Progress in Disaster Science 25 (2025) 100405

Contents lists available at ScienceDirect

Progress in Disaster Science

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

Machine learning for human mobility during disasters: A systematic literature review

Jonas Gunkel [a,*], Max Mühlhäuser [b], Andrea Tundis [a]

[a] Institute for the Protection of Terrestrial Infrastructures, German Aerospace Center (DLR), Rathausallee 12, St. Augustin 53757, Germany
[b] Department of Computer Science, Technische Universität Darmstadt (TU Darmstadt), Hochschulstrasse 10, 64289 Darmstadt, Germany

ARTICLE INFO

Keywords:
Human mobility
Disaster mobility
Disaster response
Machine learning
Deep Learning

ABSTRACT

Understanding and predicting human mobility during disasters is crucial for effective disaster management. Knowledge about population locations can greatly enhance rescue missions and evacuations. Realistic models that reflect observable mobility patterns and volumes are crucial for estimating population locations. However, existing models are limited in their applicability to disasters, as they are typically restricted to describing regular mobility patterns. Machine learning models trained to capture patterns observable in provided training data also face this limitation. The necessity of large amounts of training data for machine learning models, coupled with the scarcity of data on mobility in disasters, often constrains the feasibility of their training. Various strategies have been developed to overcome this issue, which we present and discuss in this systematic literature review. Our review aims to support and accelerate the synthesis of novel approaches by establishing a knowledge base for future research. This review identified a condensed field of related contributions exhibiting high methodology and context diversity. We classified and analyzed the relevant contributions based on their proposed approach and subsequently discussed and compared them qualitatively. Finally, we elaborated on general challenges and highlighted areas for future research.

1. Introduction

In recent years, the economic loss and human harm caused by disasters have reached consistently high levels. Large-scale events, such as hurricanes, typhoons, and earthquakes, have caused billions of dollars in damage and claimed thousands of lives [1,2]. Moreover, these events severely threaten the vital function of critical infrastructures. Due to its spatial extent, the transportation infrastructure is especially exposed to disasters [3]. Their impact on human mobility is twofold: On the one hand, the physical components of the transportation network may be damaged and become unusable. On the other, the mobility behavior of the population is affected as people may leave their routines and behave unexpectedly [4]. Consequently, the regular spatio-temporal mobility patterns are disrupted, introducing a complexity that is challenging to comprehend. However, a systematic understanding and situation assessment of human mobility during disasters is essential for developing preventive measures or planning rescue and evacuation missions. Models that describe the mobility dynamics in such situations are necessary to gain insights into the population's location and its mobility behavior. Machine learning (ML) models can adopt this role as they can capture and reproduce patterns from observed mobility data.

The interest in ML models for predicting future mobility has raised increasingly in recent years [5], accelerated by a growing amount of mobility data. ML has been applied to various tasks in mobility modeling, e.g., predicting a person's next location, predicting crowd flows between different regions, and generating synthetic trajectories [6]. Despite the heterogeneity of ML applications for mobility, most publications have focused on regular mobility behavior, and disasters have received only limited attention. As the learned patterns of regular mobility may not be representative of disasters, the proposed models of these publications exhibit only constrained applicability to irregular situations [7,8]. Nevertheless, it can be assumed that ML is a valid and promising approach to predicting mobility during disasters: past studies revealed that mobility after an earthquake is indeed predictable [9] and frequently visited locations after a disaster are influenced by general factors such as social relationships, home, and workplace [10]. However, learning these patterns requires large amounts of mobility data for the respective situation. The general rarity of large-scale disasters and the resulting scarcity of corresponding mobility data constrain the training of ML models. Therefore, there is a great need for new strategies

* Corresponding author.
E-mail addresses: Jonas.gunkel@dlr.de (J. Gunkel), Max@tk.tu-darmstadt.de (M. Mühlhäuser), Andrea.tundis@dlr.de (A. Tundis).

https://doi.org/10.1016/j.pdisas.2025.100405
Received 29 July 2024; Received in revised form 29 December 2024; Accepted 17 January 2025
Available online 24 January 2025

or adaptations of existing ML methods to enable predictions or simulations of mobility in disasters.

To the best of our knowledge, no comprehensive overview and analysis of ML models to predict human mobility during disasters exists. To close this gap, we have conducted a literature review of the ML approaches, analyzing and discussing the proposed strategies. Our work aims to stimulate future research for this critical task by providing a foundation of knowledge, focusing on the proposed methods, their limitations, and general challenges. This study has identified areas where further research is needed, thus laying the groundwork for a successful synthesis of new ideas. Moreover, our study simultaneously provides insights into ML approaches for mobility in situations that generally face data scarcity. Our main contribution can be summarized as:

- a systematic and comprehensive analysis of existing approaches and proposed methodologies for predicting mobility during disasters,
- a discussion of the specific strategies and key challenges, as well as the opportunities and limitations of the proposed methods, and
- a foundation of knowledge and an overview of fields for future research.

The rest of the paper is structured as follows. An overview of related surveys and literature reviews is given in Section 2, followed by Section 3, precisely introducing this work's context, and Section 4, outlining the method for identifying and analyzing relevant literature. Then, Section 5 presents the findings of our review, structured in groups of similar approaches. Section 6 provides a final discussion of critical challenges and highlights research directions for future work. Conclusions are drawn in Section 7.

2. Previous related surveys and uncovered aspects

The popularity of ML approaches for capturing and predicting the dynamics of complex systems has led to an increased application of such approaches for mobility modeling. Consequently, numerous literature reviews and surveys have been published that analyze and present the contributions and advancements in this vast field of research. This section presents surveys and literature reviews related to our work and points out the limitations of these publications that this study aims to fill.

The available reviews take various perspectives, focusing on different aspects. A recent review by Luca et al. [6] provides a comprehensive overview of popular deep learning (DL) methods in human mobility and a discussion on different tasks and open challenges. Jiang & Luo [11] surveyed a subgroup of DL methods for traffic forecasting. They reviewed the application of graph neural networks, presenting the addressed tasks (e.g., traffic flow, speed or demand prediction) and a collection of open datasets. Others reviewed research on specific sub-systems of transportation. Zhu et al. [12] centered their review on railway systems from the perspective of situation perception, future state prediction, and operation optimization. In addition, Xie et al. [13] enlightened the field of urban crowd flow prediction, presenting and discussing popular machine learning methods in this field. On the contrary, Veres & Moussa [14] provided a review centering on various tasks and selected corresponding approaches. In [15], Ahmed & Diaz listed an extensive collection of open datasets and reviewed contributions addressing the tasks of passenger localization, transport mode detection, and machine learning mobility model generation. All these reviews focus on different ML methods, specific tasks, or specific aspects of mobility. In doing so, they do not sufficiently address mobility during disasters or rare events. In contrast, our work provides insights into ML approaches for predicting irregular mobility in situations that face data scarcity.

Several reviews address disasters and discuss the role of ML and big data in enhancing resilience in infrastructure systems and mobility. For example, a recent survey by Kyrkou et al. [16] explores the potential of ML for tasks such as early warning or human recognition in the different stages of emergency management. A focus on mobility data is attained by Haraguchi et al. [17] in their review on human mobility data for risk reduction and resilience. It presents different data sources on human mobility and research on the opportunities of data analytic approaches for different stages of disaster management. Additionally, Yabe et al. [18] conducted a review on the role of mobile phone location data in natural disasters and epidemics. The article provides an extensive analysis of various types of data and their application in disaster response and recovery. These contributions above investigate the role of data and ML in disaster management. However, they neglect methods to predict mobility following a disaster. This task faces high spatio-temporal complexity and is hardly comparable to other applications of ML in disaster management, such as remote sensing or human recognition.

While these surveys present and synthesize relevant approaches and perspectives, they do not thoroughly review ML methods for human mobility in disasters. As mobility during disasters exhibits peculiar characteristics (e.g., increased irregularity), and corresponding data is usually scarce, popular general methods are presumably only limitedly applicable. None of the above reviews focus on methods that can capture mobility in such extreme events. Our work aims to fill precisely this gap by analyzing, presenting, and discussing existing contributions in this field.

3. Contextual preliminaries

This section defines the scope of this review by introducing its conception of human mobility, disasters and disaster mobility.

The term *human mobility* describes the movements of human beings over a period of time. It encompasses the behavior of individuals, with the locations visited by a person reflecting their activities. In general, human mobility shows great periodicity, exhibiting recurring patterns [19]. Human mobility can be observed at the individual level, for example in the case of a person's sequence of location visitations, or at the aggregated level, such as in the case of aggregated crowd flows between a set of locations, such as points of interest (POIs).

Disasters are periods of disruption in the functioning of communities [20]. Consequently, disasters may be times in which the transportation infrastructure is partly unavailable or human behavior is disturbed to a degree that precludes continuing established patterns of human mobility. The focus of this study is on large-scale disasters that result in significant changes to human mobility patterns. Accordingly, we consider various scenarios pertinent to our research, ranging from natural disasters such as earthquakes to large-scale emergencies such as gas leakages in urban environments.

By *disaster mobility*, we refer to the response of human mobility to a disaster. Specifically, we restrict to the immediate reaction of human mobility to the occurrence of the disaster, without considering potential long-term effects. Therefore, our research is focused on the transition from typical mobility patterns to uncertain human mobility reactions to disasters.

4. Review methodology

This work presents a structured literature review (SLR) to gain a comprehensive understanding of ML approaches for predicting human mobility during disasters. The SLR is opposed to the semi-narrative and integrative reviews, which find application when a full review is hindered by an extensive body of related literature or for critically reviewing mature or emerging topics [21]. Contrarily, the SLR allows for an analysis of the current state of the art and establishes a foundation of knowledge and existing approaches for synthesizing future research. Therefore, the SLR is most suitable for identifying, analyzing, and discussing the field of research of our concern. This section presents details of our review methodology.

First, the research questions (Table 1) were defined, sharpening the scope of the review and supporting the goal of providing a comprehensive overview of the literature about ML methods for disaster mobility. The first research question (RQ1) serves to identify existing work and the state of the art methods within the field. Consequently, a comparison of the approaches is enabled. Research question two (RQ2) is designed to discern and highlight challenges that emerged from the literature discovered by RQ1. These challenges are potential topics for further research. By the third question (RQ3), methods are identified which show promising results in addressing the challenges highlighted by RQ2. These approaches may have significant potential to be further investigated for predicting human mobility in disasters.

To properly answer the aforementioned research questions, a review of the complete body of relevant literature was required. Therefore, appropriate keywords were defined that reflect the field of research framed by the defined research questions and Section 3. The three key aspects of our research – *machine learning*, *disasters*, and *mobility* – served as the cornerstone of the literature search, structuring the exploration of the field. To find a possibly complete body of relevant literature, synonyms and related terms to these main keywords were included, as shown in Table 2. Search strings were created based on these keywords and their synonyms, as depicted in Appendix A. With these search strings, literature databases, including Scopus, ACM, IEEExplore, and Web of Science, were queried by a title-abstract-keyword search to further contribute to finding a complete body of literature.

Following the literature search, the found papers were subject to a thorough filtering process. First, they were filtered by duplicates to obtain a list of unique papers. In this step, all cover letters of proceedings and other collections were removed as well. In the next step, all papers that, based on title, did not match the scope of this literature review were removed. Subsequently, all papers not written in English or not retrievable were removed. Finally, the found papers were filtered by a full review. In this last step, only papers that employed machine learning methods and addressed human mobility in disasters were retained.

As proposed by Wohlin et al. [22], a subsequent backward and forward snowball search supplemented the selected articles to ensure a comprehensive collection of all relevant literature. This snowball search aimed to identify further relevant papers that the systematic database search missed. Based on citing papers and citations of the previously reviewed papers, three additional publications were revealed. In combination, 27 papers were deemed relevant for this work. Fig. 1 depicts a graphical representation of the filtering steps.[1]

It is important to note that some publications represent extensions of previous work. That is, the approach presented by Song et al. [23] includes the former contributions [24,25], and Song et al. [10] present an extension of [26,27]. The papers presenting such extensions were considered jointly for the review to provide a concise overview of the proposed approaches. Combining such extensions decreased the set of unique contributions to 23 articles. This set constituted the final

Table 1
Presentation of the research questions. The research questions for designing the database search and conducting the review strongly focus on proposed methods and approaches.

	Research Questions
RQ1	Which existing ML methods are suitable for predicting disaster mobility?
RQ2	What are challenges and fields of future research in disaster mobility prediction with ML methods?
RQ3	Which approaches have been proposed for specifically addressing the challenges mentioned in RQ2?

[1] The work was done before December 31, 2023.

Table 2
Keywords and related terms for the literature search. The keywords for querying literature databases reflect the three topics of concern *Machine Learning*, *Disaster*, and *Mobility*.

	Keywords
Machine learning	Machine learning, deep learning, artificial intelligence, neural network, big data, data mining, urban computing
Disaster	Disaster, crisis, hazard, catastrophe, emergency, evacuation, extreme event, extreme situation
Mobility	Mobility, crowd flow, evacuation, relocation, traffic, trajectory, urban dynamics

selection of papers, which were subsequently reviewed regarding the initially formulated research questions. Consequently, the review focused on the general approaches and the proposed methodologies, the applied strategies to address the challenges peculiar to disasters, and the existing limitations of the contributions.

