Public Transport (2024) 16:99–158
https://doi.org/10.1007/s12469-023-00334-7

ORIGINAL RESEARCH

Artificial intelligence for improving public transport: a mapping study

Å. Jevinger[1,2] · C. Zhao[2,3] · J. A. Persson[1,2] · P. Davidsson[1,2]

Accepted: 19 July 2023 / Published online: 20 November 2023

Abstract

The objective of this study is to provide a better understanding of the potential of using Artificial Intelligence (AI) to improve Public Transport (PT), by reviewing research literature. The selection process resulted in 87 scientific publications constituting a sample of how AI has been applied to improve PT. The review shows that the primary aims of using AI are to improve the service quality or to better understand traveller behaviour. Train and bus are the dominant modes of transport investigated. Furthermore, AI is mainly used for three tasks; the most frequent one is prediction, followed by an estimation of the current state, and resource allocation, including planning and scheduling. Only two studies concern automation; all the others provide different kinds of decision support for travellers, PT operators, PT planners, or municipalities. Most of the reviewed AI solutions require significant amounts of data related to the travellers and the PT system. Machine learning is the most frequently used AI technology, with some studies applying reasoning or heuristic search techniques. We conclude that there still remains a great potential of using AI to improve PT waiting to be explored, but that there are also some challenges that need to be considered. They are often related to data, e.g., that large datasets of high quality are needed, that substantial resources and time are needed to pre-process the data, or that the data compromise personal privacy. Further research is needed about how to handle these issues efficiently.

Keywords Artificial intelligence · Machine learning · Public transit · Mass transit · Public transport · Literature review

✉ Å. Jevinger
ase.jevinger@mau.se

[1] Internet of Things and People Research Center, Department of Computer Science and Media Technology, Malmö University, Malmö, Sweden

[2] The Swedish Knowledge Centre for Public Transport, K2, Lund, Sweden

[3] Faculty of Engineering, Department of Technology and Society, Lund University, Lund, Sweden

1 Introduction

Several market reports predict an increase in demand for public transport (PT) services during the coming years (Market Research Future 2021; Markets and Markets 2019). Some of the main reasons mentioned for this is the growing population (and thereby growing demand) and the environmental effects of car travel, pressuring people to review their travelling behaviours. These reasons will presumably remain, even though Covid 19 appears to at least temporally have stopped the trend. Moreover, road congestions are becoming an increasingly severe problem in some cities, which calls for alternative modes of travel (Moya-Gómez and García-Palomares 2017). Consequently, many governing bodies at both local and national level are working to move travellers from private cars to PT. PT is thereby faced with requirements to become a more attractive travelling alternative which, together with growing demands, pose great challenges for the PT sector. At the same time, complexity of the transport systems and traffic behaviour increase, as the number of alternative travelling options expand and the transport systems evolve. Emerging technologies such as Artificial Intelligence (AI) open up for new opportunities to handle these, and other types of challenges (UITP 2020). These opportunities may also bring financial benefits. For instance, it has been estimated that AI will increase profitability within transportation and storage by 44% and within public services by 27%, by the year 2035 (Purdy et al. 2017).

The term Artificial Intelligence was coined in 1955 and has been defined as "the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable." (McCarthy 1998; McCarthy et al. 2006). However, over the years, many different definitions of this complex term have been suggested. Russell and Norvig (2010) identified four main categories of AI definitions: Thinking humanly, Acting humanly, Thinking rationally and Acting rationally. Since the research community still is far from a commonly agreed definition, we have, in the literature review presented in this paper, chosen to rely on the authors of the selected publications' use of the term AI.

In recent years, the amount of research on how to apply AI to PT has increased rapidly. The aim of these studies is often to extend the general understanding of the PT system and its travellers, or to make PT more efficient, leading to, e.g., cost savings and a reduced environmental impact. The intended users of the AI technology differ depending on these aims, from governmental PT authorities, via planners and operators, to individual travellers. We argue that PT has a number of characteristics that make it suitable for applying AI, e.g., a vast amount of data is continuously being generated but is not used to its full potential, and the PT systems are affected by distributed decision making involving both travellers and several PT actors. Thereby, AI may make a difference in many situations by supporting both individuals and organizational actors. Moreover, the Internet of Things (IoT) is also an enabler in this area (Davidsson et al. 2016), which could spark many changes in the PT domain when combined with AI.

This paper aims to review research on applications of AI that can be used to improve PT, i.e., to make PT better in some way. This improvement can benefit, for instance, the travellers, the actors or the environment; however, studies in which PT is only used as a means to achieve other types of goals, such as, when buses are used as probes for detecting road potholes (Sharma and Sharma 2019), are not included in the review.

Previous attempts have been made to review this topic, at least partially. For instance, Abduljabbar et al. (2019) reviewed previous research using AI within transport in general; however, little attention was paid to PT modes other than buses. Similarly, Koushik et al. (2020) conducted a review on activity-based travel behaviour studies that employ machine learning (ML) techniques, but with little focus on PT. Liyanage et al. (2019) made an environmental scan and analysis of the technological, social, and economic impacts surrounding flexible on-demand mobility, including some of the AI-based tools that have been applied within PT. Li et al. (2018b) conducted a literature review of the practice of using smart card data for estimating traveller destination, where some approaches are based on ML. Furthermore, Welch and Widita (2019) reviewed literature on sources of big data and big data applications, of which some were AI-based, applied to public transportation problems. Similarly, Ge et al. (2021) reviewed the current state of the art of public transport data sources, as well as summarized and analysed the potentials and challenges of the main data sources. They also presented an information management framework to enhance the use of the data sources. Finally, several literature review studies focus on the benefits of using agent technology within PT (Bazzan and Klügl 2014; Chen and Cheng 2010). Although these reviews provide some useful insights into how AI can be utilized within PT, to the best of our knowledge, there is no previous review focusing on exploring the potential of using AI to improve PT. This gap in literature, in combination with a significant growing interest in AI within PT, suggests that the proposed review would represent both a timely and an important contribution.

The goal of this review is to contribute to the general question: What is the potential of AI to improve PT? To answer this question, we have identified the following more concrete research questions:

1. What problems have been addressed and what are the intended benefits?
2. To what extent have these benefits been realized?
3. How has AI been used, and what are the requirements on data availability?
4. What are the main challenges faced when using AI to improve PT?

The first two questions concern the intended benefits, e.g., cost savings or service quality improvement, and application areas, e.g., a particular transport mode and task. To better understand the potential of AI for PT, it is important to comprehend how AI actually contributes in the different studies and which requirements need to be met to achieve the benefits; in particular, what data is needed. These aspects are covered by the third question. Finally, to understand the potential, we also have to understand the challenges.

The result is expected to provide increased knowledge to different types of PT actors and public authorities about the potential of AI to improve PT both in general and in specific areas, including planning and operation improvement; for instance, in terms of efficiency, reliability and safety. Furthermore, the review provides scholars with an overview of the state of the art and an indication of current knowledge gaps. Notably, the review does not intend to dig into which particular AI method is best suited to address a particular application area in the PT domain, since that would only be possible if multiple methods were applied in very similar context. Moreover, the success of using, for instance, ML is often more connected to other aspects than the chosen ML algorithm, such as, parameter tuning, data preparation, data availability, etc. (see, e.g., Lavesson and Davidsson (2006)). Instead, this review intends to identify application areas of successful implementations and real usage in business if such exist. Our approach to reach the aim of this study is not to identify every single article in the area, but rather to base our analysis on a representative sample of the literature in the domain. The next section explains the methodology used for the literature review. In Sect. 3, the results are presented and analysed. Section 4 discusses the result and concludes the study.

2 Methodology

To investigate the potential of AI to improve PT, a literature review was carried out. The aim of this review was to identify a relevant and representative sample of research, and by characterizing the research performed, synthesize an indication of the potential. An indication of potential may be that a problem was addressed by using AI; and a stronger indication is, of course, if validated benefits are reported. Notably, we view potential from a broad perspective, including, for instance, what type of benefits can be achieved, and what type of AI-technologies and data sources appear to be useful.

The literature review was conducted in a systematic manner, following most of the guidelines provided by Kitchenham and Charters (2007). The main difference between these guidelines and our review is that another approach was used to refine the selection criteria (see step 4 below), and that we did not perform any study quality assessment. A study quality assessment primarily aims at providing more refined selection criteria, and it estimates the quality differences between the studies. As mentioned above, this review aims at identifying a representative sample of the current research in this area, irrespective of quality. However, the reviewed studies were classified according to benefit validation (e.g., conceptual solution or experiments based on real data) and whether the proposed solutions have been implemented in business, which gives an indication of the quality (Kitchenham and Charters 2007). Given that our research questions are to some extent on a general level, our review has resulted in a systematic mapping study. However, some research questions are on a more specific level and thus, the study also has elements of a systematic literature review (Kitchenham and Charters 2007). In short, the literature review included the following sequential steps:

1. The research questions were specified (see Sect. 1).
2. A preliminary review protocol was developed and agreed by all researchers in the research group.
3. Potentially relevant research studies were identified based on the search strategy (see Sect. 2.1).
4. The review protocol was revised based on a study of 20 of the research papers identified in step 3 (see Sect. 2.1).
5. The primary research studies were identified based on the selection criteria (see Sect. 2.2).
6. The data were extracted (see Appendices A and B).
7. The data were analysed and synthesized (see Sect. 3).

The review protocol included the research questions, search strategy, selection criteria, selection procedure, data extraction strategy, and synthesis strategy. The data extraction strategy involved a classification framework that defined a number of categories to which the identified research studies were mapped.

The literature review was performed by a research group of four researchers. In step 4, 20 randomly selected studies were divided between the group members, amounting to 5 studies per researcher to read and to analyse. The results from this work were then discussed by the entire research group. Based on these discussions, the selection criteria and classification framework in the review protocol were revised. Section 2.1 presents the results of the agreed upon final review protocol. However, the classification framework is presented in Sect. 3, as it was further revised during the subsequent classification work with the rest of the studies, i.e., it was extended whenever a study could not be properly classified by the framework.

2.1 Review protocol

As mentioned above, the review protocol included the research questions, search strategy, selection criteria, selection procedure, data extraction strategy, and synthesis strategy. The *search strategy* was defined as follows:

- Search databases: Scopus, IEEE Xplore, and Web of Science.
- Publication year: no limit.
- Search phrase: ("AI" or "Artificial Intelligence" or "Machine Learning") and ("Public transport" or "Mass transit" or "Public Transit").
- Search fields: title and abstract, i.e., the search phrase may match the title, the abstract, or partly the title, and partly the abstract.

The main reason for selecting Scopus, IEEE Xplore, and Web of Science is that these are three of the most important databases for computer science research, and that they cover a large scope of public transport research (Bar-Ilan et al. 2007; Hoonlor et al. 2013). Moreover, IEEE Xplore and Scopus are identified as important within computer science (software engineering) by the review guidelines developed by Kitchenham and Charters (2007). The search phrase focuses on AI, ML and PT (in different nuances). ML, which traditionally is viewed as a subarea of AI, was added to the search phrase due to the recent huge practical success of ML, which has significantly contributed to the increased interest in AI (Holzinger et al. 2018). As with AI, the review relies on the authors of the selected papers' perceptions of the ML concept. Obviously, there might be more search phrases – such as, deep learning, software agents etc. – that can generate a larger number of studies; however, as stated above, the aim was to identify a representative sample. Moreover, by using this strategy, we rely on the publication authors' views on what is included in the concept of AI and ML.

The *selection criteria* were defined as follows in the review protocol:

(1) Inclusion criteria

  (a) Written in English.
  (b) Journal articles, conference papers, and bookchapters.
  (c) Describe the use of AI to improve or to understand PT.

(2) Exclusion criteria

  (a) Earlier or shorter versions of a paper if there is a refined or extended version.
  (b) Review studies.

Criterion 1c implies that all selected studies must be related to PT, including, e.g., bus, railway and bike sharing systems. Furthermore, studies in which only a part of the route involves PT are included in the review, i.e., at least some part of a route must involve PT, whereas other parts may involve other means of transport (e.g., private bicycle or car). All selected studies must describe the use of AI to improve or to understand PT. This means that studies addressing, for instance, strictly legal issues related to implementing AI in PT have not been included. Furthermore, studies describing general approaches to improve vehicles or road infrastructure are not included. Criterion 2b implies that we only focus on primary studies, not secondary studies, i.e. review studies.

In the *selection procedure*, the research studies were randomly distributed and studied by the research group members. Any ambiguity raised by a group member during this work was discussed by the entire research group. The selection procedure included the following steps:

1. Search the three databases according to the search strategy.
2. Remove all duplicates (identical papers appearing in more than one database).
3. Read all paper abstracts and include/exclude based on the selection criteria.
4. Read all full papers and include/exclude based on the selection criteria.

As mentioned above, the data *extraction strategy* involved a classification framework. The classification framework consists of a number of categories that were developed based on the research questions above, the initial review of 20 papers (step 4 above), and some literature reviews in related areas (Davidsson et al. 2005). Each category contains a number of classes, which were identified iteratively during the review, based on the content of the studies. After the selection procedure, all data relevant for the classification framework were extracted from the selected studies, i.e., for each study information was mapped to the framework. As before, all ambiguities were discussed by the entire research group. When all studies had been mapped to the classification framework, the framework categories were divided between the research group members and further studied to detect any inconsistencies. Thereafter, the information in each of the categories was analysed as part of the *synthesis strategy*. Moreover, a cross-analysis between the most relevant categories was performed.

2.2 Paper selection

The database searches cover all papers up to and including the year 2020. The initial searches resulted in 187 records in Scopus, 39 records in IEEE Xplore, and 77 records in Web of Science. Of these, 91 records were duplicates, and thereby removed. In the subsequent abstract scanning process, another 71 records were removed (24 based on criterion 1b, 46 based on criterion 1c and 1 based on criterion 2b). Thus after the abstract scanning, 141 studies remained. The full paper scanning process resulted in the removal of 30 studies (6 based on criterion 1a, 18 based on criterion 1c, 2 based on criterion 2a, and 4 based on criterion 2b). As a result, 111 studies were included in the subsequent framework classification and analysis.

3 Results

During the review, we found that AI is typically used for automating or supporting decision making, including planning. In principle, AI can be employed in all different potential steps—from structuring data to suggesting decisions or even affecting the real world by automation. Therefore, the aspects of how AI improves PT have been incorporated into the classification framework. Table 1 shows the resulting classification framework based on all reviewed papers, including categories and corresponding classes. In the subsequent subsections, the main classification results are quantified and analysed. The complete set of results are presented in Appendix A. The framework allows for double classification, when suitable.

