

Research Article

Transferability of a Machine Learning-Based Model of Hourly Traffic Volume Estimation—Florida and New Hampshire Case Study

Przemysław Sekuła ^{1,2}, Zachary Vander Laan ¹, Kaveh Farokhi Sadabadi ¹,
Krzysztof Kania ², and Sara Zahedian ¹

¹Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, USA

²Faculty of Informatics and Communication, University of Economics in Katowice, Katowice, Poland

Correspondence should be addressed to Przemysław Sekuła; psekula@umd.edu

Received 31 March 2021; Revised 19 October 2021; Accepted 3 November 2021; Published 26 November 2021

Academic Editor: Chi-Hua Chen

Copyright © 2021 Przemysław Sekuła et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper focuses on the problem of model transferability for machine learning models used to estimate hourly traffic volumes. The presented findings enable not only an increase in the accuracy of existing models but also, simultaneously, reduce the cost of data needed for training the models—making statewide traffic volume estimation more economically feasible. Previous research indicates that machine learning volume estimation models that leverage GPS probe data can provide transportation agencies with accurate estimates of hourly traffic volumes—which are fundamental for both operational and planning purposes—and do so with a higher level of accuracy than the prevailing profiling method. However, this approach requires a large dataset for model calibration (i.e., input and continuous count station data), which involves significant monetary investment and data-processing effort. This paper proposes solutions, which allow the model to be prepared using a much smaller dataset, given that a previously collected dataset, which may be gathered in a different place and time period, exists. Based on a broad selection of experiments, the results indicate that the proposed approach is capable of achieving similar model performance while collecting data for a 5 times shorter time period and utilizing 1/4 of the number of continuous count stations. These findings will help reduce the cost of preparing and maintaining the traffic volume models and render the traffic volume estimations more financially appealing.

1. Introduction

This research investigates the transferability of the artificial neural network (ANN) models applied in hourly traffic volume estimation. The introduced methodology explores the feasibility of training an ANN model to estimate the traffic volume by transferring data from another region when sufficient data are unavailable for the target network. This methodology requires ground truth observations to train the models, but previously trained models can be applied at the entire road network, regardless of the existence of the ground truth data. The estimation of traffic flow parameters is one of the fundamental needs in many of the intelligent transportation system applications as they are

used for a wide variety of purposes from the planning to the design and operational stages of highway networks [1–3]. Two of the most crucial inputs required by transportation agencies in order to calculate statewide performance measures are traffic volumes and speeds. There are many well-known methods and solutions for obtaining network-wide speed data, which are already implemented for commercial use. However, network-wide traffic volumes are much more challenging to obtain and thus remain a key missing dimension for quantifying traffic conditions, transportation system performance assessment, and cost-effective management of mobility projects and programs. The current estimation methods require continuous count station (CCS) data and knowledge regarding aggregate volume estimates

such as annual average daily traffic (AADT), making them both expensive and inaccurate in locations where CCS data are unavailable. Moreover, the accuracy of these models tends to decrease when the objective is to estimate traffic volumes with higher temporal granularity (e.g., for 15-minute intervals). These difficulties are generated by a lack of network-wide available data reflecting the spatiotemporal traffic volume changes. The I-95 Corridor Coalition founded the “Volume and Turning Movement” project in 2015, seeking to provide representative volume and turning moving data products and assess their accuracy and feasibility. During the first phase of the project, it was shown that an ANN-based volume estimation model leveraging probe data as a key input was able to improve hourly volume prediction accuracy on functional road class (FRC) 1 roads by 26% compared to the currently often-used profiling method [4]. In the second phase, this solution was employed statewide in Florida [5], and extended to the principal, major, and minor arterials. While implementing the model at the statewide level helped us to meet the project’s main goals, it required a large calibration dataset to accomplish this. In order to enable this, three months of data from 173 CCSs were collected as the ground truth for training the ANN model to learn the relations between the input variables and actual, observed traffic volumes. However, many states do not have such a robust network of CCSs and thus may not have sufficient calibration data to train a volume estimation model. Nonetheless, some patterns learned from such a large dataset may not be limited to a specific time or location and may be successfully applied elsewhere. For example, the impact of weather or traffic jams on the same types of roads should be similar regardless of the location. A natural way to circumvent the data limitations is to transfer knowledge from areas where enough data are available to areas where the amount of collected data is insufficient to apply to ML techniques. Due to the significant spatiotemporal variation in traffic volumes, previously trained models cannot provide reliable results in a new study area. However, a small dataset from the study area can be combined with large datasets obtained from other regions to train an ANN model with acceptable accuracy. This paper explores whether target geographies with a small calibration network can leverage a larger, previously collected dataset to enhance the volume estimation performance. To illustrate this, we focused on estimating traffic volumes in New Hampshire—a small state with a limited CCS network—by utilizing the previously described Florida dataset, which is four times the size.

The contribution of this article is twofold. We proposed a solution that enables traffic volume estimations even with small datasets, provided that the other available large datasets or previously trained models are leveraged. We also designed, explored, analyzed, and assessed different techniques that allow leveraging available large external datasets or previously trained models to improve traffic volume estimations when only a small dataset for a particular area and time is available. As the cost of the data is the main factor that impedes wide usage of traffic volume estimations, these contributions are of significant practical importance.

The remainder of the paper is organized into five sections. First, the literature review is presented to show how the proposed approach fits into the existing traffic volume estimation research. Afterwards, the data used for the analysis are presented, followed by a description of the model—including hyperparameter selection and details of the training procedure. Subsequently, the experiments and results are discussed and, finally, conclusions are drawn, as several extensions of this work are proposed.

2. Literature Review

The existing literature contains many approaches for estimating traffic volumes [2, 4]. One possible way to categorize these studies is to divide them into two groups of parametric and nonparametric methods [6, 7]. Parametric methods use linear and nonlinear regression or the autoregressive integrated moving average (ARIMA) and their modifications to analyze historical data. The main limitation of parametric methods is that they yield the best results in analyzing stable phenomena with linear relationships between parameters.

