Estimating Passenger Demand Using Machine Learning Models: A Systematic Review

Adjei Boateng[1], Charlse Anum Adams[1,], and Emmanuel Kofi Akowuah[2,**]*

[1]Regional Transport Research and Education Centre Kumasi, Civil Engineering Department, KNUST, Kumasi, Ghana
[2]Computer Engineering Department, College of Engineering, KNUST, Kumasi, Ghana

Abstract. This article investigated machine learning models used to estimate passenger demand. These models have the potential to provide valuable insights into passenger trip behaviour and other inferences. The estimate of passenger demand using machine learning model research and the methodologies used are fragmented. To synchronise these studies, this paper conducts a systematic review of machine learning models to estimate passenger demand. The review investigates how passenger demand is estimated using machine learning models. A comprehensive search strategy is conducted across the three main online publishing databases to locate 911 unique records. Relevant record titles, abstracts, and publication information are extracted, leaving 102 articles. Furthermore, articles are evaluated according to eligibility requirements. This procedure yields 21 full-text papers for data extraction. 3 research thematic questions covering passenger data collection techniques, passenger demand interventions, and intervention performance are reviewed in detail. The results of this study suggest that mobility records, LSTM-based models, and performance metrics play a critical role in conducting passenger demand prediction studies. The model evaluation was mostly restricted to 3 performance metrics which needs improved metric for evaluation. Furthermore, the review determined an overreliance on the long- and short-term memory model to estimate passenger demand. Therefore, minimising the limitation of the LSTM model will generally improve the estimation models. Furthermore, having an acceptable trainset to avoid overfitting is crucial. In addition, it is advisable to consider multiple metrics to have a more comprehensive evaluation.

1 Introduction

Getting to a location to participate in activities such as work, recreation, and socialisation is a necessity for human survival. Transport enables these activities to be carried out. Individuals use private transportation, usually self-owned vehicles. The public uses a system of shared transportation units provided by entities. More people patronise public transport. Therefore, its impacts are immediately felt when it is effective. The demand for transportation services continues to increase as a result of increasing urbanisation. In London, more than two billion passenger trips were made in 2009 (W. Wang et al., 2011). The increasing demand for passengers threatens the safety and quality of transportation services. Public transport operators must optimise operations by accurately estimating passenger demands, fleet, and income (Hänseler et al., 2017). Estimating passenger demand is the foundation of an efficient transportation system. It is challenging if operators do not use modern technologies and mathematical models in managing transportation (Sbai & Ghadi, 2018). The deployment of new transportation-related technology has resulted in an exponential increase in the availability of data on passenger movements. Furthermore, recent strides in machine learning research have resulted in numerous applications of machine learning (Hillel et al., 2021).

This paper provides a systematic review for the estimation of passenger demands using machine learning techniques. The review focusses on three thematic research questions covering passenger data collection techniques, passenger demand interventions, and intervention performance.

1.1 Public Transport Operation

Public transport refers to a mobility service that is used by the public. Public transport operates in modes such as buses, trains, subways, and ferries. Bus operations are the most common form of public transport around the world. Buses offer a practical and economical means of transportation for most commuters. They operate in designated lanes or share lanes with other vehicles. Commuter rail, metro rail, and light rail systems are examples of public train operations. Commuter rail systems are intended to transport passengers between suburban and urban locations, while metro, subway, and light rail systems are used primarily for transit within cities. Research has shown that public transport reduces traffic congestion and air pollution, as well as provides social and economic benefits to communities by increasing access to jobs and services (Litman & Litman, 2022; Meinardi et al., 2008).

* Corresponding authors: carladams1702@yahoo.com
, , "emekoahh@gmail.com

1.2 Previous passenger demand estimation

Passenger demand estimation studies aim to forecast the number of people who will use a particular transportation system or service in the future. These studies are crucial for transportation planners to determine the appropriate level of service and guide decisions about infrastructure investments and transportation policies. One popular method for estimating passenger demand is the gravity model, which is widely used in transportation planning. The model assumes that the number of trips between two locations is proportional to the product of their population and inversely proportional to the distance between them (Becker et al., 2018). A study in Madrid paid attention to the demand for interurban rail travel. The paper analysis is based on the estimation of disaggregated Nested Logit models using information from travellers. It analyses the competition between newly built high-speed trains with other modes. An evaluation was conducted using cost-benefit analysis to inform investment decisions. In addition, it analyses the response of the demand to various policy scenarios for high-speed train services. This analysis uses willingness to pay for improved levels of service as an indicator (Román et al., 2010). In general, passenger demand estimation studies are critical for transportation planning and can provide valuable information on factors that influence travel behaviour.

2 Research Method

2.1 Background

There are ten different types of systematic review. This review adopted an effectiveness review. Effectiveness is the degree to which an intervention, when applied appropriately, produces the desired outcome. When an effectiveness review is adopted, the PICO framework is recommended to develop a review question (Munn et al., 2018).

2.2 PICO framework for Question Development

The PICO process uses a case scenario from which a question is constructed. P represents the characteristics of the population, I represents the study intervention, C represents the comparator, and O represents the results. The case statement is as follows; To operate public transport successfully, operators should be able to anticipate passenger demand and infer journeys. However, the operation is negatively affected by the absence of automatic data collection in the service. Table 1 shows the PICO framework.

Table 1. PICO Framework for review.

Component	Remarks
Population	Passengers

Intervention	Machine learning models
Comparator	Data collection system
Outcome	Demand estimates

2.3 The Question

Therefore, the question developed for the review is as follows; What machine learning models can be used to estimate travel demands in various data collection systems?