3.1 Benefits

We have classified the benefits aimed for, or achieved, in the studies through explicit claims of benefits connected to the following: *Cost savings, Service quality improvement, Environmental, General understanding, Safety/security*. Naturally, the statements in the individual studies concerning how the results can be used to improve PT, are often on a more detailed level. Moreover, sometimes the same AI technology could be used for several purposes, e.g., improving service and reducing emissions as suggested by Mackett (1994), Prashanth et al. (2016) and Shatnawi et al. (2020). These studies have been double classified. However, such multiple benefits probably apply for more studies, although they are

Table 1 Classification framework

Categories	Classes for classification
What benefits are aimed for and in which areas?	
Benefits aimed for	Cost savings, Service quality improvement, Environmental, General understanding, Safety/security
Transport modes addressed	Any PT, PT and other modes, Bus, Railway, Bike sharing systems
Time horizons considered	Short term, Medium term, Long term
Are the benefits realized?	
Benefit validation	Conceptual/theoretical, Experiments based on artificial data, Experiments based on real data, Implemented in small scale, Implemented in large scale
Implemented in business	Broadly used, Single example, Nothing reported
How is AI used, or suggested to be used, for achieving the benefits and what are the associated requirements?	
Mechanism for achieving benefit	Current status estimate, Prediction, Planning/scheduling/resource allocation
Support or automation	Support, Automation
AI technology used	ML (Machine learning), Reasoning (Reasoning and Knowledge representation), Heuristic Search (Heuristic Search Algorithms), MAS (Multi-Agent Systems)
Data needs	Number/ID of boarding/alighting passengers at different stations, Station congestion/demand levels, Traveller characteristics/behaviour/opinions, People's positions/acceleration/etc., Surveillance on in-vehicle passengers, Fare transactions and journey searches, PT vehicle positions/acceleration/etc. and actual arrival/ departure times, Surveillance on PT system and in-vehicle conditions, Vehicle capacity, Timetables and structure of PT system, PT organization/funding/marketing/services/maintenance, Local weather conditions, Local built and natural environment as well as city regulations, Private car positions, Conditions/traffic volumes/speeds on roads, Signal light state, Road traffic incidents
Applications in PT	Supporting dispatching, Supporting diagnosis, Prediction and planning, Supporting emergency management, Supporting maintenance, Vehicle tracking, Monitoring passenger flows, Traffic monitoring, Traffic management, Security surveillance, Estimating/predicting the number of travellers in a vehicle/station, Fraud detection, Fare collection, Understanding travel patterns, Understanding/predicting the travellers' mode choice, Predicting passenger dwelling time and flow, Understanding the effects of the built environment on travel behaviour, Predicting travellers' social demographics, Estimating travel/arrival time, Supporting route choice, Balancing and availability of rental bikes, Traveller recommender systems, Improving communication with travellers, Monitoring traveller's state, Capturing travellers' opinions, Supporting PT system planning, Analysing PT systems, Funding PT systems, Generating synthetic data, Timetable scheduling support, Predicting ticket prices, Estimating real-time onboard bus ride comfort, Optimizing when to backup surveillance video files

Table 1 (continued)

Categories	Classes for classification
What are the (potential) hindrance of using AI to improve public transport?	
Challenges of using AI	Large datasets required, Data from different sources required, Incomplete datasets, Insufficient data quality, Preparing/computing datasets takes time/resources, Select most appropriate AI algorithm, Personal privacy

not explicitly stated. Hence, the classification of a study into a benefit category mainly reflects the stated motivational goals but not its entire potential areas of benefit.

Cost savings include attempts to reduce PT costs by, for instance, increasing resource utilization through improved scheduling and reduced resources while maintaining service, using, e.g., trip or mode optimization (Manivannan et al. 2020; Tekin et al. 2018). Service quality improvements include not only the optimization of line scheduling but also examples of detecting problems in PT for quick alleviation, such as bunching (Degeler et al. 2020). If the motive is connected to the environment, including ecological sustainability and energy savings, it is classified as Environmental. Within this class, studies that use AI for mode choice analysis with the purpose of reduced car driving are included (cf. Lazar et al. 2019). If AI is applied for mode choice analysis, but with the purpose of making things better in PT in several ways, it falls within the class of General understanding. The class General understanding contains PT system forecasting without a particular focus of the usage of the forecast, e.g., forecasting traveller mode choices within e-mobility (Ferrara et al. 2019) or using mobility data for mode choice predictions (Liang et al. 2019). General understanding also includes additional aspects not covered in the other classes, such as PT funding (Ubbels and Nijkamp 2002) and assessing equity (Mayaud et al. 2019). Figure 1 shows the number of studies addressing different benefits.

Our findings show that AI has mainly been used for improving service quality and general understanding. Most studies related to the General understanding class concern mode choice analysis with the purpose of making things better in PT in several ways. Given that AI in general often is viewed as a tool for making systems more efficient, one might expect that the Cost savings class, which involves resource improvement aspects, would have included more studies than we have found in our review. Notably, no study had a clear primary goal to increase revenue.

3.2 Transport modes

A broad range of transport modes are addressed in the studies. Moreover, while some studies focus on a single transport mode, e.g., conventional buses, trains or bike sharing systems, others address the combination of several transport modes, e.g., the combination of private car and railway for the same journey. We have identified five different main transport mode classes: *Bus*, *Railway*, *Any PT*, *PT and other modes* and *Bike sharing systems*. The classes *Bus*, *Railway* and *Bike sharing systems* refer to studies that focus on a single transport mode (Aditi et al. 2020; Singla et al. 2015; Tang et al. 2020), whereas studies addressing both bus and railway are double classified to the *Bus and Railway* classes. The class *Any PT* includes both studies that specifically state that their results can be applied for any type of PT mode, and studies that do not specify which type of PT they are addressing but whose results can be applied for any type of PT mode (Roulland et al. 2014; Van Egmond et al. 2003). The class *PT* and other modes include studies that focus on both PT modes and other transport modes (e.g., private car or bicycle) (Chapleau et al. 2019; Tu et al. 2016). Figure 2 presents the number of studies in each type of transport mode class. As can be seen, most of the studies address the application of *AI for bus*, followed by *Railway*, *Any PT* and *PT* and other modes. A few studies also experiment with applying AI for bike sharing systems.

3.3 Time horizons

In terms of supported time horizons, we have identified three classes: *Long term*, *Medium term* and *short term*. However, since it is often difficult to clearly assess the time horizon of the influence based on the information present in the studies, this classification should be viewed as indicative. *Long-term support* refers to studies applying AI for supporting long-term decisions, or can be used annually for the distant future, e.g., using AI for selecting the appropriate form of PT systems for cities (Liang et al. 2019; Mackett 1994; Victoriano et al. 2020). In total, 32 studies address supported decisions concerning the long-term horizon, covering a broad range of applications. *Medium-term support* refers to studies that address decisions affecting the practices of PT for the coming weeks and up to a year. A total of 34 studies have explored this type of supported decisions, where the support was provided through, for example, using AI to optimize the PT routes and the number of operational vehicles (Tekin et al. 2018). Decisions related to the short-term horizon influence the practices in real-time contexts, e.g., travel-time prediction via forecasts of the status

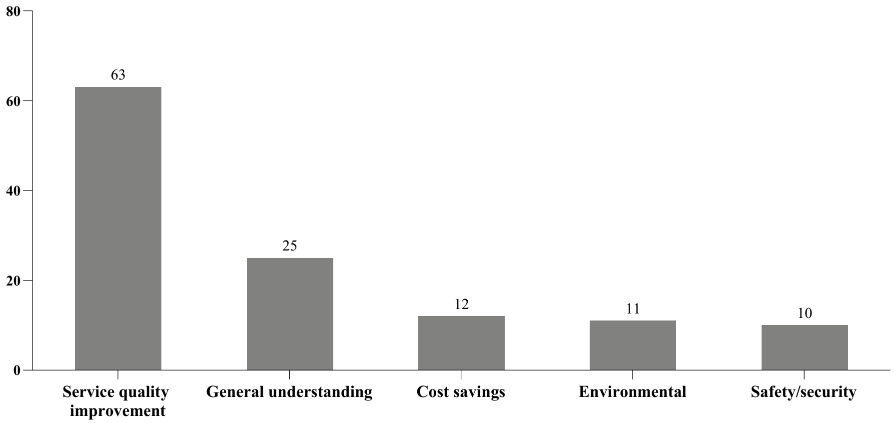


Fig. 1 Number of studies addressing different benefits

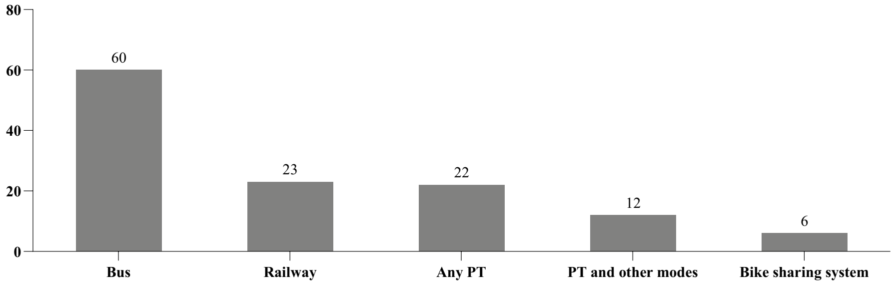


Fig. 2 Number of studies addressing different transport modes

of the buses (position of the bus, estimating the delay or detecting critical situations) in order to support and allow the traffic controller to quickly handle situations (Wei et al. 2017), as well as automation. As presented in Fig. 3, most studies focus on short-term decision support, whereas equally many focus on long-term and medium-term decisions.

3.4 Benefit validation and implemented in business

The maturity level of the AI solutions suggested in the studies is reflected by two dimensions: (1) to what extent the benefits of applying AI in PT are validated, and (2) to what extent the benefits/applications are implemented in business.

The level of validation of the application is classified to five levels: *Conceptual/theoretical, Experiments based on real data, Experiments based on artificial data, implemented in small scale*, and *Implemented in large scale*. *Conceptual/theoretical* results include studies whose results are on a conceptual/theoretical level and thus, not realized in practice. They mainly focus on exploring solutions based on artificial data and modelling (Cao et al. 2011; Dimanche et al. 2017; Sosnowska and Skibski 2018), and the results serve as contributions to potential further development. Experiments based on artificial data and Experiments based on real data include studies that show applied experiments. Those studies often first propose a conceptual/theoretical method that applies AI for solving certain PT-related problems, and then they test their methods, based on either artificial data or real data, such as registered PT vehicle positions, actual arrival/departure times, or the number of boarding/alighting passengers at different stations (Bahuleyan and Vanajakshi 2017; Berbey Alvarez et al. 2015). Very few cases are implemented in large or small scale. The main difference between the implementation in large scale and small scale is related to both the development status of the application and the scale of the implementations. The applications implemented in a small scale are at an earlier development stage, implemented in one or two empirical cases through research or pilot projects. They are tested in the real world for improving and validating the application (Sykes et al. 2019), as well as demonstrating the need for improving/developing the applications (Mackett 1994). The applications implemented in a large scale are at a later stage of the development and implemented in multiple cases in a large scale, such

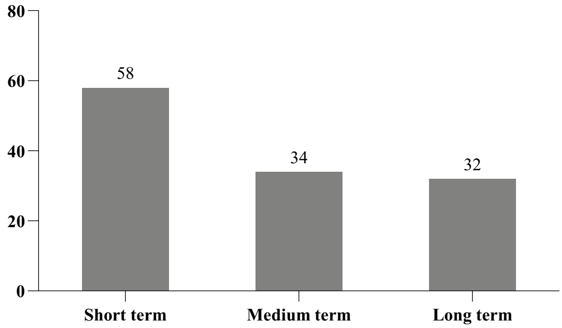


Fig. 3 Number of studies addressing different time horizons

as an expert system for station management which was tested in the railway system of Hongkong. Its feasibility was proved by the implementation which supported the validation in the next step (Chang 1996). Figure 4 shows the number of the studies of each validation category.

For the *Implemented in business* category, three classes were identified: *Broadly used*, *Single example* and *Nothing reported*. The class *Broadly used* refers to applications that have been broadly used in business for supporting certain decisions (Scemama 1995), while *Single example* includes applications that have only been used in business in a single case (Bocchetti et al. 2009; Mackett 1994). The rest of the studies either specifically state that the applications have not been used in business or they do not report anything about whether the applications were used in business or not (Bembalkar and Game 2019; Manivannan et al. 2020; Ubbels and Nijkamp 2002). These studies have been classified as *Nothing reported*. As shown in Fig. 5, a very limited number of services/applications have been reported as used in business. However, this may not indicate that the rest of the services are not used in practice, since this issue remains unreported in most studies.

3.5 Mechanism for achieving benefit

To better understand the way AI was applied in the different studies, we used a category concerning the mechanisms AI provided for achieving the benefits. Figure 6 presents the number of studies addressing these mechanisms. We have classified the studies into the following classes: *Current status estimate, Prediction, Planning/scheduling/resource allocation.* Here, the rather common case of using pattern recognition is included in the *Current status estimate* class. An example of image recognition concerns recognition of traffic signs (Sykes et al. 2019), but also other data sources, such as, smart car data can be used in the current status estimate (Zhang and Cheng 2018). Predictions concern estimates of future states, typically including an ML model, for instance, predicting the travel time based on vehicle positions (Bahuleyan and Vanajakshi 2017) or travel behaviour (Chapleau et al. 2019). The literature review only included five studies focusing on automation, all the other focused on decision support. Two of these studies were classified as Planning/scheduling/resource allocation, being a case of controlling traffic lights (Cao et al. 2011)

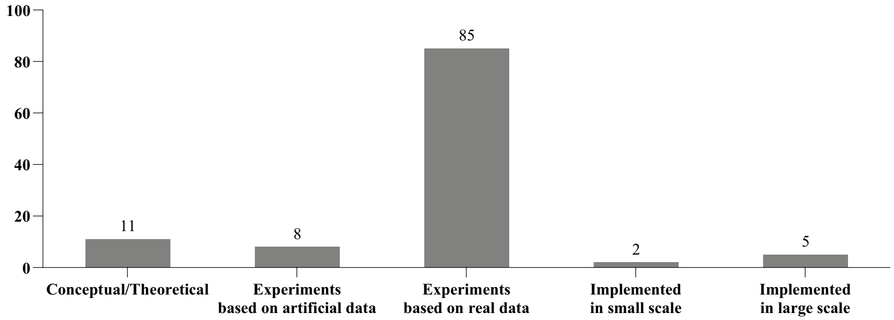


Fig. 4 Number of studies addressing different means of benefit validation

and the other about automatic control of backup of surveillance videos (Cui et al. 2020) and one study classified as *Prediction* being a case of information messaging (Genser et al. 2020), and two classified as *Current status estimate*, being a case of Covid 19 detection (Liu and Huang 2020) and one being a case of smart fare collection (Mastalerz et al. 2020). AI appears rather evenly distributed between the classes, although prediction is the most commonly used mechanism, whereas the more advanced tasks of planning/scheduling/resource allocation are the least common. Note, however, that in many studies classified as *Current status estimate* and *Prediction* the ambition of the studies are on a more advanced tasks, e.g., to plan, but this is then done with other methods, typically by a domain expert/planner.