As traffic shows features of chaotic systems with nonlinear relationships [8], the researchers’ attention turned more to nonparametric methods. They can approach any nonlinear characteristics of the traffic flow data. This group includes many different methods that are mostly concerned with AADT estimations. Reference [9] used support vector machine (SVM) models for AADT prediction, [10] employed artificial neural networks (ANN) and SVM to predict AADT volumes, [11] also employed SVM-based models, [12] and employed classification and regression trees to predict short-term AADT volumes. Decision trees were also used by [13] in intersection traffic prediction. Reference [14] proposed a solution based on genetic algorithms, and a number of publications ([15–18]) used Bayesian networks in predicting traffic flow. Comparative analyses ([19–22]) show that while parametric methods can be used under certain conditions, nonparametric methods are better in others. In addition, there are many hybrid approaches that employ various methods simultaneously, e.g., [23–26].

Among nonparametric solutions, ANNs have been growing in popularity ever since being proposed to estimate motion parameters [27]. The recent advancement in massive data availability has made it possible for ANN models to achieve better results than models based on statistical methods [28]; thus they are widely used to predict traffic flow and AADT. Reference [29] used fuzzy networks to deal with the uncertainty of spatiotemporal data features in traffic flow prediction. Reference [30] handled the problem of missing data in forecasting AADT using recurrent neural networks. Reference [31] examined the possibility of estimating AADT from a one-week dataset using the ANN-fuzzy approach. An extensive literature review related to the use of ANNs for traffic flow prediction was included in [32]. Moreover, multiple recent studies have leveraged ANNs to address the hourly traffic volume estimation. Reference [33] used an ANN to estimate hourly traffic volume using continuous count station features as direct inputs to their model.

Reference [34] estimated freeway off-ramps hourly traffic volume using an ANN. This paper illustrated that the off-ramp volume estimation can achieve acceptable accuracy when proper data are fed into the model. An exhaustive review and classification of the deep learning methods and models used in the estimation of road traffic parameters is also included in [2].

Although ANNs achieve good results, a lot of data are necessary, and meeting this requirement is usually expensive and time-consuming. According to [35], one of the fundamental problems of using predictive models is the limited possibility of using them outside the area or period of study in which it was conducted. A potential solution to the problem of insufficient data is to obtain additional data from other available sources, such as GSM stations [36], GPS data [1], social media, or applications installed voluntarily on mobile devices [37]. Although Big Data technologies and open data sources have made the data problem easier to solve than ever before, in many cases it still remains an issue [38]. The second approach assumes that it is possible to transfer data and models developed in one spatiotemporal context to another location. Transferability can be investigated in the temporal and spatial dimensions. Temporal transferability is the possibility of using observations or models trained in a given time window to estimate parameters in other periods. Spatial transferability means the possibility of using observations or models trained for one area in another location [39]. The spatiotemporal transferability of forecasting models is of considerable practical interest. It can save time and money on gathering data or preparing the model itself and overcoming the lack of a trained personnel [40]. This issue has been of interest in the context of traffic volume estimation for decades (see for instance [3, 40–43]). For machine learning models, this problem is defined as transfer learning. The transfer of data or knowledge from one domain to another is a way to solve the problem of insufficient training data, which is particularly important when using ANNs [44]. Reference [45] demonstrated that the size of the dataset is crucial for the accuracy of traffic forecasts in the ANN-based models and, in terms of results, it is more important than other factors. Hence, the problem of using data or knowledge of traffic patterns acquired in one place at different locations still needs a solution [7, 46]. The current study aims to address this problem for hourly traffic volume estimation. This paper explores the viability of using existing CCS data from Florida to help calibrate an hourly traffic volume estimation model in New Hampshire, where a much smaller calibration dataset is available. In particular, the following two specific issues were explored concerning the model performance:

Using significantly fewer CCSs in the target state (New Hampshire) to calibrate the model

Collecting calibration data from the target state for a shorter time period.

The proposed solution can be classified as transfer learning based on mapping [44], which consists of combining data sources from the source and destination domain,

thus creating a new (larger) dataset. This method is based on the assumption that although the data from both domains differ, this new set of data contains sufficient features of the target domain for accurate forecasting. Namely, it applies a strategy named fine-tuning without freezing transferred layers described in [3].

3. Data

Two datasets were used to test the impact of model transferability: Florida and New Hampshire. The Florida dataset encapsulates October–December 2016, while New Hampshire data are from June–September 2017. The Florida dataset is over four times larger than the corresponding New Hampshire dataset (that is, it contains data from four times more Continuous Count Stations), but apart from the location and size, the structure of both datasets is the same. Each dataset is organized spatially by the traffic message channel (TMC) segments and temporally at the hour level. Table 1 summarizes the ground truth values and input variables used for model training and development, which are further described below.

3.1. Ground Truth Data. Hourly traffic volumes from continuous count stations were used as a ground truth (expected output) for the neural network volume estimation models. 173 traffic sensors were used in Florida, while 42 were used in New Hampshire, and these stations were located at all types of major roads (motorways, highways, major and minor arterials). To obtain counts on the TMC network (i.e., the road network used for analysis), each unique count station and traffic direction was mapped to a corresponding TMC segment via GIS analysis.

3.2. Input Features.

Vehicle Probe Counts. The hourly aggregated vehicle probe volumes were obtained based on the raw global positioning system (GPS) data provided by a probe data vendor. The raw waypoints were initially snapped to the XD road segments by the provider. To remain consistent with other data sources, we used a bridge between the XD and TMC segment definitions to match the waypoints with the TMC segments and then aggregated the data at an hourly level. Additionally, each vehicle was associated with one of the three weight classes (class 1: below 14,000 lbs, class 2: between 14,000 and 26,000 lbs, and class 3: above 26,000 lbs), and the probe volumes for each weight class were provided separately to the model. The median of all penetration rates (based on a comparison of the vehicle probe counts with the corresponding continuous count stations) was 2.19% in Florida and 2.3% in New Hampshire.