Based on the reviewed approaches, the selected papers were assigned to different groups reflecting similar methodologies. In this context, methodologies were considered to be similar if they were based on the same learning framework and, where applicable, employed comparable learning strategies or neural network architectures. First, the approaches were grouped according to their general learning framework, i.e., *reinforcement learning*, *unsupervised learning* and *supervised learning*. The approaches following a reinforcement learning or unsupervised learning framework were not further distinguished due to their limited number of publications. Contrarily, the approaches classified as supervised learning were subdivided depending on whether *deep learning* methods or *traditional machine learning* methods were used. Then, the deep learning approaches were grouped to *recurrent neural networks* and to approaches that use networks specifically designed to capture spatio-temporal dependencies, which were labeled *spatio-temporal neural networks*. Finally, the *situation-aware approaches*, which aim to adapt to perceived situations, and those approaches that deployed a *transfer learning* strategy were distinguished accordingly. This classification resulted in a hierarchical structure containing seven final groups of publications, as presented graphically in Fig. 2. Note that some publications could not be classified exclusively to a unique group as they comprised hybrid approaches, combining different methods, or evaluated different methods as a comparison. Such publications were classified depending on their core contribution, as depicted in Table 3.

Within the derived groups, the publications were analyzed, discussed, and compared in terms of their approaches, addressed tasks and scenarios, and results. While a quantitative comparison of the respective results would enable the identification of the best-performing approach, the intricate diversity of contexts and tasks addressed posed a significant challenge, making such an endeavor unfeasible. Moreover, most datasets used for training the proposed models are not publicly available, further hindering a benchmark test of the models' performance. A summary of performance metrics is presented, where available, to address this issue. If quantitative results were not provided, qualitative information on the results is included in the presentations. The results of the analysis and comparison of the reviewed publications are presented in the following section.

5. Machine learning methods for human mobility in disasters

This section presents the different ML approaches for human mobility prediction in disasters proposed in the reviewed papers. It is structured in different subsections according to the deployed learning framework. First, the group of supervised approaches is presented, split into subsections according to Fig. 2. A presentation of the application using unsupervised learning follows. Finally, the last subsection provides an overview of related reinforcement learning approaches. Each subsection first presents the respective papers, followed by a brief discussion. This discussion compares the presented approaches and

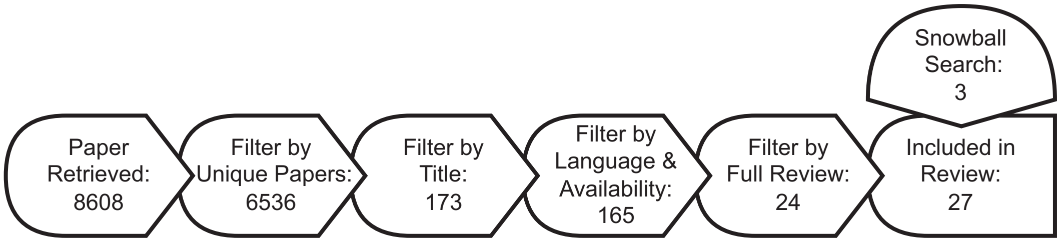

Fig. 1. Literature filtering process. The initially found 8608 papers were reduced to 24 papers during the filtering process. Supplemented with 3 papers found by a snowball search, the final set of papers comprised 27 publications.

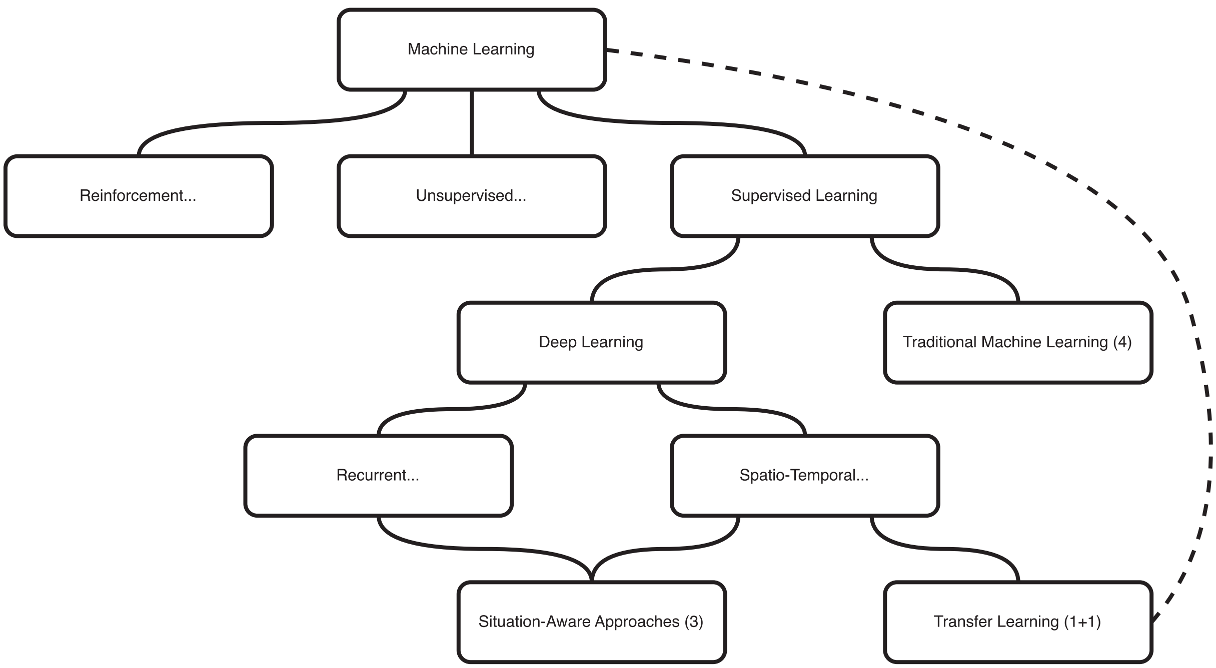

Fig. 2. Hierarchical presentation of the found approaches. Here, the numbers indicate the amount of papers within each group. Note that the *situation-aware approaches* constitute a subset of the *recurrent neural networks* and the *spatio-temporal neural networks*. Additionally, the *transfer learning* group is only partially a subgroup of the *spatio-temporal neural networks*, as it includes one approach that this framework could not adequately classify.

Table 3
The two approaches that did not allow for a unique classification were assigned to the group reflecting their core contribution.

Authors	Approach	Assigned Group
Song et al. [28]	Unsupervised & supervised learning	Unsupervised learning
Chikaraishi et al. [29]	Comparison of different models including DL models and traditional ML models	Traditional ML approaches

identifies their opportunities and most pressing limitations. Furthermore, a tabular summary is provided for each subsection, providing a distilled overview of the respective publications. The table presents general information about the addressed task, learning strategy, adopted model or architecture, and used data for most subsections. In addition, the availability of the data and code and the evaluation metrics employed are included. For those publications that provide quantitative results for experiments which specifically concern disasters, a brief summary of these results is included as well. If such information is unavailable, the respective cell is marked with "n.a." (not available). The tables in Section 5.1.4, Section 5.1.5, and Section 5.3 have a structure adapted to the respective subsection.

5.1. *Supervised learning*

This subsection concentrates on research contributions that adopt supervised learning. First, we introduce the identified traditional ML approaches which do not employ DL methods. Second, we present the reviewed DL approaches, distinguishing them depending on their architecture, as depicted in Fig. 2. We start by presenting approaches using recurrent neural networks, followed by works proposing spatio-temporal neural networks. Then, we highlight different architectures and approaches to establish a situation-awareness. Finally, methods that rely on transfer learning are presented.

5.1.1. *Traditional machine learning*

Traditional ML approaches such as support vector machines (SVMs) or decision trees (DTs) have been successfully used for a variety of tasks in mobility modeling [15]. Several works have adopted such traditional methods to solve different tasks in disaster mobility prediction. An overview of the according publications, including information on their methodology and results, is given in Table 4.

Anyidoho et al. [30] addressed predicting evacuation flows caused by a hurricane using demographic, GPS, and hurricane and flood severity data. Their model consists of two stages. At the first stage, a classifier is trained to decide whether a positive evacuation flow exists between a given origin-destination (OD) pair. The classifier is a gradient tree-boosting model with XGBoost and returns the probability of a

Table 4
Comparison of different traditional ML approaches. The selected works investigate scenarios with only limited complex spatial dependencies.

Authors	Disaster	Task	Strategy	Model	Data	Evaluation	Result	Data/Code available?
Anyidoho et al. [30]	Hurricane	Evacuation flow prediction between cities	XGBoost applied to decision trees	XGBoost applied to decision trees	GPS data, hurricane data, survey data	PDE, R^2, RMSE, CPC	Incoming evacuees PDE: 0.053, R^2: 0.843, RMSE: 1.567 OD prediction CPC: 0.544, R^2: 0.546, RMSE: 101.635	No/No
Chikaraishi et al. [29]	Heavy rain, landslide	Traffic flow & street occupancy prediction	Comparison of different models	ARIMA, vector autoregression, random forest, XGBoost, SVM, shallow & deep multi-layer perceptrons	Traffic sensor data	R^2, MAE	Flow (R^2/MAE) RF: 0.8/0.31, XGB: 0.82/0.29 Occupancy (R^2/MAE) RF: 0.93/0.18, XGB: 0.94/0.17	No/No
Khaefi et al. [31]	Volcano eruption	Evacuation destination prediction	XGBoost applied to decision trees	XGBoost applied to decision trees	Call details record data	AUC	0.77	No/No
Yabe et al. [32]	Earthquake, typhoon	Overall daily delay prediction in commuting	Multi-class classification & regression	Logistic and linear regression	GPS data	Predictive accuracy, Pearson corr.	Predictive accuracy: 64.8 %, Pearson correlation: 0.76	No/No

positive evacuation flow. If this probability surpasses 50 %, a regressor estimates the magnitude of the flow at the second stage. Again, the regressor is based on gradient tree-boosting. Their boosting strategy aims for a flexible model capable of capturing the spatio-temporal characteristics of evacuation flows. Specifically, their results show that their model has increased capabilities to learn long-range evacuations compared to traditional gravity models. Moreover, the resulting OD matrix coincides with the ground truth by more than 50 %.

Another task of disaster mobility prediction was addressed by Chikaraishi et al. [29]. They tested different methods to predict the traffic flow volume and street occupancy, i.e., the share of time a car occupies a location, after heavy rains and subsequent landslides on a highway in Japan. The evaluated methods include ARIMA, vector autoregression, random forests, SVM, XGBoost, and shallow and deep multi-layer perceptrons. Among these methods, XGBoost and random forest performed best in the investigated setting with a single street. However, the authors remarked that these methods may encounter difficulties extrapolating to unseen data.

Khaefi et al. [31] trained a model to predict the evacuation destination in case of volcano eruptions on the archipelago of Vanuatu. They used a decision tree model with XGBoost. Based on an island's inhabitants' call details records, the model estimates their individual evacuation destination island.

Yabe et al. [32] developed a model to predict the irregularity of commuting induced by a disaster, meaning the total temporal disruption of daily commuting. Their method can be split into two stages. The authors first employed a logistic regression, classifying whether a trip from one location to the next is started with a significant delay, on time, or significantly early. Based on the predicted class, a subsequent linear regression estimates the duration of the irregularity. In aggregation, the model returns the total duration of irregular mobility, providing an overview of the disaster's impact. Yabe et al. tested their approach with data on earthquakes and typhoons, evaluating the performance of predicting different activity times, such as leaving home or working. Their results show an accuracy of around 60 %–75 %, depending on the considered activity. While the general pattern across the activities shows high similarity among the two scenarios investigated, the results for typhoons exhibit an overall higher accuracy.

The above-presented publications investigate heterogeneous disasters and tasks following different traditional ML methodologies. In the following, the approaches are compared and discussed in relation to different aspects. The first aspect addresses the *complexity of the scenarios* investigated in the respective publications. It concerns the suitability of the presented approaches to capture the dynamics in the considered disasters and transportation systems. The second aspect focuses on the *generalizability* of the approaches. Here, the extent to which the proposed models may be applied to predict the dynamics in unseen disasters is discussed. The third aspect centers on the *data requirements* of the approaches. This aspect addresses the discrepancy between the quantity of necessary data and the limited availability of data.

5.1.1.1. Complexity of scenarios. While all these approaches appear to achieve their objectives, they are used for scenarios without complex spatial dependencies (i.e., [29,32]) or reduce the large spatial system to a linear setting (i.e., [30,31]). Presumably, these approaches are limited in their ability to predict the complex dynamics of large-scale transportation systems. Moreover, it is unlikely that traditional ML approaches can outperform DL models, which have been proven to capture deep spatio-temporal dependencies successfully. A recent survey discusses this aspect and emphasizes the superiority of DL over traditional ML methods to capture the complexity of transportation systems in different tasks, such as crowd flow or trajectory generation [6].

5.1.1.2. Generalizability. As indicated by Chikaraishi et al. [29], the proposed models may exhibit poor performance when deployed for events or locations other than those for which they were originally developed and trained. While this limitation is pertinent to a wide range of ML methods, the presented works do not include any strategies to improve the generalizability of their approaches.