3.6 AI technology

The most commonly used AI technology is *ML*, which is applied in 77% of the studies. ML is used for generalizing from large amounts of data. The result is typically a classifier that can be used to predict or estimate some unknown value, e.g., related to travel behaviour, traffic flows, land use, predictive maintenance, number of passengers in a vehicle, or estimated arrival time. Many different ML algorithms have been applied, such as Neural Networks, Support Vector Machines, K-Nearest Neighbours, Decision Trees, Random Forests, Naïve Bayes, and Rough Set Analysis. In many of the studies, different ML algorithms are tested and compared, in order to identify the one best suited for the particular aim. However, there is no clear consensus which algorithms are most appropriate for the different tasks.

Another technology often used in the reviewed studies is *Heuristic Search*[1], such as the A* algorithm, Genetic Algorithms, and Ant Colonies. These algorithms have been used for different optimization tasks, such as scheduling, routing, finding shortest paths in PT networks, and vehicle allocation, as well as, optimizing the PT system as a whole.

AI technologies for *Reasoning* are typically based on some kind of formal logic, e.g., fuzzy logic or expressed as if-then rules in an *Expert System*. They are used to

[1] Note that the term AI is often not used when addressing *Heuristic Search* methods.

Fig. 5 Number of studies used in business to different extents

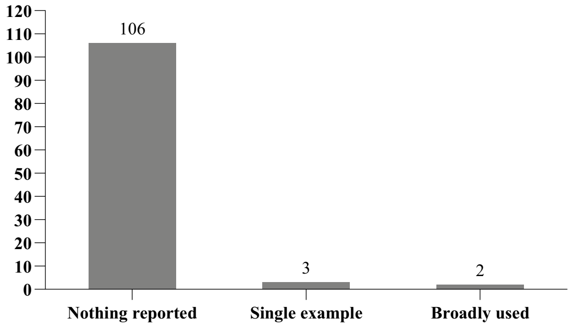
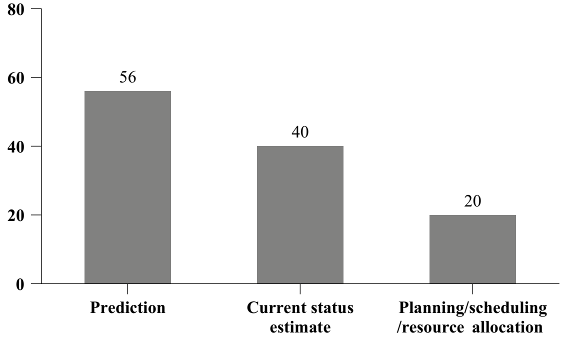


Fig. 6 Number of studies addressing different mechanisms for achieving the benefits



automatically draw conclusions based on known facts, e.g., which congestion control action to choose based on the current congestion pattern.

Multi-Agent Systems (*MAS*) is an AI technology in which several intelligent entities are typically collaborating to manage a complex task, such as traffic control. However, we found only one application of MAS (Cao et al. 2011); one reason for this could be that the term AI is not always used when describing MAS applications.

There has been a clear current trend over recent years to apply ML to solve PT problems, whereas *Reasoning* mainly was used in the early work of applying AI in PT. Figure 7 shows the number of studies addressing each type of AI technology.

3.7 Data needs

Based on the data used in the different studies, 17 subclasses were identified. To get a clearer picture of the relations between the data and the different parts/parties involved in a PT system, these 17 subclasses were then grouped into 4 main classes. This resulted in the following main classes and subclasses:

- Data connected to travellers:
  - Number/ID of boarding/alighting passengers at different stations (e.g., passenger flow between stations).
  - Station congestion/demand levels (e.g., passenger flows at a station).

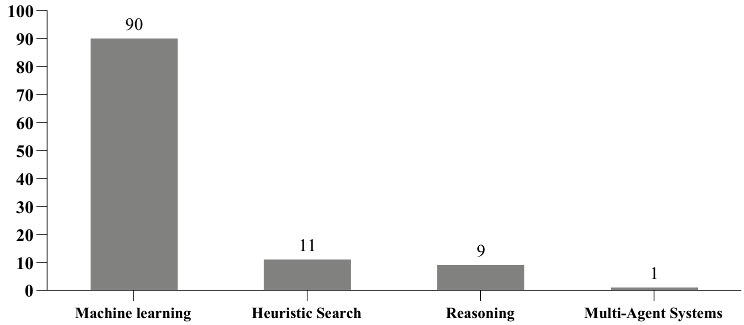


Fig. 7 Number of studies addressing different types of AI technologies

  - Traveller characteristics/behaviour/opinions (e.g., personal trip characteristics, purpose and duration of trips, or personal costs).
  - People's positions/acceleration/etc. (passenger movement data)
  - Surveillance on in-vehicle passengers (e.g., passenger position and movement inside vehicle).
  - Fare transactions and journey searches (e.g., fare records generated through different ticketing media used by the traveller, or renting and return records of bike sharing system).

- Data connected to PT system:

  - PT vehicle positions/acceleration/etc. and actual arrival/departure times (vehicle movement data).
  - Surveillance on PT system and in-vehicle conditions (e.g., video and audio surveillance data from PT system, or CO2 concentration in bus).
  - Vehicle capacity.
  - Timetables and structure of PT system (e.g., interchanges and travel times, or number of stations and vehicles along a route).
  - PT organization/funding/marketing/services/maintenance (e.g., records of frequency and size of maintenance parts orders, or PT marketing and partnerships).

- Data connected to outdoor environment:

  - Local weather conditions (e.g., weather conditions when making a trip).
  - Local built and natural environment as well as city regulations (e.g., natural environment around households, geographical positions of different bike sharing stations, or city politics and regulations).

- Data connected to roads and private cars:

  - Private car positions.
  - Conditions/traffic volumes/speeds on roads (e.g., road condition of an intersection, road congestion, or freeway speeds and volumes).
  - Signal light state (signal light state of an intersection).
  - Road traffic incidents.

Depending on the timeframe and the data collection method, the data were also classified into *historical data* or *real-time data*, as well as *sensor data*, *questionnaire data*, or *documental data*. Studies that express a need for collecting and using data in real time (or near real time) are considered as using real-time data, whereas studies that use data collected in the past are considered as using historical data. Real-time data include passenger movements, surveillance data on the PT system or in-vehicle passengers, bus positions or arrival times etc. (Elizalde-Ramírez et al. 2019; Bocchetti et al. 2009; Belapurkar et al. 2018; Prashanth et al. 2016; Borodinov and Myasnikov 2020a), whereas historical data include historical passenger flow between stations, area maps, household characteristics, etc. (Berbey et al. 2012; Hu et al. 2016; Hagenauer and Helbich 2017). Naturally, a study may use both historical and real-time data (Agafonov and Yumaganov 2019). Sensor data and questionnaire data represent data that have been collected using different types of sensors (e.g., position data, temperature data, video data) or questionnaires (e.g., concerning user behaviour, conditions and opinions). Documental data originate from different public officials, administrative officers, or other office-workers. This class includes, for instance, characteristics and locations of different Park-and-Ride stations, and timetable and structure information from the PT system (Ferrara et al. 2019; Mayaud et al. 2019). Data collected from social media are also included in this class (Kulkarni et al. 2018).

The results of the classification show that the most commonly used data belong to the following three subclasses: Number/ID of boarding/alighting passengers at different stations; PT vehicle positions/acceleration/etc. and actual arrival/departure times; and Timetables and structure of PT system (see Appendix B). This can be interpreted as the data from these subclasses being the most useful for AI applications within PT. However, this result may also reflect the accessibility of the data, i.e., this type of data is probably more easily accessed than, for instance, data related to people's positions. Therefore, data from these subclasses are more commonly used in the studies.

Figure 8 illustrates the main results of the data needs classification. As can be seen, a relatively large share of the data used are connected to the travellers. This means that AI is not only applied for applications focused solely on the PT system, but many applications also relate to the travellers and their behaviour, characteristic, needs, etc. Furthermore, many studies use data that have been collected at an earlier point in time, i.e. historical data. Even though some studies use both historical and real-time data, one conclusion that can be drawn is that AI is mostly used to provide support for decisions that do not depend on real-time data. However, another conclusion is that by opening up to more real-time data, many more AI-applications may be enabled. Figure 8 also shows that multiple studies depend on sensor data, either if it is used in real time or not, i.e., sensors are strongly needed to enable AI applications within PT.

Note that not all studies that focus on real-time support require real-time data. For instance, some of the studies use data that are unspecified by the authors (Dimanche et al. 2017; Molina 2005; Yu et al. 2018), and some use historical data for real-time support (e.g., calculating optimal route choice for the traveller based on information about timetables and structure of the PT system (Song et al. 2015) or predicting the PT delay based on historical delay and weather (Leung et al. 2020).

3.8 Applications in public transport

The application of AI technology in PT covers many different problem areas. We have identified four main classes corresponding to different application areas. Figure 9 presents the number of studies in each main class. The most studied area concerns different ways of *travel service improvement*, including the perspectives of both the operators and the travellers. The estimation of arrival time and support for route choices are the most common applications of this kind, but there are also examples including an improved understanding of the travellers' preferences and state, as well as the integration with other services such as bike sharing. The following applications related to travel service improvement have been found in the reviewed studies:

- Estimating travel/arrival time (Agafonov and Yumaganov 2019; Bahuleyan and Vanajakshi 2017; Biyani 2019; Grzenda et al. 2020; Heghedus 2017; Heghedus et al. 2019; Kyaw et al. 2019; Reddy et al. 2016; Olczyk et al. 2017; Leung et al. 2020; Pandurangi et al. 2020; Tran et al. 2020; Yang et al. 2020a; Yuan et al. 2020).

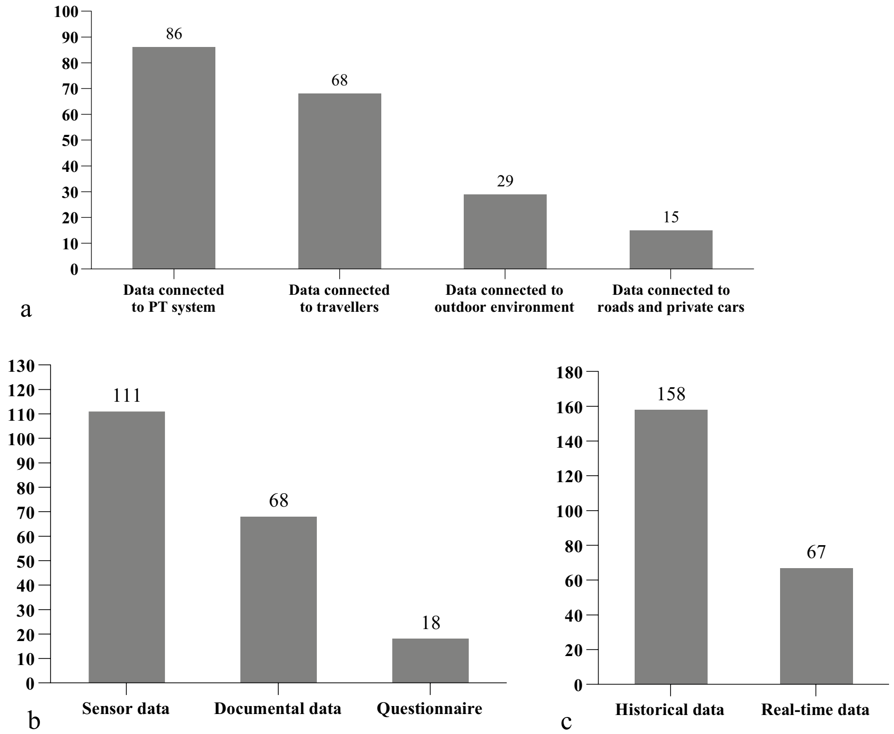


Fig. 8 Number of studies requiring **a** data from the different main classes, **b** different means of data collection, and **c** different data timeframes

- Supporting route choice (Elizalde-Ramírez et al. 2019; Nachtigall 1995; Prashanth et al. 2016; Song et al. 2015; Manivannan et al. 2020; Amrani et al. 2020).
- Balancing and availability of rental bikes (Lin et al. 2018; Singla et al. 2015; Wang and Kim 2018; Yang et al. 2020b; Bei et al. 2020).
- Capturing travellers' opinions (Kulkarni et al. 2018; Lock and Pettit 2020; Othman et al. 2019; Raflesia et al. 2018; Rahimi et al. 2020).
- Traveller recommender systems (Borodinov and Myasnikov 2019, 2020a, b).
- Improving communication with travellers (Yu et al. 2018; Kuberkar and Singhal 2020; Sykes et al. 2019; Velosa and Florez 2020).
- Monitoring traveller's state (Belapurkar et al. 2018; Liu and Huang 2020).

The second-most studied application area concerns *operations support*, including real-time support for different monitoring, diagnosis and planning tasks. These applications concern both the traffic and vehicles, as well as the travellers. For instance, some applications support traffic monitoring or vehicle tracking whereas others estimate the number of travellers in a vehicle. The following applications have been found within this area, in the reviewed studies:

- Supporting dispatching (Dimanche et al. 2017; Moreira-Matias et al. 2016; Wang et al. 2019; Degeler et al. 2020).
- Supporting diagnosis, prediction and planning (Blandin et al. 2019; Molina 2005).
- Supporting emergency management (Chang 1996).
- Supporting maintenance (Adamson et al. 2005; Killeen et al. 2019; Hermann et al. 2020).
- Vehicle tracking (Barbosa et al. 2017; Wilkowski et al. 2020).
- Monitoring passenger flows (Haq et al. 2020; Paletta et al. 2005).
- Traffic monitoring (Scemama 1995; Wei et al. 2017).
- Traffic management (Cao et al. 2011; Minea et al. 2019; Genser et al. 2020; Ayman et al. 2020).
- Security surveillance (Bocchetti et al. 2009; Rohit 2020).
- Estimating/predicting the number of travellers in a vehicle/station (Li et al. 2018a; Pasini et al. 2019; Liu et al. 2020; Skhosana et al. 2020).