Vehicle Probe Speeds. The average hourly speed estimates based on GPS data were obtained from RITIS (the regional integrated transportation information

TABLE 1: Ground truth and input variables used for model development.

Name	Category	Data type	Description
Ground truth volumes			
CCS traffic counts		Integer (≥ 0)	Hourly traffic counts from continuous count stations
Input data features			
Probe counts			
(weight class 1)	Probe counts	Integer (≥ 0)	Number of unique probe counts from weight class 1 vehicles (<14k lbs)
Probe counts (weight class 2)		Integer (≥ 0)	Number of unique probe counts from weight class 2 vehicles (14k–26k lbs)
Probe counts (weight class 3)		Integer (≥ 0)	Number of unique probe counts from weight class 3 vehicles (>26k lbs)
Probe count null flag		Binary (0, 1)	Indicator variable used to identify persistent missing probe data
Avg probe speed		Float (≥ 0)	Average (harmonic) probe speed over 1 hour time period (mph)
Reference speed		Float (≥ 0)	Estimated freeflow speed for segment (mph)
Avg travel time	Probe speeds	Float (> 0)	Avg time to travel across segment (seconds)
Probe speed null flag		Binary (0, 1)	Indicator variable used to identify persistent missing probe speed/travel times
Temperature		Float (≥ 0)	Temperature in degrees Fahrenheit
Temperature null flag		Binary (0, 1)	Indicator variable used to identify whether temperature is missing
Dewpoint		Float (≥ 0)	Dew point in degrees Fahrenheit
Dewpoint null flag		Binary (0, 1)	Indicator variable used to identify whether dewpoint is missing
Relative humidity	Weather	Float (≥ 0)	Relative humidity (%)
Relative humidity null flag		Binary (0, 1)	Indicator variable used to identify whether relative humidity is missing
Visibility		Float (≥ 0)	Visibility (miles)
Visibility null flag		Binary (0, 1)	Indicator variable used to identify whether visibility is missing
Precipitation		Float (≥ 0)	Precipitation (inches)
Precipitation null flag		Binary (0, 1)	Indicator variable used to identify whether precipitation is missing
Segment length		Float (≥ 0)	Length of segment in miles
OSM road class		Binary (0, 1)	6 one-hot encoded variables representing possible OpenStreetMap road classes
OSM lanes		Float (≥ 0)	OSM-based estimate of number of directional lanes on the road
OSM lanes null flag		Binary (0, 1)	Indicator variable used to identify whether OSM lanes is missing
Structure type	Road characteristics	Categorical	Indicates presence of a special facility (bridge, tunnel, causeway)
HPMS through lanes		Float (≥ 0)	HPMS-based estimate of number of total lanes on the road
Functional system (HPMS)		Binary (0, 1)	3 one-hot encoded variables for HPMS functional classes 1, 2, 3+
Functional road class (TMC)		Binary (0, 1)	3 one-hot encoded variables for TMC functional classes 1, 2, 3+
AADT		Float (≥ 0)	HPMS annual avg daily traffic associated with TMC segment
AADT (single)		Float (≥ 0)	Single-unit truck and bus AADT associated with TMC segment
AADT (combination)		Float (≥ 0)	Combination truck AADT associated with TMC segment
Hour	Temporal	Binary (0, 1)	24 one-hot encoded variables representing hours 0–23
Weekend		Binary (0, 1)	2 one-hot encoded variables for identifying Saturday and Sunday
Volume profile estimate		Float (≥ 0)	Time-of-day volume estimate based on TTI profiling method
Volume profile null flag	Other	Binary (0, 1)	Indicator variable used to identify whether volume profile estimate is missing

system) [47]. RITIS was created and is maintained by the Center for Advanced Transportation Technology at the University of Maryland, College Park and provides visual analytics and data query capabilities for industry-sourced probe data.

Weather Data. Weather features were extracted from all permanent weather stations using data archived by the Iowa Environmental Mesonet [48] and assigned to each TMC segment based on spatial proximity. Initial tests suggested that the most important weather features are precipitation, temperature, visibility, and humidity.

Road Characteristics. Infrastructural characteristics were extracted from both the National Performance Measurement Research Dataset (NPMRDS) TMC shapefile and the Open Street Map (OSM) road

network. As the OSM maps use a different network topology, the OSM road characteristics (road classification and number of lanes) first had to be conflated to the TMC map, which was conducted using an automated conflation algorithm developed for this purpose. The final road characteristic features used for each TMC segment include information regarding road classification, number of lanes, segment length, and reference speed, as well as historical average annual daily traffic values associated with each TMC segment (obtained from the NPMRDS TMC shapefile).

Temporal Data. Information concerning the hour of the day and the day of the week (working day/Saturday/Sunday) was also considered for each data point in order to account for temporal traffic patterns.

Other. Hourly volume profiles were derived by applying the widely used profiling method [49]. This method transforms AADT estimates derived from the highway performance monitoring system [50] into hourly volume profiles based on historic speeds available in RITIS.

4. Model

In previous research [4], it was shown that a fully connected (dense) neural network with three hidden layers yields the best volume estimation model performance. The structure of this network is presented in Figure 1 and is used for all subsequent experiments.

4.1. Training Procedure. The mean absolute error (MAE) was selected as the loss function (i.e., the function that is minimized by the learning algorithm). Although a less popular choice than the mean squared error (MSE)—which tends to place greater emphasis on higher volumes, MAE was selected because it provides a good trade-off between MAPE and R^2 performance metrics, which were used to assess the tested models and approaches. Additionally, MAE was used in previous research, therefore, the selection of MAE makes it easier to compare the results.

