optimisation methods is not carried out in this paper, but is recommended for further studies (see Sect. 6).

The performance (speed of convergence) of a GA can be further improved by elitism selection, where the elitist individual or individuals per generation will be always propagated to the next generation, regardless of whether or not they would be present in the next generation according to normal selection procedure (Katoch et al. 2021; Jebari and Madiafi 2013). However, elitism selection may lead to a risk of convergence to local optima due to decreased genetic diversity within the population of solutions (Jebari and Madiafi 2013). By exploiting widely available information

GA are applicable to virtually any problem (Goldberg 1989). Since GA are combinatorial, they are furthermore suitable for solving discrete or mixed discrete (mix of continuous and discrete design variables) optimisation problems, that usually require more computational effort (Wu and Chow 1994).

4 From raw data to capacity optimisation

Figure 4 shows schematically the procedure of our approach. We give general information about the underlying raw data of the IHPT process in the investigated German hospital in Subsect. 4.1. Subsection 4.2 deals with data preprocessing to prepare the data accordingly for the MLP model. Subsection 4.3 presents our MLP model development, as well as different validation procedures addressing the predictive performance of the MLP model. In Subsect. 4.4 the automated optimisation of the transport capacities is performed. A GA is applied to reallocate resources so that the MLP will predict reduced delayed transport cases and a validation of the optimisation performance is furthermore conducted using a simplified problem with a more limited solution space. The general aim of our approach is to create better conditions for IHPT through optimised capacity planning.

In conducting our research, we have adhered strictly to ethical guidelines and data protection standards to ensure the confidentiality and safety of patient and transporter information. The data utilised for this study were provided to us in a de-identified format, ensuring that individual patients or transporters could not be identified. Specifically, transport staff were de-identified through the use of device IDs associated with the portable devices, that receive assignments, and no personal names were included in our dataset. Furthermore, each patient was assigned a unique patient ID that prevented the identification of individuals. No medical diagnoses or health conditions were included in the data, as the information was strictly limited to logistical information, including e.g. timestamps of various activities, types of transportation vehicle, priorities, or locations (see Subsect. 4.1 for more details) involved in the IHPT process. No personal data related to individuals were processed or stored in our research database. Furthermore, our research presents only aggregated data that do not allow for any conclusions to be drawn about individual transports

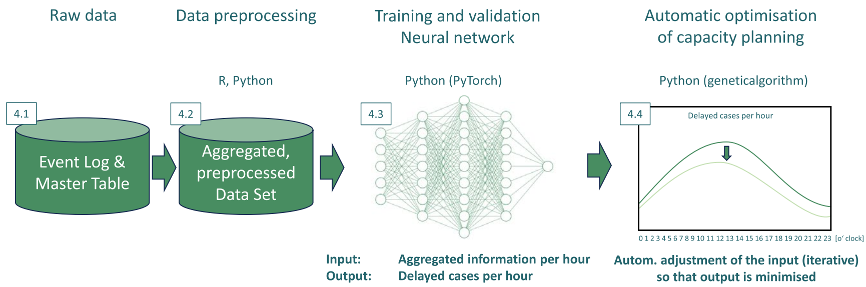

Fig. 4 Procedure of our approach: from raw data to automated resource optimisation

or individuals. Through these measures, we ensure compliance with applicable data protection regulations, maintaining the highest standards of confidentiality and integrity in handling healthcare related data. Our approach also eliminates the Hawthorne effect (marked differences in the performance of involved people due to observation and documentation of a process, see Subsect. 2.1), because an inner subset of the information that has been consistently collected over the past few years in the IHPT process is utilised.

4.1 Raw data:general information

In the hospital logistics software system, the IHPT process is logged, and specific information is associated with each transport case. Figure 5 visualises the usual IHPT process flow. In the hospital, transports take place e.g. from wards to functional areas but also in the opposite direction. There are transports between wards or between functional areas, too.

The logged information on the IHPT process is provided by the hospital in the form of two CSV files. Both datasets used for our approach encompassed a timeframe with cases where the first activity occured between January 1, 2019, and June 30, 2022 (a handful of transport cases had some activities on July 1, 2022 even though their first registered activity took place beforehand). This reflects around 3.5 years of data collection. The selected data focused solely on fully completed transports involving patients as the subjects, with no additional specialised services provided apart from transportation. One file presents an event log. The event log indicates the events, that happened within the IHPT process with corresponding activities that occurred during each transport, along with the timestamps indicating when each event and corresponding activity took place. Table 3 displays the various activities and indicates the number of transport cases in which each activity occurs. Additionally, it provides information on the overall frequency of activities, considering that they may occur multiple times within a single case or not occur at all.

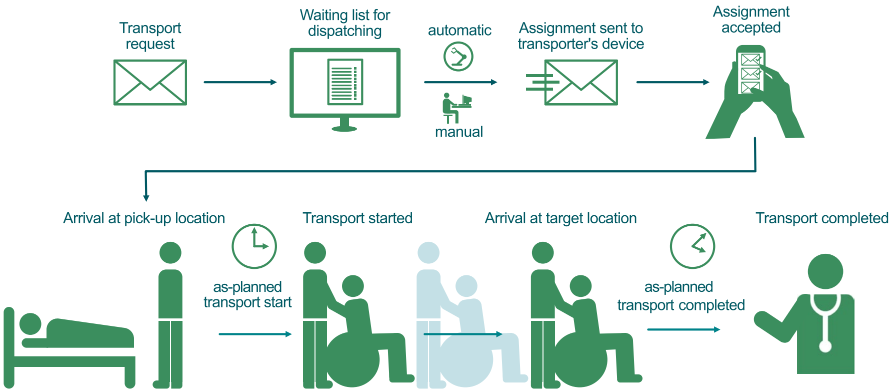

Fig. 5 Exemplary patient transport in the case study hospital

Table 3 Statistics on all activities (adopted from Kropp et al. (2024))

Activity (German - in System)	Activity (English translation)	Case count	Activity count
Auftrag abgeschlossen	Transport completed	256,266	257,205
Auftrag an Endgerät	Assignment sent to device	256,266	282,505
Auftrag angenommen	Assignment accepted	256,266	262,262
Transport begonnen	Transport started	256,266	256,556
Warteliste Kommissionierung	Waiting list for dispatching	256,266	302,271
An Abholort	Arrival at pick-up location	256,265	257,334
An Ankunftsort	Arrival at target location	256,265	256,376
Anforderung	Transport request	256,261	317,641
Vorgemerkt	Transport is pre-registered	111,906	127,579
Verfall - nicht zugestellt	Expiration - not delivered	6,021	9,714
Dispo	Transport assignment	145	153
Storniert	Transport canceled	33	39

In total, there are 2,329,635 events, each with one with an activity, timestamp and transport ID (each transport has a unique ID), per row in the event log.

The second file is a master table containing more information on the total of 256,266 cases, whereas each row represents an individual case, accompanied by 125 attribute categories that contain case-specific information. Out of these 256,266 transport cases almost 35% were completed with a delay of ten minutes or more. This threshold, above which transports are considered to be significantly delayed, was set by hospital process managers. They consider the percentage of delayed transports furthermore as critical and therefore there is a need to reduce the number of delayed IHPT cases using knowledge from the provided data. Both the event log and the master table can be linked via the transport IDs, that are present in both tables, and together they provide the raw data that needs to be preprocessed for MLP model development. Figure 6 summarises information on the raw data. There exist 125 case-related attribute categories in the master table for all of the cases. Overall, the master table contains 32,033,250 data cells (256,266 cases multiplied by 125 attribute categories), out of which 5,506,348 cells are empty or lack information (NULL/NA values). Figure 6 further classifies the attribute categories and presents the average and median number of entries with relevant information for each attribute category. By subtracting these values from the total of 256,266, the number of missing attributes on average or median per category cluster can be determined. Out of the 125 attribute categories in the original dataset 37 are related to the corresponding event timestamps (e.g. month, day of the week, calendar week, pre-calculated time differences between process steps). The other 88 attribute categories contain further general organisational information that are linked to transport cases (e.g. requesting centre, assignment centre, cost centre, last control station, first control station, priority, pick-up location, target location, pick-up house, arrival house, pick-up level, arrival level, pick-up priority, arrival priority, route id, route description, distance, remarks, service provider, operator, tour, type of transport vehicle, patient id, pick-up room number, pick-up room, arrival room number, arrival room,

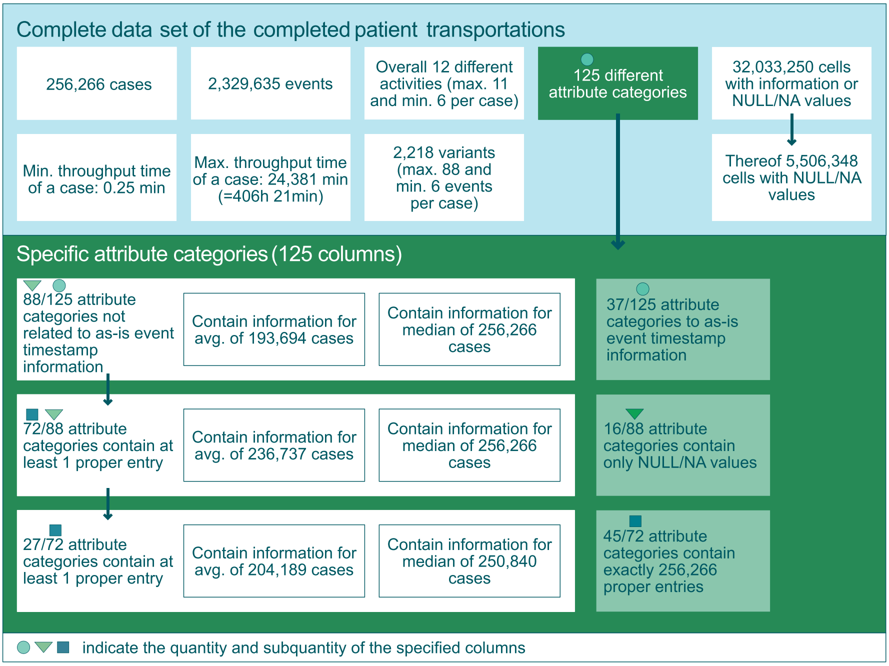

Fig. 6 Information on underlying raw data within a data collection period of around 3.5 years

complaint text, complaint category). Such attribute categories can hold specifically valuable information for deeper investigations into the root causes of process issues as organisational interrelationships can be revealed from them. However, in our use case not all transport cases have consistent information among these attributes categories. Among the 88 attribute categories with further organisational information, 16 contain no relevant information, while the remaining 72 categories include varying degrees of information. Specifically, 45 of these 72 attribute categories are completely filled with evaluable information, while the remaining 27 categories are inconsistently filled and are therefore only suitable for limited further investigations.

4.2 Preprocessing the data for MLP model training

For further investigations, the following information becomes particularly relevant as it pertains to improved resource planning: planned day of the week (Monday to Sunday) of transports as well as the as-planned and as-is starting time (specific to the hour between 0 and 23) of transports. In order to make hourly predictions with greater detail on a day of the week basis in the future, new attribute categories are being created that will be assigned to individual hours within each date and will latterly serve as input data for the MLP model.

Figure 7 shows the data preprocessing approach, using colours to illustrate the derivation and processing of specific information from the raw data (upper table)

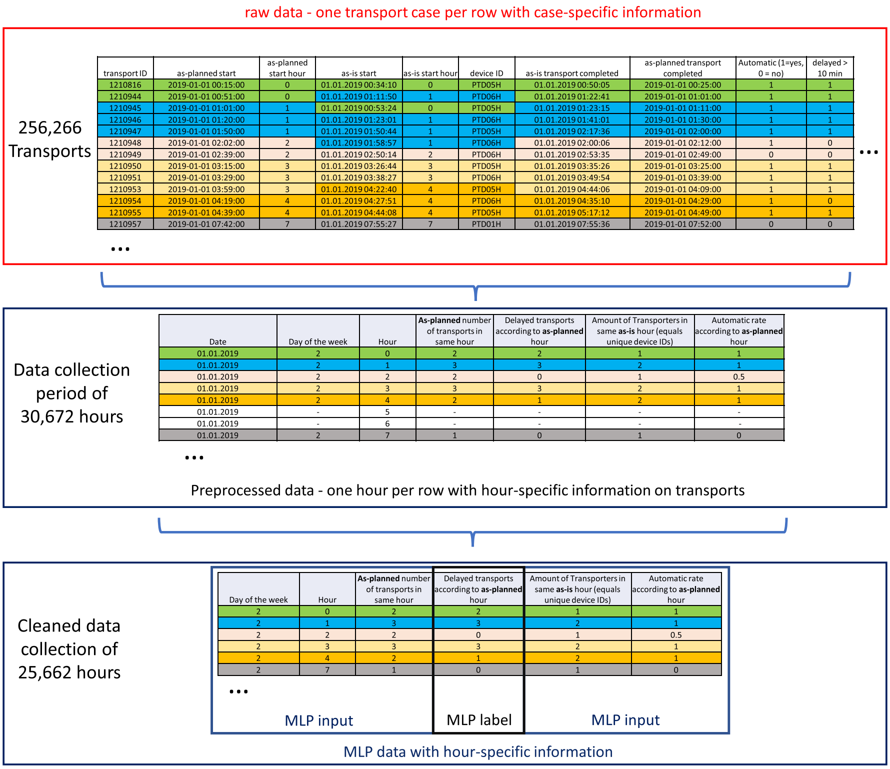

Fig. 7 Subset of the data preprocessing approach

and its placement in the individual rows of the first step preprocessed table (middle table). New attribute categories were generated using a self-developed script in R, with libraries "data.table" (Barrett et al. 2024), "dplyr" (Wickham et al. 2023) and "lubridate" (Grolemund and Wickham 2011), and Python, with libraries "numpy" (Harris 2020) and "pandas" (The pandas development team 2023; McKinney 2010). During the preprocessing *ChatGPT* (versions 3.5 and 4; https://chat.openai.com/) was used to search for and explain certain methods of the used libraries or R and python functionalities. Thus, the following information is calculated based on the raw data for each hour from January 1, 2019 to July 1, 2022 (total time span of 30,672 h):

1. "As-planned number of transports in same hour": Number of transports planned for the same date and at the same starting time (specific to the hour between 0 and 23).
2. "Delayed transports according to as-planned hour": Number of transports planned for the same date and at the same starting time (specific to the hour between 0 and 23) being delayed more than ten minutes.