5.1.1.3. Data requirements. One striking advantage of the proposed traditional approaches is their reduced need for data compared to DL approaches. Therefore, they may be less impacted by the scarcity of mobility data in disasters and thus constitute promising approaches to attaining an overview of such situations.

5.1.2. Recurrent neural networks

Recurrent neural networks (RNNs) have been proposed to model sequential relationships such as temporal dependencies over a period of time [33]. Therefore, they are a popular choice for predicting mobility, as they can learn the temporal patterns for estimating future states [34–36]. Especially long short-term memory (LSTM) [37] and gated recurrent unit (GRU) [38] networks have been used in several applications such as next location or future flow prediction [39,40]. Several approaches based on RNNs have been developed to predict mobility during a disaster. A summary of the respective publications can be found in Table 5.

For instance, Jiang et al. [41] proposed an LSTM encoder-decoder model to predict the next multiple locations of a trajectory. Their model comprises an LSTM layer that encodes the α last observed trajectory locations, followed by m LSTM layers as decoders, each producing a unique snippet of a fixed number of predicted future steps. Concatenating these snippets yields the final multi-step prediction. Using multiple decoders, the authors circumvented the error propagation problem in multi-step predictions [41]. An application of their approach to the 2011 Tōhoku earthquake in Tokyo shows a performance comparable to the application to regular weekdays. Moreover, a graphical representation of the predictions indicates that the model can clearly learn general evacuation routes.

Rahman & Hasan [42], Roy et al. [43], and Afrin et al. [44] predicted the evacuation traffic speed (Rahman & Hasan) or volume (Roy et al. & Afrin et al.), on an interstate road in Florida using LSTM networks. While Rahman & Hasan [42] included temporal metadata, as well as the upstream and downstream sensor data for a location of concern, the model by Afrin et al. [44] was trained only with traffic sensor data. Roy et al. [43] included hurricane-related tweets from Twitter in their long-term traffic volume prediction. All three approaches rely on a one-layer LSTM network to capture the sequential dependencies. Opposed to the pure LSTM network in [42,43], the model in [44] comprises an LSTM layer, followed by a flatten and dense layer and a subsequent Kalman Filter to compute the final output. All three works include a case study on traffic during Hurricane Irma. Afrin et al. [44] and Rahman & Hasan [42] presented similar results, showing notable differences in prediction errors between consecutive sensor locations. Moreover, the model of Rahman & Hasan achieved a high accuracy at constant traffic speeds, while the errors increased with strong fluctuations. Roy et al. [43] demonstrated the ability of their model to predict long-term traffic volumes up to 24 h in advance.

A different task is addressed by Mahmud et al. [45], as they incorporated a recurrent network for modeling evacuation decisions in an evacuation path planning system. They deployed a Latent RNN [46], which contains an additional latent variable, allowing the network to model more complex dependencies. This latent variable is supposed to encode the relationship between people's evacuation responses and the hurricane features. The evaluation of their approach in a hurricane scenario showed the superiority of their Latent RNN model over standard RNNs. For most points in time, their model reliably predicted the evacuation response time but clearly deviated from the ground truth directly before and after the hurricane's landfall.

Jiang et al. [47] proposed an LSTM network to predict short-term population density during earthquakes. The authors considered a population density map as an image (each pixel corresponds to a grid cell in a city). They trained their model to predict the next population density map, similar to the next frame in videos. They tested their model with different history windows: the single most recent step and the last five most recent steps. Generally, their results show the model's ability to capture short-term changes in population density. However, the results for the different history windows exhibit significant differences, as a history window of the last five steps results in an improvement of more than 40 % compared to the history window of only one step.

Fan et al. [48] adopted an ensemble of GRUs to learn the dynamics of trajectories on single days of historical data. For each day, a single stacked GRU network is trained to predict the next location of a given trajectory. These GRU networks are combined in an ensemble framework. This framework further features an online GRU learner, accounting for short-term dynamics. The online ensemble framework's purpose is to predict a trajectory's future locations as a combination of the individual networks' output. This combination is created using an attention mechanism as a gating function. This gating function decides on a linear combination of the different GRU networks. The online ensemble approach is designed to adjust dynamically to observed situations and thus is further discussed in Section 5.1.4.

All proposed models were developed for tasks that exhibit sequential dependencies of temporal (e.g., evacuation decisions [45], traffic volume [44], or traffic speed prediction [42]) or spatial extent (e.g., next location(s) [41,48] or traffic volume prediction [44]). These tasks are predestined to be solved with recurrent architectures, as presented in this section. Although there exist differences regarding the specific tasks, several aspects are similar across the respective publications.

5.1.2.1. Complexity of scenarios. As previously stated, the presented RNN approaches aim to predict sequential phenomena. However, the considered settings exhibit spatial dependencies of only limited complexity (e.g., fixed and strictly ordered sensor location [42] or single trajectories with dependence on limited historical steps [41,48]). For predicting the mobility in large-scale transportation systems, which exhibit complex spatial dependencies, incorporating these may be of great importance. The presented RNNs may be insufficient to capture these dependencies adequately. In contrast, augmenting recurrent modules with DL modules such as convolutional neural networks (CNN)

Table 5
Comparison of different works that adopt RNNs. All of these works propose approaches to predict sequential phenomena.

Authors	Disaster	Task	Strategy	Architecture	Data	Evaluation	Result	Data/Code available?
Jiang et al. [41]	Earthquake	Next locations prediction	Sequence learning	RNN encoder-decoder	GPS data	MAE, RMSE	30 min/60 min prediction MAE: 1250/2050, RMSE: 1400/2300	No/No
Rahman & Hasan [42]	Hurricane	Evacuation traffic speed prediction	Sequence learning	LSTM	Traffic sensor data	RMSE, MAE, MAPE	RMSE: 2–4, MAE: 1.5–3, MAPE: 2 %–4.5 %	No/No
Roy et al. [43]	Hurricane	Evacuation traffic volume prediction	Sequence learning	LSTM	Traffic sensor data, Twitter data	RMSE, MAPE	1 h/15 h prediction RMSE: 110/160, MAPE: 13 %/25 %	No/No
Afrin et al. [44]	Hurricane	Evacuation traffic volume prediction	Sequence learning	LSTM, Kalman filter	Traffic sensor data	MAE	MAE: 39.312	No/No
Mahmud et al. [45]	Hurricane	Evacuation decision prediction	Sequence learning	Latent RNN [46]	GPS data, road network data, weather data	RMSE, MAPE	RMSE: 0.0005, MAPE: 30 %	No/No
Jiang et al. [47]	Earthquake	Population density prediction	Sequence learning	LSTM	Mobile data, land use data	MSE	MSE: 11029/6742 (1 step/5 step history window)	No/No
Fan et al. [48]	Earthquake	Next location prediction	Online ensemble learning	GRUs, attention	GPS data	Prediction loss (cross entropy)	Prediction loss during earthquake: 0.2–0.9	No/No

or graph convolutional networks (GCN) has become a popular approach for capturing spatial and temporal dependencies simultaneously [6].

5.1.2.2. Data requirements. As stated in Section 5.1.1, the presented DL models require a substantially larger quantity of training data to produce meaningful results than traditional ML methods. In principle, data on mobility in disasters exist, such as the GPS data from Japan used by Jiang et al. [41] and Fan et al. [48], or as the mobile data used by Jiang et al. [47]. However, such data sources are often undisclosed (Table 5), which restricts reproducing the presented results and comparing the performance of different approaches. One exception is given by traffic sensor data, as used by Rahman & Hasan [42], Roy et al. [43], or Afrin et al. [44]. In many cases, such data is provided by governmental institutions. However, it should be noted that these data cannot be assumed to be representative of mobility in general, as they commonly restrict to motorized traffic on major streets and fail to represent full trips from origin to destination. Moreover, traffic sensor networks depend on power supply, which may be disrupted during disasters, resulting in incomplete data. Generally, the rarity of disasters restricts the amount of any corresponding mobility data. Consequently, all approaches face the challenge of reduced training data. Fan et al. [48] addressed this challenge by basing their model on large mobility data on regular situations and adjusting it based on limited disaster data.

5.1.2.3. Generalizability. The sequential phenomena that the presented contributions aim to predict are all linked to a specific event and a given location. Therefore, the trained models are likely to be inapplicable to other locations or events, being *location-specific* or *event-specific*, respectively. Especially the models relying on traffic sensor data (i.e., [42–44]) face the limitation of being location-specific due to the specific spatial arrangement of the sensors. This limitation also applies to the contributions aiming to predict the next location of trajectories (i.e., [41,48]). As these models are trained with sequences of coordinates, they are restricted to the corresponding areas. In contrast, these models may be applicable to other disasters with similar severity and equal spatial extent, as the resulting mobility patterns may be comparable. In particular, the work presented by Fan et al. [48] reaches beyond the event-specificity by creating a model that can adapt continuously to short-term observations. Moreover, the model developed by Mahmud et al. [45] may be applicable to other hurricanes, as the authors aim to establish a connection between general hurricane features (e.g., wind speed or distance to hurricane center) and the evacuation decision. However, if the spatial extent of a disaster deviates from previously observed ones, one cannot assume to predict the mobility in such a situation with the proposed approaches reliably.

5.1.3. Spatio-temporal neural networks

Deep learning has been used extensively to capture the complex spatio-temporal patterns in human mobility [6]. Specific network architectures, which may be referred to as spatio-temporal neural networks, have been proven particularly effective in achieving this objective. These architectures often consist of combinations of recurrent networks (e.g., LSTM, GRU) to capture sequential dependencies and convolutional networks (e.g., CNN, GCN) to discern spatial dependencies. Examples of such combinations are graph convolutional recurrent unit (GCRU) [49] or convolutional LSTM network (ConvLSTM) [50]. An overview of approaches incorporating these combinations in human mobility in disasters is presented in Table 6.

Jiang et al. [7] and Wang et al. [8] leveraged GCRUs in their proposed models. The model developed by Wang et al. [8] comprises an encoder-decoder setup based on GCRUs. Here, the encoder and the decoder share the same architecture of a two-layer GCRU network in a pyramidal structure. In this structure, the sequential outputs of the first layer are pairwise concatenated and used as input for the second layer, reducing the dimension of the inputs. In contrast, the model proposed by Jiang et al. [7] relies on a one-layer GCRU encoder. The encoder,

Table 6
Comparison of different spatio-temporal neural networks. These networks combine recurrent networks and convolutional networks to capture the spatio-temporal characteristics of human mobility.

Authors	Disaster	Task	Strategy	Architecture	Data	Evaluation	Result	Data/Code available?
Jiang et al. [7]	Typhoon, hurricane	In- and outflow prediction of regions (typhoon)/POI visitation prediction (hurricane)	Meta learning	GCRU encoder, one-shot decoder	GPS, POI visitation, Twitter, spatial metadata	RMSE, MAE, MAPE	Typhoon RMSE: 2470.5/2491, MAE: 805.8/822.1, MAPE: 9.62 %/9.87 % Hurricane RMSE: 718.6, MAE: 247.2, MAPE: 28.96 %	Yes/Yes
Wang et al. [8]	Blizzard	Multi-modal mobility volume prediction	Dynamic filter generation	Pyramidal GCRU encoder-decoder, attention	Bike-sharing, taxi trips, POI visitation	RMSE, MAE, MAPE	Blizzard NYC RMSE: 23.63, MAE: 15.79, MAPE: 33.33 % Blizzard DC RMSE: 4.103, MAE: 2.004, MAPE: 35.03 %	Yes/Yes
Jiang et al. [51]	Earthquake, typhoon	Crowd flow and density prediction	Multitask learning	Convolutional LSTM	GPS	MSE	MSE (density/flow) Earthquake: 5.549/0.102, Typhoon: 6.753/0.17	No/Yes
Rahman & Hasan [52]	Hurricane	Edge flow prediction in transportation network	Transfer learning	LSTM, GCN	Traffic detector	RMSE, MAE, R^2	RMSE: 399.69, MAE: 268.03, R^2: 0.943	No/No
Hao & Wang [53]	Hurricane	Mobility volume prediction on census tract level	Explainable AI	GCN, LSTM, attention	GPS, demographic, land use, flood vulnerability, weather	RMSE	RMSE: 0.065 (seen city and event)/0.112 (unseen event)/0.151 (unseen city)	Yes/Yes
Zhiwen et al. [54]	Typhoon	Causal impact prediction on mobility volume	Continuous treatment prediction	RNN, GCN	GPS, weather, Google trend	RMSE, MAE	RMSE: 0.012–0.118, MAE: 0.008–0.077	No/No
Wang et al. [56]	Typhoon	Traffic speed prediction	–	GCN, GRU	Car hailing data, weather data	SMAPE, MAE, RMSE	SMAPE: 7.2 %, MAE: 0.67, RMSE: 1.03	No/Yes

augmented with social, temporal, and spatial metaknowledge, is decoded in one step to reduce error accumulation [7]. Both models [7,8] incorporate strategies to establish a situation-awareness and thus are further examined in Section 5.1.4.