The model was trained with the Adam algorithm [51] proposed by Diederik P. Kingma and Jimmy Ba in 2014. Among many advantages, such as quick convergence, computational efficiency, and intuitive hyperparameters, the Adam algorithm is robust in terms of hyperparameter settings and usually does not require much hyperparameter tuning—a feature that was particularly important due to the number of experiments required. Overall, we trained 546 models, which made it impossible to separately tune the hyperparameters manually for each model. Thus, during initial tests, we discovered that the default hyperparameters ($\alpha = 0.01, \beta_1 = 0.9, \beta_2 = 0.999$) work reasonably well and achieve strong model performance. Additionally, experiments showed that tuning hyperparameters around the default values did not significantly change the results—only the speed of convergence. Furthermore, due to the implementation of Dropout [52] after each hidden layer, the models turned out to be resistant to overfitting. Based on these initial findings, we decided to use the default hyperparameters for all training procedures and train the networks longer than required (i.e., to avoid tuning at the expense of some efficiency). The sample loss plots for the models with the smallest and largest datasets are presented in Figure 2. Both train and validation losses do not significantly decrease in the final few epochs, thus demonstrating that the models were trained long enough, and further training would not have improved the accuracies. On the other hand, the validation losses do not increase with time, which shows that the models do not overfit.

The smallest datasets contained only 1 week of New Hampshire data, while the largest one contained all 3 months of both New Hampshire and Florida data. The charts come from the initial (tuning) experiments, where the New Hampshire data were split into train and validation datasets.

4.2. Partitioning Training and Validation Data. Initial tests were first performed to verify the structure of the network and tune the network hyperparameters. These tests were performed for each experiment with a fixed split of New Hampshire data allocated into the training and validation parts. To avoid data leakage different, continuous count stations were used for training and validation sets. Additionally, the split was made taking into consideration the functional road class (each FRC was represented in the same proportions in the training and validation sets). Next, after determining all the hyperparameters, the full crossvalidation procedure was employed. Each model was trained using data for 41 New Hampshire Continuous Count Stations and tested on the 42nd station. This procedure was repeated 42 times to ensure that all NH data were included in the test dataset. While being time consuming, this approach allowed us to avoid data leakage and to take full advantage of the given datasets.

4.3. Evaluating Model Accuracy. During each iteration of cross-validation, model performance is quantified at the test location via the following error metrics: R^2 , MAPE (mean absolute percentage error), and EMFR (Error to Maximum Flow Ratio), a process that is repeated during each experiment. These metrics were also used in previous research [4, 5], which renders further comparisons easier. R^2 , MAPE, and EMFR metrics are presented in the following equations, where y_i denotes an actual volume, \hat{y}_i is a volume estimate, \bar{y} stands for the sample average, y_{\max} represents the maximum observed traffic volume, and n is the number of data points used to compute the metric:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1)$$

$$\text{MAPE} = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \right) \times 100, \quad (2)$$

$$\text{EMFR} = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_{\max}} \right| \right) \times 100. \quad (3)$$

R^2 represents the fraction by which the variance of the errors is less than the variance of the dependent variable. In other words, it shows the percent of variance explained. MAPE expresses accuracy as a relation between an absolute error and the real observed value. MAPE is widely used for traffic volume estimations, mainly due to its ease of interpretation. However, MAPE has a few flaws. Namely, in the case of small volumes, the MAPE can be very high even if the absolute error is relatively small. Due to this, MAPE is highly affected by the time periods when the traffic volumes are small, whilst for planning and operational purposes, the time periods with high traffic volumes are usually much more important. EMFR is used in order to deal with this problem. EMFR is defined as the relation between an absolute error and the maximum observed value. Thus, if the absolute error is much smaller than the maximum observed traffic, the

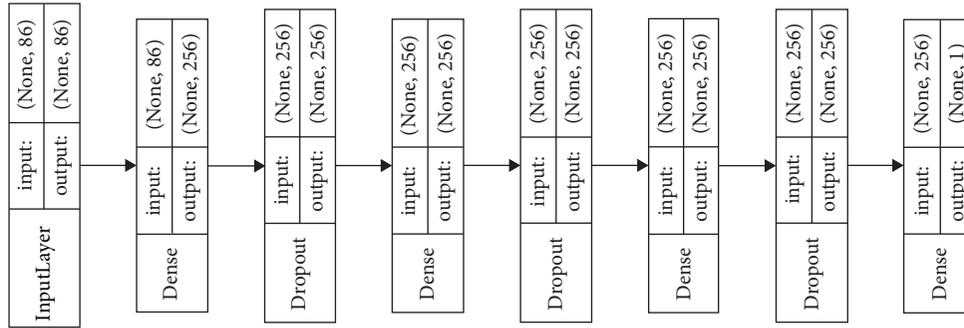


FIGURE 1: A fully connected (dense) ANN was employed to estimate traffic volumes.

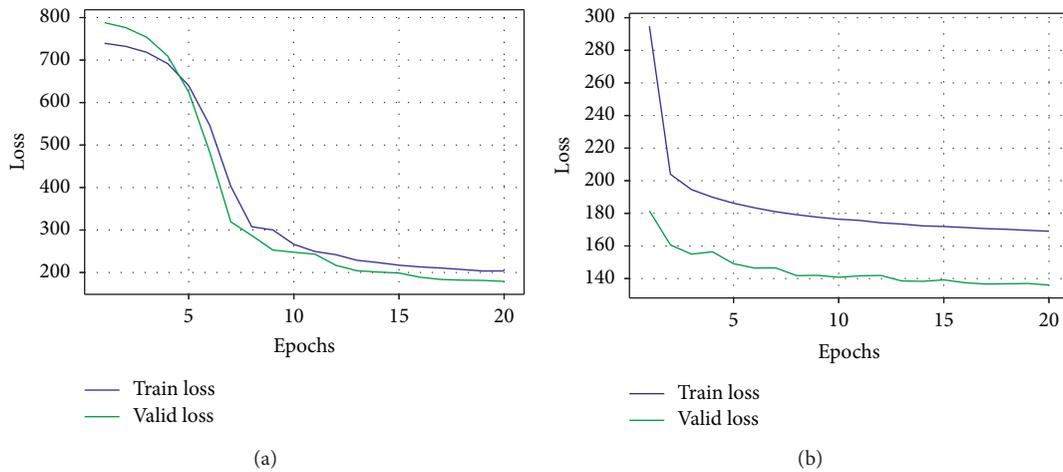


FIGURE 2: The behavior of the train and validation loss functions for the smallest and largest tested datasets. (a) Smallest dataset and (b) largest dataset.

value of EMFR is also small, regardless of the current traffic. On the other hand, if the observed traffic is close to the maximum, the values of EMFR are close to the values of MAPE. EMFR has also a practical meaning, in the above-mentioned VTM project, this metric was used to define the model quality thresholds (less than 10% EMFR was considered “satisfactory,” while less than 5% EMFR was considered “very good”).