3 Search Methods

3.1 Search Strategy

Adopting a search strategy is to find papers that will be useful for review. Cross-ref, Scopus, and Google Scholar search engines are used to search for articles to provide comprehensive coverage of the field. The exact search procedure is used on each database. This review focusses on papers with trip inferences. Therefore, only papers with titles directly related to trip inference are included. The following initial phrases are tested: travel pattern, trip inference, passenger density, travel demand, and origin-destination. The phrases are limited to the title, which helps the researcher filter out irrelevant papers. To select papers that discuss machine learning techniques, only papers with one or more selected phrases related to machine learning are selected. The following initial phrases are tested: machine learning, neural network, decision tree, computer vision, artificial intelligence, random forest, boosting, support vector, deep vision, and image processing. The initial search was not limited to a specific time frame.

The Google Scholar search for papers containing at least one of the above machine learning-related phrases alongside at least one of the trip-inference phrases across all relevant fields returns 17,700 results as of August 2022. The search in the Scopus database returned more than 1,240 results. Whilst Crossref returns 1,730 results.

3.2 Selection Criteria

The setting of inclusion and exclusion criteria is a standard and required practise when designing systematic research protocols. It helps the researcher produce reliable and repeatable results. The following inclusion criteria are determined for articles found in the search to be included in the study (Booth Andrew et al., 2016):

3.2.1 Inclusion Criteria:

1. Studies that employ one or more machine learning techniques for predictive trips or demands.

2. Studies on the estimation of passenger demands in transport modes.

3. Studies that investigate density passenger densities from deep vision systems, google data, or use non-intrusive methods.

4. Studies that investigate the passenger data from collection systems.

3.2.2 Exclusion Criteria:

The following exclusion criteria are determined for the articles found in the search to be included in the study:

1. All related papers or documents were published before 2018.

2. Studies in peer-reviewed journals or conference proceedings not written in English.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline is used for paper selection. First, duplicates are eliminated from the search records. Subsequently, the title and abstract of the articles are checked against the eligibility criteria.

Finally, the remaining full-text articles are assessed for eligibility (Liberati et al., 2009). A paper with multiple datasets and methodologies is treated as a separate study for analysis.

3.3 Data extraction strategy

A list of attributes is created to extract relevant data without bias from each study. Furthermore, it limits subjectivity in the data extraction process. The attributes are specific, objective, quantifiable, or categorical. The studies are carefully reviewed and every attribute is noted and tallied. The attributes are shown in table 2.

Table 2. Research Question and Corresponding attributes for data extraction.

No.	Description
1	**What are the methods for collecting passenger data?**
a	Transport system
b	Information collected
c	Device used for collection
d	Nature of dataset and size
e	Time Aggregation
2	**Which functions have been used for determining passenger demands and where?**
a	Intervention used
b	Classification of intervention
c	Unit of analysis
d	Geographical area studies have been conducted
3	**How is the intervention performance determined?**
a	Data Split Ratio
b	Performance metrics used

3.4 Study selection

The following search terms are used to carry out the search strategy outlined in Section 3.1.

Scopus: (TITLE ("trip inference" OR "origin destination" OR "travel pattern" OR "passenger density" OR "travel demand")) AND TITLE-ABS-KEY ("machine learning" OR "neural network" OR "decision tree" OR "computer vision" OR "random forest" OR "boosting" OR "support vector" OR "deep vision" OR "image processing"))

Google Scholar: ("trip inference" OR "origin destination" OR "travel pattern" OR "passenger density" OR "travel demand") AND ("machine learning" OR "neural network" OR "decision tree" OR "computer vision" OR "random forest" OR "boosting" OR "support vector" OR "deep vision" OR "image processing").

Semantic Scholar: ("trip inference" OR "origin destination" OR "travel pattern" OR "passenger density" OR "travel demand") AND ("machine learning" OR "neural network" OR "decision tree" OR "computer vision" OR "random forest" OR "boosting" OR "support vector" OR "deep vision" OR "image processing").

Due to the restriction on search length in Google Scholar, this search is divided into two separate searches, with the results combined. The search was carried out on 20/12/2019 on all three databases. Figure 1 shows a PRISMA flow chart of the study selection process.

A total of 2,095 records were returned from the search. There were 346 records returned from Crossref, 635 records from Scopus, and 1,114 records from Google Scholar. Duplicates are then removed, leaving 911 records to be screened. The total number of records after removing duplicates is more than the results obtained from two of the databases, showing that there were results from Crossref/Scopus which were not returned with the Google Scholar search. The remaining records are then screened to determine whether they meet the eligibility criteria outlined in Section 2.4.2. During the selection, 809 articles are excluded for relevance based on their title and abstract.

The full text is obtained for the remaining 102 articles for further review. Of these, another 81 are excluded based on the selection criteria. This process leaves 21 selected articles for data extraction.

4 Results and Discussion

4.1.1 Selected papers

The final number of papers used for the data extraction is 21. Table 3 provides a unique identifier for each paper and year of publication.

Table 3. Papers selected for review.

No	Paper	Year
S1	Clustering-Based Travel Pattern for Individual Travel Prediction of Frequent Passengers by Using Transit Smart Card	2022
S2	Continuous Time and Multi-Level Graph Representation Learning for Origin-Destination Demand Prediction	2022
S3	Short-Term Holiday Travel Demand Prediction for Urban Tour Transportation: A Combined Model Based on STC-LSTM Deep Learning Approach	2022
S4	Short-Term Prediction of Passenger Demand in Multi-Zone Level: Temporal Convolutional Neural Network with Multi-Task Learning	2019
S5	AI-based neural network models for bus passenger demand forecasting using smart card data	2022
S6	Exploring temporal variability in travel patterns on public transit using big smart card data	2022
S7	Analysis of Bus Trip Characteristic Analysis and Demand Forecasting Based on GA-NARX Neural Network Model	2020
S8	A deep spatio-temporal fuzzy neural network for passenger demand prediction	2019
S9	Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modelling	2019
S10	Passenger Demand Forecasting with Multi-Task Convolutional Recurrent Neural Networks	2019
S11	Deep learning-based origin-destination prediction via contextual information fusion	2022
S12	Prediction of Urban Taxi Travel Demand by Using Hybrid Dynamic Graph Convolutional Network Model	2022
S13	Contextualized Spatial-temporal Network for Taxi Origin-Destination Demand Prediction	2019
S14	A GAN framework-based dynamic multi-graph convolutional network for origin–destination-based ride-hailing demand prediction	2022
S15	A novel deep recurrent neural network for Short-term travel demand forecasting under on-demand ride services	2019
S16	Online Metro Origin-Destination Prediction via Heterogeneous Information Aggregation	2019
S17	A Spatiotemporal Analysis of Taxis Demand: A Case Study in the Manizales City	2021
S18	Taxi Hotspots Identification through Origin and Destination Analysis of Taxi Trips using K-means Clustering and H-indexing	2021
S19	Travel Destination Prediction Based on Origin-Destination Data	2020
S20	Prediction of City-Scale Dynamic Taxi Origin-Destination Flows Using a Hybrid Deep Neural Network Combined with Travel Time	2019
S21	Predictability of short-term passengers' origin and destination demands in urban rail transit	2022