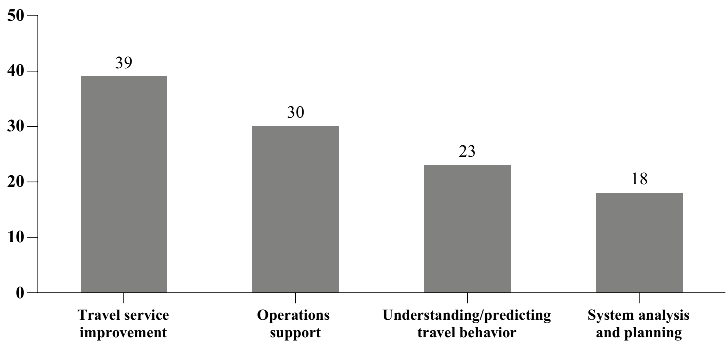


Fig. 9 Number of studies addressing different application areas of AI technology in PT

- Estimating real-time on-board bus ride comfort (Nguyen et al. 2021).
- Fraud detection (Claiborne and Gupta 2018).
- Optimizing when to backup surveillance video files (Cui et al. 2020).
- Fare collection (Mastalerz et al. 2020).

The third application area concerns the *understanding/predicting travel behaviour*, such as travel patterns, mode choice, and how the built environment affects travel behaviour. The following applications have been found within this area, in the reviewed studies:

- Understanding travel patterns (Berbey et al. 2012; Ghaemi et al. 2015; Jung and Sohn 2017; Kedia et al. 2017; Sun and Yang 2018; Tang et al. 2020; Xue et al. 2014; Yu et al. 2015; Zhang et al. 2019; Shalit et al. 2020).
- Understanding/predicting the travellers' mode choice (Chapleau et al. 2019; Ferrara et al. 2019; Hagenauer and Helbich 2017; Lazar et al. 2019; Liang et al. 2019; Niklas et al. 2020; Tu et al. 2016; Victoriano et al. 2020; Zhou et al. 2019).
- Understanding the effects of the built environment on travel behaviour (Deng and Yan 2019; Hu et al. 2016).
- Predicting passenger dwelling time and flow (Berbey Alvarez et al. 2015).
- Predicting travellers' social demographics (Zhang and Chen 2018).

A slightly less studied application area concerns support for long-term PT *system analysis and planning*. This area includes, for instance, support for determining the appropriate PT system (e.g., for PT system planning in a city), optimizing or simulating timetable scheduling, and evaluating a PT system, from different aspects. The following applications have been found within the area, in the reviewed studies:

- Supporting PT system planning (Degeler et al. 2020; Leprêtre et al. 2019; Mackett 1994, 1996; Roulland et al. 2014; Shatnawi et al. 2020; Ullón et al. 2020; Van Egmond et al. 2003).
- Timetable scheduling support (Bembalkar and Game 2019; Othman and Tan 2018; Tan et al. 2011; Tekin et al. 2018; Xie et al. 2004).
- Analysing PT systems (Mayaud et al. 2019; Sosnowska and Skibski 2018).
- Generating synthetic data (Golubev et al. 2016).
- Predicting ticket prices (Aditi et al. 2020; Branda et al. 2020).

3.9 Challenges of using AI

In most of the reviewed studies, the challenges encountered when applying the AI methods were not discussed at all. The challenges that actually have been identified mainly concern data availability and quality. For many AI methods, in particular ML, large datasets, often from different sources, are needed to get good results, and they may require substantial efforts to collect and be difficult to store (Mastalerz et al. 2020; Nguyen et al. 2021; Reddy et al. 2016; Roulland et al. 2014; Tran et al. 2020).

Moreover, some key types of data can be difficult to obtain at all, and sometimes the quality of the data is too low to be useful (Kulkarni et al. 2018; Mastalerz et al. (2020; Nguyen et al. 2021; Rahimi et al. (2020; Shalit et al. (2020; Tran et al. (2020; Yuan et al. (2020). Also, there is often a need to pre-process the data before it can be used, which, together with the processing of the data, may require a substantial amount of resources; in particular, if time is a limiting factor (Jung and Sohn 2017; Sykes et al. 2019; Tran et al. (2020). Much data related to PT concern individual travellers in one way or another, which impose the challenge of how to avoid compromising the personal privacy of the travellers (Ferrara et al. 2019). Finally, to apply AI methods often requires significant AI knowledge and skills, e.g., it could be difficult to select the most suitable AI method for the problem at hand (Ghaemi et al. 2015; Wang and Kim 2018).

3.10 Cross analysis and trends

The reviewed studies cover a broad variety of AI applications for PT, particularly in terms of benefit areas, transport modes, data needs, and the type of applied AI technologies. After having classified the studies according to the characteristics of each category above, we carried out a cross analysis to synthesize the knowledge across all the categories. This section presents the main cross analysis results. Moreover, a histogram of the publication years of the selected studies is presented.

Regarding which applications use which AI technologies, all four application areas identified in Sect. 3.8 were represented among the applications using *Machine learning*. *Heuristic Search* and *Reasoning* were also used for most of the identified application areas, apart from *Understanding/predicting travel behaviour* and *Travel service improvement*. Moreover, by analysing the classification from the time horizon and addressed benefits perspectives, a number of relations could be extracted. Studies concerning the *Long-term* time horizon mainly contribute to the *General understanding* and *Environmental* benefits, whereas benefits of improving *Safety/security* are not addressed in this time horizon at all. The benefits of the long-term planning can be summarized into the following five areas: (1) transport planning and urban planning in relation to build environment and land use planning; (2) future PT route planning based on the understanding of predicted mode choices related to the individual characteristics, e.g., household size, demographic information, trip attributes; (3) optimization of PT fairs and funding mechanism; (4) developing policy indications for reforming in the regulatory framework of the organizations; (5) supporting the improvement of the maintenance system, for instance, by improving the re-ordering strategies for a multiplicity of different supply parts in relation to their usage and consumption rates, hence a better forecasting of the item consumption can be reached. The studies concerning *medium-term* time horizon primarily intend to achieve *Service quality improvement, General understanding* and *Cost saving* related to: (1) improving the general understanding of the PT system and travellers; (2) optimizing the service attributes such as scheduling and route planning; (3) evaluation of the services; and (4) predicting the electronic fare fraud by detecting indication of fraud in fare transaction records. The studied *real-time*-related solutions support all types of benefits, to different extents. The supported decisions mainly focus on improving the service quality for the passengers in the form of reducing travel time and transferring time, which is attained, for example, by providing a method for identifying a transit station for passengers to switch from private to public transport along the traveller's trajectory, or by improving the efficiency of bicycle distribution from a bike sharing system. Real-time horizon support is also explored by developing methods for improving the PT information system to enable passengers to access the information more easily and to plan their trips based on more dynamic and accurate information. As for the different levels of maturity, it is worth noting that the implemented services are real-time and long-term-related supported services, while none of the medium-term-related services has been implemented.

As for data needs, most of the services aiming for *Service quality improvement* and *Safety/security* require real-time data, as opposed to services aiming for *General understanding, Environmental* benefits or *Cost savings*, which most often do not require real-time data. These results indicate that many of the former services provide real-time decision support, whereas most of the latter do not. Furthermore, we conclude that access to historical data is a requirement for most of the AI-based services within PT. In particular, all applications, except one, in this study aiming for *Cost savings* and *General understanding* require historical data. Figure 10 shows the relationships between the main classes of data needs and the AI technologies used in the studies. As can be seen, studies using *Machine Learning* and *Heuristic Search* need data from all classes, where *Data connected to the PT system* is most extensively used. Studies focussing on Reasoning need *Data connected to travellers* most extensively, whereas *Multi-Agent system* studies need *Data connected to outdoor environment* and *Data connected to roads and private cars* to an equal extent. However, the number of studies in these latter two AI-classes are too few to actually draw any conclusions.

Finally, to illustrate the change over time concerning research interests in this area, a histogram of the publication years of the studies included in this literature review, is presented in Fig. 11. Surprisingly, only two studies were published in the years 2006 to 2010. As can be seen, all application areas have gained increased research interests over the past 10 years. Travel service improvement has increased the most and is currently also receiving the most research attention, followed by *Understanding/predicting travel behaviour*. These results indicate a clear focus on the traveller, and less focus on operational and systems support.

4 Discussion and conclusions

This study reviewed more than 87 scientific publications which describe a broad variety of applications of AI in PT, particularly in terms of what benefits they aim for, which transport modes they concern, which data sources are used, and what type of AI technologies are applied. The method used to select what articles to include have some limitations. For instance, some relevant articles are not indexed in any of the three databases but meet the other criteria (e.g., Kumar et al. (2014); Palacio (2018); Shakeel et al. (2019)). Others do not use "Public transport" or "Mass transit" or "Public Transit" in the title or abstract (e.g. Toqué et al. (2017)). Similarly, some articles do not use "AI" or "Artificial Intelligence" or "Machine Learning" in the title or abstract (e.g. Berlingerio et al. (2013)). However, we believe that our selection of articles is representative with respect to the work in the field of AI for public transport.

The review shows that the interest of using AI in PT appears to be growing, given the rapidly increasing number of studies during the last couple of years. This trend strengthens the hypothesis that AI may have great potential to improve PT. Further, the reviewed studies propose several types of AI solutions for different application areas, tasks and decision makers (including travellers). This wide scope of AI usage in this domain can also be seen as an indication of great potential. Finally, a large portion of the studies use real data, which points at the possibility of actually obtaining data for AI-applications. However, very few of the research studies provide solid evidence of improvements of PT, since they are mainly experimental and without results from being implemented in business (at least not reported at the date of publication). This might indicate that AI has been applied in real life only to a small extent; however, perhaps a more likely explanation may be that the extent of AI implementation is not reflected in scientific publications, since most business companies have no interest in writing scientific publications. Nevertheless, given the variety of application areas and the growing interest, we draw the conclusion that there is great potential for using AI to improve PT.

The review also shows that the studies almost exclusively deal with AI for decision support, and not automation. In the near future, however, autonomous PT vehicles will most probably represent a major application area within PT and automation. The reason why these studies have not been captured by our search phrase is probably that they are focussed on transport vehicles in general and not dedicated to PT. Nevertheless, our results indicate that AI in the near future has a significant role to play in supporting human decision makers (and not replacing them). Furthermore, the studies are rather evenly distributed between support for real-time decision making and support for planning (medium term and long term). The purpose is mostly to increase the quality of service of PT services or to increase the knowledge about traveller behaviour, and, to a much lesser extent, achieve direct cost savings. Actually, no study had the primary

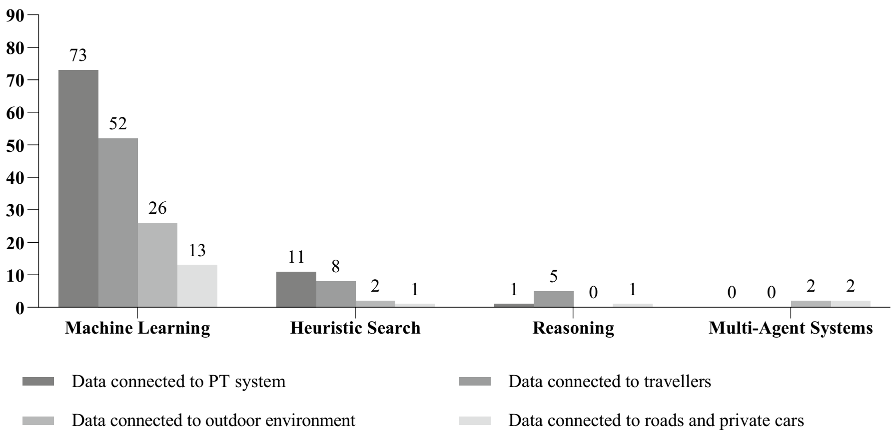


Fig. 10 Relations between Data needs and AI technologies

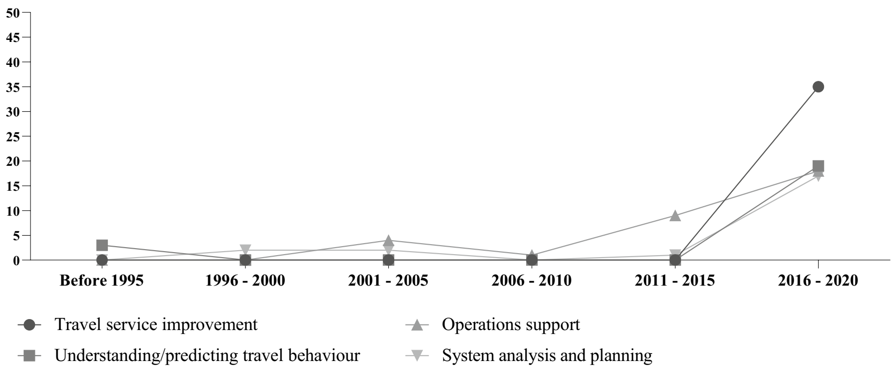


Fig. 11 Number of studies for different periods of years

goal to increase revenue, but this might have been a secondary goal and/or a potential effect. AI is applied for several modes of transport, where train and bus are the dominant modes, but also other transport modes are considered in some studies, as well as the entire PT system. We found that there are three main mechanisms that the AI solutions contribute to, where the most frequent is prediction, followed by estimating the current state and resource allocation, including planning and scheduling. The data used in the AI applications are largely sensor data concerning the traveller and/or the PT system. It is noteworthy that historical data of several types are also used to a great extent and that many applications do not require real-time data at all. Finally, only a few studies discussed challenges that were encountered when applying the AI methods. The challenges that actually were identified mainly concern data availability and data quality, which we think is a good pointer to future research needs; in particular, which data is the most important, how to make the data available and how to improve data quality.