3. “Amount of Transporters in same as-is hour (equals unique device IDs)”: Number of transporters that are active for the same actual date and at the same actual time (specific to the hour between 0 and 23). This information is generated based on the timestamp associated with the “Transport started” activity, in conjunction with the device ID managed by each transporter.
4. “Automatic rate according to as-planned hour”: Automation rate for the same planned date and at the same planned time (specific to the hour between 0 and 23). This rate indicates the extent to which transports scheduled at the same time were finally manually or automatically assigned.

Afterwards, rows without information and additionally, the date column are then excluded from the training of the MLP model, as the model’s focus is solely on day of the week and hour of the day. The lower table of Fig. 7 presents the remaining and final preprocessed data (total of 25,662 rows with hourly information) including the schematic input and label data used for training and evaluating the MLP model. Like this, on average, around 153 samples are available for each hour on each day of the week (25,662 samples divided by the product of 7 days of the week times 24 h). To avoid the risk of not having enough training data available, we also decided not to aim at more granular predictions for days of the week in specific months or calendar weeks. This would require a larger dataset during MLP model development for similarly accurate predictions. In addition to the four derived information per day of the week and hour (see lower table lower table of Fig. 7), further organisational attribute categories from the raw data could be included to make more specific statements. However, the aim of this paper is to show that the use of MLP in combination with a GA can theoretically achieve optimisations in the planning of transport capacities, so that a coarse level of predictions, i.e. for days of the week, is sufficient for our purposes.

4.3 Model developement

The MLP model development is done by using the python libraries “PyTorch” (Paszke et al. 2019), “NumPy” (Harris 2020), “pandas” (The pandas development team 2023; McKinney 2010) and “scikit-learn” (Pedregosa 2011). Data visualisations are conducted using the python library “Matplotlib” (Hunter 2007). During the development of the MLP model *ChatGPT* (versions 3.5 and 4) was used to search for and explain certain methods of the used libraries or python functionalities.

In order to model the process of the IHPT we utilise an MLP model that is able to conduct a regression. With this model we aim to predict the resulting delayed cases by specific given input data (see Fig. 7).

4.3.1 MLP model specifications

The procedure of the MLP (see Subsect. 3.1) model development is shown in Fig. 8. Firstly the entire preprocessed dataset (from Subsect. 4.2) is divided into training set, validation set and test set, with proportions of 80%, 10%, 10%.

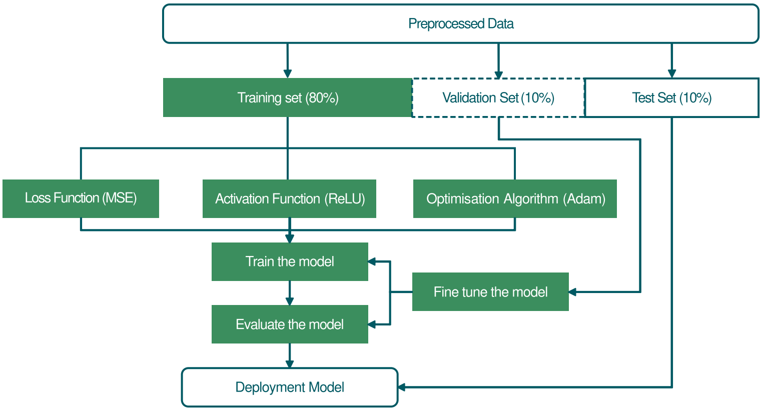

Fig. 8 MLP development steps in our approach

The mean squared error (*MSE*) is selected to be the loss function. The *MSE*, that is a commonly used metric for the performance evaluation of regression models, is calculated according to equation 1 (Russell and Norvig 2021; Goodfellow 2016).

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{1}$$

where m is the number of samples, y_i are the actual values and $\hat{y}_i$ are the predicted values.

For the MLP model we choose one input layer, three hidden layers and one output layer. A schematic visualisation of our model is shown in Fig. 9. In our case we have 5 neurons in the input layer, then 64, 256 and 64 neurons in the hidden layers as well as one neuron in the output layer. The hidden layers with their neurons are derived through trial-and-error procedures where this setup showed a *MSE* values compared

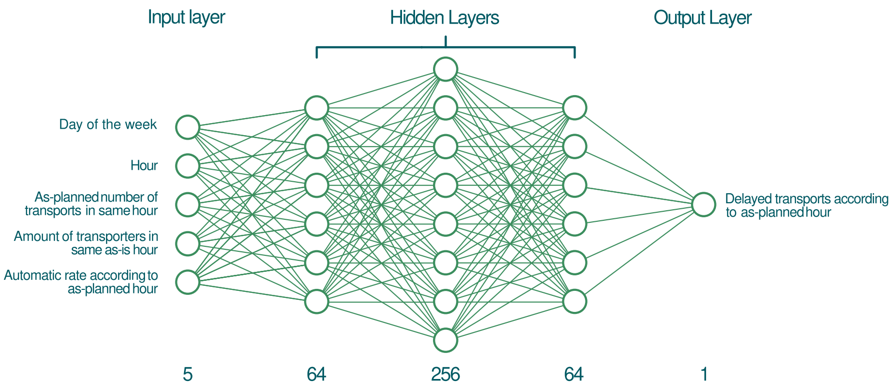

Fig. 9 Structure of the developed MLP with one input layer, three hidden layers and one output layer. Per layer the number of neurons is indicated. The input data per input neuron and the output data for the output neuron is shown

to other attempted setups. ReLU (see Subsect. 3.1) is chosen as activation function (in the hidden layers). Attempts with other activation functions such as Sigmoid, Tanh or Softplus showed no significant improvements or deteriorations.

We utilise the Adaptive Moment Estimation (Adam) as optimisation algorithm within the MLP training, that was recommended in Nas and Koyuncu (2019) for hourly patient arrival data that is characterised by a lot of variation. It is for first-order gradient-based optimisation of stochastic objective functions based on adaptive estimates of lower-order moments (Kingma and Ba 2017). During the training process the parameters of the model are updated in 500 epochs. After the parameters are updated in each epoch, the validation loss is monitored along with the training loss. In order to avoid overfitting, an early-stop method is applied when the validation loss stops falling. Figure 10 shows for different dataset sizes how the model training and validation losses evolve. In comparison with a preprocessed dataset of only 2019, a preprocessed dataset of 2019 to 2020 and a preprocessed dataset of

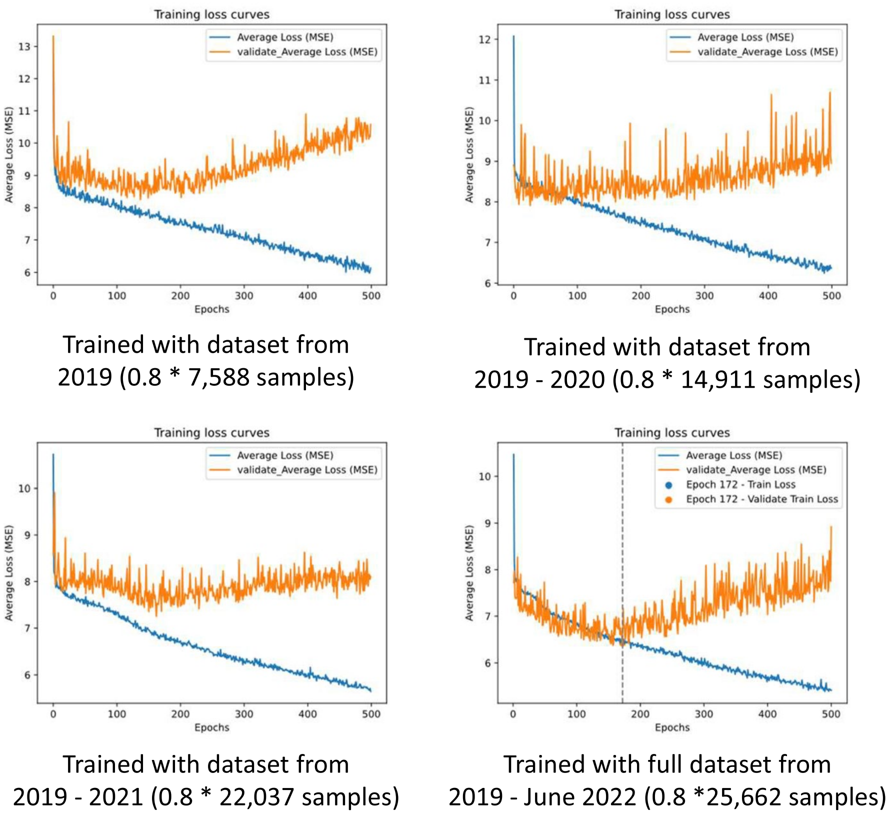

Fig. 10 Training and validation loss per training epoch for different dataset sizes. The full preprocessed dataset showed over the whole training the best results. Through early stopping the MLP model at epoch 172 is considered the best as training and validation loss are both relatively low

2019 to 2021, the complete preprocessed dataset from 2019 to June 2022 (derived from Subsect. 4.2) shows, as expected, the best results. Through early stopping the MLP model at epoch 172 is considered as best fitting model, as training and validation loss are both relatively and similarly low. Both are at a *MSE* of around 6.4. We use the test dataset, that showed similar loss results as the validation set, to verify the generalisation ability of the model. From figure 10 it can be interpreted that the inclusion of additional data in the training would further improve the model, as the *MSE* achieved for the training and validation data with more data is in lower ranges. Even though the complete dataset of 25,662 samples contains only discrete values as delayed transport cases in every single sample, the MLP predicts continuous values. However, the fact that our model can handle continuous values both as output and as input information is intentional, so that it can also deal with e.g. average scenarios.

4.3.2 Further MLP validation

Next to *MSE* we investigate further performance metrics on our model. A performance evaluation metric in terms of variance explanation ability of regression models can be the coefficient of determination R^2 (Miles 2014; Ozili 2022; AlDahoul 2021; Muloiwa et al. 2023). R^2 measures in a model how good the variance in the actual values y_i is explained by the input variables in the set of samples and is shown in equation 2 (Miles 2014; Mikut 2008; Muloiwa et al. 2023).

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{m} (y_i - \overline{y})^2} \tag{2}$$

where m is the number of samples, y_i are the actual values, $\hat{y}_i$ are the predicted values, and $\overline{y}$ is the mean of the actual values y_i. Thereby, the improvement of the regression estimate of the model is evaluated against the simple mean of the actual values in the sample set (Mikut 2008). The values of R^2 range usually between 0.0 (no relationship) and 1.0 (deterministic relationship) (Mikut 2008).

For the performance evaluation of models with multiple input variables there exists also a so called *adjusted* R^2, because in literature an undesired increase in R^2 could be observed with increasing number of input variables (Miles 2014). The *adjusted* R^2 is to be calculated according to equation 3 (Miles 2014).

$$adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - k - 1} \tag{3}$$

where R^2 is calculated according to equation 2, N is the sample size, and k is the number of input variables in the model. The higher R^2 is, the better a model is able to explain the variance in the actual data (Miles 2014). A model with a R^2 of 1.0, which is the highest possible value, is able to explain all of the variance in the set of samples (Miles 2014). A value of 0.0 (or a negative R^2, that could be theoretically observed) means that the predictive performance of a model is just as accurate (or theoretically even less accurate, if R^2 is negative) as the average value of the set of samples (Ozili 2022). In science domains, where researchers deal with objects, molecules, materials, or atoms whose properties are known and whose behaviour are predictable and do not change over time, it is reasonable to expect a high R-squared in the developed models (Ozili 2022). In models dealing with human behaviour, which can change over time, an acceptable value of R^2 for a prediction model is already considered to be between 0.5 and upwards towards 1.0 (Ozili 2022). The mentioned characteristics of or R^2 are also valid for *adjusted* R^2 (Ozili 2022). As R^2 underlies linear assumptions, R^2 can tend to mislead performance evaluations related to non-linear models and should not be used as sole performance metric (Spiess and Neumeyer 2010). In Spiess and Neumeyer (2010) it is shown on exemplary non-linear models in the pharmacological and biochemical domain that using R^2 (and also *adjusted* R^2) could indicate the best performing model in only around 28-43% (depending on the underlying experimental noise in the data) of the experimental evaluations. Nevertheless, R^2 (and also *adjusted* R^2) is frequently used in the literature to evaluate ANNs, and we consider it only as an additional performance metric to the *MSE* from Subsect. 4.3.1 and further metrics that will be presented in this subsection.