5. Experimental Results and Discussion

5.1. Experiment 1: Model Comparison. The goal of the first group of experiments was to explore the possibility of using the Florida dataset for New Hampshire predictions. During the experiments, four different approaches were employed and compared. The outline of the experiments is shown in Figure 3.

The “base” model was trained with the New Hampshire data only; no Florida data were used. The detailed results are presented in Table 2 in the *Base* columns. The mean and median R^2 s for this approach were 0.72 and 0.82, while the mean and median MAPEs were 43.3% and 26.9%, and the mean and median EMFRs were 8.12% and 7.04%, respectively. The typical (i.e., median) model performance is good

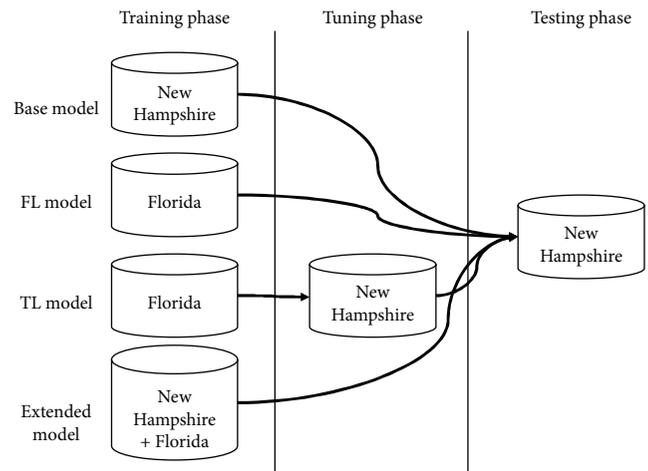


FIGURE 3: The outline of the experiments—Stage 1.

and is comparable with the results that were obtained for the state of Florida [5] in the previous stage of the project (median R^2 : 0.83, median MAPE: 25%). They are also much better than the currently used profiling method (median R^2 : 0.60, median MAPE: 55.6%). However, the problem with the

TABLE 2: The detailed results for different approaches.

	R2				MAPE (%)				EMFR (%)			
	Base	Fl	TL	Ext	Base	Fl	TL	Ext	Base	Fl	TL	Ext
Mean	0.72	0.34	0.77	0.81	43.3	61.0	34.4	34.9	8.12	12.61	7.36	7.15
Std	0.49	1.30	0.30	0.13	47.6	43.8	26.1	29.0	4.03	7.52	2.75	2.23
min	-2.46	-8.31	-1.08	0.06	16.2	27.9	14.1	16.0	4.44	7.52	4.27	4.67
25 per.	0.76	0.54	0.76	0.78	22.7	40.8	20.8	21.4	5.85	8.20	5.93	5.79
50 per.	0.82	0.70	0.85	0.84	26.9	49.0	26.9	25.5	7.04	9.95	6.67	6.68
75 per.	0.88	0.80	0.89	0.88	37.6	59.6	34.0	34.2	8.40	13.96	7.90	7.69
max	0.94	0.92	0.95	0.94	311	286	149	182	28.3	48.0	20.8	20.0

Bold values indicate mean and median values.

“base” model is the outliers, which can be seen in Figure 4—which shows the R^2 metrics with and without outliers—and in Table 2. High standard deviations, a large difference between mean and median R^2 s and MAPE s and unacceptable results for the worst continuous count station R^2 : -2.74 and MAPE: 326%) emphasize the problem.

Base stands for the model trained on New Hampshire data only, Fl is the model trained on Florida data and tested with NH data, TL stands for transfer learning (the model was trained on the Florida dataset, and then fine-tuned with NH dataset), and *Ext* stands for extended dataset (both Florida and New Hampshire datasets were included into the training set).

Base stands for the model trained on New Hampshire data only, *Florida* is the model trained on Florida data and tested with NH data, TL stands for transfer learning (the model was trained on the Florida dataset, and then fine-tuned with the NH dataset), and *Extended* stands for the extended dataset (both Florida and New Hampshire datasets were included in the training set).

Time series charts (a, b, and c) present the real (blue) and estimated (red) traffic volumes. For an ideal model, these two lines should overlap. Hexabin charts (d, e, and f) present the relations between the real and the estimate traffic volumes. The color of hexagons corresponds to the number of points within each hexagon. As long as the estimates are close to the real values, the points are located on the diagonal line.

Additionally, we selected three locations with EMFR equal to 5%, 10%, and 25% and plotted both time series and hexabin charts for each location (Figure 5). The main purpose of this figure is to facilitate the interpretation of the presented metrics.

The second approach involved training the model with only Florida data, and then testing on the New Hampshire dataset. The detailed results are shown in Table 2 in the *Fl* columns. Although the results are still better than for the profiling method, they are the worst from all the tested models, in terms of both mean and median R^2 s, MAPE s, and EMFR s. This suggests that it is difficult to simply transfer the model from one area (state) and time period, use it in another state and time period, and maintain a high level of estimation accuracy.