4.1.2 Publication Source

Table 4 provides details of all the journals and conferences/proceedings from which the papers were selected. The articles come from a wide spread of publications, with a total of 13 different journals and 8 different conferences featured.

Table 4. Summary of published sources.

Publication	Type
Transportation Research Record	Conference
ACM SIGKDD Conference on Knowledge Discovery and Data Mining	Conference
KSCE Journal of Civil Engineering	Journal
IEEE Transactions on Intelligent Transportation Systems	Journal
Journal of Urban Management	Journal
Environment and Planning B: Urban Analytics and City Science	Journal
IEEE Access: Big Data Technology & applications in intelligent transportation	Journal
2019 SIAM International Conference on Data Mining (SDM)	Conference
ACM SIGKDD Conference on Knowledge Discovery and Data Mining	Conference
PAKDD 2019: Advances in Knowledge Discovery and Data Mining	Conference
International Journal of Multimedia Tools and Applications	Journal
Journal of Sensors	Journal
IEEE Transactions on Intelligent Transportation Systems	Journal
Journal Information Sciences	Journal
IOP Conference Series: Materials Science and Engineering	Conference
IEEE Transactions on Pattern Analysis and Machine Intelligence	Journal
CIARP 2019: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications	Conference
Journal of Physics	Journal

Conference on Complex, Intelligent, and Software-Intensive Systems	Journal
IEEE Access	Journal
Journal of Transport	Journal

4.1.3 Publication Year

Figure 1 shows the distribution of the articles from 2019 to 2022. There is a recent spike in interest in the publication of demand estimation using machine learning models. 48% of the articles were published in the recent year 2022. The remaining 52% of the publications were done before 2022. This demonstrates a rising interest in research for demand estimation using machine learning models.
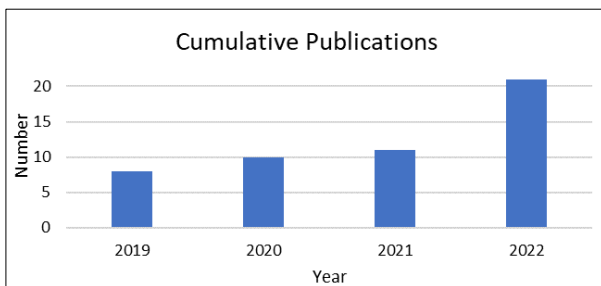
Fig. 1. Cumulative Publication of papers.

4.1.4 What are the methods for collecting passenger data?

The following sections present an overview of the methods for collecting passenger data used in the selected articles.

Question 1.a What type of transport system has passenger data been collected?

Based on the responses to Q1a, the transportation systems are grouped into taxis, buses, and metro/rail. 62% of the transportation systems were taxis, 19% were buses, and 19% were metro/rail systems.

Question 1.b What device/instrumentation was used to collect the data?

From the review, 57% of the data were recovered from mobility records or databases, 19% from smart cards, then 19% from automatic fare collection systems, and 5% from image captured records. All studies conducted in a metro/rail system collected information from smart cards except (Han et al., 2022a; Q. Zhao et al., 2022). Researchers resorted to smart cards and automatic fare collection systems because it generates the desired data at less cost, since it requires minimal installation, personnel, and man hours (Ait-Ali & Eliasson, 2022).

Question 1.c What type of information was collected for the study?

The information collected from devices or through instruments is shown in Table 5. 19% of the studies used the tap-in tap-out information of smart cards for analysis. Smart cards have a large data volume capacity,

broad coverage, and high authenticity. Immutable ID allows the researcher to obtain long-term travel information about passengers, which offers the potential to mine travel patterns and travel predictions. Therefore, researchers prefer tap-in tap-out information because it considers complete trip attributes in addition to real-time prediction (Ye & Ma, 2022; Z. Zhao et al., 2018). A significant number of studies used information on boarding and exiting, which is about 43%. This information was from mobility records/databases or smart cards. Passenger request information was about 24%. The information usually contained request ID, pick-up time, pick-up coordinates, drop-off time, and drop-off coordinates (Liang et al., 2019). S11, S12, S18, S19, and S20 combine passenger requests with vehicle GPS, metrological data, or coded road network. These kinds of studies constituted about 25% of the studies.

Information was classified into passenger-orientated information and transport unit-orientated information. Passenger-orientated studies are studies where the primary data used for the analysis were the passenger data, while vehicle-orientated studies analysis was based on the transport unit. Table 5 shows the information collected and the orientation of the study. 90% of the studies were in the passenger-oriented category and 4% were transported unit oriented. S12 was the only study to focus on both passenger and transport unit-orientated categories. Data from vehicle-orientated studies were applied to taxi services, while data from passenger-orientated studies were mostly applied to metro, rail, and bus services. The researcher was able to assess vehicular data because they were in an open source dataset from the operators (Huang et al., 2022). The availability of passenger data on smart cards influenced the focus of transport system studies. In transport systems such as metro, rail and bus systems, the desired passenger data was obtained through smart cards, which includes information such as travel patterns, frequency of use, and duration of trips. These data are analysed to gain insight into passenger behaviour and preferences, which informs decisions related to the design and operation of these systems. As a result, studies on these transport systems tend to be passenger-orientated, as they are based on readily available data.