Figure 11 shows the Confusion Matrix of the MLP Model over the complete dataset (25,662 samples) from which R^2 and *adjusted* R^2 can be derived. Evaluating real and predicted delayed cases (continuous values are not rounded for this calculation) over the complete dataset results in both a R^2 and a *adjusted* R^2 of 0.794 (*adjusted* R^2 is slightly lower from the fourth decimal onwards). The performance of the MLP model in terms of variance explanation ability is thus in a good range, according to the acceptable bounds as per Ozili (2022).

Also, a metric called mean error bias (*MEB*), can be calculated, according to Hernández-Orallo (2013), in equation 4. It indicates whether a model tends to

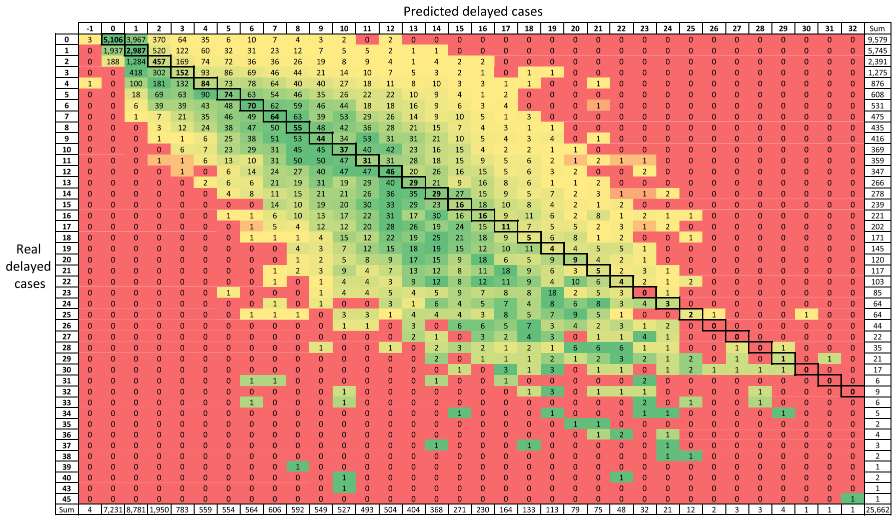

Fig. 11 Confusion Matrix of the MLP model. Predicted delayed cases, as the MLP model outputs continuous values, are for simplified visualisation rounded to the nearest integer value (values exactly halfway between integer values are rounded to the nearest even value)

over- or underestimate with its predictions (Hernández-Orallo 2013). For the developed MLP model the *MEB* is around −0.056. This indicates that the MLP model predictions $\hat{y}_i$ only slightly underestimates the actual values y_i in the complete dataset. In contrast, a positive *MEB* would indicate, that a model tends to overestimate with its predictions (Hernández-Orallo 2013). Furthermore, Figure 11 and also *MEB* calculations for dataset subpartitions show that the MLP model tends to underestimate more with its predictions when the values of real delayed cases are in higher regions. This may be due to the higher frequency of low values of real delayed cases (i.e. a high frequency of actual values y_i with a low value) in the dataset that can lead to a biased model. For real delayed cases of e.g. zero (see the first row of figure 11), the MLP model even overestimates slightly with its predictions.

$$MEB = \frac{\sum_{i=1}^{m} (\hat{y}_i - y_i)}{m} \tag{4}$$

where m is the number of samples, $\hat{y}_i$ are the predicted values and y_i are the actual values.

As we want to reconfigure the capacity planning on a day of the week level, we evaluate the performance of our MLP model for average day of the week delayed cases predictions. Table 4 compares the average delayed cases per day of the week, that occurred in reality, with the delayed cases that are predicted by the MLP model using average day of the week input tensors. Depending on the day of the week, the total predicted delayed cases deviate more or less from reality, but with an average relative deviation of around 3.9% per day of the week the predictions of the MLP model are within a satisfactory range. Furthermore, the hourly course of the predicted delayed cases over the individual days of the week also comes close to the original course. In Subsect. 4.4, the example of Monday (see Fig. 14 in Subsect. 4.4) will be shown in more detail along with the optimisation process of capacity planning.

To provide further validation, we adopt the method in Torgo and Ribeiro (2009), which introduces the generalisation of precision, recall and F score for regression problems. These performance metrics are commonly used in ML classification problems but can be transferred to regression problems, too Davis and Goadrich (2006);

Table 4 Real vs. predicted delayed cases for average days of the week

Day of the week	Avg. real delayed cases	Avg. predicted delayed cases	Relative deviation
Monday	106.1429	107.5983	1.37%
Tuesday	84.4208	81.9533	2.92%
Wednesday	92.3770	86.4243	6.44%
Thursday	77.7650	82.5056	6.10%
Friday	91.6284	87.8406	4.13%
Saturday	15.2033	15.2692	0.43%
Sunday	12.2802	12.9949	5.82%
			Avg. 3.89%

Torgo and Ribeiro (2009). This method has been applied by the authors of Torgo and Ribeiro (2009) to evaluate the performance of different ML models in a regression problem in the economic domain. The core idea is to divide the data into two categories, “target events” (the samples which are considered relevant) and “non-events” (the samples which are considered not or less relevant) (Torgo and Ribeiro 2009). This makes it possible to align performance metrics on the ability of the model to make predictions in the relevant ranges (Torgo and Ribeiro 2009). Subsequently, the precision, recall and F score can be calculated (Torgo and Ribeiro 2009).

The standard event-driven classification has a positive class, here representing the previously defined “target events” and a negative class, here representing the previously defined “non-events” (Torgo and Ribeiro 2009). Table 5 is a general confusion matrix of a classification problem, where the precision and recall of a model can be calculated according to equations 5 and 6 (Torgo and Ribeiro 2009; Flach 2003).

$$Precision = \frac{TP}{POS} \tag{5}$$

$$Recall = \frac{TP}{PPOS} \tag{6}$$

where *TP* stands for True Positive, *POS* stands for Actual Positive, *PPOS* stands for Predicted Positive (Flach 2003).

Our general goal was to develop a model that allows accurate predictions of delayed cases in the common operation of the hospital under investigation.

Therefore, the data is divided into two categories according to the “Delayed transports according to as-planned hour”: The first category represents the majority of the data, and is designated as the positive class, thus representing the usual operational behaviour in the hospital. The remaining data is considered as negative class (outliers, that occurred rarely in the historical data), which are not relevant for the transport capacity planning. A relevance function $\phi()$ is needed to divide the dataset into these two classes (Torgo and Ribeiro 2009). The choice of this relevance function $\phi()$ is domain-dependent and not always easy to set (Torgo and Ribeiro 2009). In Torgo and Ribeiro (2009), the relevance function $\phi()$ is defined dependent on the ML model label and prediction. In our case, the relevance function is thus dependent on the real and predicted “Delayed transports according to as-planned hour” (MLP label, see Subsect. 4.2). The mathematical expression of our chosen relevance function $\phi()$, that weights the positive class with 1 (relevant) and the negative class with 0 (not relevant), is presented in equation 7 (adapted from Torgo and Ribeiro (2009)).

Table 5 Standard Confusion matrix of classification problem (according to Torgo and Ribeiro (2009); Flach (2003))

	Predicted positive	Predicted negative
Actual positive	TP	FN
Actual negative	FP	TN

TP, true positive; FN, false negative; FP, false positive; TN, true negative

$$\phi(Y) = \begin{cases} 0 \text{ if } Y \geq t_E \\ 1 \text{ if } Y < t_E \end{cases} \tag{7}$$

where t_E is the relevance threshold, dependent on the real "Delayed transports according to as-planned hour" and Y can be either the real or predicted "Delayed transports according to as-planned hour" (Torgo and Ribeiro 2009). By introducing this concept of relevance, the regression problem is transformed into a binary classification problem (Torgo and Ribeiro 2009).

With equation 7, the definition of precision and recall for a regression problem are as shown in equations 8 - 9 (Torgo and Ribeiro 2009).

$$Recall = \frac{\sum_{\phi(y_i) \leq t_E} \alpha(\hat{y}_i, y_i) \cdot \phi(y_i)}{\sum_{\phi(y_i) \leq t_E} \phi(y_i)} \tag{8}$$

$$Precision = \frac{\sum_{\phi(\hat{y}_i) \leq t_E} \alpha(\hat{y}_i, y_i) \cdot \phi(\hat{y}_i)}{\sum_{\phi(\hat{y}_i) \leq t_E} \phi(\hat{y}_i)} \tag{9}$$

where $\phi(\hat{y}_i)$ is the relevance function dependent on the predicted "Delayed transports according to as-planned hour" and $\phi(y_i)$ is the relevance function dependent on the real "Delayed transports according to as-planned hour" (see equation 7) and $\alpha(\hat{y}_i, y_i)$ is the accuracy of prediction (loss between $\hat{y}$ and y) defined according to equation 10 (Torgo and Ribeiro 2009).

$$\alpha(\hat{y}, y) = I(L(\hat{y}, y) \leq t_L) \tag{10}$$

where $I(\,)$ is a indicator function given 1 if its argument is true and 0 otherwise. It is dependent on a tolerance threshold t_L that is to be defined and reflects an admissible error within a loss function $L()$ (e.g. the absolute or squared deviation). Following (Torgo and Ribeiro 2009) the loss function $L()$ is chosen to reflect the absolute deviation between $\hat{y}$ and y. The definition of the F score is shown in equation 11 (Torgo and Ribeiro 2009).

$$F = \frac{(\beta^2 + 1) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{11}$$

where $0 \leq \beta \leq 1$, determines the relative importance of recall to precision (Torgo and Ribeiro 2009). For an equal weighting of precision and recall β is chosen to be one. For this setup the F score is also known as F1 score (Russell and Norvig 2010; Ruiz-Sepúlveda et al. 2009).

Now everything except the relevance threshold t_E and the tolerance threshold t_L are defined. To choose t_E we consult Figure 12, that visualizes the cumulative frequency of the real "Delayed transports according to as-planned hour".

The top 4.88% of the data are considered as outliers and mapped to the negative class, when t_E is set to 17. For this t_E we calculate the precision with different tolerance thresholds t_L. Figure 13 shows the results. With increasing tolerance the

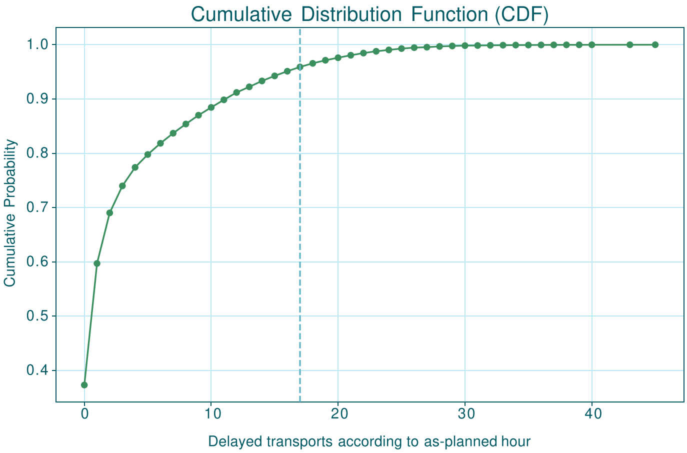

Fig. 12 Cumulative distribution of real "Delayed transports according to as-planned hour" in complete dataset (25,662 hourly information)

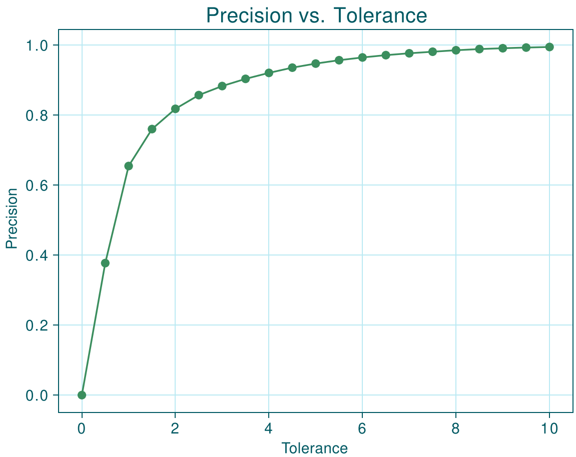

Fig. 13 Relationship of precision of the MLP model and tolerance threshold t_L

positive gradient of the precision decreases. A similar behaviour is observed for recall and F1 score.

For example, a tolerance t_L of two leads to a precision of around 0.81. This means that the MLP model predicts with 81% probability the correct "Delayed transports according to as-planned hour" with a tolerance of two delayed transports when outliers of 17 and more delayed transports are not considered. The recall with a tolerance t_L of two is around 0.83 and F1 score is around 0.82. The reached precision, recall and F1 score (and thus indirectly the *MSE* from Subsect. 4.3.1) of the MLP model are rated not yet good enough for practical implementation by the hospital process managers. They propose a minimum precision, recall and F1 score of 0.9 with a tolerance t_L of one. A tolerance of one is argued by the hospital managers because it equals a range of two delayed cases (i.e. up to overestimating and underestimating the real delayed cases by one delayed case), as previous investigations revealed that on average a transporter can more or less conduct three transports per hour (it takes a transporter 20 min from the acceptance of a transport assignment to transport completion). With this in mind, a tolerance of one around the real amount of delayed transports and thus an allowed range of two delayed transports reflects less than what could possibly be conducted on time by an additional FTE. From a purely economic point of view, a misprediction of the number of delayed transports that leads to a range of over- and understaffing of more than 1 FTE (e.g. a tolerance of two leads to an allowed range of four delayed cases of the predictions, within a maximum of two over and two under the real delayed cases, which is more than what an average FTE can handle) is seen as not precise enough for practical implementation by the hospital process managers. Our current MLP model reaches only a precision of 0.64, a recall of 0.66 and an F1 score of 0.65, when the tolerance t_L is set to one. Therefore, the MLP model still needs to be improved before it can be used in practice. Nevertheless, in the following further theoretical capacity planning evaluations using our MLP model will be conducted.