To further analyse our findings, we only find one similar study that our result can compare to; an international study from the industry perspective made by UITP (UITP, 2020). Similar to this review, the UITP study identifies challenges connected to the availability of large datasets and data quality. Additionally, the UITP study identifies challenges connected to general knowledge and capacity of deploying AI, as well as establishing commitment from top management to drive the change for utilizing AI. These challenges were not documented by the publications included in our review.

Although studies have identified challenges connected to data needs and data preprocessing, surprisingly few explicit indications were given concerning the need for labelling data. Potentially, this characterizes the domain, where a lot of data exists. However, the need for labelling data, or supporting supervised learning in general, may increase when new applications of AI for PT are being developed and deployed. That may engage more domain experts in developing AI for supporting their more complex decision making than the ones needed without AI support nowadays.

Appendix

Appendix A: Literature review papers

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Adamson et al. (2005)	Medium term	Cost savings, Service quality improvement	Prediction	Any PT	Experiments based on real data	Nothing reported	PT organization/funding/marketing/services/maintenance	ML	Nothing reported	Supporting maintenance
Aditi et al. (2020)	Short term	Cost savings	Prediction	Railway	Experiments based on real data	Nothing reported	Fare transactions and journey searches	ML	Nothing reported	Predicting ticket prices
Agafonov and Yumaganov (2019)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	Conditions/traffic volumes/speeds on roads, Timetables and structure of PT system, PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Estimating travel/arrival time
Amrani et al. (2020)	Short term, Long term	Service quality improvement	Planning/scheduling/resource allocation	Railway	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations	ML	Nothing reported	Supporting route choice

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Ayman et al. (2020)	Short term	Environmental	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Surveillance data on PT system and in-vehicle conditions, Local weather conditions, Local build and natural environment as well as city regulations, Conditions/traffic volumes/speeds on roads	ML	Nothing reported	Traffic management
Bahuleyan and Vanajakshi (2017)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Conditions/traffic volumes/ speeds on roads	ML	Nothing reported	Estimating travel/arrival time
Barbosa et al. (2017)	Short term	Service quality improvement	Current status estimate	Bus	Experiments based on artificial data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Vehicle tracking

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Bei et al. (2020)	Short term	Cost savings	Prediction	Bike sharing systems	Experiments based on real data	Nothing reported	Fare transactions and journey searches	ML	Nothing reported	Balancing and availability of rental bikes
Belapurkar et al. (2018)	Short term	Safety/security	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Surveillance on in-vehicle passengers, Surveillance on PT system and in-vehicle conditions	ML	Nothing reported	Monitoring traveller state
Bembalkar and Game (2019)	Medium term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	Timetables and structure of PT system, PT vehicle positions/acceleration/etc. and actual arrival/departure times, Fare transactions and journey searches	ML	Nothing reported	Timetable scheduling support
Berbey et al. (2012)	Medium term	Service quality improvement	Prediction	Any PT	Experiments based on artificial data	Nothing reported	Number/ID of boarding/alighting passengers at different stations	Reasoning	Nothing reported	Understanding travel patterns

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Berbey Alvarez et al. (2015)	Medium term, Long term	Cost savings	Prediction	Railway	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Station congestions/ demand levels	Reasoning	Nothing reported	Predicting passenger dwelling time and flow
Blandin et al. (2019)	Short term, Medium term	Service quality improvement	Planning/ scheduling/ resource allocation	Bus, Railway	Experiments based on artificial data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Station congestions/ demand levels, People's positions/ acceleration/etc., PT vehicle positions/acceleration/ etc. and actual arrival/ departure times	ML	Nothing reported	Supporting diagnosis, prediction and planning
Bocchetti et al. (2009)	Short term	Safety/security	Current status estimate	Railway	Implemented in large scale	Single example	Surveillance on PT system and in-vehicle conditions	ML	Nothing reported	Security surveillance

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Borodinov and Myasnikov (2019)	Short term	Service quality improvement	Prediction	Bus, Railway	Experiments based on real data	Nothing reported	Timetables and structure of PT system, People's positions/acceleration/etc., PT vehicle positions/acceleration/etc. and actual arrival/departure times, Fare transactions and journey searches	ML	Nothing reported	Traveller recommend-er systems
Borodinov and Myasnikov (2020a)	Short term	Service quality improvement	Current status estimate	Bus, Railway	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Traveller recommend-er systems
Borodinov and Myasnikov (2020b)	Short term	Service quality improvement	Current status estimate	Bus, Railway	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Traveller recommend-er systems
Branda et al. (2020)	Long term	General understanding	Prediction	Bus	Experiments based on real data	Nothing reported	Fare transactions and journey searches	ML	Nothing reported	Predicting ticket prices
Cao et al. (2011)	Short term	Environmental	Planning/scheduling/resource allocation	Bus	Conceptual/Theoretical	Nothing reported	Conditions/traffic volumes/speeds on roads, Local weather conditions, Signal light state	MAS	Nothing reported	Traffic management

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Chang (1996)	Short term	Safety/security	Current status estimate	Railway	Implemented in large scale	Nothing reported	Surveillance on PT system and in-vehicle conditions	Reasoning	Nothing reported	Supporting emergency management
Chapleau et al. (2019)	Long term	Environmental, General understanding	Prediction	PT and other modes	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions	ML	Nothing reported	Understanding/predicting the traveller mode choice
Claiborne and Gupta (2018)	Medium term	Safety/security	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Fare transactions and journey searches	ML	Nothing reported	Fraud detection
Cui et al. (2020)	Short term	Safety/security	Planning/scheduling/resource allocation	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Optimizing when to backup surveillance video files
Degeler et al. (2020)	Short term, Medium term	Service quality improvement	Prediction	Bus, Railway	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Supporting PT system planning, Traffic management

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Deng and Yan (2019)	Long term	General understanding	Current status estimate	PT and other modes	Experiments based on real data	Nothing reported	Local built and natural environment as well as city regulations	ML	Nothing reported	Understanding the effects of the built environment on travel behaviour
Dimanche et al. (2017)	Short term	Service quality improvement	Current status estimate	Railway	Conceptual/Theoretical	Nothing reported	Unspecified	ML	Nothing reported	Supporting dispatching
Elizalde-Ramírez et al. (2019)	Short term	Service quality improvement	Planning/scheduling/resource allocation	Any PT	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, People's positions/acceleration/etc., Local built and natural environment as well as city regulations, Timetables and structure of PT system	Heuristic Search	Nothing reported	Supporting route choice
Ferrara et al. (2019)	Medium term, Long term	General understanding	Prediction	PT and other modes	Experiments based on real data	Nothing reported	Private car positions, Local built and natural environment as well as city regulations	ML	Personal privacy	Understanding/predicting the traveller mode choice

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Genser et al. (2020)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Conditions/traffic volumes/speeds on roads, Signal light state	ML	Nothing reported	Traffic management
Ghaemi et al. (2015)	Medium term, Long term	General understanding	Prediction	Any PT	Conceptual/Theoretical	Nothing reported	Number/ID of boarding/alighting passengers at different stations	Reasoning	Select most appropriate AI algorithm	Understanding travel patterns
Golubev et al. (2016)	Long term	General understanding	Prediction	Any PT	Experiments based on artificial data	Nothing reported	People's positions/acceleration/etc.	Heuristic Search	Nothing reported	Generating synthetic data
Grzenda et al. (2020)	Short term	Service quality improvement	Current status estimate	Railway	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Estimating travel/arrival time

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Hagenauer and Helbich (2017)	Long term	General understanding	Prediction	PT and other modes	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions, Local weather conditions, Local built and natural environment as well as city regulations	ML	Nothing reported	Understanding/predicting the traveller mode choice
Ullón et al. (2020)	Medium term	Service quality improvement	Current status estimate	Bus	Experiments based on artificial data	Nothing reported	Surveillance on in-vehicle passengers	ML	Nothing reported	Monitoring passenger flows
Heghedus (2017)	Short term, Medium term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Estimating travel/arrival time
Heghedus et al. (2019)	Long term	General understanding	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Estimating travel/arrival time
Herrmann et al. (2020)	Short term	Cost savings	Planning/scheduling/resource allocation	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Supporting maintenance

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Hu et al. (2016)	Long term	General understanding	Prediction	Any PT	Conceptual/Theoretical	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Local built and natural environment as well as city regulations	ML	Nothing reported	Understanding the effects of the built environment on travel behaviour
Jung and Sohn (2017)	Medium term	General understanding	Prediction	Bus	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Local built and natural environment as well as city regulations	ML	Preparing/computing datasets takes time/resources	Understanding travel patterns
Kedia et al. (2017)	Long term	Environmental	Prediction	Bus	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions	Reasoning	Nothing reported	Understanding travel patterns
Killeen et al. (2019)	Medium term	Cost savings	Prediction	Bus	Conceptual/theoretical	Nothing reported	Surveillance on PT system and in-vehicle conditions, PT organization/funding/marketing/services/maintenance	ML	Nothing reported	Supporting maintenance

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Kuberkar and Singhal (2020)	Short term	Service quality improvement	Prediction	Any PT	Conceptual/theoretical	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Improving commu-nica-tion with travellers
Kulkarni et al. (2018)	Medium term	Service quality improvement	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions	ML	Insufficient data quality, Incomplete datasets	Capturing traveller opinions
Kyaw et al. (2019)	Short term	Service quality improvement	Current status estimate	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Estimating travel/arrival time
Lazar et al. (2019)	Medium term, Long term	Environmental, General understanding	Prediction	PT and other modes	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions	ML	Nothing reported	Understanding/predicting the traveller mode choice
Leprêtre et al. (2019)	Long term	Service quality improvement	Planning/scheduling/resource allocation	Bus	Experiments based on artificial data	Nothing reported	Local built and natural environment as well as city regulations	ML	Nothing reported	Supporting PT system planning

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Leung et al. (2020)	Short term	Service quality improvement	Prediction	Railway	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system, Local weather conditions	ML	Nothing reported	Estimating travel/arrival time
Li et al. (2018a)	Medium term, Long term	Service quality improvement	Current status estimate	Bus	Experiments based on real data	Nothing reported	Surveillance on PT system and in-vehicle conditions	ML	Nothing reported	Estimating the number of travellers in a vehicle
Liang et al. (2019)	Long term	General understanding	Prediction	PT and other modes	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions	ML	Nothing reported	Understanding/predicting the traveller mode choice
Lin et al. (2018)	Short term	Service quality improvement	Prediction	Bike sharing systems	Experiments based on real data	Nothing reported	Local weather conditions, Local built and natural environment as well as city regulations, Fare transactions and journey searches	ML	Nothing reported	Balancing and availability of rental bikes
Liu and Huang (2020)	Short term	Safety/security	Current status estimate	Any PT	Implemented in large scale	Broadly used	Traveller characteristics/behaviour/opinions	ML		Monitoring traveller state

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Liu et al. (2020)	Medium term	Service quality improvement, Safety/security	Prediction	Bus	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Estimating/predicting the number of travellers in a vehicle/station
Lock and Pettit (2020)	Medium term	Service quality improvement	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Traveller characteristics, behaviour and opinions, PT vehicle positions and actual arrival/departure times, Timetables and structure of the PT system	ML	Nothing reported	Capturing traveller opinions
Mackett (1994)	Long term	Environmental, Service quality improvement	Planning/scheduling/resource allocation	Any PT	Implemented in small scale	Single example	Unspecified	Reasoning	Nothing reported	Supporting PT system planning
Mackett (1996)	Long term	Environmental	Planning/scheduling/resource allocation	Bus, Railway	Experiments based on real data	Nothing reported	Unspecified	Reasoning	Nothing reported	Supporting PT system planning

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Manivannan et al. (2020)	Long term	Cost savings	Planning/scheduling/resource allocation	PT and other modes	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions, People's positions/acceleration/etc.	ML	Nothing reported	Supporting route choice
Mastalerz et al. (2020)	Short term	Service quality improvement	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, People's positions/acceleration/etc., PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Insufficient data quality, Large datasets required	Fare collection
Mayaud et al. (2019)	Long term	General understanding	Current status estimate	PT and other modes	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions, Local built and natural environment as well as city regulations, Timetables and structure of PT system	ML	Nothing reported	Analysing PT systems
Minea et al. (2019)	Short term	Service quality improvement	Current status estimate	Bus	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations	ML	Nothing reported	Traffic management

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Molina (2005)	Short term	Service quality improvement	Current status estimate, Prediction	Bus	Implemented in large scale	Nothing reported	Unspecified	Reasoning	Nothing reported	Supporting diagnosis, prediction and planning
Moreira-Matias et al. (2016)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Supporting dispatching
Nachtigall (1995)	Short term, Medium term	Service quality improvement	Planning/scheduling/resource allocation	Railway	Experiments based on real data	Nothing reported	Timetables and structure of PT system	Heuristic Search	Nothing reported	Supporting route choice
Nguyen et al. (2021)	Short term	Service quality improvement	Current status estimate	Bus	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions, PT vehicle positions/acceleration/ etc. and actual arrival/departure times	ML	Insufficient data quality, Large datasets required	Estimating real-time onboard bus ride comfort
Niklas et al. (2020)	Long term	General understanding	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Timetables and structure of the PT system, Local built and natural environment as well as city regulations	ML	Nothing reported	Understanding/predicting the traveller mode choice

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Olczyk and Galuszka (2017)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Estimating travel/arrival time
Othman and Tan (2018)	Medium term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	Local weather conditions, Local built and natural environment as well as city regulations, Timetables and structure of PT system	ML	Nothing reported	Timetable scheduling support
Othman et al. (2019)	Medium term	Service quality improvement	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Traveller characteristics, behaviour and opinions	ML	Nothing reported	Capturing travellers opinions
Paletta et al. (2005)	Short term, Medium term, Long term	Service quality improvement	Current status estimate	Bus	Experiments based on real data	Nothing reported	Station congestions/demand levels	ML	Nothing reported	Monitoring passenger flows
Paliwal and Biyani (2019)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Estimating travel/arrival time

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Pandurangi et al. (2020)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Local weather conditions, Conditions/traffic volumes/speeds on roads	ML	Nothing reported	Estimating travel/arrival time
Pasini et al. (2019)	Short term	Service quality improvement	Current status estimate, Prediction	Railway	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations	ML	Nothing reported	Estimating the number of travellers in a vehicle
Prashanth et al. (2016)	Short term	Environmental, Service quality improvement	Prediction	PT and other modes	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Conditions/traffic volumes/speeds on roads, Local weather conditions, Road traffic incidents, Timetables and structure of PT system	ML	Nothing reported	Supporting route choice
Raflesia et al. (2018)	Long term	General understanding	Current status estimate	Railway	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions	ML	Nothing reported	Capturing traveller opinions