Jiang et al. [51] addressed the challenge of simultaneously predicting crowd density and flow as a multi-step prediction. Their model consists of an encoder-decoder network based on stacked ConvLSTM networks. In the initial step, the dimension of the crowd flow data is reduced to support an equal treatment of crowd density and crowd flow tensors. This task is done by a CNN autoencoder, which compresses the relevant information in a low-dimensional embedding. The actual encoder-decoder model comprises four stacked ConvLSTM layers. In the first layer, flow and density data from a fixed number of preceding time steps are encoded independently. The resulting encodings are combined in the second layer to produce a single vector representation of the input data. The decoding process is conducted in a reverse order to the encoding. The third layer decodes the vector representation to produce joint representations of flow and density. The last layer then splits these representations and produces the final multi-step prediction. This approach was evaluated for an earthquake and a typhoon in Tokyo. A visual comparison of the graphical presentation of the predicted time series and the ground truth demonstrates that the model can predict disturbed mobility. However, the graphs indicate a tendency for the model to overestimate crowd density during disruptions slightly.

Other works consider synthesizing GCN and LSTM networks. For instance, the model proposed by Rahman & Hasan [52] is based on a GCN to capture the spatial dependencies, with the underlying graph representing the transportation network. The convolution operations on the graph are computed using a dynamic adjacency matrix. It represents the time-dependent evolution of the connectivity (i.e., travel time) between the graph nodes by adapting the edge weights to the observed traffic situation. An LSTM layer further processes the GCN output and returns the traffic flow on each edge of the graph. This model is embedded in a transfer learning framework, further presented in Section 5.1.5.

Hao & Wang [53] chose a different approach, predicting the frequencies of mobility flows between census tracts during extreme weather. In their model, various census tract-level metadata are first embedded by a dense layer. The resulting representation of a census tract is fed into an LSTM layer that encodes the mobility frequencies for each census tract at each time step. Subsequently, these encodings are processed by an attention layer, enabling the capture of potential delays between the occurrence of weather hazards and human response. In the next step, a GCN layer aggregates this information from neighbored census tracts. This aggregated information is combined with the encodings to compute the output. Different experiments were conducted to analyze the performance of the model in comparison to different model variations. This comparison shows that the full model outperformed most of the chosen variations. Moreover, experiments were conducted to evaluate the model's performance when applied to previously unseen cities or hurricanes rather than those observed during training. The results demonstrate that the performance declined by approximately 50 % or 33 %, respectively. Given that the approach does not include any strategies to adapt to unseen events or locations, these results arouse hope toward a general machine learning model for disaster mobility.

The original RNN, in combination with a two-layer GCN, was proposed by Zhiwen et al. [54] to predict the causal impact of typhoons on mobility volume. Their model aims to minimize the influence of confounders by interpreting this task as a continuous treatment problem [55]. Here, the treatment is the weather hazard, and the treatment's effect is the weather's impact on human mobility. The confounder represents observed situations by encoding spatial metadata, google search trends, and past mobility. Their model comprises an RNN layer, followed by a GCN to compute an encoding, representing the confounders. A subsequent softmax layer represents the conditional density of the weather hazard, given the confounder. This layer tightly links the confounder encoding to the treatment, eliminating undesired features of the observed situation to estimate the causal effect of extreme weather on mobility. The authors conducted experiments on the responses of walking, car, and train traffic affected by a typhoon. The results show that train-based mobility strongly responds to winds, whereas car-based mobility correlates with rainfall. Moreover, the results indicate that the decision to stay at home is not strongly influenced by rainfall or wind.

Wang et al. [56] combined GRUs with GNNs to predict the traffic speed during hurricanes. Using car-hailing data, the authors constructed a graph representing the road network. The nodes of this graph contain the up- and downstream traffic speed and meteorological data. For this graph, the authors combined a GNN with a GRU in an encoder-decoder setup to predict the traffic speed for several future hours. Furthermore, they used the traffic speed predictions to estimate the resilience of the traffic network under consideration.

The presented spatio-temporal neural networks employ architectures based on RNNs combined with different convolutional networks. Additionally, several other aspects are common to the approaches covered in this section.

5.1.3.1. Complexity of scenarios. The combination of recurrent and convolutional networks benefits from its increased capability to capture deep dependencies in large spatio-temporal systems. Consequently, these models enable to consider scenarios with large-scale transportation systems such as entire cities [8,51,56] or countries [7,54]. However, the enhancement of the models to capture spatial dependencies may be accompanied by an increase in complexity, i.e., an increase in the number of model parameters that must be optimized during training.

5.1.3.2. Data requirements. The increase in model complexity necessitates a greater quantity of training data. This need for training data opposes the naturally scarce data on disasters. Consequently, the data scarcity may constrain the models' expressiveness, thereby impeding their performance. The works presented by Jiang et al. [7] and Wang et al. [8] specifically address this limitation. Their strategy involves learning general mobility patterns and adapting to currently observed situations. Therefore, they can rely on more training data without being dependent on extensive disaster mobility data. Similarly, Rahman & Hasan [52] circumvented the lack of disaster mobility data by pre-training a model on general mobility data and fine-tuning it with limited data to reflect the disaster.

5.1.3.3. Generalizability. Reducing the amount of required mobility data on disasters not only circumvents the scarcity of such data but may also improve the model's performance in unseen events. For instance, the approaches by Jiang et al. [7] and Wang et al. [8] mentioned above may be applicable to a variety of disasters that happen in the exact location. However, these approaches are based on the assumption that mobility during a disaster may be represented as a combination of mobility patterns from previously observed situations, even regular situations. This assumption is further discussed in Section 5.1.4. Furthermore, learning deep spatial features adapts a model specifically to the location of consideration. Therefore, the spatio-temporal networks are likely to be highly location-specific and generalize poorly to unseen locations. This limitation is reflected in the results presented by Hao & Wang [53], which demonstrate that their proposed model performed better for unseen events than for unseen locations.

5.1.4. Situation-aware approaches

As discussed before, ML approaches for human mobility are typically limited to reproducing patterns present in the training data. To address this limitation, dynamically adapting to observed situations has been incorporated in various models. During the training process, these models store situation-specific information. In order to predict future states, the models first perceive the current situation. Subsequently, they adapt dynamically to accommodate the perception. For instance, specific components, such as the weights of particular modules, are modified. By doing so, these models establish a situation-awareness that increases their flexibility, possibly leading to improved performance in unseen situations. An overview of such approaches in disaster mobility prediction is given in Table 7. This table builds upon the information presented in Table 5 and Table 6 and only provides details pertinent to this section.

Jiang et al. [7] proposed a model based on meta learning. In general, meta learning can be described as "learning to learn" [57] and refers to gaining experience during training, often from a distribution of related tasks. In meta learning, there are various methods to acquire that experience. Here, Jiang et al. [7] followed a model-based memory augmentation [58], where memory storage is established during training. As described in Section 5.1.3, their model is based on GCRUs and is split into two networks that attain the role of teacher and student, respectively. The *teacher network* learns spatial, temporal, and social metaknowledge and builds a memory pool that stores the learned metaknowledge during training. Given observed social metadata, an according state from the memory pool is queried for augmenting the data processed by the *student network*. The student network consists of a GCRU encoder that depends on the teacher network in two respects: On the one hand, the GCRU is parameterized by spatial metaknowledge from the teacher network, which adapts the graph convolutions to observed spatial metadata. On the other hand, the human mobility data is augmented with social metaknowledge processed by the teacher network, representing the population's situation perception. A subsequent decoder generates the final output of the model. To test their model, Jiang et al. [7] mined Twitter data as a social covariate representing the disaster context and the population's situation perception. They conducted experiments on mobility during a typhoon's landfall in different prefectures in Japan and a hurricane's landfall in several counties in Florida. Their results show that their approach outperformed the chosen baselines without a memory pool, indicating the model's capability to detect anomalous situations and adjust the output accordingly. Additionally, they performed a case study on typhoon Hagibis. The study demonstrates that their model reliably estimates the pattern of short-term human outflow from the affected prefectures before and during the landfall.

A similar approach is taken by Wang et al. [8], who employed a dynamic filter generator to adapt the model based on its perception of the current mobility. Their model is trained to generate a memory pool of parameter sets depending on the mobility states observed in the training data. For predicting future mobility states, this memory pool is queried by an encoding of current mobility data and returns an initial set of parameters. A dynamic filter network [59] further processes these parameters, which returns a set of weights. This resulting set of weights parameterizes a GCRU encoder-decoder module that predicts the mobility demand. Here, the current situation is perceived solely through mobility and time data. An evaluation of this approach for taxi demand during a blizzard shows that the model detected the immediate drop in demand and adjusted its prediction accordingly.

Fan et al. [48] adopted a different perspective on situation-aware mobility prediction. They proposed an online ensemble learning approach. In ensemble learning, multiple base learners are combined to form a more robust prediction with enhanced generalization capabilities [60]. Fan et al. [48] trained an ensemble model of base GRU predictors with data from distinct single days, learning mobility patterns in historical data of only the respective days. In addition, their model is supplemented by an online GRU network based only on recently observed data. Thus, this online network introduces a short-term prediction of the current mobility. Finally, a gating function based on a one-layer GRU network computes a linear combination of the different predictors, yielding the model's output. Applying this approach to the 2011 Tōhoku earthquake in Tokyo, the authors found that the online GRU network was assigned a significantly higher weight than the others, reflecting the unprecedented nature of the earthquake. This result further highlights the discrepancy between mobility during regular situations and disasters.

The approaches presented above aim to acquire a situation perception upon which the models adjust their output dynamically. Although the approaches aim at a similar goal, they differ significantly in their methodologies. While Jiang et al. [7] leveraged various metadata to establish the perception, the model by Wang et al. [8] perceives the current situation solely through mobility and time data. Moreover, the teacher network proposed by Jiang et al. [7] only appends and informs the student network in the meta learning framework, whereas the dynamic filter generator approach by Wang et al. [8] intervenes deeper by adapting the weights of the encoder-decoder setup directly. In contrast, the ensemble learning model by Fan et al. [48] follows a different strategy. Instead of adapting upon querying a stored state, this model combines different networks that store knowledge from different situations. Still, all presented approaches may benefit from increased flexibility and robustness, as well as their ability to adapt to observed situations. Moreover, as the different methods build memory from previously observed situations, they inherit an increased capability to learn from rare events. In the previous sections, these presented approaches have already been discussed in terms of *complexity*, *data requirements*, and *generalizability*. For this section, the *uniqueness of disasters* emerges as an additional aspect that requires discussion. Here, the particularity and discrepancy of mobility in different disasters and its relevance to the presented approaches are examined.

5.1.4.1. Uniqueness of disasters. The methodologies presented in this section are based on the assumption that the mobility patterns in an unseen situation can be inferred from the knowledge gained from past mobility events. In particular, this assumption also encompasses the possibility of extrapolating patterns of disaster mobility from regular mobility patterns. This assumption may conflict with the irregularity of mobility during disasters. Depending on the severity of a disaster, patterns of regular mobility may be insufficient to describe patterns of disaster mobility. To address this limitation, Fan et al. [48] included an online learning predictor in their ensemble learning framework, enabling the incorporation of recent observations of disaster mobility. This strategy may enhance the model's capacity to adapt to a specific disaster under consideration, as the resemblance of currently observed patterns to past ones may decrease proportional to their temporal distance, especially in disasters. Moreover, the uniqueness of disasters is linked to the generalizability discussed in the previous sections. A model can only predict future mobility satisfactorily if the event under

Table 7
Comparison of different approaches aiming to train a situation-aware model. These models benefit from their ability to actively adapt to observed situations by relying on stored information. This table presents an extension for the information about the considered approaches given in Table 5 and Table 6. Consequently, it does only contain information that is relevant to this section.

Authors	Strategy	Approach
Jiang et al. [7]	Meta learning	Teacher-student network, memory pool. The teacher network establishes a situation perception and passes this information to the mobility encoder-decoder.
Wang et al. [8]	Dynamic filter generation	Memory pool, weight adaption. The filter (weights of the encoder-decoder network) is computed based on the situation perception.
Fan et al. [48]	Ensemble learning	Ensemble of networks trained on single days. An attention mechanism combines these multiple predictors based on the perceived situation.

consideration shares central properties with the events in the training data.

5.1.5. Transfer learning

Transfer learning (TL) refers to leveraging learned knowledge from data on a *source domain* to enhance the performance of a model for a related *target domain* [61]. In the context of mobility, TL has been employed to estimate the mobility in an unseen city or during an unseen event based on sufficient data from another city or similar event [62,63]. Similarly, for disasters, this strategy enables predicting future mobility or generating trajectories during disasters as a transfer from one city to another or as a transfer from regular to disastrous events. However, to date, only a limited number of related works exist, resulting in only two publications that have been deemed relevant for this section of the review. In Table 8, an overview of contributions based on TL for predicting human mobility in disasters can be found.

Rahman & Hasan [52] investigated pre-training a model with extensive data on regular situations and adapting it to the disaster by fine-tuning it with scarce corresponding mobility data. As presented previously in Section 5.1.3, their model consists of LSTM networks and a GCN, where the underlying graph represents the transportation network. This model is pre-trained with extensive traffic sensor data on regular traffic in Florida. To predict evacuation traffic, a gating function is trained with limited evacuation data while keeping the pre-trained parameters frozen. This gating function learns the relevance of different traffic flow features for the evacuation context and reweights the output of the pre-trained model. Enhanced with an encoding of current evacuation demand features, the gating function's output gives the final prediction as flows on the graph's edges. Rahman & Hasan conducted a case study on mobility during the evacuation before a hurricane. They found that the predictive performance for the evacuation performance was significantly increased by including the trained gating functions, as MAE and RMSE decreased by at least 74 %. However, compared to the performance for regular mobility, the model showed an increase in RMSE and MAE of more than 75 %, emphasizing the challenge of transfer across different events with scarce data.