Table 5. Summary of Information Collected for Analysis

Data Information Collected	Passenger	Vehicle
Tap in - Tap out	Yes	No
Crowd flow images and external features	Yes	No
Alighting and boarding	Yes	No
Passenger Request	Yes	No
Passenger request, Vehicle GPS	Yes	Yes

Passenger request, metrological data	Yes	No
Passenger request, taxi GPS data, Road network Passenger	Yes	Yes

Question 1.d What is the Nature of the Dataset, Size, Collection Duration, and aggregation?

The nature, size, period, and temporal aggregation of the collected data are shown in table 6. The nature of the data for most studies was transactional requests or data from smart card cards. This is because it was less costly to obtain transactional data than to conduct manual data surveys. Furthermore, smart card systems and service requests were readily available at locations where the studies were carried out (Sun et al., 2020). The size of the data set ranged from 10,000 to more than 700 million trips. The size of the data sets is correlated with the spatial area under study. For S5, 694,000 records were collected on 18 routes and 1,781 bus stops. While S6 had about 769 million records with 42,000 bus stops and 325 subway stations. S1, S2, S6, S7, S13 and S15 collected more than ten million records from complete networks. S4, S5, S16, S19, and S20 collected records of less than 10 million transactions, these studies were carried out on selected routes or a transport service with a limited time frame for data collection.

Although time is continuous, various time intervals were adopted for the studies. This adoption of the time interval is an essential part of each study. S19 did not state the temporal aggregation for the study. S19 was excluded from the temporal aggregation analysis. Data were temporally aggregated in 5 min, 10 min, 15 min, 20 min 25 min, 30 min, 60 min, and 24 hours, except for S1 and S7 which used 24 hours and 10 minutes, respectively. 21% and 37% of the studies used 15min and 60min, respectively, for temporal aggregation of the data. S21 used all categories of aggregation in the study. The research adopted varying time intervals based on population dynamics. Furthermore, the adopted interval is relatively stable in order not to conceal the time-varying laws of passenger demand. The review showed that 15 min and 60 min were the preferred time intervals for data aggregation (Giraldo-Forero et al., 2019; Yang et al., 2022).

The duration of data collection was from 5 to 365 days. Most studies collected their data in 30 days. Studies such as S6, S11, S12, S13, S14 and S19 data were collected over 120 days. The data was readily available data from automatic collection systems and open-source databases, which made collection easier. Therefore, researchers resorted to collecting long-duration data for their studies.

Table 6. Nature, size of the dataset, Data Duration, and temporal aggregation of the data

No	Nature	Dataset Size	Duration	Temporal Aggregation
S1	trips	96,065,229	31 days	24 hours
S2	Transaction	279,227,618 38,498,427	56 days 181 days	30 minutes
S3	Transactions	2,808 On peak 2,493 Off-peak	55 days	15 minutes
S4	Car calling requests	120,000 per day	90 days	20 minutes
S5	Transactions	693,689	28 days	15 minutes
S6	Transaction	769,413,168	120 days	60 minutes
S7	Passenger Flow	10,658,558	5 days	10 minutes
S8	Service requests	5.24 million	30 days	60 minutes
S9	Service Request	Not stated	31 days 30 days	60 minutes
S10	Service Request	1.5mil mobile user	60 days	60 minutes
S11	Transaction	160,000,000	365 days	60 minutes
S12	transaction	12,179,185	180 days	15 minutes
S13	Service Request	132,000,000	365 days	30 minutes
S14	Service Request	4.9 million records per day	90 days	20 minutes
S15	taxi trips	78 million trips	163 days	10 minutes
S16	Daily Ridership	8.820 million 2.35 million	90 days 31 days	15 minutes
S17	Service Request	70,9531 requests	120 days	60 minutes
S18	Taxi GPS	28,590 taxis	7 days	10 minutes
S19	Travel records	1,495,814	212 days	Not Stated
S20	Trip	10,000 trips	10 days	60 minutes
S21	Transactions	10,658,558	30 days 61 days	5, 10, 15, 20, 25, 30, 60,120 min

4.1.5 What functions have been used to determine passenger demands and where?

The following sections present an overview of the methods for collecting passenger data used in the selected articles.

Question 2.a Which intervention and class of intervention was used in the study?

Based on the responses to Question 2.a, the interventions were categorised according to the type of

model in addition to the snippet of the model. The known types of models are supervised, unsupervised, semi-supervised, and reinforcement learning. The identified model class was further classified into subtypes. Table 7 shows the interventions, the type of model, and the comments. 76% of the studies conducted the research using a supervised regression model. The remaining percentage was shared between the unsupervised clustering model and the unsupervised dimension reduction model at 19% and 5%, respectively. Forecasting traffic demand using deep neural networks has attracted widespread interest. Most studies that used neural network models combined several machine learning models as their intervention. (Han et al., 2022b). Recurrent neural networks are suitable for forecasting studies. However, it is limited in using information from a distant past and also performs poorly for long-term memory (Liyanage et al., 2022). Therefore, S3, S5, S8, S9, S10, S11, S12, S13, S14, S15 and S20 combined long-short-term memory (LSTM) models with other models for their studies. Hybrid LSTM studies constituted 73% of supervised regression model studies. LSTM models are generally identified to outperform RNNs in time-series data forecasting (Yeon et al., 2019). Therefore, the LSTM-based model is preferred for predicting and forecasting series.