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Rahimi et al. (2020)	Short term	Service quality improvement, Safety/security	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions	ML	Incomplete datasets	Capturing traveller opinions
Reddy et al. (2016)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Large datasets required	Estimating travel/arrival time
Rohit (2020)	Short term	Safety/security	Current status estimate	Bus	Experiments based on real data	Nothing reported	Surveillance data on PT system and in-vehicle conditions	ML	Nothing reported	Security surveillance
Roulland et al. (2014)	Medium term	Cost savings	Prediction	Any PT	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations	ML	Data from different sources required	Supporting PT system planning
Scemama (1995)	Short term	Service quality improvement	Current status estimate, Planning/scheduling/resource allocation	PT and other modes	Implemented in large scale	Broadly used	Conditions/traffic volumes/speeds on roads	Reasoning	Nothing reported	Traffic monitoring

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Shalit et al. (2020)	Medium term	General understanding	Prediction	Any PT	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Timetables and structure of PT system, Local build and natural environment as well as city regulations	ML	Incomplete datasets	Understanding travel patterns
Shatnawi et al. (2020)	Medium term	Environmental, Service quality improvement	Planning/scheduling/resource allocation	Bus	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions, Timetables and structure of the PT system, Local built and natural environment as well as city regulations	Search	Nothing reported	Supporting PT system planning
Singla et al. (2015)	Short term	Cost savings	Planning/scheduling/resource allocation	Bike sharing systems	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/opinions, Fare transactions and journey searches	Heuristic Search	Nothing reported	Balancing and availability of rental bikes

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Skhosana et al. (2020)	Short term	Cost savings, Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations	ML	Nothing reported	Estimating/predicting the number of travellers in a vehicle/station
Song et al. (2015)	Short term	Cost savings	Planning/scheduling/resource allocation	Bus	Experiments based on artificial data	Nothing reported	Timetables and structure of PT system	Heuristic Search	Nothing reported	Supporting route choice
Sosnowska and Skibski (2018)	Long term	General understanding	Current status estimate	Bus	Conceptual/Theoretical	Nothing reported	Timetables and structure of PT system	Heuristic Search	Nothing reported	Analysing PT systems
Sun and Yang (2018)	Medium term	Service quality improvement	Planning/scheduling/resource allocation	Bus, Railway	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Traveller characteristics/behaviour/opinions	ML	Nothing reported	Understanding travel patterns
Sykes et al. (2019)	Short term	Service quality improvement	Current status estimate	Bus	Implemented in small scale	Nothing reported	Local built and natural environment as well as city regulations, Timetables and structure of PT system	ML	Preparing/computing datasets takes time/resources	Improving commu-nica-tion with travellers

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Tan et al. (2011)	Medium term	Service quality improvement	Planning/scheduling/resource allocation	Bus	Experiments based on real data	Nothing reported	Station congestions/demand levels, Vehicle capacity, Timetables and structure of PT system	Heuristic Search	Nothing reported	Timetable scheduling support
Tang et al. (2020)	Medium term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Local weather conditions	ML	Nothing reported	Understanding travel patterns
Tekin et al. (2018)	Medium term	Cost savings	Planning/scheduling/resource allocation	Bus, Railway	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Vehicle capacity, Timetables and structure of PT system	Heuristic Search	Nothing reported	Timetable scheduling support
Tran et al. (2020)	Short term	Environmental, Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions and actual arrival/departure times, Conditions, traffic volumes and speeds on roads	ML	Insufficient data quality, Large datasets required, Preparing/ computing datasets takes time/resources	Estimating travel/arrival time

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Tu et al. (2016)	Long term	General understanding	Prediction	PT and other modes	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Traveller characteristics/behaviour/opinions, Timetables and structure of PT system	ML	Nothing reported	Understanding/predicting the traveller mode choice
Ubbels and Nijkamp (2002)	Long term	General understanding	Current status estimate	Any PT	Experiments based on real data	Nothing reported	PT organization/funding/marketing/services/maintenance	ML	Nothing reported	Funding PT systems
Ullón et al. (2020)	Medium term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Supporting PT system planning
Van Egmond et al. (2003)	Long term	General understanding	Current status estimate	Any PT	Experiments based on real data	Nothing reported	Local built and natural environment as well as city regulations, PT organization/funding/marketing/services/maintenance	ML	Nothing reported	Supporting PT system planning

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Velosa and Florez (2020)	Short term	Service quality improvement	Current status estimate	Any PT	Experiments based on artificial data	Single example	Timetables and structure of PT system, Local build and natural environment as well as city regulations	ML	Nothing reported	Improving communication with travellers
Victoriano et al. (2020)	Long term	General understanding	Current status estimate	PT and other modes	Experiments based on real data	Nothing reported	Traveller characteristics/behaviour/ opinions	ML	Nothing reported	Understanding/ predicting the traveller mode choice
Wang and Kim (2018)	Short term	Service quality improvement	Prediction	Bike sharing systems	Experiments based on real data	Nothing reported	Fare transactions and journey searches	ML	Select most appropriate AI algorithm	Balancing and availability of rental bikes
Wang et al. (2019)	Short term	Service quality improvement	Current status estimate	Bus	Experiments based on real data	Nothing reported	Station congestions/ demand levels, Surveillance on PT system and in-vehicle conditions, PT vehicle positions/acceleration/etc. and actual arrival/departure times	ML	Nothing reported	Supporting dispatching

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Wang et al. (2020)	Long term	Environmental	Planning/scheduling/resource allocation	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Surveillance data on PT system and in-vehicle conditions, Local weather conditions	ML	Long term	Environmental
Wei et al. (2017)	Short term	Service quality improvement	Current status estimate, Prediction	Bus	Experiments based on real data	Nothing reported	Conditions/traffic volumes/speeds on roads	Heuristic Search	Nothing reported	Traffic monitoring
Wilkowski et al. (2020)	Short term	Safety/security	Current status estimate	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Surveillance on PT system and in-vehicle conditions	ML	Nothing reported	Vehicle tracking
Xie et al. (2004)	Medium term	Service quality improvement	Planning/scheduling/resource allocation	Bus	Conceptual/Theoretical	Nothing reported	Station congestions/demand levels, Timetables and structure of PT system	Heuristic Search	Nothing reported	Timetable scheduling support
Xue et al. (2014)	Long term	Service quality improvement	Prediction	Bus, Railway	Conceptual/Theoretical	Nothing reported	Number/ID of boarding/alighting passengers at different stations	ML	Nothing reported	Understanding travel patterns

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Yang et al. (2020a)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Estimating travel/arrival time
Yang et al. (2020b)	Short term	Service quality improvement	Prediction	Bike sharing systems	Experiments based on real data	Nothing reported	Fare transactions and journey searches, Local weather conditions	ML	Nothing reported	Balancing and availability of rental bikes
Yu et al. (2018)	Short term	Service quality improvement	Prediction	Bus	Conceptual/Theoretical	Nothing reported	Unspecified	ML	Nothing reported	Improving communication with travellers
Yu et al. (2015)	Medium term	General understanding	Prediction	Bus, Railway	Conceptual/Theoretical	Nothing reported	Number/ID of boarding/alighting passengers at different stations	ML	Nothing reported	Understanding travel patterns

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Yuan et al. (2020)	Short term	Service quality improvement	Prediction	Bus	Experiments based on real data	Nothing reported	PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system, Conditions/traffic volumes/speeds on roads	ML	Incomplete datasets	Estimating travel/arrival time
Zhang et al. (2019)	Medium term	General understanding	Current status estimate	Bus, Railway	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, PT vehicle positions/acceleration/etc. and actual arrival/departure times, Timetables and structure of PT system	ML	Nothing reported	Understanding travel patterns

Paper	Time horizon	Benefit area	Benefit mechanism	Transport mode	Validation	Implemented in business	Data needs	AI technology	Challenges	Applications in PT
Zhang and Cheng (2018)	Medium term, Long term	General understanding	Current status estimate	Bus, Railway	Experiments based on real data	Nothing reported	Number/ID of boarding/alighting passengers at different stations, Traveller characteristics/behaviour/opinions	ML	Nothing reported	Predicting traveller social demographics
Zhou and Wang (2019)	Long term	General understanding	Prediction	Bike sharing systems	Experiments based on real data	Nothing reported	Fare transactions and journey searches, Local built and natural environment as well as city regulations	ML	Nothing reported	Understanding/predicting the traveller mode choice

Appendix B: Number of papers using data from the different main classes, subclasses, time frames and data collection methods

The table below shows the resulting classification of the data needs, with the four main classes furthest to the left and one subclass at each row. Since each study may use data from several subclasses, the sum of studies (198) does not correspond to the number of studies included in the review (111). Moreover, since each study using data from a particular subclass may collect this data by several (sometimes unclear) means and use it in several timeframes, the corresponding sums of studies using data from the different timeframes (158 + 67) and from the different collection means (111 + 18 + 68) do not correspond to the total sum of studies in (198).

	Subclass	Papers	Historical data	Real-time data	Sensor data	Questionnaire	Documental data
Data connected to travellers	Number/ID of boarding/alighting passengers at different stations[a,b]	23	21	5	19	2	
	Station congestion/demand levels[a,b]	6	5	2	3	1	
	Traveller characteristics/behaviour/opinions	20	19	2	1	12	7
	People's positions/acceleration/etc[b]	6	4	5	6		
	Surveillance on in-vehicle passengers	2	1	1	2		
	Fare transactions and journey searches[a,b,c]	11	10	2	5		8
	Sum	55	50	11	28	15	10
Data connected to PT system	PT vehicle positions/acceleration/etc. and actual arrival/departure times[a,b]	37	20	28	36		

	Subclass	Papers	Historical data	Real-time data	Sensor data	Question-naire	Documental data
	Surveillance on PT system and in-vehicle conditions	10	5	5	10		
	Vehicle capacity	2	2				2
	Timetables and structure of PT system	33	33				33
	PT organization/funding/marketing/services/maintenance	4	4				4
	Sum	60	42	24	31	0	28
Data connected to outdoor environment	Local weather conditions[b]	11	9	4	9	1	1
	Local built and natural environment as well as city regulations[b,c]	18	17	2	6	2	12
	Sum	22	19	5	12	3	9
Data connected to roads and private cars	Private car positions	1	1		1		
	Conditions/traffic volumes/speeds on roads[b,c]	11	6	8	11		1
	Signal light state[b]	2	1	2	2		
	Road traffic incidents[a]	1		1			
	Sum	10	3	7	9	0	1
	Total sum	198	158	67	111	18	68

[a]One or several papers have unlcear data collection

[b]One or several papers use both real-time and historical data

[c]One or several papers use both sensor and documental data

Acknowledgements This research has been partially funded by K2 - The Swedish Knowledge Centre for Public Transport, and has been partially carried out within the project AI Driven Mobility funded by VINNOVA.

Author contribution All authors have contributed equally to the literature review and the construction of the classification framework. All authors have contributed to the writing of the paper. Å.J. has taken the main responsibility of the final editing. All authors have read and agreed to the submitted version of the manuscript.

Funding Open access funding provided by Malmö University. This research has been partially funded by K2 - The Swedish Knowledge Centre for Public Transport, and has been partially carried out within the project AI Driven Mobility funded by VINNOVA.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

Abduljabbar R, Dia H, Liyanage S, Bagloee SA (2019) Applications of artificial intelligence in transport: an overview. Sustainability 11(1):189, 1–24. https://doi.org/10.3390/su11010189

Adamson K, Campbell P, Orsoni A (2005) Hybrid Decision Support Based on Knowledge Discovery and AI Techniques for the Management of Maintenance Services in the Public Transport Sector. Proceedings of 2005 IEEE Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 674–678. https://doi.org/10.1109/IDAACS.2005.283071

Aditi AD, Dureja A, Abrol S, Dureja A (2020) Prediction of Ticket Prices for Public Transport Using Linear Regression and Random Forest Regression Methods: A Practical Approach Using Machine Learning. In: Batra, U., Roy, N., Panda, B. (eds) Data Science and Analytics, Springer, 140–150. https://doi.org/10.1007/978-981-15-5827-6_12

Agafonov AA, Yumaganov AS (2019) Performance comparison of machine learning methods in the bus arrival time prediction problem. Proceedings of CEUR Workshop, 57–62. https://doi.org/10.18287/1613-0073-2019-2416-57-62

Amrani A, Pasini K, Khouadjia M (2020) Enhance Journey Planner with Predictive Travel Information for Smart City Routing Services. 2020 Forum on Integrated and Sustainable Transportation Systems (FISTS), 304–308. https://doi.org/10.1109/FISTS46898.2020.9264859

Ayman A, Wilbur M, Sivagnanam A, Pugliese P, Dubey A, Laszka A (2020) Data-Driven Prediction of Route-Level Energy Use for Mixed-Vehicle Transit Fleets. 2020 IEEE International Conference on Smart Computing, 41–48. https://doi.org/10.1109/SMARTCOMP50058.2020.00026

Bahuleyan H, Vanajakshi LD (2017) Arterial path-level travel-time estimation using machine-learning techniques. J Comput Civil Eng 31(3):04016070. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000644

Bar-Ilan J, Levene M, Lin A (2007) Some measures for comparing citation databases. J Informetr 1(1):26–34. https://doi.org/10.1016/j.joi.2006.08.001

Barbosa R, Cardoso DO, Carvalho D, França FM (2017) A neuro-symbolic approach to GPS trajectory classification. Proceedings of European Symposium on Artificial Neural Networks, 411–416

Bazzan AL, Klügl F (2014) A review on agent-based technology for traffic and transportation. Knowl Eng Rev 29(3):375. https://doi.org/10.1017/S0269888913000118

Bei Y, Ge Y, Zhang D (2020) A machine learning based shared bikes scheduling method. Proceedings of the 2020 4th International Conference on Cloud and Big Data Computing, 32–36. https://doi.org/10.1145/3416921.3416938

Belapurkar N, Harbour J, Shelke S, Aksanli B (2018) Building Data-Aware and Energy-Efficient smart spaces. IEEE Internet of Things J 5(6):4526–4537. https://doi.org/10.1109/JIOT.2018.2834907

Bembalkar R, Game P (2019) Infrastructure cost reduction of Municipal Public Transport using machine learning. Int J Sci Technol Res 8(12):2104–21074

Berbey A, Galán R, Bobi SJD, Caballero R (2012) A fuzzy logic approach to modelling the passengers' flow and dwelling time. WIT Trans Built Environ 128:359–369. https://doi.org/10.2495/UT120311