A different perspective on TL was adopted by Fan et al. [64]. Methodologically, their work can be assessed as transfer learning in a more expansive sense. The authors developed an approach that aims to translate observed trajectories during a disaster across cities. The first stage of their methodology involves establishing a cell-to-cell matching between grids in two cities based on trajectory data obtained under regular conditions. Initially, this matching is estimated based on the cities' population densities over a given period. Subsequently, the matching of cells is updated iteratively by consulting a trajectory similarity metric and a subset of trajectories for each city. The final cell-to-cell matching can serve as a translator of trajectories between cities. To translate observed trajectories, the authors proposed a Hidden Markov Model. Here, the Viterbi algorithm is adopted to generate trajectories from sampled emission and transition probabilities of new trajectories. The authors demonstrated the applicability of their approach to disasters by simulating the impact of an earthquake in Osaka based on trajectories during an earthquake in Tokyo. First, the cell-to-cell matching was generated using large amounts of trajectory data on normal conditions for both cities. Second, limited data on trajectories during the earthquake in Tokyo served as the basis for generating trajectories in Osaka. The simulation indicates a significant disruption immediately after the occurrence and an increased number of trajectories leaving the city in the following hours.

The proposed models demonstrate the potential of TL in disaster mobility prediction. They are promising approaches, particularly in disasters where data is usually scarce. That is, as these methods directly address the *generalizability* of mobility predictions. Nevertheless, such approaches may be limited by their underlying assumptions, which are opposed by the *uniqueness of disasters*.

5.1.5.1. Generalizability. TL offers the opportunity to transfer captured patterns from the source domain to the target domain. Consequently, this strategy represents a promising approach to simulating mobility in unprecedented disasters. A meaningful and realistic simulation may be achieved by directly incorporating the relationship between the source and the target domains in the training process. However, this strategy is constrained by its potential inability to achieve universal generalizability, as the resulting models are typically tailored to the specific target domain and situation of concern.

5.1.5.2. Data requirements. The principal advantage of adopting TL for mobility predictions is the reduced need for data on the target domain. Therefore, scarce mobility data on disasters may no longer present a restriction for training ML models. For instance, the approach proposed by Fan et al. [64] does not require any mobility data on disasters in the considered target domain. However, leveraging extensive data on the source domain transfers mobility patterns to the target domain. Consequently, the central properties of the source and target domains must be similar to predict the dynamics in the target domain realistically.

5.1.5.3. Uniqueness of disasters. Similarly to the contributions presented in Section 5.1.4, transferring mobility patterns between cities or events is based on the assumption that central properties of mobility are common to the source and target domain. In this section, this assumption is expressed in two ways. The first version states that mobility during disasters can be modeled as a combination of regular mobility patterns. This version is identical to the assumption for the situation-aware approaches presented in Section 5.1.4 and applies particularly to the work of Rahman & Hasan [52]. As previously discussed in Section 5.1.4, this assumption may conflict with the irregularity of mobility during disasters. The second version concerns the mobility transfer across cities (i.e., as in [64]) and expresses a general resemblance of

Table 8
Comparison of different approaches using transfer learning. These approaches capture patterns in a data-rich source domain, enabling predictions in a data-poor target domain. As no other table reports [64], this table presents the complete set of columns. For the sake of completeness, the corresponding information about [52] is included.

Authors	Disaster	Task	Setting	Transfer Strategy	Model	Data	Evaluation	Result	Data/Code available?
Rahman & Hasan [52]	Hurricane	Edge flow prediction in transportation network	Transfer from regular situation to evacuation situation	Train gating function to distill relevant features of pre-trained model	LSTM, GCN	Traffic detector data	RMSE, MAE, R^2	RMSE: 399.69, MAE: 268.03, R^2: 0.943	No/No
Fan et al. [64]	Earthquake	Trajectory generation in a new city	Transfer from one city to another	Region to region correspondence	HMM, Gibbs sampling	GPS data	Population variation covariance, CityEMD [65], average speed	n.a.	No/No

mobility between different cities. More specifically, transferring mobility patterns observed in one city to another postulates that general mobility patterns are common in different cities. In addition, simulating mobility events (e.g., disasters) in the target domain that were initially observed in the source domain is based on the assumption that there are common changes of mobility in such events for different cities. A limiting factor for such a strategy may lie in the uniqueness of mobility patterns in different cities and their changes in disasters. However, by explicitly learning a correspondence between the regions of the respective cities, Fan et al. [64]) partially addressed this limitation.

5.2. *Unsupervised learning*

Only Song et al. [28] used unsupervised learning as the main component in their proposed framework. As indicated in Table 3, their framework consists of unsupervised and supervised learning. However, the component based on unsupervised learning was deemed the main contribution, resulting in an assignment of the approach to the group of unsupervised learning. A condensed presentation of the work can be found in Table 9.

The proposed model comprises a deep belief network (DBN) [66] to compute deep representations of mobility data. DBNs are based on stacked Restricted Boltzmann Machines that are trained layer-wise in an unsupervised way to learn deep features inherent in the data. Song et al. [28] adopted two parallel DBNs, one trained on trajectory data and the other trained on data describing the behavior context of the observed mobility. In the proposed model, these DBNs are combined by regression layers to learn joint representations for mobility and its behavior context. Finally, an output layer is trained in a supervised manner to compute the mobility prediction. This output contains a sequence of mobility behavior, i.e., the context of the next location such as *home*, *shelter*, or *workplace*, and the corresponding locations. The authors conducted experiments with earthquake mobility data and compared the performance with previous works [10,23], which are presented in Section 5.3. They found significantly increased predictive accuracy and concurrence of generated trajectories with their ground truth. Additionally, the results show a decline in performance as the earthquake's intensity increases, demonstrating a correlation between the severity of disasters and the irregularity of mobility.

The presented approach aims to enhance the prediction by learning deep features of disaster mobility. More general mobility features may be learned by decoupling the behavior context from the respective trajectories, representing the humans' motivation and decisions. Consequently, this approach is a direct attempt to improve the *generalizability* of such predictive models.

5.2.1.1. Generalizability. Learning deep *behavior context features* may distill general behavior during a disaster. Due to their disentanglement from specific locations, these features may encode location-invariant information that is common to different disasters and locations. However, the generalizability of such features can be assumed to be limited. This is because the extent of mobility change may vary considerably depending on the severity of the disaster. Therefore, learned behavior context features are presumably only common to disasters of similar severity and extent. Furthermore, the learned *trajectory features* are strictly linked to a specific location. Consequently, their expressiveness is limited to this specific location and the specific spatial extent of the considered scenario. Nevertheless, the disentanglement of trajectories and respective behavior contexts is a promising approach as it may enable the capture of general mobility information.

5.3. *Reinforcement learning*

In reinforcement learning (RL), an agent learns how to behave in its environment by receiving feedback on its performed actions in the form of reward values [67]. The agent optimizes its actions by maximizing the rewards using a trial-and-error approach. To accomplish this task, the reward function, a key component that evaluates the actions of the agent, is required in RL. Inferring such a reward function from observed data is the task in inverse reinforcement learning (IRL). The combination of IRL and RL provides a framework for simulating mobility in an agent-based manner. The selected publications that aim to simulate human mobility during disasters following an RL or IRL approach are summarized in Table 10.

Fan et al. [68] proposed a framework to simulate the potential impact of a flood on mobility based on trajectory data. Their framework consists of reward function inference using IRL, reward shaping to reflect the flood situation, and trajectory generation using deep RL. First, the authors adopt a k-nearest neighbor ($k-$NN) regression to generate OD pairs for trajectories. For a given location, k-NN regression identifies the k most similar origins of observed trajectories for estimating the location's corresponding destination as the average of the k origins' destinations. For each pair, maximum entropy inverse reinforcement learning [69] is adopted to estimate a reward function from extensive trajectory data. This reward function evaluates each possible step on a city grid. Subsequently, the rewards are adjusted to ensure robust learning. Additionally, a flood scenario is incorporated by imposing penalties on rewards for severely affected areas. Finally, Fan et al. [68] employed a deep RL approach, using the prioritized experience replay algorithm [70] for computing the behavioral policy. Following this policy, possible trajectories of vehicles can be predicted for each OD pair under the influence of a flood. Fan et al. [68] tested their method for a flood scenario in the Houston metropolitan area. They found that the differences between predicted and actual locations were minor compared to the length of trips. More than half of the predicted destinations differed by less than three miles from the ground truth. Moreover, the length of the predicted trips showed high accuracy, exhibiting a mean percentage error of below 5 % for travel distances.

Song et al. [23] combined IRL with a Bayesian approach to predict human trajectories after an earthquake. Their model is based on a mobility graph constructed using trajectory data and collaborative learning [71]. Trajectories on this graph are modeled as Markov Decision Processes, which take the nodes of the constructed mobility graph as states. To represent the decisions of changing locations as observed in the data, a cost function for the trajectories is computed. For choosing the transitions between states, represented as edges between nodes, Song et al. [23] proposed maximum entropy inverse reinforcement

Table 9
Summary of the covered approach for modeling mobility in disasters using unsupervised learning. Here, unsupervised learning was adopted to learn deep representations of mobility features.

Authors	Disaster	Task	Strategy	Architecture	Data	Evaluation	Result	Data/Code available?
Song et al. [28]	Earthquake	Next behavior and locations prediction	Deep feature representations	DBNs	GPS, earthquake, disaster report, road structure, POI	Prediction accuracy, log-likelihood, expected distance error	Prediction acc. (behavior/ mobility) small earthquakes: 92.32 %/81.58 %, large earthquakes: 83.19 %/75.27 %	No/No

Table 10
Comparison of identified reinforcement learning approaches. The approaches consist of two stages: In the first stage, a reward a cost function is inferred. Following this function, trajectories are generated in the second step.

Authors	Disaster	Task	Strategy	Data	Evaluation	Result	Data/Code available?
Fan et al. [68]	Flood	Trajectory generation for flood scenario	IRL & deep RL	GPS, flood information	RMSE, MPE	Travel distance MPE: 4.29 %, RMSE: <0.1 for 80 % of trajectories	No/No
Song et al. [23]	Earthquake	Next location prediction	IRL & Bayesian inference	GPS	Predictive accuracy, log-likelihood, expected distance error, Jaccard similarity coefficient	Prediction accuracy: 50 % - 60 %, log-likelihood: −4, distance error: 0.04	No/No
Song et al. [10]	Earthquake	Next behavior and location prediction	Baum-Welch-Algorithm, Bayesian inference & IRL	GPS, earthquake, disaster report, road structure, POI	Predictive accuracy, log-likelihood, expected distance error, matching of trajectories [69], 90 % matching of trajectories [69], log-probability [69]	Behavior prediction accuracy: 60 % - 70 %, log-likelihood: −3, distance error: 0.03 Mobility prediction matching: 82.75 %, 90 % matching: 61.69 %, log-prob: −6.17	No/No

learning [69] to infer a cost parameter for each possible transition. This parameter enables the generation of trajectories by estimating the final destination of a partially observed trajectory with Bayesian inference. Here, a region-specific prior, deduced from the mobility graph, and the likelihood of the partially observed trajectory, calculated using the cost parameter, are combined to give the probability of the respective destination. The authors tested their approach with trajectories in the Greater Tokyo Area during and after the 2011 Tōhoku earthquake. They found that their model achieved an accuracy of 50 % to 60 % for next locations predictions. For the aggregated population flow, they found a notably higher accuracy, ranging from 86 % to 92 %. Moreover, in an earlier version of this work, Song et al. [24] divided the training process into different stages of the disaster, resulting in a significant increase in performance compared to training jointly over the entire disaster period.

In [10], Song et al. proposed to extend the previous work by modeling mobility as a Hidden Markov Model (HMM). Here, the observable states represent the locations of a trajectory, and the hidden states represent the underlying behavior, such as staying at home or going to work. For inferring the required probability distributions for the HMM, the authors adopted the Baum-Welch-Algorithm [72], yielding initial hidden state probabilities, hidden state transition probabilities, and output probabilities of observable states. Given a partially observed trajectory, the next hidden state is then estimated using Bayesian inference and a particle filter approach [73], enabling the prediction of a person's state in the subsequent step. The proceeding sequence of cells that a person traverses to reach this state is created using maximum entropy inverse reinforcement learning [69] on a mobility graph, as previously described for [23]. Finally, the approach described in [74] is used to obtain the actual movement of a person as the most likely route from a given origin to the desired destination. Additionally, the authors presented an adaption of their model, aiming to increase its ability to generalize across different disasters. It is based on an HMM, where the hidden states represent the context of a location (e.g., home or working location), and the observable states represent different disaster information, as well as travel time and distance of observed trips. The methodology for generating trajectories using the new HMM generally remains as described above, but it incorporates the possibility of including different modes of transportation for the actual mobility generation. This new model no longer depends on specific locations but on underlying activities and general disaster properties. As such, Song et al. [10] showed that it exhibits increased performance for predicting mobility in the Greater Tokyo Area after an earthquake, compared to both the former model and the model proposed in [23]. The authors found an accuracy exceeding 60 % for predicting the next locations 20 days after the disaster.