The third approach was based on transfer learning and the fine tuning procedure. It was similar to the second one, but this time the model trained with Florida data was fine-tuned with New Hampshire data. The results for this

scenario are presented in both Figure 4 and Table 2 (column TL). During the training process, consecutively one, two, or three (all) hidden layers of the model were unfrozen, i.e., the weights of these layers were changed in the fine-tuning procedure. Due to the fact that the model did not overfit, the best results were obtained for all three unfrozen layers. The results presented in the papers are from the network with all the layers unfrozen. The transfer learning based models behave better than the “base” New Hampshire model, for all the metrics, except the median MAPE s, which are equal for both approaches. The transfer learning approach results also in a significant reduction in outliers, which may be noticed in both Figure 4 and Table 2.

Finally, we added the Florida dataset to the training data. For the crossvalidation procedure, each training set for this scenario contained all Florida data, and data for 41 of 42 New Hampshire Continuous Count Stations. The detailed results are shown in Table 2 in the columns *Extended* and in Figure 4. The results indicate that not only is the typical model performance superior to the other approaches (in terms of mean and median R^2 s and MAPE s this approach is the best for three out of four metrics) but it also best deals with outliers.

5.2. Experiment 2: Impact of Dataset Size. The goal of the second set of experiments was to check if it is possible to use both the transfer learning and the extended dataset approaches, as explained in the previous subsection with a dataset, which covers a relatively small time period. These experiments have important practical implementations—the cost of the data depends on the size of the dataset. For these experiments, we compared three approaches—the “base” approach based on the model trained on New Hampshire, the transfer learning approach based on the model pre-trained on the Florida dataset and fine-tuned on the New Hampshire dataset, and the “extended” model—based on using merged Florida and New Hampshire datasets for training.

This time, instead of using all the New Hampshire data in the training procedure, we reduced the time scope of the data; we trained the models using the NH dataset reduced to twelve, eleven, ten, and so on down to one week and (for the extended approach) the entire dataset from Florida. During the training, we repeated the full crossvalidation procedure, and each of the models was tested on a full three months

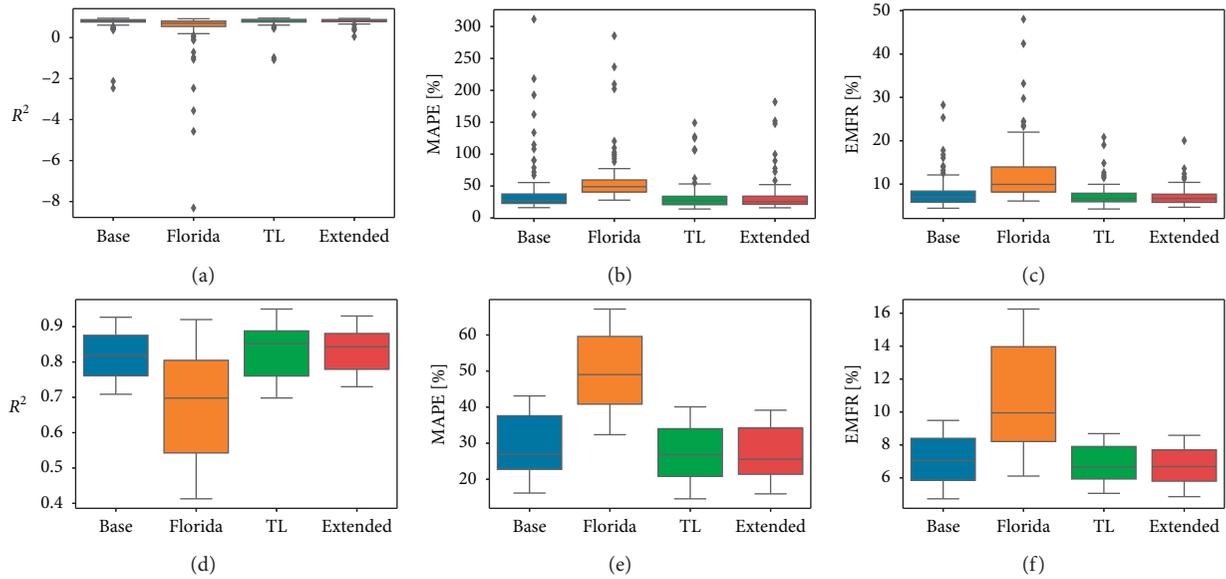


FIGURE 4: R^2 , MAPE, and EMFR metrics for different approaches with and without outliers. (a) R^2 with outliers, (b) MAPE with outliers, (c) EMFR with outliers, (d) R^2 without outliers, (e) MAPE without outliers, and (f) EMFR without outliers.