Table 7. Intervention and Class of Intervention

No	Intervention Used	Type of Model	Class of Model Remarks
S1	DBSCAN algorithm	Unsupervised clustering	Density-based method
S2	Continuous-Time and Multi-Level Graph Representation	Unsupervised clustering	Neural network with multi-level memories for nodes
S3	Long Short-Term Memory model	Supervised regression	LSTM with Spatio temporal correlation
S4	End-to-end multitask learning temporal convolutional neural network	Supervised regression	Temporal Convolutional Neural Network with Multi-Task Learning
S5	LSTM and BiLSTM deep learning models	Supervised regression	Bidirectional LSTM
S6	K means, multivariate ordinary least square regression	Unsupervised clustering	Customer segregation method
S7	Genetic Algorithm-Optimised NARX neural network model	Supervised regression	ARIMA with genetic algorithm neural network
S8	Spatio-Temporal Fuzzy Neural Network	Supervised regression	LTSM with fuzzy logic
S9	Graph Convolution	Supervised regression	Multi-task LSTM
S10	Multi-Task Convolutional Recurrent Neural Networks	Supervised regression	LSTM with CNN
S11	Auxiliary-tasks Enhanced Spatio-Temporal Network	Supervised regression	LSTM with CNN and GCN
S12	Hybrid Dynamic Graph Convolutional Network Model	Supervised regression	LSTM with GRU
S13	Contextualised spatial-Temporal Network	Supervised regression	ConvoLSTM and global correlation index
S14	Dynamic Multi-Graph Convolutional Network with Generative Adversarial Network structure	Supervised regression	LSTM with Multiple Graph Convolution
S15	Stacked bidirectional long short-term memory neural network	Supervised regression	LSTM stacked bidirectional
S16	Heterogeneous Information Aggregation Machine	Supervised regression	Unified neural network module. Made up Long Short term GCGRU
S17	Generalized Additive models	Supervised regression	Regressive model with non-linear smoothing
S18	K-means, Clustering and H-indexing	Unsupervised clustering	
S19	Destination Prediction Algorithm	Supervised regression	LightGBM
S20	Convolutional LSTM	Supervised regression	LSTM with CNN
S21	Temporal Pearson Coefficients and Approximate Entropy	Unsupervised Dimension Reduction	unsupervised statistical technique

Question 2.b What was the unit of analysis?

S1 uses travel patterns to improve the accuracy of passenger travel information prediction. It achieved its objectives by analysing the degree of accuracy of the proposed model. S2 and S7 analyse demand predictions by aggregating data into different time zones and updated transactions. The researcher minimises the abundance of results by aggregating the data. S3, S8 and S15 analysed the prediction of short-term demand and the effect of weather conditions, such as air temperature and the air quality index. The research determined this from the passenger correlation flow and showed that the population is attracted to the same type of facility. Furthermore, the shorter the distance between the bus

top or the numerous bus routes to the stop, the higher the degree of correlation of tourist flow. S4 analysis Spatio-temporal dynamic time-warping test where correlation between zones was evaluated based on short-term passenger demand data. S5 article analysis predictions of varying temporarily aggregated short-term data. S6 analyses the variability of the travel pattern. To investigate the variability, probability density functions were developed and curves representing different groups of passengers were fitted and evaluated. S9, S11, S12, S13, S14 and S16 analyse the number of demands from passengers in each region. S11 also incorporates an objective function as part of its analysis. S10 analyses the prediction of passenger demand over a fixed temporal interval. Spatial analysis of S17 and S21 considers the requested demand and attributes the intensity of the demand to land use. They also perform a temporal analysis, in which the demand behaviour is analysed over time. Further analysis was performed in which the observed demand in specified periods was critically analysed.

S18 generates various clusters; it uses the elbow method to find the optimised number of clusters. Due to spatial restrictions, a high number of clusters is also used to enable the study to determine hotspots. S19 analysed the distance travelled by passengers; the study also used an evaluation metric and the performance of the model for its analysis. S20 analyses the correlation between travel frequency and travel time by comparing travel frequency and travel time in different time intervals. It should be noted that the following articles S4, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S20, and S21 analyse the performance and comparison of the model implemented in the study.

Table 8. Summary of the analysis

No.	Summary Analysis
S1	Cluster Travel Pattern and the regularity of passenger travel
S2	Predict the departure time and demand of passengers
S3	Predict forecast short-term holiday travel demand
S4	Predict the short-term passenger demand in a multizone level
S5	Predict the ticketing system based on bus passenger demand.
S6	Investigate the temporal variability in travel patterns for passengers
S7	Analyse the demand and temporal variability of passenger flow
S8	Predict passenger demand including the interactions of all known important factors
S9	Obtain mobility patterns to predict passenger demands from one region to another

S10	Predict passenger demand
S11	Predict passenger travel demands from one region to another.
S12	Improve prediction accuracy to capture the characteristics of urban travel demand.
S13	Predict temporal variability of taxi demand between region pairs
S14	Investigate origin-destination-based demand prediction
S15	Predict short-term travel demand based on historical data and other information
S16	Exploit heterogeneous information to learn the evolutionary patterns of ridership
S17	Present spatio-temporal and temporal analysis of demand flow in taxi requests
S18	Determining Hotspots by measuring the clusters' h-index
S19	Predict the travel destination only depending on the departure time and coordinates
S20	Predict passenger travel demands, as well as manage taxi operations and scheduling
S21	Reflect the inherent time-space correlations and complexity of the passenger flow

Question 2.c In what geographical areas have these studies been conducted?

Figure 2 shows the geographical locations where the selected studies have been conducted. Most studies have been conducted in Asia (60%) and the United States (30%). This can be attributed to the open data policies in these regions. The data required for these studies are readily available and in the appropriate format. S18 did not state the location where the research was conducted. Public transportation in Africa is usually unregulated and its mode of operation presents a challenge in collecting information on passenger demand (Kumar & Barrett, 2008).
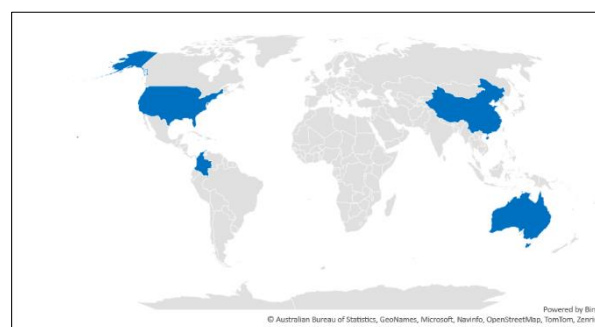
Fig. 2. Geographical locations where selected studies have been done.