Our results show in general, that it is feasible to use precision, recall and F1 score to evaluate the performance of regression models. Furthermore, by setting different relevance thresholds t_E and evaluating the respective performance metrics, it is possible to gain a deeper insight into the performance of the model within specific regions of interest.

4.4 Optimising the resource planning

In this subsection the resource planning will be optimised theoretically by adapting an input tensor so that the output of the MLP model (which reflects the delayed cases) will be minimised. As an example the input tensor of a mean Monday will be utilised as original tensor (see Fig. 14 left part) that is to be altered to find an optimisation of the delayed cases, that are predicted to occur. Figure 14 (right part) shows on the one hand the real delayed cases per hour on an average Monday and also the amount of delayed cases per hour that the MLP model predicts based on the original tensor. In reality on average there were in total around 106 delayed cases on Mondays whereas the MLP predicts a total of around 108 cases. As the total predicted delayed cases differs less than 2% from the real delayed cases and as the trend of the predictions across all hours apparently corresponds well to reality, the reliability of the model to reflect an average Monday is considered good, and an optimisation of the input tensor will be carried out in the following.

For the optimisation a GA (see also Subsect. 3.2) is utilised, as it intents to anneal points of optima in complex problems. The GA explores the MLP with numerous solution attempts involving multiple new combinations of discrete variables in the input tensor to finally reach an optimised adaption of the original input tensor (see Fig. 14 left part) and lead to an optimisation of the overall predicted delayed cases compared to the historical Monday average (see Fig. 14 right part). For this purpose the package "geneticalgorithm" (Solgi 2020) in python was used. Further used python libraries are "PyTorch" (Paszke et al. 2019) and "NumPy" (Harris 2020).

	Day of the week	Hour	As-planned number of transports in same hour	Amount of trans-porters in same as-is hour	Automatic rate according to as-planned hour
tensor([[	1.0000,	0.0000,	1.5165,	0.8242,	1.0000],
[	1.0000,	1.0000,	1.2527,	0.6813,	1.0000],
[	1.0000,	2.0000,	0.8846,	0.5989,	1.0000],
[	1.0000,	3.0000,	0.8022,	0.5714,	1.0000],
[	1.0000,	4.0000,	0.6264,	0.4121,	1.0000],
[	1.0000,	5.0000,	0.6154,	0.4835,	0.9578],
[	1.0000,	6.0000,	3.3626,	1.1209,	0.2268],
[	1.0000,	7.0000,	27.5714,	6.5549,	0.1566],
[	1.0000,	8.0000,	28.3242,	6.6374,	0.1405],
[	1.0000,	9.0000,	30.6978,	6.7363,	0.1538],
[	1.0000,	10.0000,	32.4011,	6.8956,	0.1475],
[	1.0000,	11.0000,	29.2253,	6.9780,	0.1511],
[	1.0000,	12.0000,	31.0385,	6.8846,	0.1556],
[	1.0000,	13.0000,	26.9890,	7.6538,	0.1695],
[	1.0000,	14.0000,	18.5549,	7.2747,	0.2985],
[	1.0000,	15.0000,	9.6538,	2.3791,	0.7746],
[	1.0000,	16.0000,	5.9341,	1.9396,	0.9728],
[	1.0000,	17.0000,	4.0879,	1.6978,	1.0000],
[	1.0000,	18.0000,	3.4451,	1.5330,	1.0000],
[	1.0000,	19.0000,	3.0220,	1.5714,	1.0000],
[	1.0000,	20.0000,	2.7198,	1.3846,	0.9994],
[	1.0000,	21.0000,	2.5275,	0.9231,	1.0000],
[	1.0000,	22.0000,	2.5055,	1.0000,	0.9940],
[	1.0000,	23.0000,	2.1264,	0.8407,	1.0000]])

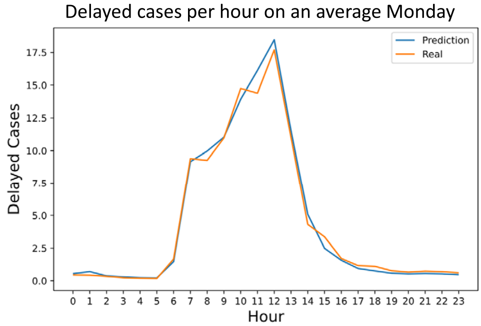

Fig. 14 Left: original input tensor with mean values for an average Monday. Right: real and predicted (through MLP model) delayed cases per hour on an average Monday

Data visualisations are conducted using the python library "Matplotlib" (Hunter 2007). During the development of the optimisation procedure involving the GA, *ChatGPT* (versions 3.5 and 4) was used to search for and explain certain methods of the used libraries or python functionalities.

4.4.1 Optimising the number of transporters per hour

Initially, the aim is to optimise the number of delayed cases, i.e. minimise them, by simply adjusting the number of transporters per hour. To run the GA an objective function is defined. This function takes an adapted input tensor and calculates the objective value according to equation 12. The objective value consists of two components: The first part in equation 12 is defined by the sum of the MLP predictions (these reflect the sum of the delayed cases per day of the week, here Monday). Secondly there is a constraint part in the objective function, wherein there are constraint penalty terms defined according to other desired goals, that are to be reached in addition to minimising the delayed cases.

To optimise the number of transporters per hour two constraint penalties are defined. The first constraint penalty is determined by computing the difference between the sum of the values in the fourth column of the adapted tensor and the sum of the values in the fourth column of the original tensor (see equation 13). This ensures that the sum of the number of transporters per hour, which are adjusted during the optimisation process, does not deviate greatly upwards from the original tensor. If the difference is negative (sum of the number of transporters per hour after optimisation is less than before), the penalty is set to zero. If fewer transporters are in operation in total and the delayed cases are still minimised, this is considered good, but the constraint penalty should not decrease the objective value as minimising the delayed transports is the main goal. Second, there is a constraint penalty if negative delayed cases are predicted by the model (see equation 14).

The meaning of the other variables within the objective function and the constraint penalties are defined shown in equations 15 - 21.

$$\text{objective1} = \sum_{i=1}^{m} MLP(x_i) + w_1 \cdot CP_1 + w_2 \cdot CP_2 \tag{12}$$

where,

$$CP_1 = \text{ConstraintPenalty1} = max\left(\left(\sum_{i=1}^{m}(y_i^4) - \sum_{i=1}^{m}(z_i^4)\right), 0.0\right) \tag{13}$$

$$CP_2 = \text{ConstraintPenalty2} = -\left(\sum_{i=1}^{m}(min(MLP(x_i), 0.0))\right) \tag{14}$$

with

$$m = \text{overall no. of rows in the adapted and original tensor} \tag{15}$$

$$MLP = \text{Multilayer Perceptron (predicts delayed cases)} \tag{16}$$

$$x_i = \text{row i of the adapted tensor} \tag{17}$$

$$w_1 = \text{weighting factor of } CP_1 \tag{18}$$

$$w_2 = \text{weighting factor of } CP_2 \tag{19}$$

$$y_i^4 = \text{element i of column 4 of the adapted tensor} \tag{20}$$

$$z_i^4 = \text{element i of column 4 of the original tensor} \tag{21}$$

m is 24 in our case, since there is one row for all hours from 0 to 23 in both the adapted input tensor and the original tensor. w_1 is set to 3.5 and w_2 is set to 150. These weighting factors are chosen subjectively to reflect the authors' view of the importance of the constraint penalties. The event log showed, that on average, it took a transporter 20 min from the acceptance of a transport assignment to transport completion. Thus, around three transports per hour can be fulfilled on average per transporter. A weighting factor w_1 of 3.5 expresses that one more transporter can be added in the capacity planning if this leads to at least 3.5 less delayed transports. This is more than what would be possible by conservatively assuming that all three average feasible transports would be completed on time by one more transporter in that hour. Thus, in terms of the objective function, the GA will only be able to identify a better solution if an additional transporter reduces more than 3.5 delayed transports in one hour (see CP_1 in equation 13). With a value of 150, w_2 is set large enough to ensure that no delayed cases are predicted by the model (see CP_2 in equation 14).

After the objective function is set, the 24 elements in the fourth column of the Monday tensor (see Fig. 14) will be considered as integer variables, that are to be optimised by the GA. Possible integer values are set to be between one and 12, as on average the active transporters per hour on Mondays were in all hours under eight transporters and furthermore for the optimisation it is to be considered that in every hour there should be at least one active transporter. The parameters of the GA are set through trial-and-error to the following and the procedure of the GA is presented in algorithm 1:

- 'max_num_iteration': 2,000,
- 'population_size': 10,000,
- 'mutation_probability': 0.1,
- 'elit_ratio': 0.01,
- 'crossover_probability': 0.5,
- 'parents_portion': 0.3,
- 'crossover_type': 'uniform',
- 'max_iteration_without_improv': None.

Algorithm 1 GA to minimize the objective value

```
Initialize population with 10,000 random individuals
while iteration ≤ 2,000 do
    Evaluate objective value of each individual
    Sort individuals by objective value in ascending order
    Select the top 1% of individuals as elites
    Select 30% of the population as parents
    Perform uniform crossover (see documentation of python package "geneticalgorithm" [Solgi 2020]) on selected parents to produce offspring
    Apply mutation to offspring with a probability of 10%
    Combine elites, parents, and offspring into new population
    Increment iteration counter
end while
```

At each iteration, the objective value is evaluated based on the best found input tensor. Overall, the GA aims to optimise the objective function, which includes both the MLP model predictions and constraint penalties. The optimisation process iteratively updates the input tensor to minimise the objective value by adjusting the tensor values within the specified constraints.

Figure 15 (left part) shows the finally derived input tensor that leads to an optimisation. The MLP model predicts with the new tensor a sum of delayed cases for

Day of the week	Hour	As-planned number of transports in same hour	Amount of trans-porters in same as-is hour	Automatic rate according to as-planned hour
tensor([[1.0000,	0.0000,	1.5165,	1.0000,	1.0000],
[1.0000,	1.0000,	1.2527,	1.0000,	1.0000],
[1.0000,	2.0000,	0.8846,	1.0000,	1.0000],
[1.0000,	3.0000,	0.8022,	1.0000,	1.0000],
[1.0000,	4.0000,	0.6264,	1.0000,	1.0000],
[1.0000,	5.0000,	0.6154,	1.0000,	0.9578],
[1.0000,	6.0000,	3.3626,	2.0000,	0.2268],
[1.0000,	7.0000,	27.5714,	2.0000,	0.1566],
[1.0000,	8.0000,	28.3242,	2.0000,	0.1405],
[1.0000,	9.0000,	30.6978,	2.0000,	0.1538],
[1.0000,	10.0000,	32.4011,	2.0000,	0.1475],
[1.0000,	11.0000,	29.2253,	12.0000,	0.1511],
[1.0000,	12.0000,	31.0385,	12.0000,	0.1556],
[1.0000,	13.0000,	26.9890,	11.0000,	0.1695],
[1.0000,	14.0000,	18.5549,	8.0000,	0.2985],
[1.0000,	15.0000,	9.6538,	3.0000,	0.7746],
[1.0000,	16.0000,	5.9341,	4.0000,	0.9728],
[1.0000,	17.0000,	4.0879,	1.0000,	1.0000],
[1.0000,	18.0000,	3.4451,	1.0000,	1.0000],
[1.0000,	19.0000,	3.0220,	1.0000,	1.0000],
[1.0000,	20.0000,	2.7198,	1.0000,	0.9994],
[1.0000,	21.0000,	2.5275,	1.0000,	1.0000],
[1.0000,	22.0000,	2.5055,	1.0000,	0.9940],
[1.0000,	23.0000,	2.1264,	1.0000,	1.0000]])

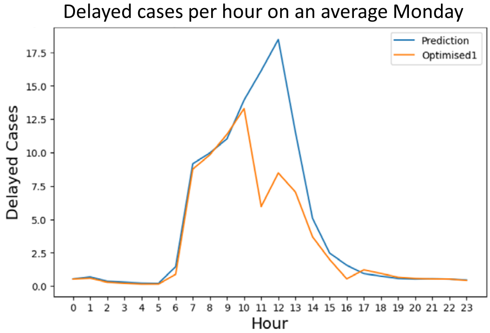

Fig. 15 Left: best found solution for the input tensor with adjusted number of Transporters per hour (see Fig. 14 for comparison to the original tensor for an average Monday) leading to less delayed cases. Right: predicted (through MLP model) delayed cases per hour on an average Monday before and after input tensor optimisation (see left) through the GA

a Monday of around 79 delayed cases. Before optimisation an average Monday was predicted to have around 108 delayed cases. At the same time, the 73.6 transporter hours were merely redistributed so that the optimised tensor now contains a total of 73 transporter hours (due to the requirement to use only integer transporter numbers per hour). Figure 15 (right part) shows the predicted (through MLP model) delayed cases per hour on an average Monday before and after input tensor optimisation through the GA. As the sum of transporters, as wanted, stays quite the same, the sum of delayed transports as well as the objective function value are minimised. In summary, it is theoretically possible to achieve around 27% less delayed transports merely by reallocating transport capacities.