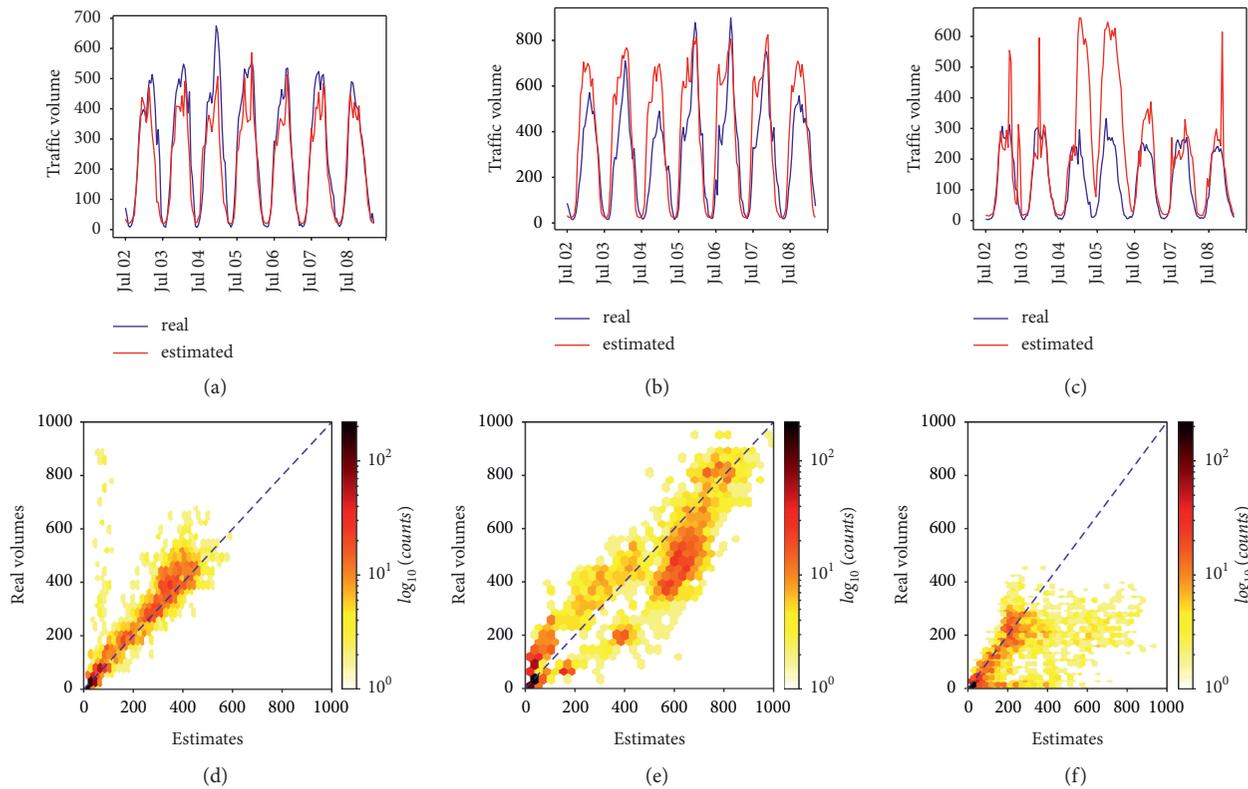


FIGURE 5: Examples of time series and hexabin plots for different EMFR values. (a) EMFR = 5%, (b) EMFR = 10%, (c) EMFR = 25%, (d) EMFR = 5%, (e) EMFR = 10%, and (f) EMFR = 25%.

New Hampshire dataset that corresponded to the tested continuous count station. The detailed results are presented in Table 3, while a comparison of the models' behavior is illustrated in Figure 6.

The size of the ribbons is proportional to the standard deviation ($0.2 * \text{std}$). Due to the large values of the standard deviation in case of the "base" model, plotting the entire standard deviation makes the figures difficult to read.

TABLE 3: The detailed results regarding the size of the New Hampshire dataset.

	Base model											
	R2				MAPE (%)				EMFR (%)			
	1 w	5 w	9 w	3 m	1 w	5 w	9 w	3 m	1 w	5 w	9 w	3 m
Mean	0.08	0.63	0.69	0.72	61.9	51.9	46.9	43.3	11.46	8.75	8.27	8.12
Std	3.65	1.00	0.69	0.49	100.5	66.7	55.3	47.6	9.13	5.05	4.46	4.03
min	-23.9	-5.98	-3.80	-2.46	23.4	15.3	16.6	16.2	6.27	4.92	4.62	4.03
25 per	0.53	0.71	0.75	0.76	33.9	24.6	24.1	22.7	8.08	6.25	6.06	5.85
50 per	0.71	0.81	0.83	0.82	43.3	31.0	27.6	26.9	9.44	7.60	7.13	7.04
75 per	0.80	0.87	0.87	0.88	51.1	40.6	39.0	37.6	11.50	8.80	8.17	8.40
max	0.90	0.93	0.95	0.94	798	424	375	311	68.8	38.6	32.7	28.3

	Transfer learning											
	R2				MAPE (%)				EMFR (%)			
	1 w	5 w	9 w	3 m	1 w	5 w	9 w	3 m	1 w	5 w	9 w	3 m
Mean	0.76	0.78	0.78	0.77	35.5	36.1	37.3	34.4	7.88	7.48	7.39	7.36
Std	0.19	0.23	0.27	0.30	24.4	28.8	34.7	26.1	2.44	2.62	2.71	2.75
min	-0.25	-0.53	-0.81	-1.08	15.1	15.6	15.3	14.1	5.01	4.51	5.34	4.27
25 per	0.72	0.76	0.75	0.76	23.5	21.4	20.7	20.8	6.29	5.84	5.78	5.93
50 per	0.81	0.84	0.85	0.85	30.2	28.5	28.3	26.9	7.21	6.89	6.72	6.67
75 per	0.87	0.88	0.89	0.89	36.2	37.7	35.7	34.0	8.37	7.94	8.06	7.90
max	0.94	0.95	0.95	0.95	166	178	191	149	17.3	19.2	20.0	20.8

	Extendedsnmodel											
	R2				MAPE (%)				EMFR (%)			
	1 w	5 w	9 w	3 m	1 w	5 w	9 w	3 m	1 w	5 w	9 w	3 m
Mean	0.77	0.80	0.80	0.81	36.5	36.0	32.9	34.9	7.83	7.28	7.34	7.15
Std	0.13	0.11	0.13	0.13	23.6	28.4	19.9	29.0	1.72	1.67	2.28	2.23
Min	0.10	0.32	0.23	0.06	17.3	16.2	16.0	16.0	4.91	5.03	4.70	4.67
25 per	0.74	0.77	0.75	0.78	23.4	22.6	22.0	21.4	6.55	6.06	5.85	5.79
50 per	0.80	0.82	0.84	0.84	29.7	26.5	28.0	25.5	7.76	6.98	6.64	6.68
75 per	0.86	0.81	0.88	0.88	37.0	34.3	36.1	34.2	8.77	7.93	7.77	7.69
max	0.92	0.94	0.94	0.94	146	190	138	182	12.5	12.3	17.2	20.0

Bold values indicate mean and median values.