While following the same goal of simulating mobility in disasters, the approaches presented above differ in methodology. On the one hand, Fan et al. [68] presented a pipeline combining IRL and RL. On the other, Song et al. [23] and Song et al. [10] leveraged IRL to infer the cost function but predicted the mobility following a Bayesian approach or by choosing the most likely route, respectively. As such, the methods in [10,23] do not rely on any DL strategy or classical reinforcement learning to generate a sequence of states. Contrarily, to simulate trajectories, Fan et al. [68] adopted a deep Q-network to train the agent's policy in a large-scale environment. Nevertheless, several aspects of these publications are similar, as will be discussed in the following paragraphs.

5.3.1.1. Complexity of scenarios. The large-scale environments considered in the presented works represent a challenge from a computational perspective. As IRL is usually an iterative process that exhibits disproportionate growth of computational expense [75], inferring a reward function for mobility in a large-scale transportation system is expensive. This expense may even be more significant when each possible origin and destination pair is considered individually.

5.3.1.2. Uniqueness of disasters. As reinforcement learning bears the opportunity to consider actions on an agent level, it enables taking human interdependencies into account, e.g., the tendency to move in crowds during disasters [10]. Song et al. [10] and Song et al. [23] partially incorporated such interdependencies as they included route popularity in their mobility graph construction. With this exception, the selected publications neglect such direct mutual dependencies to a vast extent, potentially missing a promising opportunity.

5.3.1.3. Generalizability. Reward functions inferred from expert demonstrations are better transferable to other environments than computed policies, which often are unstable to small changes in the environment [75]. In the case of individual mobility, the reward function aims to account for human decision-making and may, therefore, generalize across similar situations. Previous research on trajectory simulation leveraged this property, exploring the possibility of transferring a reward function between cities [76]. Fan et al. [68] considered this idea differently for approaching 'what-if' analyses by manipulating the reward function according to a specific disaster. While providing an initial estimation of possible mobility during disasters, it is open to discussion of how realistic such simulations are. It is important to note that the realism of mobility simulations may decrease as the severity of disasters increases. In such cases, mobility decisions are influenced by a more complex set of factors than just road network availability or regional flood depth. Song et al. [10] addressed this limitation as they aimed to generalize across different events and locations by focusing on general, location-independent mobility features. However, this approach might conflict with the uniqueness of different disasters and, therefore, depend on the heterogeneity of disasters included in the training data.

6. Discussion

Our review identified and analyzed a heterogeneous body of ML approaches for human mobility prediction in disasters. Despite its relevance and the profound potential impact of these approaches on revolutionizing disaster management, the field of study comprises only a limited amount of work to date. Generally, the identified approaches exhibit a rich diversity, from traditional ML models to sophisticated DL methods. However, DL approaches dominate the field, primarily leveraging common DL architectures and mechanisms such as GCNs, LSTMs, and attention, reflecting the current trend in the field.

Following the employed methods, the addressed tasks and scenarios also vary in their extent and complexity, as elaborated in Section 5. This variety is because the respective models have specific capabilities or limitations in capturing different mobility phenomena. For instance, the RNNs presented in Section 5.1.2 are predestined to predict sequential processes such as the locations of a trajectory. In contrast, the spatio-temporal networks elaborated in Section 5.1.3 often outperform RNNs in capturing spatially distributed phenomena, such as crowd flows in a transportation system. Moreover, certain approaches excel in directly addressing the specific challenges posed by disasters The subsequent discussion will address these challenges, highlighting the identified strategies to address them, and noting the remaining limitations.

6.1. *Challenges & limitations*

We discuss the most pressing challenges and limitations in the field of question based on the discussions that completed the subsections of Section 5. The *complexity of scenarios*, as discussed in Section 5, does not center on disasters but on mobility prediction in general. Therefore, this aspect is not included in this section. The remaining challenges discussed below apply generally to each of the approaches reviewed.

6.1.1. *Data-induced constraints*

One significant challenge pertains to the availability of data. While datasets on mobility are publicly available in large numbers [77], public datasets that include mobility during large-scale disasters are rare. Only a fraction of the presented literature used publicly available data, as most datasets were provided by a third-party company (Fig. 3). Moreover, the *general rarity of large-scale disasters* limits the amount of existing training data in general. This limitation contrasts the need for *large amounts of training data* with the increasing complexity of DL models.

Strategies exist to leverage mobility data on regular situations as a partial substitute for mobility data on disasters. For instance, such strategies are given by the *situation-aware approaches* (Section 5.1.4) or by adopting *transfer learning* (Section 5.1.5). However, these strategies rely on assumptions that may be infeasible for certain situations.

Furthermore, the type of data available imposes constraints on the applicability of different models. It is not possible to employ approaches that focus on individual-level mobility when only aggregated mobility data is available. For example, aggregated flow data hinders the training of an RL model, which necessitates individual-level trajectories. In general, the further aggregation of data results in the further abstraction of the mobility represented. Data on the absolute volumes of location visitations or in- and outflows of POIs, for instance, exhibit only greatly reduced spatial dependencies and therefore limit the possibility of training a model for predicting complex spatio-temporal mobility patterns.

6.1.2. *Generalizability*

In order to meaningfully impact disaster management, developed models must be especially capable of predicting mobility in unprecedented future events. Therefore, these models need to *generalize to other locations and disasters*. However, mobility in different locations and disasters will likely exhibit different extents. Consequently, applying corresponding models to situations not covered by the training data may be unsuitable.

Several approaches have been discussed in Section 5, which distinctly address this challenge. The proposed approaches range from models that dynamically adapt to short-term observations of mobility, as those presented in Section 5.1.4, over TL strategies that transfer mobility patterns from one city to another, as discussed in Section 5.1.5, to an RL method that manipulates an inferred reward function to represent a disaster, as seen in Section 5.3.

Although these approaches are promising contributions, significant limitations exist to their generalizability. On the one hand, the situation-aware approaches are presumably *restricted to known situations*. Therefore, they may encounter difficulties predicting mobility that deviates significantly from the training data. Moreover, given their reliability on the short-term situation perceptions, these approaches are *limited to now-casting*, or forecasting with a *short time horizon*. On the other hand, transferring disaster mobility patterns between cities *restricts to the specific disaster and its severity*. Furthermore, a manipulated reward function may be *highly sensitive* to the unknown impact and spatial extent of a disaster, potentially resulting in a lack of realism and expressiveness in the corresponding simulation. Therefore, the proposed approaches are expected to be insufficient to predict or simulate the mobility in a city of concern with an unprecedented disaster.

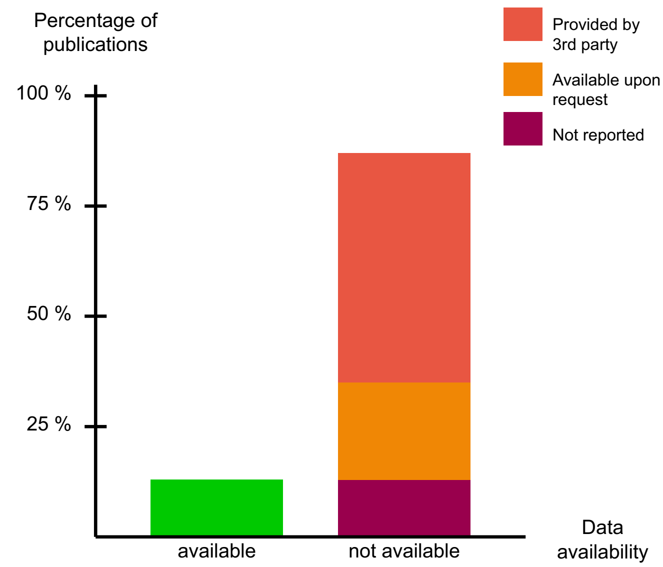

Fig. 3. Percentages of availability of data. The majority of the used datasets were obtained from 3rd parties. While the authors of 13 % of the contributions disclosed the data they used, the authors of the remaining 87 % did not provide their datasets. Of these datasets, 21.7 % are available upon request. An additional 13 % of the publications did not report the data source.

6.1.3. *Uniqueness of disasters*

Disasters usually occur as *short-term events* of limited duration during which the mobility may be *disrupted significantly*. Therefore, predicting mobility during disasters involves predicting mobility patterns that exhibit a high degree of irregularity compared to regular situations. However, when investigating a potential disaster that has not happened before in a location of concern, one can only resort to data on similar disasters, potentially from another location, or to mobility data on regular situations. By doing so, one implicitly assumes that the mobility patterns of past disasters or even regular situations resemble those expected in the disaster of concern. In other words, the *mobility patterns of different locations or events are assumed to be similar*. This assumption introduces an *inductive bias*, which is likely to limit the expressiveness and the degree of realism of the corresponding model's output.

Instead, the impact of a disaster on mobility may be influenced by a large number of different factors such as the disaster's severity, its spatial extent, the people's preparedness, or location-specific factors. As these factors may vary between different disasters, their effect on human mobility may differ greatly, even when the disasters appear to be similar (e.g., two floods: one with official warning in advance and one without). Therefore, this uniqueness of disasters and the disasters' unique effects on mobility pose a limitation on predictive models for disasters.

As a concluding remark, the presented challenges and limitations are inseparably intertwined. For instance, the uniqueness of disasters represents a significant hurdle for the generalizability of models, as mobility in dissimilar events or disasters may not be predictable by a single model. Furthermore, the scarcity of data on disaster mobility complicates the development of a model that generalizes well across different locations and disasters, as the heterogeneity of scenarios observable in the data is limited. Accordingly, capturing the unique mobility patterns in different disasters could be improved if more data on disaster mobility were available. Therefore, the availability of disaster mobility data from a variety of locations and disasters could significantly contribute to overcoming the aforementioned challenges to a considerable extent.

6.2. Future work

Following these challenges and limitations, we present different areas for future work. The challenges of data scarcity and generalizability imply the need for further research on capturing complex patterns from limited data and successfully transferring patterns of mobility in disasters between different cities and situations. Moreover, increasing a model's degree of realism by further incorporating social dynamics, such as the interaction between people, is an understudied field that demands additional investigation.

6.2.1. Increasing the generalizability

As discussed in the previous section, developing a model that generalizes across locations and disasters constitutes a significant challenge. By focusing on techniques to increase these two aspects of generalizability, future work has the potential to significantly improve the simulation of unprecedented disasters, thereby profoundly impacting the field of disaster modeling. Potential approaches may be motivated or adopted from an increasing body of literature on models that aim to recreate mobility patterns in unseen locations. Corresponding approaches may leverage spatio-temporal adaptions from a source to a target city (e.g., [78–80]). Another promising approach may be to adopt methods from graph matching [81] to specifically translate observed mobility patterns during a disaster across cities. Alternatively, further investigating how to exploit reinforcement learning to simulate mobility during a disaster and location of concern based on observed mobility data from a similar but different context (e.g., from a different city, or from the same city with a disaster of different spatial extent) is an interesting research direction (e.g., [76]).

6.2.2. Incorporating mutual influences

The presented approaches in Section 5.3 which simulate individual trajectories of the population consider the respective route choices mostly isolated, neglecting collective behavior in groups of people. It is crucial to investigate how to include the possibility of dependence between agents' actions in simulations, as this may significantly enhance the realism of the model by introducing emergent and collective behavior observed in disasters. Multi-agent reinforcement learning, which enables an agent's action to be dependent on other agents' actions, might serve this purpose. Investigating how to model the simulation of disaster mobility as such a multi-agent reinforcement learning problem is an exciting direction for future work.

6.2.3. Foundation models as few-shot learners

Recently, foundation models such as pre-trained large language models (LLMs) have been proposed increasingly to learn general patterns from sparse data, including time series data [82]. Their ability to capture complex patterns from a limited number of samples makes them suitable for learning mobility patterns from sparse data. First publications have leveraged LLMs to learn mobility patterns in rare events where data is scarce, such as sports venues [83]. Exploring the potential of LLMs for a few-shot prediction of mobility during disasters is an exciting field that can be addressed by future work.

6.2.4. Application in disaster management

In order to have a meaningful impact, developed models must provide an actual benefit for disaster management. Generally, developed models may serve as a basis for *what-if simulations*, up to their generalizability. Future work may analyze the potential of such disaster mobility models to improve evacuation, resource allocation and rescue mission planning. Furthermore, investigating how these models may be included in holistic disaster management systems to introduce human behavior in multi-infrastructure models, such as digital twins [84,85], is an interesting field for future research.

These directions present a non-exhaustive list as motivation for further research. In the future, we plan to investigate how to translate mobility patterns between cities with DL approaches to address the challenge of *generalizability*.