4.4.2 Further resource optimisations

In addition to the transporters per hour, other parameters can also be optimised. Therefore, another attempt is made to adjust the input tensor in order to minimise delayed cases.

In addition to the number of transporters per hour, now also the automation rate may be adjusted every hour, as well as the number of planned transports per hour. Regarding automation, the rate should generally always be either exactly 1 or exactly 0 after being adapted. This is based on the findings from preliminary analyses, that showed a mixed operation is inefficient.

The total number of transports to be planned on an average Monday should not be lower than the number without tensor adjustment. During the observation period, approximately 270 transports were planned on an average Monday. This boundary is incorporated in a third constraint penalty enhancing the two mentioned constraint penalties of Subsect. 4.4.1. The new objective function, where the GA aims to optimise three columns of the original tensor, is now given in equation 22. The third constraint penalty is defined according to equation 23. The meaning of the other variables within the objective function and the constraint penalties are defined shown in equations 24 - 26. If the sum of the number of planned transports per hour after optimisation is higher than before, the penalty is set to zero (see equation 23). If more transports are planned in total and the delayed cases are still minimised, this is considered good, but the constraint penalty should not decrease the objective value as minimising the delayed transports is the main goal. The weighting w_3 of the third constraint penalty is subjectively set to 3.5 according to the authors' view on the importance of the constraint penalty. This means that if one less transport was to be planned per day, this must result in at least 3.5 fewer delayed cases. This is about ten times higher than the average delayed case decrease of almost 0.35 (due to an average delay rate of almost 35% over all transport cases, see Subsect. 4.1) a transport conducted less would lead to. Thus, the GA can understand that fewer transports to be planned through tensor adjustment are only valid, if this results in a significant (tenfold) reduction of delayed cases and thus decrease of the objective value. The other variables to be set stay as in Subsect. 4.4.1.

$$\text{objective3} = \sum_{i=1}^{m} MLP(x_i) + w_1 \cdot CP_1 + w_2 \cdot CP_2 + w_3 \cdot CP_3 \tag{22}$$

where in addition to equations 13 - 21 from Subsect. 4.4.1,

$$CP_3 = \text{Constraint Penalty3} = max\left(\left(\sum_{i=1}^{m}(z_i^3) - \sum_{i=1}^{m}(y_i^3)\right), 0.0\right) \tag{23}$$

with

$$w_3 = \text{weighting factor of } CP_3 \tag{24}$$

$$z_i^3 = \text{element i of column 3 of the original tensor} \tag{25}$$

$$y_i^3 = \text{element i of column 3 of the adapted tensor} \tag{26}$$

Furthermore only adjustments in the number of planned transports by the GA will be allowed between the core operating hours of the hospital (between hours 7 and 17). Due to the low expected number of transport assignments outside the core operating hours, adjustments have been excluded in these periods and the values just stay as the historical average Monday values. We defined the range for the adjusted number of planned transports in the core operating hours to possibly include all integers from zero to 50. Up to about 32 transports per hour were planned for Mondays on average (see Fig. 14 at hour 10). Furthermore, in the historical dataset, just a few samples were found, that included over 50 planned transports (the maximum was 59 planned transports) in a respective hour. So, the range between zero and 50 was considered meaningful to provide the GA with a sufficient degree of freedom for optimisation.

The parameters of the GA stay the same as in Subsect. 4.4.1. Figure 16 (left part) shows the finally derived input tensor that leads to an optimisation with adaptions in columns three to five in comparison to the original tensor. Figure 16 (right part) shows the predicted (through MLP model) delayed cases per hour on an average Monday before and after input tensor optimisation through the GA. While the number of planned transports stays at approximately 270 and the number of transporters is at 73, the MLP predicts with the new tensor a sum of delayed cases for a Monday of around 63 delayed cases. This reflects a reduction of around 42% in delayed cases compared to 108 predicted delayed cases on the historical average Monday (see Fig. 14) by reallocating the number of transporters per hour, the number of planned transports per hour and the delay rate per hour.

4.4.3 GA validation

A limited solution space is created to validate the functionality of the GA. The more degrees of freedom the GA has, the more exponentially increasing solutions there will be in general. Thus we aim to evaluate the performance of the GA with a problem, where there is a limited, comprehensible number of solutions. In total 100 final solutions will be generated by the GA (following the complete procedure from algorithm 1, see Subsect. 4.4.1) and it is analysed how these compare to all possible solutions in the limited solution space.

The assumption is, that on an average Monday it is only possible to adjust the transporters between the full hours of seven and 16 o'clock (i.e. a total of ten hourly slots). A maximum of 60 transporters (73.6 transporter hours on an

Day of the week	Hour	As-planned number of transports in same hour	Amount of trans-porters in same as-is hour	Automatic rate according to as-planned hour

```
tensor([[ 1.0000,  0.0000,  1.5165,  2.0000, 1.0000],
        [ 1.0000,  1.0000,  1.2527,  2.0000, 1.0000],
        [ 1.0000,  2.0000,  0.8846,  1.0000, 1.0000],
        [ 1.0000,  3.0000,  0.8022,  1.0000, 1.0000],
        [ 1.0000,  4.0000,  0.6264,  1.0000, 1.0000],
        [ 1.0000,  5.0000,  0.6154,  1.0000, 1.0000],
        [ 1.0000,  6.0000,  3.3626,  3.0000, 0.0000],
        [ 1.0000,  7.0000, 12.0000,  7.0000, 1.0000],
        [ 1.0000,  8.0000, 49.0000, 12.0000, 1.0000],
        [ 1.0000,  9.0000, 42.0000,  3.0000, 0.0000],
        [ 1.0000, 10.0000, 44.0000,  3.0000, 0.0000],
        [ 1.0000, 11.0000,  8.0000,  2.0000, 1.0000],
        [ 1.0000, 12.0000,  7.0000,  3.0000, 1.0000],
        [ 1.0000, 13.0000, 50.0000,  3.0000, 0.0000],
        [ 1.0000, 14.0000,  6.0000,  5.0000, 0.0000],
        [ 1.0000, 15.0000, 14.0000,  7.0000, 0.0000],
        [ 1.0000, 16.0000,  9.0000,  4.0000, 0.0000],
        [ 1.0000, 17.0000,  4.0000,  3.0000, 1.0000],
        [ 1.0000, 18.0000,  3.4451,  2.0000, 0.0000],
        [ 1.0000, 19.0000,  3.0220,  2.0000, 1.0000],
        [ 1.0000, 20.0000,  2.7198,  1.0000, 1.0000],
        [ 1.0000, 21.0000,  2.5275,  2.0000, 1.0000],
        [ 1.0000, 22.0000,  2.5055,  2.0000, 1.0000],
        [ 1.0000, 23.0000,  2.1264,  1.0000, 1.0000]])
```

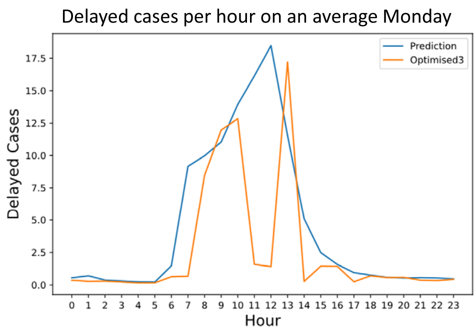

Fig. 16 Left: best found solution for the input tensor with adjusted number of Transporters per hour, number of planned transports per hour and delay rate per hour (see Fig. 14 for comparison to the original tensor for an average Monday) leading to less delayed cases. Right: predicted (through MLP model) delayed cases per hour on an average Monday before and after input tensor optimisation (see left) through the GA

average Monday minus 13.7 transporter hours from the non-adjustable hour slots of an average Monday equals 59.9 transporter hours) should then be distributed across ten specific chosen slots out of the 24 h on a day. Each slot should have at least one transporter and a maximum of 12 transporters (possible integer values in the GA are set to be between one and 12 like in Subsect. 4.4.1). In addition, the number of transporters from one of these ten slots to the next slot should not differ by more than two transporters, for example:

- A slot that starts with one transporter can have one, two or three transporters in the next slot.
- A slot that starts with five transporters can have three, four, five, six or seven transporters in the next slot, and so on.

An objective function that penalises not meeting these requirements is shown in equation 27. In contrast to the objective function from equation 12, see Subsect. 4.4.1, a new penalty CP_4 (or “ConstraintPenalty4”, see equation 27 and equation 28) is added. Figure 17 visualises the problem and main conditions that the GA is facing.

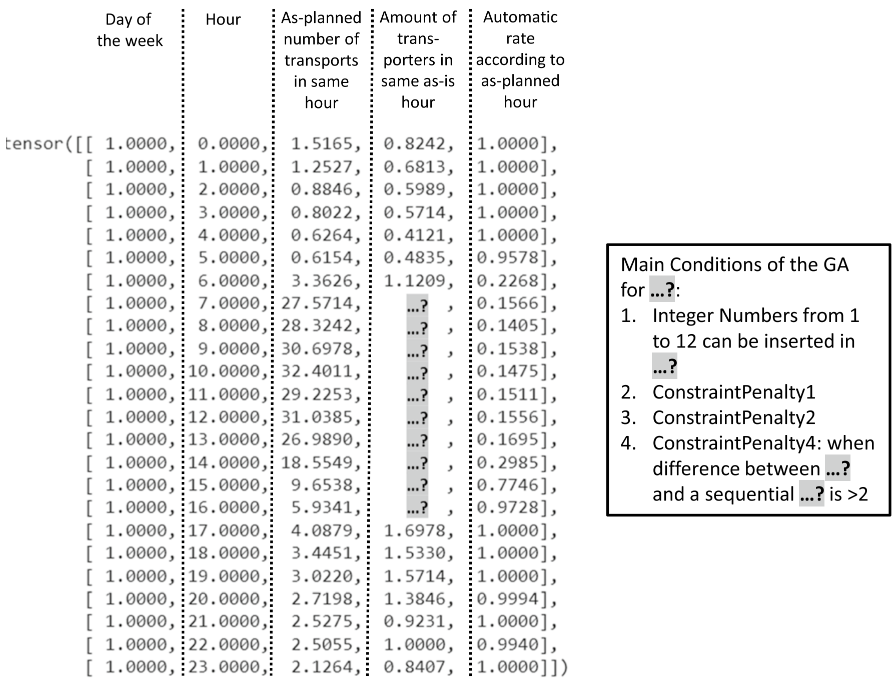

Fig. 17 Problem with limited solutions, that the GA aims to solve. The goal is to minimise the objective function

$$\text{objective\_limited} = \sum_{i=1}^{m} MLP(x_i) + w_1 \cdot CP_1 + w_2 \cdot CP_2 + w_4 \cdot CP_4 \tag{27}$$

where in addition to equations 13 - 21 from Subsect. 4.4.1,

$$CP_4 = \text{Constraint Penalty4} = \sum_{i=8}^{16} \begin{cases} \left| y_i^4 - y_{i+1}^4 \right|, & \text{if } \left| y_i^4 - y_{i+1}^4 \right| > 2 \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

with

$$w_4 = \text{weighting factor of } CP_4 \tag{29}$$

The weighting w_4 of the new constraint penalty CP_4 is subjectively set to 150, to ensure that the GA does not develop solutions, that allow next hourly time slots to differ by more than two transporters and thus lie outside the selected limited solution space. The weightings w_1 and w_2 stay as defined in Subsect. 4.4.1.

Firstly, a specially created recursive algorithm is used to try all possible numbers of transporters that fulfil the conditions for each slot. The final result, how many solutions there are to distribute a maximum of 60 transporters to ten slots, whereby the distribution fulfils the specified conditions, is 5,589,997 solutions.

If these 5,589,997 solutions are assessed with the objective function (see equation 27), the following values of the objective function are obtained:

- min 105.0527,
- max 396.6024,
- average 180.8667,
- median 143.2295.

For the solutions in which the total number of transporters is 60, which is also the maximum possible, the objective function is always slightly higher than the number of delayed transport cases. This is because, according to equation 27, more than the original 59.9 transporters are penalised in a solution that is evaluated through the objective function. In these instances the objective function is greater than the number of predicted delayed cases by 0.23 in absolute numbers.

It should be noted that only 69 of these 5,589,997 solutions lead to equal or better objective function values than the average Monday with an objective function value of 107.5983. The fact that there are so few solutions that lead to any improvement at all is partly due to that only integer numbers are allowed, and partly due that the difference between two consecutive hours in the transporter numbers must not be greater than two, which severely limits the possible improving solutions. The best solution with an objective value of 105.0527 is visualised in Figure 18.