4.1.6 How is the performance of the intervention determined?

The following sections present how the intervention performance has been used in the selected articles.

Question 3.a What is the data split ratio used in the studies?

To ensure that the model used can generalise to unseen data, the model is trained on a diverse dataset. Data are split into sets to identify and correct overfitting issues, therefore improving the overall performance of the machine learning model. The training set split between 70% and 80% was 83% of the selected articles. S3, S10, S13, S17, S19, and S20 split more than 80% of their data as a training set and did not consider a validation set. S4. S2, S4, S12, S14, and S16 with a training set below 80% further split the training set into a training and validation set. Splitting a data set into training and testing sets is an important step in machine learning as it also helps to evaluate the performance of a model. The 80/20 data split is a common choice because it strikes a good balance between having enough data to train a model and having enough data to test it. Choosing a small test set presents a risk that the model is unable to generalise well to new data. Models can learn faster with a large training set, thus reducing the overall training time. There is no hard and fast rule on data splitting. However, different data split ratios may be more appropriate depending on the specific problem, the amount of data available and the complexity of the model. It is a good practise to experiment with different data split ratios and evaluate the model's performance before making a final decision (Aurélien Géron, 2019; Raschka & Mirjalili, 2019).

Question 3.b What performance metrics were used?

Performance metrics are useful for evaluating the performance of models because they provide a standardised way to compare different models. Table 9 shows the performance metrics used for the evaluation of the models used in the studies. Three discrete performance metrics were used in the selected studies. The root mean square error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) were the metrics commonly used to evaluate the models. They provide a quantitative measure of how well a model performs in terms of predicting the output variable compared to the actual values. RMSE is a measure of the average difference between the predicted and actual values, but with a higher weight given to larger errors. It is useful when you want to penalise larger errors more heavily. MAE is a measure of the average difference between the predicted and actual values, with equal weight given to all errors. It is useful when you want to give equal importance to all errors. MAPE is a measure of the average percentage difference between predicted and actual values. It is useful when you want to evaluate the accuracy of a model in terms of percentage error, which can be helpful in situations where the magnitude of the error is important. S1, S2, S6, S7, S9, S13, S14, S19, and S21 employed different metrics or combined with the commonly used metric.

S1 used only the prediction accuracy as a metric, where the prediction accuracy is equal to the number of correctly predicted travel data divided by the total number of travel data. The correct prediction referred to the difference between the predicted time and the actual value as less than 1 hour. For S2, the Pearson Correlation Coefficient (PCC) was added to MAE and RSME as a metric. S2 and S21 adopted PCC because the results are linear, the variables are quantitative, normally distributed, and have no outliers. S6 adopted F1 score, precision and recall as its metrics. Recall and precision metrics are used to evaluate the performance of a classification or information retrieval model. Recall measures the percentage of relevant instances that were correctly retrieved by the model out of all relevant instances in the data set. It is used to evaluate the completeness of the results produced by the model. Precision measures the percentage of retrieved instances that are relevant from all retrieved instances. It is used to evaluate the accuracy of the results produced by the model. Then the F1 score was applied, which is the harmonic mean of precision and recall. It symmetrically represents both precision and recall in one metric. S7 adopted the mean error, a fitness function, and the mean standard error as a performance metric for its model. MSE measures the precision of the estimates since it reflects the magnitude of the error to expect in the estimates. A smaller MSE indicates that the estimates are more precise and have less variability, while a larger MSE suggests that the estimates are less precise and have more variability. The MSE values were associated with the neural network training effect, where a lower value indicated better neural network training. The fitness function was adopted because the study involved the use of a genetic algorithm. The function is used to guide the optimisation process. The reciprocal value of the mean squared error is used as a fitness function. ME sums up the variances and divides the result by sample population. The variance is the difference between the measured value and the true value. S9 adopted the RSME and the symmetric mean absolute percentage error (SMAPE), which measure the accuracy based on relative errors. The study justified the choice of SMAPE because it was not scale sensitive. S13 and S14, in addition to MAE and RSME, include the R2 score as a performance metric that indicates how well the data fit the regression model. The R2 score is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. The score or value derived from this metric is independent of the context because it measures the goodness of fit of the model. MAE, RSME, and MAPE have primarily featured metrics in the selected papers. MAE, RSME and MAPE were found to be used in 57%, 67% and 52% of the studies, respectively. It suggests that these metrics are commonly used in model evaluation and machine learning studies. The other metrics were selected based on the preference and the objective of the study.

Table 9. Summary of the performance metrics for evaluation.

No.	MAE	RSME	MAPE	Others
S1	-	-	-	prediction accuracy
S2	MAE	RSME	-	Pearson Correlation Coefficient
S3	MAE	RSME	MAPE	-
S4	MAE	RSME	MAPE	-
S5	MAE	RSME	MAPE	-
S6	-	-	-	Score, Precision, and Recall
S7	-	-	-	ME, MSE, fitness Function
S8	MAE	RSME	-	-
S9	-	RSME	-	SMAPE
S10	MAE	RSME	-	-
S11	MAE	RSME		-
S12	MAE	RSME	MAPE	-
S13	-	RSME	MAPE	R2
S14	MAE	RSME	MAPE	R2
S15	MAE	RSME	MAPE	
S16	-	-	MAPE	-
S17	MAE	RSME	MAPE	-
S19	-	-	-	Score
S20	MAE	RSME	MAPE	-
S21	-	-	MAPE	Accuracy