7. Conclusion

This study presented a comprehensive review and analysis of ML methods and approaches for predicting human mobility in disasters. Our review included a systematic search for all relevant literature and a deep analysis and comprehensive presentation of the identified contributions. We grouped the proposed approaches hierarchically and compared them qualitatively. While a quantitative comparison would enable the approach with the best performance, the heterogeneity of the scenarios and tasks prohibited such an endeavor. Moreover, we thoroughly discussed all selected publications' limitations and open challenges. In contrast to regular circumstances, disasters introduce distinct particularities to mobility prediction. In this context, we identified the scarcity of disaster mobility data, the uniqueness of disasters, and the generalizability of models as the most significant challenges and limitations. While a significant proportion of publications adopted common network architectures to predict disaster mobility, several strategies have been proposed to specifically address these challenges. Notable strategies that emerged in our review include employing transfer learning, equipping networks with situation-awareness (the ability to perceive and adapt to the current situation), and distilling location-invariant behavior features. We proposed a set of potential research directions based on the identified open challenges, limitations, and research gaps in the existing literature. Future work should focus on enhancing the generalizability of models and developing approaches to learn complex mobility patterns from limited data. These directions have the potential to significantly advance the field of disaster mobility prediction.

Although the field of research in question has the potential to have a significant impact, it has yet to receive the attention it deserves, as evidenced by the limited number of relevant publications. Moreover, the significant diversity of approaches indicates that the potential for further research is yet to be fully exploited. Efforts to do so are inevitable to produce trustworthy models which can be used in practice. In conclusion, our review represents a comprehensive synthesis and foundation of knowledge that is of great value for both new and established researchers, enriching the existing literature and guiding future research directions. As such, it contributes to facilitating further research and the development of new approaches.

CRediT authorship contribution statement

Jonas Gunkel: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Max Mühlhäuser:** Supervision. **Andrea Tundis:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research work was funded from the German Aerospace Center (DLR) and it was conducted in the context of the urbanModel project in cooperation with the LOEWE center emergenCITY (Program of Hesse State Ministry for Higher Education, Research and the Arts)

Appendix A. Appendices

A.1. Query strings

A.1.1. IEEE xplore

((“All Metadata”: Machine AND “All Metadata”: Learning) OR (“All Metadata”: Deep AND “All Metadata”: Learning) OR (“All Metadata”: Artificial AND “All Metadata”: Intelligence) OR (“All Metadata”: Neural AND “All Metadata”: Network) OR (“All Metadata”: Big AND “All Metadata”: Data) OR (“All Metadata”: Data AND “All Metadata”: Mining) OR (“All Metadata”: Urban AND “All Metadata”: Computing)) AND (“All Metadata”: Disaster OR “All Metadata”: Crisis OR “All Metadata”: Hazard OR “All Metadata”: Catastrophe OR “All Metadata”: Emergency OR “All Metadata”: Evacuation OR (“All Metadata”: Extreme AND “All Metadata”: Event) OR (“All Metadata”: Extreme AND “All Metadata”: Situation)) AND (“All Metadata”: Mobility OR (“All Metadata”: Crowd AND “All Metadata”: Flow) OR “All Metadata”: Evacuation OR “All Metadata”: Relocation OR “All Metadata”: Traffic OR “All Metadata”: Trajectory OR (“All Metadata”: Urban AND “All Metadata”: Dynamics))

A.2. Web of science

AB = (((Machine AND Learning) OR (Deep AND Learning) OR (Artificial AND Intelligence) OR (Neural AND Network) OR (Big AND Data) OR (Data AND Mining) OR (Urban AND Computing)) AND (Disaster OR Crisis OR Hazard OR Catastrophe OR Emergency OR Evacuation OR (Extreme AND (Event OR Situation))) AND (Mobility OR (Crowd AND Flow) OR Evacuation OR Relocation OR Traffic OR Trajectory OR (Urban AND Dynamics))) OR TI = ((((Machine AND Learning) OR (Deep AND Learning) OR (Artificial AND Intelligence) OR (Neural AND Network) OR (Big AND Data) OR (Data AND Mining) OR (Urban AND Computing)) AND (Disaster OR Crisis OR Hazard OR Catastrophe OR Emergency OR Evacuation OR (Extreme AND (Event OR Situation))) AND (Mobility OR (Crowd AND Flow) OR Evacuation OR Relocation OR Traffic OR Trajectory OR (Urban AND Dynamics)))) OR AK = ((((Machine AND Learning) OR (Deep AND Learning) OR (Artificial AND Intelligence) OR (Neural AND Network) OR (Big AND Data) OR (Data AND Mining) OR (Urban AND Computing)) AND (Disaster OR Crisis OR Hazard OR Catastrophe OR Emergency OR Evacuation OR (Extreme AND (Event OR Situation))) AND (Mobility OR (Crowd AND Flow) OR Evacuation OR Relocation OR Traffic OR Trajectory OR (Urban AND Dynamics)))).

A.3. Scopus

TITLE-ABS-KEY(((Machine AND Learning) OR (Deep AND Learning) OR (Artificial AND Intelligence) OR (Neural AND Network) OR (Big AND Data) OR (Data AND Mining) OR (Urban AND Computing)) AND (Disaster OR Crisis OR Hazard OR Catastrophe OR Emergency OR Evacuation OR (Extreme AND (Event OR Situation))) AND (Mobility OR (Crowd AND Flow) OR Evacuation OR Relocation OR Traffic OR Trajectory OR (Urban AND Dynamics))).

A.4. ACM digital library

Abstract:(((machine AND learning) OR (deep AND learning) OR (artificial AND intelligence) OR (neural AND network) OR (big AND data) OR (data AND mining) OR (urban AND computing)) AND (disaster OR crisis OR Hazard OR catastrophe OR emergency OR evacuation OR (extreme AND (event OR situation))) AND (mobility OR (crowd AND flow) OR evacuation OR relocation OR traffic OR trajectory OR (urban AND dynamics))) OR title:((((machine AND learning) OR (deep AND learning) OR (artificial AND intelligence) OR (neural AND network) OR (big AND data) OR (data AND mining) OR (urban AND computing)) AND (disaster OR crisis OR Hazard OR catastrophe OR emergency OR evacuation OR (extreme AND (event OR situation))) AND (mobility OR (crowd AND flow) OR evacuation OR relocation OR traffic OR trajectory OR (urban AND dynamics)))) OR keyword:((((machine AND learning) OR (deep AND learning) OR (artificial AND intelligence) OR (neural AND network) OR (big AND data) OR (data AND mining) OR (urban AND computing)) AND (disaster OR crisis OR Hazard OR catastrophe OR emergency OR evacuation OR (extreme AND (event OR situation))) AND (mobility OR (crowd AND flow) OR evacuation OR relocation OR traffic OR trajectory OR (urban AND dynamics)))).

Data availability

No data was used for the research described in the article.

References

[1] Our World in Data. Global Damage Costs From Natural Disasters. https://ourworldindata.org/grapher/damage-costs-from-natural-disasters; 2024. accessed 7 February 2024.

[2] Our World in Data. Number of Deaths From Disasters. https://ourworldindata.org/explorers/natural-disasters?time=1978..latest&facet=none&hideControls=true&Disaster+Type=All+disasters&Impact=Deaths&Timespan=Annual&Per+capita=false&country=~OWID_WRL; 2024. accessed 7 February 2024.

[3] Pitilakis K, Argyroudis S, Kakderi K, Selva J. Systemic vulnerability and risk assessment of transportation systems under natural hazards towards more resilient and robust infrastructures. Transp Res Proc 2016;14:1335–44. https://doi.org/10.1016/j.trpro.2016.05.206.

[4] Provitolo D, Dubos-Paillard E, Muller J-P. Emergent human behaviour during a disaster: Thematic versus complex systems approaches. In: Emergent Properties in Natural and Artificial Complex Systems EPNACS 2011 Within ECCS'11, European Conference on Complex System, Vienna, Austria; 2011. https://shs.hal.science/halshs-02118076.

[5] Zantalis F, Koulouras G, Karabetsos S, Kandris D. A review of machine learning and iot in smart transportation. Fut Intern 2019;11(4):94. https://doi.org/10.3390/fi11040094.

[6] Luca M, Barlacchi G, Lepri B, Pappalardo L. A survey on deep learning for human mobility. ACM Comput Surv 2021;55(1):1–44. https://doi.org/10.1145/3485125.

[7] R. Jiang, Z. Wang, Y. Tao, C. Yang, X. Song, R. Shibasaki, S.-C. Chen, M.-L. Shyu, Learning social meta-knowledge for nowcasting human mobility in disaster, in: Proceedings of the ACM Web Conference 2023, WWW'23, Association for Computing Machinery, New York, NY, USA, 2023, 2655–2665. doi:https://doi.org/10.1145/3543507.3583991.

[8] Wang Z, Jiang R, Xue H, Salim FD, Song X, Shibasaki R. Event-aware multimodal mobility nowcasting. Proc AAAI Conf Artif Intell 2022;36(4):4228–36. https://doi.org/10.1609/aaai.v36i4.20342.

[9] Lu X, Bengtsson L, Holme P. Predictability of population displacement after the 2010 Haiti earthquake. Proc Natl Acad Sci 2012;109(29):11576–81. https://doi.org/10.1073/pnas.1203882109.

[10] Song X, Zhang Q, Sekimoto Y, Shibasaki R, Yuan NJ, Xie X. Prediction and simulation of human mobility following natural disasters. ACM Trans Intell Syst Technol 2016;8(2):1–23. https://doi.org/10.1145/2970819.