Now the GA is used to find an optimised solution for the problem. To evaluate the GA performance, the complete GA procedure (see algorithm 1 from Subsect. 4.4.1) is repeated 100 times and then all generated optimised solutions are compared to the 5,589,997 possible solutions. In contrast to the parameters of the GA from before (see Subsect. 4.4.1 and 4.4.2), the parameter "population_size"

Day of the week	Hour	As-planned number of transports in same hour	Amount of trans-porters in same as-is hour	Automatic rate according to as-planned hour
1.0000	0.0000	1.5165	0.8242	1.0000
1.0000	1.0000	1.2527	0.6813	1.0000
1.0000	2.0000	0.8846	0.5989	1.0000
1.0000	3.0000	0.8022	0.5714	1.0000
1.0000	4.0000	0.6264	0.4121	1.0000
1.0000	5.0000	0.6154	0.4835	0.9578
1.0000	6.0000	3.3626	1.1209	0.2268
1.0000	7.0000	27.5714	2.0000	0.1566
1.0000	8.0000	28.3242	2.0000	0.1405
1.0000	9.0000	30.6978	4.0000	0.1538
1.0000	10.0000	32.4011	6.0000	0.1475
1.0000	11.0000	29.2253	8.0000	0.1511
1.0000	12.0000	31.0385	10.0000	0.1556
1.0000	13.0000	26.9890	10.0000	0.1695
1.0000	14.0000	18.5549	8.0000	0.2985
1.0000	15.0000	9.6538	6.0000	0.7746
1.0000	16.0000	5.9341	4.0000	0.9728
1.0000	17.0000	4.0879	1.6978	1.0000
1.0000	18.0000	3.4451	1.5330	1.0000
1.0000	19.0000	3.0220	1.5714	1.0000
1.0000	20.0000	2.7198	1.3846	0.9994
1.0000	21.0000	2.5275	0.9231	1.0000
1.0000	22.0000	2.5055	1.0000	0.9940
1.0000	23.0000	2.1264	0.8407	1.0000

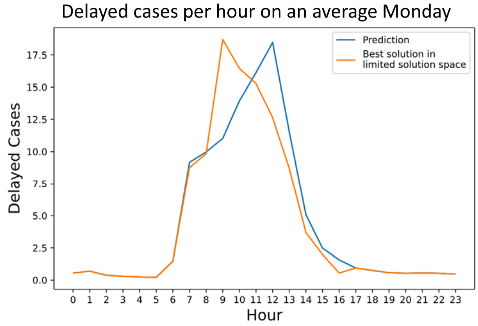

Fig. 18 Left: Input tensor with mean values for an average Monday with adapted no. of transporters in ten hours (best solution in limited solution space). Right: Predicted (through MLP model) delayed cases per hour on an average Monday before and after inserting the best input tensor (best solution in limited solution space, see left)

is decreased from 10,000 to ten to accelerate the 100 independent GA runs, each of which goes through 2,000 iterations. A total of 66 unique final solutions were generated by the GA within in the 100 final solutions.

In the following the aggregated results of the objective function value of the 100 final GA solutions are presented:

- min 106.5049,
- max 117.3433,
- average 109.1974,
- median 108.1439.

In 33 out of the 100 runs the GA led to improvements in the objective function. In the other 67 runs the GA did not lead to an improvement in the objective function. However, as mentioned before, there are rarely improvement solutions in the chosen limited solution space. But still, due to the following reasons the GA performs in a good manner. With a minimum objective function value of 106.5049 in the 100 runs, the GA was better than 99.99% of all possible solutions. With an average objective function value of 109.2 in the 100 runs, the GA was better than 99.96% of all possible solutions. Even with a maximal objective function value of 117.3433 in the 100 runs, the GA was better than 96.94% of all possible solutions. This illustrates the effective functionality of the GA for the limited solution space. At this point, the functionality of the GA is considered validated, although the performance may deteriorate with a different and, above all, larger solution space. However, with an almost infinite number of solutions, which is possible depending on the considered problem, the validation is extremely complex and time-consuming, which will not be pursued in greater depth at this point. The main purpose of this work is to demonstrate the general mode of operation for combining the MLP model and the GA for optimised IHPT capacity planning, and this is hereby confirmed.

5 Discussions

Table 6 summarises the results of the optimisation procedure. In addition to the mean real delayed cases on an average Monday, the predictions of the MLP for an average Monday are depicted. Furthermore, the improvements with respect to the delayed cases are presented for a sole adjustment of transporters (see also Subsect. 4.4.1) as well as for an adjustment involving transporters, planned transports, and automation rates (see also Subsect. 4.4.2).

It is noticeable that the GA reallocates the variables (number of transporters, number of planned transports and automation rate per hour) in such a way that many delayed cases result in a few hours of the day and few delays occur in the remaining hours (see the last two columns of Table 6). Also, there are partly larger jumps in hourly variables, which could be difficult to implement in practice. For example, Fig. 16 (left part) from Subsect. 4.4.2 shows that at hour 12 there should be seven transports planned, at hour 13 there should be 50 transports planned, at hour 14 there should be six transports planned and at hour 15 there should be 14 transports planned. Also the number of active transporters in the hours seven, eight and nine is reallocated to seven, then 12 and then three transporters in Fig. 16 (left part). When there are these significant variations in the needed hourly capacity of transporters, they may, however, be available for other tasks within the hospital during times when they are not engaged in IHPT. Also, whether the transport disposition from automatic (“1”) to manual (“0”) and vise versa can be switched on an hourly basis (see left part of Fig. 16) is questionable. In general, the GA can produce solutions in certain hours that are widely different from the behaviour that the MLP model could have encountered during model development with the historical IHPT data, and which may therefore lead to questionable predictions. From a practical view, further investigations on these scenarios need to be conducted. Nonetheless, the mentioned peaks and discrepancies can be leveled out by the implementation of further constraints in the objective functions, that the GA aims to optimise. However, this does require practical knowledge and would certainly compromise the optimisation potential.

The chosen example from Subsect. 4.4 can be critically viewed, too. Data from an average Monday in a data frame of 3.5 years were used for the original input tensor that is to be optimised. Instead, it is also possible to use an average Monday or other days of the week from more recent weeks. An average day of the week that is to be optimised from a more recent period could be more meaningful for current planning than an average day of the week calculated from the past 3.5 years. However, the initial task is to determine which historical periods are more reflective of current and future periods than the average long-term dataset, thus enabling even better optimisation approaches.

The results presented in Subsect. 4.4 varied more or less in each attempt with the GA. The GA identifies improvements, but it is impossible to ascertain that the

Table 6 Comparison of delayed cases depending on input data

Time		Delayed cases			
Day of the week	Hour	Mean real delayed cases	Predicted delayed cases with mean input data	Optimised delayed cases with adaption of transporters	Optimised delayed cases with adaption of transporters, planned cases and automation rate
1	0	0.4505	0.5477	0.5362	0.3597
1	1	0.4341	0.7026	0.6032	0.2814
1	2	0.3407	0.3759	0.2994	0.2994
1	3	0.2363	0.3023	0.2262	0.2262
1	4	0.1923	0.2384	0.1606	0.1606
1	5	0.1758	0.2214	0.1648	0.1610
1	6	1.6703	1.4700	0.8934	0.6324
1	7	9.3956	9.1645	8.7574	0.6712
1	8	9.2637	9.9757	9.8353	8.4722
1	9	10.9780	11.0323	11.3663	11.9613
1	10	14.7527	13.9322	13.2942	12.8311
1	11	14.3901	16.1261	5.9639	1.5842
1	12	17.6923	18.4657	8.4750	1.4016
1	13	11.0275	11.5347	7.0498	17.1950
1	14	4.3132	5.0831	3.7008	0.2763
1	15	3.3626	2.4829	2.0003	1.4365
1	16	1.7088	1.5726	0.5529	1.4202
1	17	1.1648	0.9449	1.2315	0.2467
1	18	1.1099	0.7564	0.9545	0.7011
1	19	0.7692	0.5794	0.6714	0.5669
1	20	0.6648	0.5321	0.5830	0.5830

Table 6 (continued)

Time		Delayed cases			
Day of the week	Hour	Mean real delayed cases	Predicted delayed cases with mean input data	Optimised delayed cases with adaption of transporters	Optimised delayed cases with adaption of transporters, planned cases and automation rate
1	21	0.7363	0.5570	0.5480	0.3688
1	22	0.6978	0.5344	0.5344	0.3362
1	23	0.6154	0.4659	0.4290	0.4290
Delayed cases per day		106.1429	107.5983	78.8316	62.6020
Planned cases per day		269.8846	269.8846	269.8846	270.4066
Transporters per day		73.5769	73.5769	73	73

optimal solution has been reached, as the GA works in a trial-and-error manner and can also converge to local optima. However, the presented use case involves a complex problem with multiple optimisation variables. Nevertheless, the GA has shown to be computational practical in achieving significant improvements in a complex environment characterised by an objective function containing an MLP model for an IHPT capacity planning problem. Subsection 4.4.3 furthermore proved for a smaller problem instance that the GA performed reliably good.

All in all, the investigations from Subsect. 4.4 have demonstrated, that under conditions of consistent or similar quantities of planned transport assignments and transporters, resource reallocation through a GA led to a theoretical reduction ranging from 27% to 42% in delayed cases for the example of an average Monday. In the same manner, the capacity planning for other days of the week can also be optimised. Comparing these potentials with the IHPT literature included in Sect. 1 and 2, although mostly different areas in IHPT are subject to improvement in literature, we can see that the theoretical improvement potentials of our approach are in the higher region. Haldar et al. (2019), Jaroon (2018), Séguin et al. (2019), Kergosien et al. (2011), Fröhlich Von Elmbach et al. (2019), Fiegl and Pontow (2009), Bärmann et al. (2024) reached average IHPT improvements of up to 20%, (Gopal 2016; Turan et al. 2011; Elmbach et al. 2015; Vancroonenburg et al. 2016; Hanne et al. 2009; Meephu et al. 2023) of up to between 21% and 36%. The transferability of the potential of our approach to the real world is given by the reached performance of the MLP model (see Subsect. 4.3). For average day of the week scenarios and concerning the R^2, the results are in an acceptable region (see Subsect. 4.3.2). Precision, recall and F1 score from Subsect. 4.3.2 (and thus indirectly the *MSE* from 4.3.1) are considered not yet good enough for a practical implementation of the MLP model by hospital process managers. Different approaches to improving the performance of MLP models can be, for example, a model development with more data, but also further filtering of outliers or the development of several specialised prediction models. One model for core hospital working hours (during the day from Monday to Friday) and another model for other hours, which are then trained, validated and tested with different datasets.

Also, depending on the chosen weighting factors of the constraint penalties in the objective functions and the possible variable ranges that the GA can choose from, as defined in Sect. 4.4, the improvement potentials revealed by the GA approach can be different. These parameters need to be further discussed and possibly adapted with domain experts, also by performing a detailed quantitative sensitivity analysis, in order to obtain meaningful improvement ideas for IHPT capacity planning. Furthermore, the effects of a potential implementation in practice should be thoroughly evaluated and compared.

In terms of generalisability and transferability, the preprocessing (see Subsect. 4.2) can be similarly carried out at other hospitals without major effort. In total, more than 370 hospitals in Germany, Austria, Switzerland, Italy and Sweden use the software from which the datasets were extracted within this study. However, the MLP input and label information (see figure 7 in Subsect. 4.2) used here can certainly also be extracted and aggregated from other software systems, as long as similar information is stored. The MLP training, validation and testing (see Subsect. 4.3) must then be carried out on an individual basis in order to develop an accurately functioning MLP model for other hospital environments. Parameters must be set and evaluated again like conducted in Subsect. 4.3.1 and 4.3.2. The GA (see Subsect. 4.4) will also be applicable to the newly designed ANN. But its functionality should be confirmed again, similar to Subsect. 4.4.3. Furthermore, re-calibrating the setup and boundaries of adaptable variables, as well as weights and constraint penalties in the objective functions to be optimised, is needed.

In summary, the approach presented in Sect. 4 can be adapted to different clinical and, more generally, logistical or other environments where capacity planning is required. Often only certain parameters need to be adapted. However, the overall concept, namely the combination of data preparation, process prediction by an ANN and subsequent optimisation by a metaheuristic algorithm to find improved solutions for processes, can be quickly applied.

6 Conclusion and future work

This paper has demonstrated the feasibility of establishing a more efficient resource allocation in capacity planning using historical IHPT data. To achieve this, we aggregated and preprocessed transport-related raw data (information on 256,266 transport cases) into hourly-based data (information on 25,662 h) within an observation period of approximately 3.5 years from January 2019 to mid 2022 (covers actually 30,672 h, but hours with missing data were filtered out). Therefore the number of planned transports, the number of active transporters, and the automation rate of transport disposition as well as the number of delayed transports for each hour in the observation period are calculated to train a MLP model. While certain performance metrics of the developed MLP model are within a range, that needs to be further improved for practical implementation, mainly by increasing the volume of training data, predictions for hourly delayed cases are already feasible for individual days of the week. By incorporating data from additional observation periods, it might be possible in the future to achieve predictions that are specific to calendar weeks, months, or even specific days in a year. More specific predictions require a more extensive data availability so that respective observations can be sufficiently trained into the model.