5 Conclusion

The review investigates how passenger demand is estimated with a collection system using machine learning models. It examines three thematic research questions that cover passenger data collection techniques, passenger demand interventions, and intervention performance. A comprehensive search strategy is carried out across the three main online publishing databases to locate unique 911 records. Relevant record titles, abstracts, and publication information are extracted, leaving 102 articles. In addition, the articles are evaluated according to the eligibility requirements. This procedure yields 21 full-text papers for data extraction. For data collection techniques, approximately 57% of the data was recovered from mobility records or open databases. This is because the data were readily available, in an appropriate format, over a substantial period, and less expensive to obtain. Traffic data was mostly aggregated in multiples of 15-minute intervals because it provided sufficient time resolution to capture changes in traffic patterns while minimising noise. Furthermore, it becomes more manageable and easier to compare and analyse the aggregated data (Giraldo-Forero et al., 2019; Yang et al., 2022). Demand forecasting using machine learning models has attracted widespread interest. The common intervention for passenger demand was supervised regression model. 73% of the studies usually combined LSTM models with other models for their research. Making LSTM-based models is preferred for predicting and forecasting series (Liyanage et al., 2022; Yeon et al., 2019). 83% of the articles split the training set between 70% and 80%. Splitting a data set prevents overfitting and allows one to evaluate the performance of a model. MAE, RSME, and MAPE were featured as performance metrics in the selected articles. MAE, RSME and MAPE were found to be used in 57%, 67% and 52% of the studies, respectively. It suggests that these metrics are commonly used in model evaluation and machine learning studies. Despite the exhaustive nature of this document, the following limitations must be acknowledged. Only three Internet databases were used to search for relevant materials. Furthermore, this review did not evaluate grey and unpublished literatures. However, it is expected that the selected papers will cover new methodologies. The approach to the review is intended to be as objective as possible. Data extraction and discussion are carried out by the first author under the supervision of the coauthors. All results have been double-checked, but the authors are responsible for any residual inaccuracies.

The results of this study suggest that mobility records, LSTM-based models, and performance metrics play a critical role in performing passenger demand prediction studies. Besides, the review determined an overreliance on the long- and short-term memory model to estimate passenger demand. Therefore, minimising the limitation of the LSTM model will generally improve the estimation models. Additionally, having an acceptable train set to avoid overfitting is crucial. Furthermore, it is advisable to consider multiple metrics to have a more comprehensive evaluation.

References

1. Ait-Ali, A., & Eliasson, J. (2022). The value of additional data for public transport origin–destination matrix estimation. Public Transport, 14(2), 419–439. https://doi.org/10.1007/S12469-021-00282-0

2. Aurélien Géron. (2019). Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media, 851. https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

3. Bai, L., Yao, L., Kanhere, S. S., Yang, Z., Chu, J., & Wang, X. (2019). Passenger demand forecasting with multi-task convolutional recurrent neural networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11440 LNAI, 29–42. https://doi.org/10.1007/978-3-030-16145-3_3/COVER

4. Becker, K., Terekhov, I., Niklaß, M., & Gollnick, V. (2018). A global gravity model for air passenger demand between city pairs and future interurban air mobility markets identification. 2018 Aviation Technology, Integration, and Operations Conference. https://doi.org/10.2514/6.2018-2885

5. Booth Andrew, Sutton Anthea, & Papaioannou Diana. (2016). Systematic Approaches to a Successful Literature Review. Systematic Approaches to a Successful Literature Review, 1–336. https://www.google.co.uk/books/edition/Systematic_Approaches_to_a_Successful_Li/JD1DCgAAQBAJ?hl=en&gbpv=0&kptab=overview

6. Duan, Z., Zhang, K., Chen, Z., Liu, Z., Tang, L., Yang, Y., & Ni, Y. (2019). Prediction of city-scale dynamic taxi origin-destination flows using a hybrid deep neural network combined with travel time. IEEE Access, 7, 127816–127832. https://doi.org/10.1109/ACCESS.2019.2939902

7. Giraldo-Forero, A. F., Garcia-Lopez, S., Rodriguez-Marin, P. A., Martinez, J., Céspedes-Villar, Y. R., Cardona, O., Acosta, J. C., & Trujillo, L. C. (2019). A Spatiotemporal Analysis of Taxis Demand: A Case Study in the Manizales City. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11896 LNCS, 514–524. https://doi.org/10.1007/978-3-030-33904-3_48/COVER

8. Han, L., Ma, X., Sun, L., Du, B., Fu, Y., Lv, W., & Xiong, H. (2022a). Continuous-Time and Multi-Level Graph Representation Learning for Origin-Destination Demand Prediction. Dl.Acm.Org, 22, 516–524. https://doi.org/10.1145/3534678.3539273

9. Han, L., Ma, X., Sun, L., Du, B., Fu, Y., Lv, W., & Xiong, H. (2022b). Continuous-Time and Multi-Level Graph Representation Learning for Origin-Destination Demand Prediction. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 516–524. https://doi.org/10.1145/3534678.3539273

10. Hillel, T., Bierlaire, M., Elshafie, M. Z. E. B., & Jin, Y. (2021). A systematic review of machine learning classification methodologies for modelling passenger mode choice. Journal of Choice Modelling, 38, 100221. https://doi.org/10.1016/J.JOCM.2020.100221

11. Huang, Z., Zhang, W., Wang, D., & Yin, Y. (2022). A GAN framework-based dynamic multi-graph convolutional network for origin–destination-based ride-hailing demand prediction. Information Sciences, 601, 129–146. https://doi.org/10.1016/J.INS.2022.04.024

12. Kumar, A., & Barrett, F. (2008). Stuck in Traffic: Urban Transport in Africa Africa Infrastructure Country Diagnostic.