[11] Jiang W, Luo J. Graph neural network for traffic forecasting: a survey. Expert Syst Appl 2022;207:117921. https://doi.org/10.1016/j.eswa.2022.117921.
[12] Zhu L, Chen C, Wang H, Yu FR, Tang T. Machine learning in urban rail transit systems: a survey. IEEE Trans Intell Transp Syst 2023:1–26. https://doi.org/10.1109/tits.2023.3319135.
[13] Xie P, Li T, Liu J, Du S, Yang X, Zhang J. Urban flow prediction from spatiotemporal data using machine learning: a survey. Inform Fus 2020;59:1–12. https://doi.org/10.1016/j.inffus.2020.01.002.
[14] Veres M, Moussa M. Deep learning for intelligent transportation systems: a survey of emerging trends. IEEE Trans Intell Transp Syst 2020;21(8):3152–68. https://doi.org/10.1109/tits.2019.2929020.
[15] Ahmed DB, Diaz EM. Survey of machine learning methods applied to urban mobility. IEEE Access 2022;10:30349–66. https://doi.org/10.1109/access.2022.3159668.
[16] Kyrkou C, Kolios P, Theocharides T, Polycarpou M. Machine learning for emergency management: a survey and future outlook. Proc IEEE 2023;111(1):19–41. https://doi.org/10.1109/jproc.2022.3223186.
[17] Haraguchi M, Nishino A, Kodaka A, Allaire M, Lall U, Kuei-Hsien L, et al. Human mobility data and analysis for urban resilience: a systematic review. Environ Plann B Urban Analyt City Sci 2022;49(5):1507–35. https://doi.org/10.1177/23998083221075634.
[18] Yabe T, Jones NK, Rao PSC, Gonzalez MC, Ukkusuri SV. Mobile phone location data for disasters: a review from natural hazards and epidemics. Comput Environ Urban Syst 2022;94:101777. https://doi.org/10.1016/j.compenvurbsys.2022.101777.
[19] González MC, Hidalgo CA, Barabási A-L. Understanding individual human mobility patterns. Nature 2008;453(7196):779–82. https://doi.org/10.1038/nature06958.
[20] Mentges A, Halekotte L, Schneider M, Demmer T, Lichte D. A resilience glossary shaped by context: reviewing resilience-related terms for critical infrastructures. Int J Disast Risk Reduct 2023;96:103893. https://doi.org/10.1016/j.ijdrr.2023.103893.
[21] Snyder H. Literature review as a research methodology: an overview and guidelines. J Bus Res 2019;104:333–9. https://doi.org/10.1016/j.jbusres.2019.07.039.
[22] Wohlin C, Kalinowski M, Romero Felizardo K, Mendes E. Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. Inf Softw Technol 2022;147:106908. https://doi.org/10.1016/j.infsof.2022.106908.
[23] Song X, Zhang Q, Sekimoto Y, Shibasaki R. Intelligent system for urban emergency management during large-scale disaster. Proc AAAI Conf Artif Intell 2014;28(1). https://doi.org/10.1609/aaai.v28i1.8758.
[24] Song X, Zhang Q, Sekimoto Y, Horanont T, Ueyama S, Shibasaki R. Modeling and probabilistic reasoning of population evacuation during large-scale disaster. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'13. New York, NY, USA: Association for Computing Machinery; 2013. p. 1231–9. https://doi.org/10.1145/2487575.2488189.
[25] Song X, Zhang Q, Sekimoto Y, Horanont T, Ueyama S, Shibasaki R. Intelligent system for human behavior analysis and reasoning following large-scale disasters. IEEE Intell Syst 2013;28(4):35–42. https://doi.org/10.1109/mis.2013.35.
[26] Song X, Zhang Q, Sekimoto Y, Shibasaki R. Prediction of human emergency behavior and their mobility following large-scale disaster. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'14. ACM; 2014. https://doi.org/10.1145/2623330.2623628.
[27] Song X, Zhang Q, Sekimoto Y, Shibasaki R, Yuan NJ, Xie X. A simulator of human emergency mobility following disasters: knowledge transfer from big disaster data. Proc AAAI Conf Artif Intell 2015;29(1). https://doi.org/10.1609/aaai.v29i1.9237.
[28] Song X, Shibasaki R, Yuan NJ, Xie X, Li T, Adachi R. Deepmob: learning deep knowledge of human emergency behavior and mobility from big and heterogeneous data. ACM Trans Inf Syst 2017;35(4). https://doi.org/10.1145/3057280.
[29] Chikaraishi M, Garg P, Varghese V, Yoshizoe K, Urata J, Shiomi Y, et al. On the possibility of short-term traffic prediction during disaster with machine learning approaches: an exploratory analysis. Transp Policy 2020;98:91–104. https://doi.org/10.1016/j.tranpol.2020.05.023.
[30] Anyidoho PK, Ju X, Davidson RA, Nozick LK. A machine learning approach for predicting hurricane evacuee destination location using smartphone location data, computational urban. Science 2023;3(1). https://doi.org/10.1007/s43762-023-00102-0.
[31] Khaefi MR, Prahara PJ, Rheza M, Alkarisya D, Hodge G. Predicting evacuation destinations due to a natural hazard using mobile network data. In: 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS); 2018. p. 1–6. https://doi.org/10.1109/ICICOS.2018.8621662.
[32] Yabe T, Tsubouchi K, Sudo A, Sekimoto Y. Predicting irregular individual movement following frequent mid-level disasters using location data from smartphones. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL'16. ACM; 2016. https://doi.org/10.1145/2996913.2996929.
[33] Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: Lstm cells and network architectures. Neural Comput 2019;31(7):1235–70. https://doi.org/10.1162/neco_a_01199.
[34] Belhadi A, Djenouri Y, Djenouri D, Lin JC-W. A recurrent neural network for urban long-term traffic flow forecasting. Appl Intell 2020;50(10):3252–65. https://doi.org/10.1007/s10489-020-01716-1.
[35] J. Feng, Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, D. Jin, Deepmove: Predicting human mobility with attentional recurrent networks, in: Proceedings of the 2018 World Wide Web Conference, WWW'18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018, 1459–1468. doi:https://doi.org/10.1145/3178876.3186058.
[36] Yang D, Fankhauser B, Rosso P, Cudre-Mauroux P. Location prediction over sparse user mobility traces using rnns: Flashback in hidden states!. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, International Joint Conferences on Artificial Intelligence Organization; 2020. p. 2184–90. https://doi.org/10.24963/ijcai.2020/302.
[37] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
[38] Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2014. https://doi.org/10.3115/v1/d14-1179. Association for Computational Linguistics.
[39] Kong D, Wu F. Hst-lstm: A hierarchical spatial-temporal long-short term memory network for location prediction. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization; 2018. p. 2341–7. https://doi.org/10.24963/ijcai.2018/324.
[40] Wang S, Shao C, Zhang J, Zheng Y, Meng M. Traffic flow prediction using bi-directional gated recurrent unit method. Urban Inform 2022;1(1). https://doi.org/10.1007/s44212-022-00015-z.
[41] Jiang R, Song X, Fan Z, Xia T, Chen Q, Miyazawa S, et al. Deepurbanmomentum: an online deep-learning system for short-term urban mobility prediction. Proc AAAI Conf Artif Intell 2018;32(1). https://doi.org/10.1609/aaai.v32i1.11338.
[42] Rahman R, Hasan S. Short-term traffic speed prediction for freeways during hurricane evacuation: A deep learning approach. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC); 2018. p. 1291–6. https://doi.org/10.1109/ITSC.2018.8569443.
[43] Roy KC, Hasan S, Culotta A, Eluru N. Predicting traffic demand during hurricane evacuation using real-time data from transportation systems and social media. Transp Res Part C Emerg Technol 2021;131:103339. https://doi.org/10.1016/j.trc.2021.103339.
[44] Afrin T, Aragon LG, Lin Z, Yodo N. An integrated data-driven predictive resilience framework for disaster evacuation traffic management. Appl Sci 2023;13(11). https://doi.org/10.3390/app13116850.
[45] Mahmud S, Shen H, Zhang Foutz YN, Anton J. Esep: Data-driven emergency and safe evacuation driving path planning during natural catastrophes. In: 2022 IEEE International Conference on Big Data (Big Data); 2022. p. 1641–50. https://doi.org/10.1109/BigData55660.2022.10020410.
[46] Chung J, Kastner K, Dinh L, Goel K, Courville A, Bengio Y. A recurrent latent variable model for sequential data. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, NIPS'15. Cambridge, MA, USA: MIT Press; 2015. p. 2980–8.
[47] Jiang F-Z, Zhong L, Thilakarathna K, Seneviratne A, Takano K, Yamada S, et al. Supercharging crowd dynamics estimation in disasters via spatio-temporal deep neural network. In: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA); 2017. p. 184–92. https://doi.org/10.1109/DSAA.2017.11.
[48] Fan Z, Song X, Xia T, Jiang R, Shibasaki R, Sakuramachi R. Online deep ensemble learning for predicting citywide human mobility. Proc ACM Interact Mob Wearable Ubiquit Technol 2018;2(3). https://doi.org/10.1145/3264915.
[49] Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. In: Proceedings of the 6th International Conference on Learning Representations, ICLR 2018; 2018. OpenReview.net, https://openreview.net/forum?id=SJiHXGWAZ.
[50] Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-k, Woo W-c. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS'15. Cambridge, MA, USA: MIT Press; 2015. p. 802–10.
[51] Jiang R, Song X, Huang D, Song X, Xia T, Cai Z, et al. Deepurbanevent: A system for predicting citywide crowd dynamics at big events. In: Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD'19. New York, NY, USA: Association for Computing Machinery; 2019. p. 2114–22. https://doi.org/10.1145/3292500.3330654.
[52] Rahman R, Hasan S. A deep learning approach for network-wide dynamic traffic prediction during hurricane evacuation. Transp Res Part C Emerg Technol 2023;152:104126. https://doi.org/10.1016/j.trc.2023.104126.
[53] Hao H, Wang Y. Modeling dynamics of community resilience to extreme events with explainable deep learning. Nat Hazards Rev 2023;24(2):04023013. https://doi.org/10.1061/NHREFO.NHENG-1696.
[54] Zhiwen Z, Wang H, Fan Z, Shibasaki R, Song X. Assessing the continuous causal responses of typhoon-related weather on human mobility: An empirical study in Japan. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM'23. New York, NY, USA: Association for Computing Machinery; 2023. p. 3524–33. https://doi.org/10.1145/3583780.3615513.
[55] Shi C, Blei D, Veitch V. Adapting neural networks for the estimation of treatment effects. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc; 2019. https://proceedings.neurips.cc/paper_files/paper/2019/file/8fb5f8be2aa9d6c64a04e3ab9f63feee-Paper.pdf.
[56] Wang H-W, Peng Z-R, Wang D, Meng Y, Wu T, Sun W, et al. Evaluation and prediction of transportation resilience under extreme weather events: a diffusion

graph convolutional approach. Transp Res Part C Emerg Technol 2020;115: 102619. https://doi.org/10.1016/j.trc.2020.102619.

[57] Hospedales TM, Antoniou A, Micaelli P, Storkey AJ. Meta-learning in neural networks: a survey. IEEE Trans Pattern Anal Mach Intell 2021. https://doi.org/10.1109/tpami.2021.3079209.

[58] Santoro A, Bartunov S, Botvinick M, Wierstra D, Lillicrap T. Meta-learning with memory-augmented neural networks. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, ICML'16; 2016. p. 1842–50. http://proceedings.mlr.press/v48/santoro16.html.

[59] Jia X, De Brabandere B, Tuytelaars T, Gool LV. Dynamic filter networks. In: Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, editors. Advances in Neural Information Processing Systems. vol. 29. Curran Associates, Inc; 2016. https://proceedings.neurips.cc/paper_files/paper/2016/file/8bf1211fd4b7b94528899de0a43b9fb3-Paper.pdf.

[60] Yang Y, Lv H, Chen N. A survey on ensemble learning under the era of deep learning. Artif Intell Rev 2022;56(6):5545–89. https://doi.org/10.1007/s10462-022-10283-5.

[61] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proc IEEE 2021;109(1):43–76. https://doi.org/10.1109/jproc.2020.3004555.

[62] Chen Y, Gu J, Zhuang F, Lu X, Sun M. Exploiting hierarchical correlations for cross-city cross-mode traffic flow prediction. In: 2022 IEEE International Conference on Data Mining (ICDM). IEEE; 2022. https://doi.org/10.1109/icdm54844.2022.00103.

[63] Y. Huang, X. Song, S. Zhang, J. J. Yu, Transfer learning in traffic prediction with graph neural networks, in: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), IEEE, 2021. doi:https://doi.org/10.1109/itsc48978.2021.9564890.

[64] Fan Z, Song X, Shibasaki R, Li T, Kaneda H. Citycoupling: bridging intercity human mobility. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp'16. ACM; 2016. p. 718–28. https://doi.org/10.1145/2971648.2971737.

[65] Fan Z, Song X, Shibasaki R, Adachi R. Citymomentum: An online approach for crowd behavior prediction at a citywide level. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp'15. ACM; 2015. https://doi.org/10.1145/2750858.2804277.

[66] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput 2006;18(7):1527–54. https://doi.org/10.1162/neco.2006.18.7.1527.

[67] Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. J Artif Intell Res 1996;4:237–85. https://doi.org/10.1613/jair.301.

[68] Fan C, Jiang X, Mostafavi A. Evaluating crisis perturbations on urban mobility using adaptive reinforcement learning. Sustain Cities Soc 2021;75. https://doi.org/10.1016/j.scs.2021.103367.

[69] Ziebart BD, Maas AL, Bagnell JA, Dey AK. Maximum entropy inverse reinforcement learning. In: Fox D, Gomes CP, editors. Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008. AAAI Press; 2008. p. 1433–8. http://www.aaai.org/Library/AAAI/2008/aaai08-227.php.

[70] Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay. In: Bengio Y, LeCun Y, editors. Proceedings of the 4th International Conference on Learning Representations, ICLR; 2016. http://arxiv.org/abs/1511.05952.

[71] Wei L-Y, Zheng Y, Peng W-C. Constructing popular routes from uncertain trajectories. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'12. ACM; 2012. https://doi.org/10.1145/2339530.2339562.

[72] Baum LE, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. Ann Math Stat 1970; 41(1):164–71. https://doi.org/10.1214/aoms/1177697196.

[73] Doucet A, Godsill SJ, Andrieu C. On sequential Monte Carlo sampling methods for bayesian filtering. Stat Comput 2000;10(3):197–208. https://doi.org/10.1023/A:1008935410038.

[74] R. Simmons, B. Browning, Y. Zhang, V. Sadekar, Learning to predict driver route and destination intent, in: 2006 IEEE Intelligent Transportation Systems Conference, IEEE, 2006. doi:https://doi.org/10.1109/itsc.2006.1706730.

[75] Arora S, Doshi P. A survey of inverse reinforcement learning: challenges, methods and progress. Artif Intell 2021;297:103500. https://doi.org/10.1016/j.artint.2021.103500.

[76] Pang Y, Tsubouchi K, Yabe T, Sekimoto Y. Intercity simulation of human mobility at rare events via reinforcement learning. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, SIGSPATIAL'20. ACM; 2020. https://doi.org/10.1145/3397536.3422244.

[77] Yin X, Wu G, Wei J, Shen Y, Qi H, Yin B. Deep learning on traffic prediction: methods, analysis, and future directions. IEEE Trans Intell Transp Syst 2022;23(6): 4927–43. https://doi.org/10.1109/tits.2021.3054840.

[78] Jin Y, Chen K, Yang Q. Selective cross-city transfer learning for traffic prediction via source city region re-weighting. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD'22. ACM; 2022. https://doi.org/10.1145/3534678.3539250.

[79] Wang S, Miao H, Li J, Cao J. Spatio-temporal knowledge transfer for urban crowd flow prediction via deep attentive adaptation networks. IEEE Trans Intell Transp Syst 2022;23(5):4695–705. https://doi.org/10.1109/tits.2021.3055207.

[80] Fang Z, Wu D, Pan L, Chen L, Gao Y. When transfer learning meets cross-city urban flow prediction: Spatio-temporal adaptation matters. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization; 2022. https://doi.org/10.24963/ijcai.2022/282.

[81] Zhang Z, Lu Y, Zheng W, Lin X. A comprehensive survey and experimental study of subgraph matching: trends, unbiasedness, and Interaction. Proc ACM Manage Data 2024;2(1):1–29. https://doi.org/10.1145/3639315.

[82] Gruver N, Finzi M, Qiu S, Wilson AG. Large language models are zero-shot time series forecasters. In: Oh A, Neumann T, Globerson A, Saenko K, Hardt M, Levine S, editors. Advances in Neural Information Processing Systems. vol. 36. Curran Associates, Inc.; 2023. p. 19622–35.

[83] Liang Y, Liu Y, Wang X, Zhao Z. Exploring Large Language Models for Human Mobility Prediction Under Public Events. 2023. https://doi.org/10.48550/ARXIV.2311.17351.

[84] Gebhard T, Sattler BJ, Gunkel J, Marquard M, Tundis A. Improving the resilience of socio-technical urban critical infrastructures with digital twins: challenges, concepts, and modeling. Sustain Analyt Model 2024:100036. https://doi.org/10.1016/j.samod.2024.100036.

[85] Fan C, Zhang C, Yahja A, Mostafavi A. Disaster City digital twin: a vision for integrating artificial and human intelligence for disaster management. Int J Inf Manag 2021;56:102049. https://doi.org/10.1016/j.ijinfomgt.2019.102049.