Berbey Alvarez A, Merchan F, Calvo Poyo FJ, Caballero George RJ (2015) A fuzzy logic-based Approach for Estimation of Dwelling Times of Panama Metro Stations. Entropy 17(5):2688–2705. https://doi.org/10.3390/e17052688

Berlingerio M, Calabrese F, Di Lorenzo G, Nair R, Pinelli F, Sbodio ML (2013) AllAboard: A System for Exploring Urban mobility and optimizing Public Transport using Cellphone Data. In: Blockeel H, Kersting K, Nijssen S, Železný F (eds) Machine learning and knowledge Discovery in Databases. Lecture Notes in Computer Science, vol 8190. Springer, Berlin, Heidelberg, pp 663–666. https://doi.org/10.1007/978-3-642-40994-3_50

Biyani P (2019) To each route its own ETA: A generative modeling framework for ETA prediction. Proceedings of 2019 IEEE Intelligent Transportation Systems Conference (ITSC). arXiv preprint arXiv:1906.09925

Blandin S, Wynter L, Poonawala H, Laguna S, Dura B (2019) FASTER: Fusion AnalyticS for public Transport Event Response. Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, 1404–1412

Bocchetti G, Flammini F, Pragliola C, Pappalardo A (2009) Dependable integrated surveillance systems for the physical security of metro railways. 2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), 1–7. https://doi.org/10.1109/ICDSC.2009.5289385

Borodinov AA, Myasnikov VV (2019) Analysis of the preferences of public transport passengers in the task of building a personalized recommender system. Proceedings of CEUR Workshop Proceedings. 198–205

Borodinov AA, Myasnikov VV (2020a) Evaluating classifiers to determine user-preferred stops in a personalized recommender system. Twelfth International Conference on Machine Vision (ICMV 2019), 11433, 114330 N. https://doi.org/10.1117/12.2556536

Borodinov AA, Myasnikov VV (2020b) Method of Determining User Preferences for the Personalized Recommender Systems for Public Transport Passengers. International Conference on Analysis of Images, Social Networks and Texts, 341–351. https://doi.org/10.1007/978-3-030-39575-9_34

Branda F, Marozzo F, Talia D (2020) Ticket sales prediction and dynamic pricing strategies in Public Transport. Big Data Cogn Comput 4(4):36. https://doi.org/10.3390/bdcc4040036

Cao X, Dong D, Zeng X (2011) Application of Agent in Bus Signal Priority Intersection. Proceedings of 2011 Tenth International Symposium on Autonomous Decentralized Systems, 276–280. https://doi.org/10.1109/ISADS.2011.37

Chang CS (1996) Re-engineering the Station management processes in Hong Kong Mass Transit Railway Corporation. WIT Trans Built Environ 20:269–278. https://www.witpress.com/elibrary/wit-transactions-on-the-built-environment/20/8882

Chapleau R, Gaudette P, Spurr T (2019) Application of machine learning to two large-sample Household travel surveys: a characterization of travel modes. Transp Res Rec 2673(4):173–183. https://doi.org/10.1177/0361198119839339

Chen B, Cheng HH (2010) A review of the applications of agent technology in traffic and transportation systems. IEEE Trans Intell Transp Syst 11(2):485–497. https://doi.org/10.1109/TITS.2010.2048313

Claiborne J, Gupta A (2018) Machine Learning Classifiers for Predicting Transit Fraud. Proceedings of AMCIS 2018. https://aisel.aisnet.org/amcis2018/DataScience/Presentations/37

Cui L, Su D, Zhou Y, Zhang L, Wu Y, Chen S (2020) Edge learning for surveillance video uploading sharing in public transport systems. IEEE Trans Intell Transp Syst 22:1–10

Davidsson P, Henesey L, Ramstedt L, Törnquist J, Wernstedt F (2005) An analysis of agent-based approaches to transport logistics. Transp Res Part C: Emerg Technol 13(4):255–271. https://doi.org/10.1016/j.trc.2005.07.002

Davidsson P, Hajinasab B, Holmgren J, Jevinger Ã, Persson JA (2016) The Fourth Wave of Digitalization and Public Transport: Opportunities and Challenges. Sustainability 8(12):1248. https://doi.org/10.3390/su8121248

Degeler V, Heydenrijk-Ottens L, Luo D, van Oort N, van Lint H (2020) Unsupervised approach towards analysing the public transport bunching swings formation phenomenon. Public Transp 13:533–555. https://doi.org/10.1007/s12469-020-00251-z

Deng Y, Yan Y (2019) Propensity score weighting with generalized boosted Models to explore the Effects of the built environment and residential self-selection on travel behavior. Transp Res Rec 2673(4):373–383. https://doi.org/10.1177/0361198119837153

Dimanche V, Goupil A, Philippot A, Riera B, Urban A, Gabriel G (2017) Massive Railway Operating Data Visualization; a Tool for RATP Operating Expert. IFAC-PapersOnLine 50(1):15841–15846. https://doi.org/10.1016/j.ifacol.2017.08.2324

Elizalde-Ramírez F, Nigenda RS, Martínez-Salazar IA, Ríos-Solís Y (2019) Travel plans in Public Transit Networks using Artificial Intelligence Planning Models. Appl Artif Intell 33(5):440–461. https://doi.org/10.1080/08839514.2019.1582859

Ferrara M, Liberto C, Nigro M, Trojani M, Valenti G (2019) Multimodal choice model for e-mobility scenarios. Transp Res Procedia 37:409–416. https://doi.org/10.1016/j.trpro.2018.12.210

Ge L, Sarhani M, Voß S, Xie L (2021) Review of transit data sources: Potentials, challenges and complementarity. Sustainability 13(20):11450. https://doi.org/10.3390/su132011450

Genser A, Ambühl L, Yang K, Menendez M, Kouvelas A (2020) Time-to-Green predictions: A framework to enhance SPaT messages using machine learning. Paper presented at the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 1–6, https://doi.org/10.1109/ITSC45102.2020.9294548

Ghaemi MS, Agard B, Nia VP, Trépanier M (2015) Challenges in spatial-temporal data analysis targeting public transport. Proceedings of Symposium on Information Control in Manufacturing, 442–447

Golubev A, Chechetkin I, Parygin D, Sokolov A, Shcherbakov M (2016) Geospatial Data Generation and Preprocessing Tools for Urban Computing System Development1. Procedia Comput Sci 101:217–226. https://doi.org/10.1016/j.procs.2016.11.026

Grzenda M, Kwasiborska K, Zaremba T (2020) Hybrid short term prediction to address limited timeliness of public transport data streams. Neurocomputing 391:305–317. https://doi.org/10.1016/j.neucom.2019.08.100

Hagenauer J, Helbich M (2017) A comparative study of machine learning classifiers for modeling travel mode choice. Expert Syst Appl 78:273–282. https://doi.org/10.1016/j.eswa.2017.01.057

Haq EU, Huarong X, Xuhui C, Wanqing Z, Jianping F, Abid F (2020) A fast hybrid computer vision technique for real-time embedded bus passenger flow calculation through camera. Multimed Tools Appl 79(1):1007–1036. https://doi.org/10.1007/s11042-019-08167-y

Heghedus C (2017) PhD Forum: Forecasting Public Transit Using Neural Network Models. 2017 IEEE International Conference on Smart Computing (SMARTCOMP), 1–2. https://doi.org/10.1109/SMARTCOMP.2017.7947031

Heghedus C, Chakravorty A, Rong C (2019) Neural Network Frameworks. Comparison on Public Transportation Prediction. 2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 842–849. https://doi.org/10.1109/IPDPSW.2019.00138

Herrmann P, Puka E, Skoglund TR (2020) Machine Learning-based Update-time Prediction for Battery-friendly Passenger Information Displays. 2020 IEEE 8th International Conference on Smart City and Informatization (iSCI), 49–59. https://doi.org/10.1109/iSCI50694.2020.00016

Holzinger A, Kieseberg P, Weippl E, Tjoa AM (2018) Current advances, trends and challenges of machine learning and knowledge extraction: From machine learning to explainable AI. International Cross-Domain Conference for Machine Learning and Knowledge Extraction, 1–8. https://doi.org/10.1007/978-3-319-99740-7_1

Hoonlor A, Szymanski BK, Zaki MJ (2013) Trends in computer science research. Commun ACM 56(10):74–83. https://doi.org/10.1145/2500892

Hu N, Legara EF, Lee KK, Hung GG, Monterola C (2016) Impacts of land use and amenities on public transport use, urban planning and design. Land Use Policy 57:356–367. https://doi.org/10.1016/j.landusepol.2016.06.004

Jung J, Sohn K (2017) Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data. IET Intel Transp Syst 11(6):334–339. https://doi.org/10.1049/iet-its.2016.0276

Kedia AS, Sowjanya D, Salini PS, Jabeena M, Katti BK (2017) Transit shift response analysis through fuzzy rule based-choice model: a case study of indian Metropolitan City. Transp Dev Econ 3(1):8. https://doi.org/10.1007/s40890-017-0038-9

Killeen P, Ding B, Kiringa I, Yeap T (2019) IoT-based predictive maintenance for fleet management. Procedia Comput Sci 151:607–613

Kitchenham B, Charters SM (2007) Guidelines for performing systematic literature reviews in software engineering. Technical report, Ver. 2.3 EBSE

Koushik AN, Manoj M, Nezamuddin N (2020) Machine learning applications in activity-travel behaviour research: a review. Transp Rev 40(3):288–311. https://doi.org/10.1080/01441647.2019.1704307

Kuberkar S, Singhal TK (2020) Factors influencing adoption intention of AI powered chatbot for public transport services within a smart city. Int J Emerg Technol 11(3):948–958

Kulkarni G, Abellera L, Panangadan A (2018) Unsupervised classification of online community input to advance transportation services. Proceedings of 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 261–267. https://doi.org/10.1109/CCWC.2018.8301704

Kumar V, Kumar BA, Vanajakshi LD, Subramanian SC (2014) Comparison of Model Based and Machine Learning Approaches for Bus Arrival Time Prediction. Presented at the Transportation Research Board 93rd Annual Meeting Transportation Research Board, 14-2518.

Kyaw T, Oo NN, Zaw W (2019) Building Travel Speed Estimation Model for Yangon City from Public Transport Trajectory Data. In: Zin TT, Lin JC-W (eds) Big Data Analysis and Deep Learning Applications AISC. Springer, Berlin, pp 250–257. https://doi.org/10.1007/978-981-13-0869-7_28

Lavesson N, Davidsson P (2006) Quantifying the impact of learning algorithm parameter tuning. The 21th National Conference on Artificial Intelligence (AAAI), Vol. 1, 395–400

Lazar A, Ballow A, Jin L, Spurlock CA, Sim A, Wu K (2019) Machine Learning for Prediction of Mid to Long Term Habitual Transportation Mode Use. Proceedings of 2019 IEEE International Conference on Big Data (Big Data), 4520–4524. https://doi.org/10.1109/BigData47090.2019.9006411

Leprêtre F, Fonlupt C, Verel S, Marion V (2019) Combinatorial Surrogate-Assisted Optimization for Bus Stops Spacing Problem. International Conference on Artificial Evolution (Evolution Artificielle), 42–52. https://doi.org/10.1007/978-3-030-45715-0_4

Leung CK, Elias JD, Minuk SM, de Jesus ARR, Cuzzocrea A (2020) An Innovative Fuzzy Logic-Based Machine Learning Algorithm for Supporting Predictive Analytics on Big Transportation Data. 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–8. https://doi.org/10.1109/FUZZ48607.2020.9177823

Li T, Fong S, Yang L (2018a) Counting Passengers in Public Buses by Sensing Carbon Dioxide Concentration: Data Collection and Machine Learning. Proceedings of the 2018 2nd International Conference on Big Data and Internet of Things, 43–48. https://doi.org/10.1145/3289430.3289461

Li T, Sun D, Jing P, Yang K (2018b) Smart card data mining of public transport destination: a literature review. Information 9(1):18. https://doi.org/10.3390/info9010018

Liang L, Xu M, Grant-Muller S, Mussone L (2019) Household travel mode choice estimation with large-scale data—An empirical analysis based on mobility data in Milan. Int J Sustain Transp 1–16:70

Lin F, Jiang J, Fan J, Wang S (2018) A stacking model for variation prediction of public bicycle traffic flow. Intell Data Anal 22(4):911–933. https://doi.org/10.3233/IDA-173443

Liu Q, Huang Z (2020) Research on intelligent prevention and control of COVID-19 in China's urban rail transit based on artificial intelligence and big data. J Intell Fuzzy Syst 39:9085–9090

Liu W, Tan Q, Wu W, Abulkasim H (2020) Forecast and early warning of regional bus passenger flow based on machine learning. Math Probl Eng. https://doi.org/10.1155/2020/6625435

Liyanage S, Dia H, Abduljabbar R, Bagloee SA (2019) Flexible mobility on-demand: an environmental scan. Sustainability 11(5):1262. https://doi.org/10.3390/su11051262

Lock O, Pettit C (2020) Social media as passive geo-participation in transportation planning – how effective are topic modeling & sentiment analysis in comparison with citizen surveys? Geo-spatial Inform Sci 23(4):275–292. https://doi.org/10.1080/10095020.2020.1815596

Mackett RL (1994) Determining appropriate public transport system for a city. Transportation Research Record, 44–44. Retrieved 09.30.2020, from http://onlinepubs.trb.org/Onlinepubs/trr/1994/1451/1451.pdf#page=50

Mackett RL (1996) Modelling the implications of new public transport technology: an approach using artificial intelligence. In: Hayashi Y, Roy J (eds) Transport, Land-Use and the Environment. Springer, US, pp 297–315

Manivannan MS, Kavitha R, Srikanth R, Narayanan V (2020) Suggesting alternate traffic mode and cost optimization on traffic-related impacts using machine learning techniques. intelligent computing in engineering. Springer, Singapore

Market Research Future (2021) Public Transport Market Research Report: Information by Type (Bus, Light Rail, Regional Taxi, Metro and Tram), Application (City and Rural) and Region - Forecast till 2027. Report ID: MRFR/AM/7205-CR

Markets and Markets (2019) Railway System Market by System Type, Transit Type, Application & Region - Global Forecast to 2025. Report ID: 4763771