Using a GA, the capacity resources of an average Monday were adjusted in a manner that led to a reduction in daily delayed cases, while maintaining a consistent sum of planned transports and aggregated number of transporters per hour on a day. To achieve this, firstly just the transporters were redistributed throughout the day. In a second attempt next to a transporter redistribution also adjustments were made to the distribution of planned transports and the automation rate of transport assignment disposition. We have already achieved theoretical improvements with 27% to 42% fewer delayed cases in IHPT just by reorganising resources, without adding new ones. Here, the combination of a MLP model that can predict delayed cases when given specific IHPT process information and a GA that can adapt the input information to minimise the predicted delayed cases has proven effective. However, there is potential for further improvement both in the performance of the ANN and the use of an appropriate metaheuristic algorithm to reliably find the best solution for capacity planning. The practical implementation of the proposed planning is pending. As mentioned in Subsect. 2.3, there are no predictive performance metrics of IHPT processes in literature, that we could compare our MLP model performance to. But our predictive performance metrics for delayed cases in IHPT will serve as a benchmark for further research. Moreover, we will continue to develop and compare other ANN models for the IHPT process. Filtering the training data when developing the MLP model can help to create models that are more accurate in certain situations (e.g. when the number of as-planned transports per hour is relatively high). Initial approaches using RNN and CNN architectures on our data (beyond the investigations of this paper) have also shown slightly improved *MSE* values, but need to be further investigated and compared. Metaheuristic algorithms other than GA, such as swarm-intelligence algorithms (see Subsect. 3.2), should also be used to improve the reorganisation of IHPT capacity planning and their performance is to be evaluated and compared. If in the future additional hourly attributes of the IHPT process, like e.g. transport priorities, pick-up or arrival locations, are incorporated into the MLP model training, along with the in our use case incorporated hourly information, the MLP model could generate more specific predictions and thus, offering more variables for adjustment within optimisation of delayed cases through the GA. Possible other attributes, that could be potentially used for the training of a MLP as well as solving resource optimisation problems can be found in Subsect. 4.1. However, a more extensive data collection is also here a prerequisite, ensuring that specific information is recurrently present throughout the MLP training process to allow for accurate predictions. Last but not least, the overall improvement potential for the processes under investigation will certainly vary from one application to another and should be examined and compared in a broader scale in order to draw reliable, general conclusions.

Funding Open Access funding enabled and organized by Projekt DEAL. Partial financial support was received from the use case hospital that will be held anonymous within this paper.

Data availability Research data are not shared.

Declarations

Conflict of interest The authors have no Conflict of interest to declare that are relevant to the content of this article.

References

Abiodun OI et al. (2019) Comprehensive review of artificial neural network applications to pattern recognition. IEEE Access 7:158820–158846. https://doi.org/10.1109/ACCESS.2019.2945545

Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H (2018) State-of-the-art in artificial neural network applications: A survey. Heliyon 4(11):e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Abohamad W, Ramy A, Arisha A (2017) A hybrid process-mining approach for simulation modeling. In: WSC'17. WK Chan, A D'Ambrogio, G Zacharewicz, N Mustafee, G Wainer, and EH Page (ed.). IEEE, Piscataway, NJ. pp. 1527–1538. ISBN: 978-1-5386-3428-8. https://doi.org/10.1109/WSC.2017.8247894

Agostinelli S, Covino F, D'Agnese G, de Crea C, Leotta F, Marrella A (2020) Supporting governance in healthcare through process mining: a case study. IEEE Access 8:186012–186025. https://doi.org/10.1109/access.2020.3030318

Aguirre S, Parra C, Sepúlveda M (2017) Methodological proposal for process mining projects. Int J Bus Process Integr Manag 8(2):102. https://doi.org/10.1504/IJBPIM.2017.083793. (**ISSN: 1741-8763**)

Akbari V, Sadati İ, Salman FS, Shiri D (2023) Minimizing total weighted latency in home healthcare routing and scheduling with patient prioritization. OR Spectr 45:807–852. https://doi.org/10.1007/s00291-023-00713-3. (**ISSN: 0171-6468**)

Al-Waeli AH, Kazem HA, Yousif JH, Chaichan MT, Sopian K (2019) Mathematical and neural network models for predicting the electrical performance of a PV/T system. Int J Energy Res. https://doi.org/10.1002/er.4807

AlDahoul N et al. (2021) Suspended sediment load prediction using long short-term memory neural network. Sci Rep 11(1):7826. https://doi.org/10.1038/s41598-021-87415-4. (**ISSN: 2045-2322**)

Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J (2013) Artificial neural networks in medical diagnosis. J Appl Biomed 11(2):47–58. https://doi.org/10.2478/v10136-012-0031-x. (**ISSN: 1214-021X**)

Andrews R, Wynn MT, Vallmuur K, ter Hofstede AHM, Bosley E (2020) A comparative process mining analysis of road trauma patient pathways. Int J Environ Res Public Health. https://doi.org/10.3390/ijerph17103426

Antunes BBP, Manresa A, Bastos LSL, Marchesi JF, Hamacher S (2019) A solution framework based on process mining, optimization, and discrete-event simulation to improve queue performance in an emergency department. In: Di Francescomarino C, Dijkman R, Zdun U (eds.) Business process management workshops. Vol. 362. Springer eBook Collection. Springer International Publishing and Imprint Springer, Cham. pp. 583–594. https://doi.org/10.1007/978-3-030-37453-2_47 ISBN: 978-3-030-37452-5

Badakhshan P, Alibabaei A (2020) Using process mining for process analysis improvement in pre-hospital emergency. In: ICT for an Inclusive World, Baghdadi Y, Harfouche A, Musso M (eds.). Vol. 35. Lecture notes in information systems and organisation. Springer International Publishing and Imprint Springer, Cham. pp. 567–580.https://doi.org/10.1007/978-3-030-34269-2_39

Barrett T, et al.. (2024) data.table: extension of 'data.frame'. R package version 1.14.6, https://Rdatatable.gitlab.io/data.table, https://github.com/Rdatatable/data.table. https://r-datatable.com

Beaudry A, Laporte G, Melo T, Nickel S (2010) Dynamic transportation of patients in hospitals. OR Spectr 32:77–107. https://doi.org/10.1007/s00291-008-0135-6. (**ISSN: 0171-6468**)

Beckmann U, Gillies DM, Berenholtz SM, Wu AW, Pronovost P (2004) Incidents relating to the intra-hospital transfer of critically ill patients. Intensiv Care Med 30(8):1579–1585. https://doi.org/10.1007/s00134-004-2177-9

Blum C, Roli A (2003) Metaheuristics in combinatorial optimization. ACM Comput Surv 35(3):268–308. https://doi.org/10.1145/937503.937505. (**ISSN: 0360-0300**)

Bonabeau E, Dorigo M, Theraulaz G (1999) Swarm intelligence: from natural to artificial systems. Santa Fe Institute studies in the sciences of complexity. Oxford University Press, New York, NY. ISBN: 9780195131581. https://permalink.obvsg.at/AC02609731

Bouabdallah MN, Fondrevelle J, Rached M, Bahroun Z (2013) Organization and management of hospital patient transportation system. In: 2013 international conference on control, decision and information technologies (CoDIT 2013). IEEE, Piscataway, NJ, pp. 125–130. https://doi.org/10.1109/CoDIT.2013.6689531

Bärmann A, Martin A, Müller A, Weninger D (2024) A column generation approach for the lexicographic optimization of intra-hospital transports. OR Spectr. https://doi.org/10.1007/s00291-023-00741-z

Camargo M, Dumas M, González-Rojas O (2021) Discovering generative models from event logs: data-driven simulation vs deep learning. PeerJ Comput Sci 7:577. https://doi.org/10.7717/peerj-cs.577

Canjels KF, Imkamp MSV, Boymans TAEJ, Vanwersch RJB (2021) Improving the arthrosis care process at maastricht umc+: unraveling complex and noncomplex cases by data and process mining. In: vom Brocke J, Mendling J, Rosemann M (eds) Business process management cases. Springer, Berlin and Heidelberg, pp 139–151. https://doi.org/10.1007/978-3-662-63047-1_11

Davis J, Goadrich M (2006) The relationship between Precision-Recall and ROC curves". In: Proceedings of the 23rd international conference on machine learning—ICML '06. ACM Press, Pittsburgh. pp. 233–240. https://doi.org/10.1145/1143844.1143874

Dumitru C, Maria V (2013) Advantages and disadvantages of using neural networks for predictions. Ovidius Univ Ann Ser Econ Sci 13(1):444–449. https://stec.univ-ovidius.ro/html/anale/ENG/cuprins%20rezumate/volum2013p1.pdf

Elmbach AFV, Boysen N, Briskorn D, Mothes S (2015) Scheduling pick-up and delivery jobs in a hospital to level ergonomic stress. In: IIE transactions on healthcare systems engineering. Vol. 5(1), pp. 42–53 https://doi.org/10.1080/19488300.2014.996837. http://www.tandfonline.com/doi/abs/10.1080/19488300.2014.996837

Fashoto SG, Akinnuwesi B, Owolabi O, Adelekan D (2016) Decision support model for supplier selection in healthcare service delivery using analytical hierarchy process and artificial neural network. Afr J Bus Manag 10(9):209–232. https://doi.org/10.5897/AJBM2016.8030

Fiegl C, Pontow C (2009) Online scheduling of pick-up and delivery tasks in hospitals. J Biomed Inform 42(4):624–632. https://doi.org/10.1016/j.jbi.2009.02.003

Flach PA (2003) The geometry of ROC space: understanding machine learning metrics through ROC isometrics. In: Proceedings of the 20th international conference on machine learning (ICML-03), pp. 194–201. https://aaai.org/papers/ICML03-028-the-geometry-of-roc-space-understanding-machine-learning-metrics-through-roc-isometrics/

Fröhlich Von Elmbach A, Scholl A, Walter R (2019) Minimizing the maximal ergonomic burden in intra-hospital patient transportation. Eur J Oper Res 276(3):840–854. https://doi.org/10.1016/j.ejor.2019.01.062

Gardner MW, Dorling S (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ 32(14–15):2627–2636

Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Boston (**ISBN: 0201157675**)

Goodfellow I (2016) Deep learning. In: Bengio Y, Courville A(ed.) Adaptive computation and machine learning. The MIT Press, Cambridge.https://zbmath.org/?q=an:1373.68009

Gopal K (2016) Modeling and optimization of hospital transportation system. PhD thesis. University of Akron. http://rave.ohiolink.edu/etdc/view?acc_num=akron1481314351566885

Grolemund G, Wickham H (2011) Dates and times made easy with lubridate. J Stat Softw 40(3):1–25. https://doi.org/10.18637/jss.v040.i03

Guneri AF, Gumus AT (2008) The usage of artificial neural networks for finite capacity planning. Int J Ind Eng Theory Appl Pract 15(1):16–25. https://doi.org/10.23055/IJIETAP.2008.15.1.58

Gul M, Guneri AF (2015) Forecasting patient length of stay in an emergency department by artificial neural networks. J Aeronaut Space Technol 8(2):43–48. https://doi.org/10.7603/s40690-015-0015-7

Haldar R, Gupta D, Pandey H, Srivastava S, Mishra P, Agarwal A (2019) Patient transportation delays and effects on operation theatres' efficiency: a study for problem analysis and remedial measures. Anesth Essays Res 13(3):554–559. https://doi.org/10.4103/aer.AER_75_19. (**ISSN: 0259-1162**)

Hanne T, Melo T, Nickel S (2009) Bringing robustness to patient flow management through optimized patient transports in hospitals. Interfaces 39(3):241–255. https://doi.org/10.1287/inte.1080.0379

Harris CR et al. (2020) Array programming with NumPy. Nature 585(7825):357–362. https://doi.org/10.1038/s41586-020-2649-2

Haykin SS (1999) Neural networks: a comprehensive foundation, 2nd edn. Prentice Hall, Upper Saddle River. ISBN: 978-0-13-273350-2

Hendrich AL, Nelson L (2005) Intra-unit patient transports: time, motion, and cost impact on hospital efficiency. Nurs Econ 23(4):157. https://pubmed.ncbi.nlm.nih.gov/16189980/

Hernández-Orallo J (2013) ROC curves for regression. Pattern Recogn 46(12):3395–3411. https://doi.org/10.1016/j.patcog.2013.06.014. (**ISSN: 00313203**)

Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9(3):90–95. https://doi.org/10.1109/MCSE.2007.55

Izadifar M, Abdolahi F (2006) Comparison between neural network and mathematical modeling of supercritical CO2 extraction of black pepper essential oil. J Supercrit Fluids 38(1):37–43. https://doi.org/10.1016/j.supflu.2005.11.012

Janga Reddy M, Nagesh Kumar D (2020) Evolutionary algorithms, swarm intelligence methods, and their applications in water resources engineering: a state-of-the-art review. H2Open J 3(1):135–188. https://doi.org/10.2166/h2oj.2020.128

Jaroon K (2018) Efficiency improvement of patients' transportation at the super tertiary hospital in Southern Thailand. Nurs J Minist Public Health 27:104–120. https://he02.tci-thaijo.org/index.php/tnaph/article/view/112269

Jebari K, Madiafi M (2013) Selection methods for genetic algorithms. Int J Emerg Sci 3(4):333–344

Kallrath J (2005) Online storage systems and transportation problems with applications. Applied optimization. Kluwer Academic Publishers, Boston. https://doi.org/10.1007/0-387-23485-3

Katoch S, Chauhan SS, Kumar V (2021) A review on genetic algorithm: past, present, and future. Multimed Tools Appl 80(5):8091–8126. https://doi.org/10.1007/s11042-020-10139-6. (**ISSN: 1380-7501**)

Kergosien Y, Lenté C, Piton D, Billaut JC (2011) A tabu search heuristic for the dynamic transportation of patients between care units. Eur J Oper Res 214(2):442–452. https://doi.org/10.1016/j.ejor.2011.04.033

Kingma DP, Ba J (2017) Adam: a method for stochastic optimization. https://doi.org/10.48550/ARXIV.1412.6980. arXiv:1412.6980

Klein TL, Thielen C (2024) arXiv Preprint arXiv. Improving patient transport in hospitals: a literature review of operations research methods. https://doi.org/10.48550/ARXIV.2404.03282

Kropp T, Faeghi S, Lennerts K (2023) Evaluation of patient transport service in hospitals using process mining methods: patients' perspective. Int J Health Plan Manag 38(2):430–456. https://doi.org/10.1002/hpm.3593