13. Li, W., Guan, H., Han, Y., Zhu, H., & Wang, A. (2022). Short-Term Holiday Travel Demand Prediction for Urban Tour Transportation: A Combined Model Based on STC-LSTM Deep Learning Approach. KSCE Journal of Civil Engineering, 26(9), 4086–4102. https://doi.org/10.1007/S12205-022-2051-8/METRICS

14. Liang, X., Wang, G., Renqiang Min, M., Qi, Y., & Han, Z. (2019). A Deep Spatio-Temporal Fuzzy Neural Network for Passenger Demand Prediction. Proceedings of the 2019 SIAM International Conference on Data Mining. https://epubs.siam.org/terms-privacy

15. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ (Clinical Research Ed.), 339. https://doi.org/10.1136/bmj.b2700

16. Litman, T. A., & Litman, T. (2022). www.vtpi.org Info@vtpi.org 250-508-5150. www.vtpi.org

17. Liu, L., Qiu, Z., Li, G., Wang, Q., Ouyang, W., & Lin, L. (2019). Contextualized Spatial-Temporal Network for Taxi Origin-Destination Demand Prediction. IEEE Transactions on Intelligent Transportation Systems, 20(10), 3875–3887. https://doi.org/10.1109/TITS.2019.2915525

18. Liyanage, S., Abduljabbar, R., Dia, H., & Tsai, P. W. (2022). AI-based neural network models for bus passenger demand forecasting using smart card data. Journal of Urban Management, 11(3), 365–380. https://doi.org/10.1016/J.JUM.2022.05.002

19. Meinardi, S., Nissenson, P., Barletta, B., Dabdub, D., Sherwood Rowland, F., & Blake, D. R. (2008). Influence of the public transportation system on the air quality of a major urban center. A case

study: Milan, Italy. Atmospheric Environment, 42(34), 7915–7923. https://doi.org/10.1016/J.ATMOSENV.2008.07.046

20. Miao, H., Fei, Y., Wang, S., Wang, F., & Wen, D. (2022). Deep learning based origin-destination prediction via contextual information fusion. Multimedia Tools and Applications, 81(9), 12029–12045. https://doi.org/10.1007/S11042-020-10492-6/METRICS

21. Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic review should i conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Medical Research Methodology, 18(1), 1–9. https://doi.org/10.1186/S12874-017-0468-4/TABLES/1

22. Raschka, S., & Mirjalili, V. (2019). Python machine learning : machine learning and deep learning with python, scikit-learn, and tensorflow 2. 741.

23. Román, C., Espino, R., & Carlos Martín, J. (2010). Analyzing Competition between the High Speed Train and Alternative Modes. The Case of the Madrid-Zaragoza-Barcelona Corridor. Journal of Choice Modelling, 3(1), 84–108. www.jocm.org.uk

24. Street, M., Rachman, H. O., Ridwan, K., Li, X., Li, C., & Yu, H. (2021). Taxi Hotspots Identification through Origin and Destination Analysis of Taxi Trips using K-means Clustering and H-indexing. Journal of Physics: Conference Series, 1997(1), 012006. https://doi.org/10.1088/1742-6596/1997/1/012006

25. Sun, F., Wang, X. L., Zhang, Y., Liu, W. X., & Zhang, R. J. (2020). Analysis of bus trip characteristic analysis and demand forecasting based on GA-NARX neural network model. IEEE Access, 8, 8812–8820. https://doi.org/10.1109/ACCESS.2020.2964689

26. Wang, Y., Wo, T., Yin, H., Xu, J., Chen, H., & Zheng, K. (2019). Origin-destination matrix prediction via graph convolution: A new perspective of passenger demand modeling. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1227–1235. https://doi.org/10.1145/3292500.3330877

27. Wei, W., & Yan, X. (2019). A novel deep recurrent neural network for Short-term travel demand forecasting under on-demand ride services. IOP Conference Series: Materials Science and Engineering, 688(3), 033022. https://doi.org/10.1088/1757-899X/688/3/033022

28. Yang, F., Shuai, C., Qian, Q., Wang, W., He, M., He, M., & Lee, J. (2022). Predictability of short-term passengers' origin and destination demands in urban rail transit. Transportation, 1–27. https://doi.org/10.1007/S11116-022-10313-9/METRICS

29. Ye, P., & Ma, Y. (2022). Clustering-Based Travel Pattern for Individual Travel Prediction of Frequent Passengers by Using Transit Smart Card. Https://Doi.Org/10.1177/03611981221111355, 036119812211113. https://doi.org/10.1177/03611981221111355

30. Yeon, K., Min, K., Shin, J., Sunwoo, M., & Han, M. (2019). Ego-Vehicle Speed Prediction Using a Long Short-Term Memory Based Recurrent Neural Network. International Journal of Automotive Technology, 20(4), 713–722. https://doi.org/10.1007/S12239-019-0067-Y/METRICS

31. Zhang, K., Liu, Z., & Zheng, L. (2020). Short-Term Prediction of Passenger Demand in Multi-Zone Level: Temporal Convolutional Neural Network with Multi-Task Learning. IEEE Transactions on Intelligent Transportation Systems, 21(4), 1480–1490. https://doi.org/10.1109/TITS.2019.2909571

32. Zhao, J., Kong, W., Zhou, M., Zhou, T., Xu, Y., & Li, M. (2022). Prediction of Urban Taxi Travel Demand by Using Hybrid Dynamic Graph Convolutional Network Model. Sensors 2022, Vol. 22, Page 5982, 22(16), 5982. https://doi.org/10.3390/S22165982

33. Zhao, Q., Zhu, J., Zhu, J., -, al, Qu, H., Sako, M., Gödel, M., Lehmberg, D., Brydon, R., Bosina, E., & Köster, G. (2022). Toward learning dynamic origin-destination matrices from crowd density heatmaps. Iopscience.Iop.Org, 53401. https://doi.org/10.1088/1742-5468/ac6255

34. Zhao, X., Cui, M., & Levinson, D. (2022). Exploring temporal variability in travel patterns on public transit using big smart card data. Https://Doi.Org/10.1177/23998083221089662, 50(1), 198–217. https://doi.org/10.1177/23998083221089662

35. Zhao, Z., Koutsopoulos, H. N., & Zhao, J. (2018). Individual mobility prediction using transit smart card data. Transportation Research Part C: Emerging Technologies, 89, 19–34. https://doi.org/10.1016/J.TRC.2018.01.022