Mastalerz MW, Malinowski A, Kwiatkowski S, Śniegula A, Wieczorek B (2020) Passenger BIBO detection with IoT support and machine learning techniques for intelligent transport systems. Procedia Comput Sci 176:3780–3793. https://doi.org/10.1016/j.procs.2020.09.009

Mayaud JR, Tran M, Nuttall R (2019) An urban data framework for assessing equity in cities: comparing accessibility to healthcare facilities in Cascadia. Comput Environ Urban Syst 78:101401. https://doi.org/10.1016/j.compenvurbsys.2019.101401

McCarthy J (1998) What is artificial intelligence? Technical Report. Stanford University

McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. AI Mag 27:12–12. https://doi.org/10.1609/aimag.v27i4.1904

Minea M, Dumitrescu C, Chiva I-C, Artificial Intelligence (2019) Unconventional Public Transport Anonymous Data Collection employing. 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 1–6. https://doi.org/10.1109/ECAI46879.2019.9041957

Molina M (2005) An Intelligent Assistant for Public Transport Management. In: Huang D-S, Zhang X-P, Huang G-B (eds) Advances in Intelligent Computing, LNCS. Springer, Berlin, pp 199–208

Moreira-Matias L, Cats O, Gama J, Mendes-Moreira J, de Sousa JF (2016) An online learning approach to eliminate Bus bunching in real-time. Appl Soft Comput 47:460–482. https://doi.org/10.1016/j.asoc.2016.06.031

Moya-Gómez B, García-Palomares JC (2017) The impacts of congestion on automobile accessibility. What happens in large european cities? J Transp Geogr 62:148–159. https://doi.org/10.1016/j.jtrangeo.2017.05.014

Nachtigall K (1995) Time depending shortest-path problems with applications to railway networks. Eur J Oper Res 83(1):154–166. https://doi.org/10.1016/0377-2217(94)E0349-G

Nguyen T, Nguyen-Phuoc DQ, Wong YD (2021) Developing artificial neural networks to estimate real-time onboard bus ride comfort. Neural Comput Appl 33:5287–5299. https://doi.org/10.1007/s00521-020-05318-3

Niklas U, von Behren S, Soylu T, Kopp J, Chlond B, Vortisch P (2020) Spatial factor—using a Random Forest classification model to measure an internationally comparable urbanity index. Urban Sci 4(3):36. https://doi.org/10.3390/urbansci4030036

Olczyk A, Galuszk A (2017) Cloud-based machine learning for bus arrival time prediction. Proceedings of Carpathian Logistic Congress, 173–177

Othman MSB, Tan G (2018) Machine learning aided simulation of public transport utilization. 2018 IEEE/ACM 22nd International Symposium on Distributed Simulation and Real Time Applications (DS-RT), 1–2. https://doi.org/10.1109/DISTRA.2018.8601011

Othman N, Hussin M, Mahmood RAR (2019) Sentiment evaluation of Public Transport in Social Media using Naïve Bayes Method. Int J Eng Adv Technol 9:2305–2308

Palacio SM (2018) Machine Learning Forecasts of Public Transport Demand: A Comparative Analysis of Supervised Algorithms Using Smart Card Data. XREAP WP. https://doi.org/10.2139/ssrn.3165303

Paletta L, Wiesenhofer S, Brandle N, Sidla O, Lypetskyy Y (2005) Visual surveillance system for monitoring of passenger flows at public transportation junctions. Proceedings of 2005 IEEE Intelligent Transportation Systems, 2005, 862–867. https://doi.org/10.1109/ITSC.2005.1520163

Pandurangi A, Byrne C, Anderson C, Cui E, McArdle G (2020) Design and development of an application for predicting bus travel times using a segmentation approach. In: Proceedings of the 6th international conference on geographical information systems theory, applications and management (GISTAM), pp 72–80. https://doi.org/10.5220/0009393800720080

Pasini K, Khouadjia M, Same A, Ganansia F, Oukhellou L (2019) LSTM encoder-predictor for short-term train load forecasting. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 535–551. https://link.springer.com/chapter/10.1007/978-3-030-46133-1_32

Prashanth TL, Tamilselvan AK, Chandrodaya S (2016) Multimodal transport model: Enhancing collaboration among mobility sharing schemes by identifying an optimal transit station. 2016 International

Conference on Internet of Things and Applications (IOTA), 286–291. https://doi.org/10.1109/IOTA.2016.7562739

Purdy M, Daugherty P (2017) How AI boosts industry profits and innovation. Accenture Ltd, Dublin, Ireland

Raflesia SP, Lestarini D, Rodiah D, Firdaus, (2018) Opinion mining using machine learning approach: case study of light rail transit development in Indonesia. Indones J Electr Eng Comput Sci 11(2):791–796. https://doi.org/10.11591/ijeecs.v11.i2.pp791-796

Rahimi MM, Naghizade E, Stevenson M, Winter S (2020) Service quality monitoring in confined spaces through mining Twitter data. J Spat Inform Sci 21:229–261.

Reddy KK, Kumar BA, Vanajakshi L (2016) Bus travel time prediction under high variability conditions. Curr Sci, 700–711. Retrieved 05, 2019, from http://www.jstor.org/stable/24908545

Rohit MH, Computer Vision (2020) An IoT based System for Public Transport Surveillance using real-time Data Analysis and Computer Vision. 2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC), 1–6. 10.1109/ICAECC50550.2020.9339485

Roulland F, Ulloa L, Mondragon A, Niemaz M, Bouchard G, Ciriza V (2014) Learning mobility user choice and demand models from public transport fare collection data, 1–5. 21st World Congress on Intelligent Transport Systems, ITSWC

Russell SJ, Norvig P (2010) Artificial intelligence-a modern approach, third international edition. Pearson Education London, London

Scemama G (1995) CLAIRE: an independent, AI-based supervisor for congestion management. Traffic Eng Control 36(11):604–612

Shakeel N, Baig F, Saddiq MA (2019) Modeling Commuter's sociodemographic characteristics to predict public transport usage frequency by applying supervised machine learning method. Transp Tech Technol 15:1–7. https://doi.org/10.2478/ttt-2019-0005

Shalit N, Fire M, Ben-Elia E (2020) Imputation of Missing Boarding Stop Information in Smart Card Data with Machine Learning Methods. Intelligent Data Engineering and Automated Learning-IDEAL 2020, 17–27. https://doi.org/10.1007/978-3-030-62362-3_3

Sharma SK, Sharma RC (2019) Pothole detection and warning system for Indian roads. In: Kumar M, Pandey RK, Kumar V (eds) Advances in interdisciplinary Engineering, LMNE. Springer, Singapore, pp 511–519

Shatnawi N, Al-Omari AA, Al-Qudah H (2020) Optimization of Bus stops locations using GIS techniques and Artificial Intelligence. Procedia Manuf 44:52–59. https://doi.org/10.1016/j.promfg.2020.02.204

Singla A, Santoni M, Bartók G, Mukerji P, Meenen M, Krause A (2015) Incentivizing Users for Balancing Bike Sharing Systems, 723–729, Twenty-Ninth AAAI Conference on Artificial Intelligence

Skhosana M, Ezugwu A, Rana N, Abdulhamid SI (2020) An Intelligent Machine Learning-Based Real-Time Public Transport System. Lecture Notes in Computer Science, 649–665, International Conference on Computational Science and Its Applications, 20th International Conference

Song M, Weng X, Yao S, He Q (2015) Path selection of urban public transportation based on artificial intelligence ant colony algorithm. Int J Simul-Syst Sci Technol 16(11):16

Sosnowska J, Skibski O (2018) Path Evaluation and Centralities in Weighted Graphs-An Axiomatic Approach. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 3856–3862

Sun S, Yang D (2018) Identifying public transit commuters based on both the smartcard data and survey data: a case study in Xiamen, China. J Adv Transp 2018:9693272. https://doi.org/10.1155/2018/9693272

Sykes J-D, Fleur RS, Norkulov D, Dong Z, Amineh RK (2019) Conscious GPS: A System to Aid the Visually Impaired to Navigate Public Transportation. 2019 IEEE 40th Sarnoff Symposium, 1–6. https://doi.org/10.1109/Sarnoff47838.2019.9067826

Tan D, Wang J, Liu H, Wang X (2011) The optimization of bus scheduling based on genetic algorithm. Proceedings of 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), 1530–1533. https://doi.org/10.1109/TMEE.2011.6199499

Tang T, Liu R, Choudhury C (2020) Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data. Sustain Cities Soc 53:101927. https://doi.org/10.1016/j.scs.2019.101927

Tekin S, Köfteci S, Aydin MM, Yildirim MS (2018) Trip optimization for public transportation systems with linear goal programming (LGP) method. Sigma: J Eng Nat Sci/Mühendislik ve Fen Bilimleri Dergisi 36(4):921–933

Toqué F, Khouadjia M, Come E, Trepanier M, Oukhellou L (2017) Short & long term forecasting of multimodal transport passenger flows with machine learning methods. Presented at the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), pp. 560–566. https://doi.org/10.1109/ITSC.2017.8317939

Tran L, Mun M, Lim M, Yamato J, Huh N, Shahabi C (2020) DeepTRANS: a deep learning system for public bus travel time estimation using traffic forecasting. Proceedings of the VLDB Endowment, 13, 2957–2960. doi: https://doi.org/10.14778/3415478.3415518

Tu Q, Weng J-C, Yuan R-L (2016) Impact analysis of public transport fare adjustment on travel mode choice for travelers in Beijing. 16th COTA International Conference of Transportation Professional, 850–863

Ubbels B, Nijkamp P (2002) Unconventional funding of urban public transport. Transp Res Part D: Transp Environ 7(5):317–329. https://doi.org/10.1016/S1361-9209(01)00027-X

UITP Asia Pacific Centre for Transport Excellence CTE (2020) Artificial Intelligence in Mass Public Transport. Executive Summary. Retrieved 05. 2020, from https://cms.uitp.org/wp/wp-content/uploads/2020/08/UITP-AP-CTE-AI-in-PT-Executive-Summary-Dec-2018_0.pdf

Ullón HR, Ugarte LF, Mariotto FT, Lacusta E, de Almeida MC (2020) Data-driven solution for planning bus routes of the public transport in UNICAMP. In: Proceedings of the 33rd international conference on efficiency, cost, optimization, simulation and environmental impact of energy systems (ECOS), pp 2097–2108

Van Egmond P, Nijkamp P, Vindigni G (2003) A comparative analysis of the performance of urban public transport systems in Europe. Int Soc Sci J 55(176):235–247. https://doi.org/10.1111/1468-2451.55020144

Velosa F, Florez H (2020) Edge solution with machine learning and open data to interpret signs for people with visual disability. ICAI Workshops. https://ceur-ws.org/Vol-2714/icaiw_waai_2.pdf

Victoriano R, Paez A, Carrasco J-A (2020) Time, space, money, and social interaction: using machine learning to classify people's mobility strategies through four key dimensions. Travel Behav Soc 20:1–11. https://doi.org/10.1016/j.tbs.2020.02.004

Wang B, Kim I (2018) Short-term prediction for bike-sharing service using machine learning. Transp Res Procedia 34:171–178. https://doi.org/10.1016/j.trpro.2018.11.029

Wang W, Liu J, Yao B, Jiang Y, Wang Y, Yu B (2019) A data-driven hybrid control framework to improve transit performance. Transp Res Part C: Emerg Technol 107:387–410. https://doi.org/10.1016/j.trc.2019.08.017

Wang S, Lu C, Liu C, Zhou Y, Bi J, Zhao X (2020) Understanding the energy consumption of battery electric buses in urban public transport systems. Sustainability 12(23):10007. https://doi.org/10.3390/su122310007

Wei Y, Song N, Ke L, Chang M-C, Lyu S (2017) Street object detection/tracking for AI city traffic analysis. 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, 1–5. https://doi.org/10.1109/UIC-ATC.2017.8397669

Welch TF, Widita A (2019) Big data in public transportation: a review of sources and methods. Transp Reviews 39(6):795–818. https://doi.org/10.1080/01441647.2019.1616849

Wilkowski A, Mykhalevych I, Luckner M (2020) City Bus Monitoring Supported by Computer Vision and Machine Learning Algorithms, Automatio. Springer International Publishing, Berlin, pp 326–33

Xie S-Y, Gao S, Xu B (2004) Study of an optimum scheduling algorithm about buses in city intelligent transport systems. Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826), 5, 2795–2799. https://doi.org/10.1109/ICMLC.2004.1378507

Xue M, Wu H, Chen W, Ng WS, Goh GH (2014) Identifying tourists from public transport commuters. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1779–1788. https://doi.org/10.1145/2623330.2623352

Yang Y, Heppenstall A, Turner A, Comber A (2020a) Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems. Comput Environ Urban Syst 83:101521. https://doi.org/10.1016/j.compenvurbsys.2020.101521

Yang C, Ru X, Hu B (2020b) Route temporal-spatial information based residual neural networks for bus arrival time prediction. J Harbin Inst Technol (New series) 27(4):31–39. https://doi.org/10.11916/j.issn.1005-9113.2018007

Yu L, Wu W, Li X, Li G, Ng WS, Ng S-K, Huang Z, Arunan A, Watt HM (2015) iVizTRANS: Interactive visual learning for home and work place detection from massive public transportation data. 2015 IEEE Conference on Visual Analytics Science and Technology (VAST), 49–56. https://doi.org/10.1109/VAST.2015.7347630

Yu D, Ding M, Wang C (2018) A design for a public transport information service in China. International Conference of Design, User Experience, and Usability, 435–444. https://doi.org/10.1007/978-3-319-91806-8_34

Yuan Y, Shao C, Cao Z, He Z, Zhu C, Wang Y, Jang V (2020) Bus Dynamic Travel Time Prediction: using a deep feature extraction framework based on RNN and DNN. Electronics 9(11):1876. https://doi.org/10.3390/electronics9111876

Zhang Y, Chen G (2018) Inferring social-demographics of travellers based on smart card data. Proceedings of 2nd International Conference on Advanced Research Methods and Analytics, 55–62 https://doi.org/10.4995/CARMA2018.2018.8310

Zhang T, Wang J, Cui C, Li Y, He W, Lu Y, Qiao Q (2019) Integrating geovisual analytics with machine learning for human mobility pattern discovery. ISPRS Int J Geo-Inf 8(10):434. https://doi.org/10.3390/ijgi8100434

Zhou X, Wang M, Li D (2019) Bike-sharing or taxi? Modeling the choices of travel mode in Chicago using machine learning. J Transp Geogr 79:102479. https://doi.org/10.1016/j.jtrangeo.2019.102479