Kropp T, Faeghi S, Lennerts K (2024) Process mining for capacity planning and reconfiguration of a logistics system to enhance the intra-hospital patient transport. Case Study. In: J Finkelstein, R Moskovitch, and E Parimbelli (ed.) Artificial Intelligence in medicine. Vol. 14844. Lecture notes in computer science. Springer Nature Switzerland, Cham. pp. 138–150. https://doi.org/10.1007/978-3-031-66538-7_15

Kuchera D, Rohleder TR (2011) Optimizing the patient transport function at Mayo Clinic. Quality Manag Healthc 20(4):334–342. https://doi.org/10.1097/QMH.0b013e318231a84f

Lawrence J, Luedeking S (1993) Introduction to neural networks: design, theory and applications, 5th edn. California Scientific Software, Nevada City

Lenz R, Reichert M (2007) IT support for healthcare processes—premises, challenges, perspectives. Data Knowl Eng 61(1):39–58. https://doi.org/10.1016/j.datak.2006.04.007. (**ISSN: 0169023X**)

Li Y, Xiang T, Szeto WY (2021) Home health care routing and scheduling problem with the consideration of outpatient services. Transp Res Part E Logist Transp Rev 152:102420. https://doi.org/10.1016/j.tre.2021.102420. (**ISSN: 13665545**)

Maka T, Kasemset C, Phongthiya T (2022) Intra-hospital patient transportation system planning using Bilevel decision model. Logforum 18(2):237–246. https://doi.org/10.17270/J.LOG.2022.681. (**ISSN: 18952038**)

Martin N et al. (2020) Recommendations for enhancing the usability and understandability of process mining in healthcare. Artif Intell Med 109:101962. https://doi.org/10.1016/j.artmed.2020.101962

McKinney W (2010) Data structures for statistical computing in python. In: Proceedings of the 9th python in science conference. Edited by van der Walt S and Millman J, pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Meephu E, Arwatchananukul S, Aunsri N (2023) Enhancement of Intra-hospital patient transfer in medical center hospital using discrete event system simulation. PLoS ONE 18(4):e0282592. https://doi.org/10.1371/journal.pone.0282592

Mesabbah M, Abo-Hamad W, McKeever S (2019) A hybrid process mining framework for automated simulation modelling for healthcare. In: 2019 winter simulation conference (WSC). IEEE, National Harbor, MD, USA, pp. 1094–1102. https://doi.org/10.1109/WSC40007.2019.9004800

Mikut R (2008) Data mining in der medizin und medizintechnik. KIT Scientific Publishing, Karlsruhe. https://doi.org/10.5445/KSP/1000008476

Miles J (2014) R squared, adjusted R squared". In: Wiley statsref. Balakrishnan N. (ed.) Place of publication not identified: John Wiley & Sons, Inc., Hoboken https://doi.org/10.1002/9781118445112.stat06627

Mitrea CA, Lee CKM, Wu Z (2009) A comparison between neural networks and traditional forecasting methods: a case study. In: International Journal of Engineering Business Management 1. Publisher: SAGE Publications Ltd STM, p. 11. ISSN: 1847-9790. https://doi.org/10.5772/6777

Molenbruch Y, Braekers K, Caris Berghe GV (2017) Multi-directional local search for a bi-objective dial-a-ride problem in patient transportation. Comput Oper Res 77:58–71. https://doi.org/10.1016/j.cor.2016.07.020

Morid MA, Sheng ORL, Dunbar J (2021) Time series prediction using deep learning methods in healthcare. In: CoRR arXiv:2108.13461. https://doi.org/10.48550/arXiv.2108.13461

Muloiwa M, Dinka M, Nyende-Byakika S (2023) Application of artificial neural network for predicting biomass growth during domestic wastewater treatment through a biological process. Eng Life Sci 23(5):e2200058. https://doi.org/10.1002/elsc.202200058. (**ISSN: 1618-0240, 1618-2863**)

Naesens K, Gelders L (2009) Reorganising a service department: central patient transportation. Product Plan Control 20(6):478–483. https://doi.org/10.1080/09537280902938621. (**ISSN: 0953-7287**)

Nakayama DK, Lester SS, Rich DR, Weidner BC, Glenn JB, Shaker IJ (2012) Quality improvement and patient care checklists in intrahospital transfers involving pediatric surgery patients. J Pediatr Surg 47(1):112–118. https://doi.org/10.1016/j.jpedsurg.2011.10.030

Nas S, Koyuncu M (2019) Emergency department capacity planning: a recurrent neural network and simulation approach. Comput Math Methods Med. https://doi.org/10.1155/2019/4359719

Ncibi K, Sadraoui T, Faycel M, Djenina A (2017) A multilayer perceptron artificial neural networks based a preprocessing and hybrid optimization task for data mining and classification. Int J Econ Financ Manag. https://doi.org/10.12691/ijefm-5-1-3

Neto AH, Fiorelli FAS (2008) Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy Build 40(12):2169–2176. https://doi.org/10.1016/j.enbuild.2008.06.013

Nikzad M, Movagharnejad K, Talebnia F (2012) Comparative study between neural network model and mathematical models for prediction of glucose concentration during enzymatic hydrolysis. Int J Comput Appl 56(1):43–48. https://doi.org/10.5120/8859-2818

Ozili PK (2022) The acceptable R-square in empirical modelling for social science research. SSRN Electron J. https://doi.org/10.2139/ssrn.4128165

Paszke A, et al. (2019) PyTorch: an imperative style, high-performance deep learning library. In: advances in neural information processing systems, Curran Associates, Inc., 32 pp. 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pedregosa F et al. (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

Picetti E et al. (2013) Intra-hospital transport of brain-injured patients: a prospective, observational study. Neurocrit Care 18(3):298–304. https://doi.org/10.1007/s12028-012-9802-1

Piotrowski AP, Napiorkowski MJ, Napiorkowski JJ, Rowinski PM (2017) Swarm intelligence and evolutionary algorithms: performance versus speed. Inf Sci 384:34–85. https://doi.org/10.1016/j.ins.2016.12.028. (**ISSN: 0020-0255**)

Pourbafrani M, van der Aalst WMP (2023) Data-driven simulation in process mining: introducing a reference model. In: Proceedings of the 37th ECMS international conference on modelling and simulation, ECMS 2023. Edited by E Vicario. Communications of the ECMS. [Great Britain]: ECMS, pp. 411–420. ISBN: 9783937436807. https://doi.org/10.7148/2023-0411

Rajakumari K, Madhunisha M (2020) Intelligent and convolutional-neural-network based smart hospital and patient scheduling system. In: 2020 international conference on computer communication and informatics (ICCCI). IEEE, pp. 1–5. https://doi.org/10.1109/ICCCI48352.2020.9104173

de Roock E, Martin N (2022) Process mining in healthcare—an updated perspective on the state of the art. J Biomed Inform 127:103995. https://doi.org/10.1016/j.jbi.2022.103995. (**ISSN: 1532-0480**)

Ruiz-Sepúlveda A, Triviño-Rodriguez JL, Morales-Bueno R (2009) Computing a Comprehensible Model for Spam Filtering. In: Discovery Science. Gama J, Costa VS, Jorge AM, Brazdil PB (ed.) Vol. 5808. Lecture notes in computer science. Springer Berlin Heidelberg, Berlin. pp. 457–464. https://doi.org/10.1007/978-3-642-04747-3_39

Russell SJ, Norvig P (2010) Artificial intelligence: a modern approach, 3rd edn. Artificial intelligence. Prentice-Hall, Upper Saddle River

Russell SJ, Norvig P (2021) Artificial intelligence: a modern approach. Fourth edition, global edition. Always learning. Pearson, Boston

Salami ES, Ehetshami M, Karimi-Jashni A, Salari M, Nikbakht Sheibani S, Ehteshami A (2016) A mathematical method and artificial neural network modeling to simulate osmosis membrane's performance. Model Earth Syst Environ 2(4):1–11. https://doi.org/10.1007/s40808-016-0261-0

Schmid V, Doerner KF (2014) Examination and operating room scheduling including optimization of intrahospital routing. Transp Sci 48(1):59–77. https://doi.org/10.1287/trsc.1120.0452. ISSN: 0041-1655, 1526-5447

Sharma P, Kaur M (2013) Classification in pattern recognition: a review. Int J Adv Res Comput Sci Softw Eng 3(4):298

Solgi RM (2020) geneticalgorithm: python library. https://pypi.org/project/geneticalgorithm

Spiess AN, Neumeyer N (2010) An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. BMC Pharmacol 10(1):6. https://doi.org/10.1186/1471-2210-10-6

Stefanini A, Aloini D, Dulmin R, Mininno V (2017) Service reconfiguration in healthcare systems: the case of a new focused hospital unit. In: Cappanera P, Li J, Matta A, Sahin E, Vandaele NJ, Visintin F (eds) Health care systems engineering, vol 210. Springer, Cham, pp 179–188. https://doi.org/10.1007/978-3-319-66146-9_16

Séguin S, Villeneuve Y, Blouin-Delisle CH (2019) Improving patient transportation in hospitals using a mixed-integer programming model. Oper Res Health Care 23:100202. https://doi.org/10.1016/j.orhc.2019.100202. (**ISSN: 22116923**)

Talbi EG (2009) Metaheuristics: from design to implementation. Wiley, Hoboken

The Pandas Development Team (2023) Pandas-dev/pandas: pandas. Version latest. https://doi.org/10.5281/zenodo.3509134

Torgo L, Ribeiro R (2009) Precision and recall for regression. In: Gama J, Costa VS, Jorge AM, Brazdil PB (ed.) Discovery Science. Vol. 5808. Lecture notes in computer science. Springer, Berlin Heidelberg, pp. 332–346. https://doi.org/10.1007/978-3-642-04747-3_26

Turan B, Schmid V, Doerner KF (2011) Models for intra-hospital patient routing. In: 2011 3rd IEEE international symposium on logistics and industrial informatics (LINDI 2011). IEEE, Piscataway, NJ. pp. 51–60. https://doi.org/10.1109/LINDI.2011.6031162

Ulrich RS, Zhu X (2007) Medical complications of intra-hospital patient transports: implications for architectural design and research. HERD Health Environ Res Des J 1(1):31–43. https://doi.org/10.1177/193758670700100113. (**ISSN: 1937-5867**)

van Hulzen G, Martin N, Depaire B, Souverijns G (2022) Supporting capacity management decisions in healthcare using data-driven process simulation. J Biomed Inform 129:104060. https://doi.org/10.1016/j.jbi.2022.104060. (**ISSN: 1532-0480**)

van der Aalst WMP (2016) Process mining: data science in action. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4

Vancroonenburg W, Esprit E, Smet P, Berghe GV (2016) Optimizing internal logistic flows in hospitals by dynamic pick-up and delivery models. In: Proceedings of the 11th international conference on the practice and theory of automated timetabling (PATAT 2016), pp. 371–383. https://www.patatconference.org/patat2016/files/proceedings/paper_29.pdf

West PM, Brockett PL, Golden LL (1997) A comparative analysis of neural networks and statistical methods for predicting consumer choice. In: Marketing Science 16.4 Publisher: INFORMS, pp. 370–391. ISSN: 0732-2399. https://doi.org/10.1287/mksc.16.4.370

Wickham H, François R, Henry L, Müller K, Vaughan D (2023) dplyr: a grammar of data manipulation. https://dplyr.tidyverse.org

Wu SJ, Chow PT (1994) Genetic algorithms for solving mixed-discrete optimization problems. J Frankl Inst 331(4):381–401. https://doi.org/10.1016/0016-0032(94)90004-3. (**ISSN: 00160032**)

Xiao L, Zhen L, Laporte G, Baldacci R, Wang C (2022) A column generation-based heuristic for a rehabilitation patient scheduling and routing problem. Comput Oper Res 148:105970. https://doi.org/10.1016/j.cor.2022.105970

Yang S et al. (2017) A data-driven process recommender framework. KDD Proc Int Conf Knowl Discov Data Min 2017:2111–2120. https://doi.org/10.1145/3097983.3098174

Yang S, Li J, Tang X, Chen S, Marsic I, Burd RS (2017) Process mining for trauma resuscitation. IEEE Intell Inform Bull 18(1):15–19 (**ISSN: 2168-8222**)

Yazır OA, Koç Ç, Yücel E (2023) The multi-period home healthcare routing and scheduling problem with electric vehicles. OR Spectr 45:853–901. https://doi.org/10.1007/s00291-023-00721-3. (**ISSN: 0171-6468**)

Zhou Z, Wang Y, Li L (2014) Process mining based modeling and analysis of workflows in clinical care—a case study in a Chicago outpatient clinic. In: 2014 IEEE 11th international conference on networking, sensing and control (ICNSC 2014). Piscataway, NJ: IEEE, pp. 590–595. ISBN: 978-1-4799-3106-4. https://doi.org/10.1109/ICNSC.2014.6819692

Zhu Y, Wang M, Yin X, Zhang J, Meijering E, Hu J (2023) Deep learning in diverse intelligent sensor based systems. Sensors 23(1):62. https://doi.org/10.3390/s23010062

Authors and Affiliations

Tobias Kropp[1] · **Yuhao Gao**[1] · **Kunibert Lennerts**[1]

✉ Tobias Kropp
tobias.kropp@kit.edu

Yuhao Gao
yuhao.gao@kit.edu

Kunibert Lennerts
kunibert.lennerts@kit.edu

[1] Institute of Technology and Management in Construction, Karlsruhe Institute of Technology, Gotthard-Franz-Str. 3, 76131 Karlsruhe, Baden-Württemberg, Germany