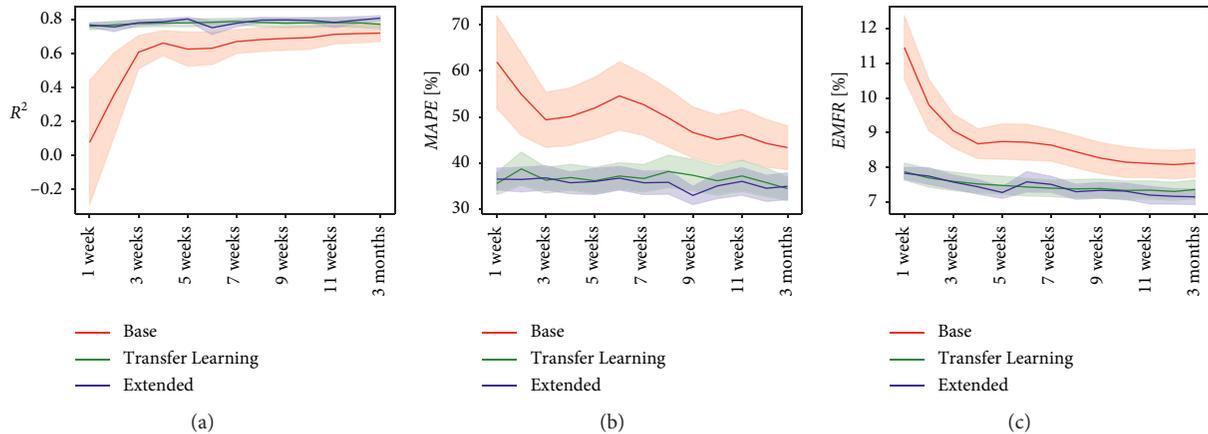


FIGURE 6: Mean (a) R^2 , (b) MAPE, and (c) EMFR metrics depending on the size of the dataset.

The accuracy of the “base” model depends primarily on the size of the New Hampshire training dataset. All the metrics show that the performance of the model diminishes when the size of the dataset is reduced. On the other hand, the accuracy of the “transfer learning” and “extended” models do not significantly change with a change in the size of The New Hampshire training dataset. It shows that by leveraging a larger dataset collected elsewhere, it is possible to achieve good results with as little as one week of data.

Moreover, the results for both the “transfer learning” and the “extended” approaches with 1 week of New Hampshire training data are better than the results of the “base” model trained on three months of data, thus emphasizing on the practical usefulness of the proposed solutions.

The “Extended” approach turned out to be slightly more accurate than the one based on transfer learning. However, the differences are not that significant, and the transfer learning-based approach has two advantages. First, due to

TABLE 4: The detailed results for different models.

	Mean			Median (%)		
	EMFR (%)	MAPE (%)	R^2	EMFR	MAPE (%)	R^2
Dense	6.29	23.6	0.882	6.22	20.8	0.882
RF	6.77	28.7	0.839	6.39	26.4	0.875
LR	11.61	63.8	0.556	8.77	52.9	0.736
BR	9.85	60.9	0.708	8.77	50.4	0.779
PR	6.9e7	7.05e8	-9.3e12	7.43	36.6	0.841

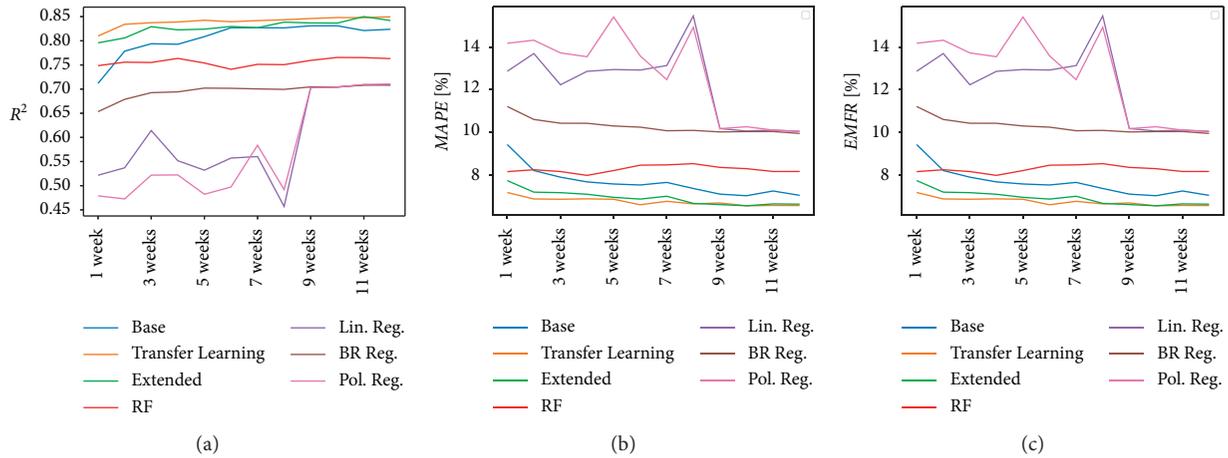
FIGURE 7: Median (a) R^2 , (b) MAPE, and (c) EMFR metrics depending on the size of the dataset.

TABLE 5: The detailed results for different models—one week long training dataset.

	Mean			Median (%)		
	EMFR (%)	MAPE (%)	R^2	EMFR	MAPE (%)	R^2
Base	11.46	61.9	0.078	9.44	43.3	0.712
TL	7.88	35.1	0.761	7.21	30.2	0.810
Extended	7.84	36.4	0.769	7.76	29.7	0.796
RF	9.85	42.9	0.627	8.18	34.0	0.749
LR	4.1e12	4.4e13	-1.0e24	12.87	75.7	0.522
BR	12.62	71.7	0.429	11.22	56.4	0.653
PR	3.1e12	3.3e13	-5.8e23	14.17	88.0	0.479

the smaller dataset, the training process is much faster for the transfer learning-based approach than for the “extended” one. Additionally, it is possible to employ transfer learning even without direct access to a larger dataset. For transfer learning, only the previously trained model is necessary. This is important due to data licensing. Some vendors offer data for “limited usage” only. It means that the data purchased by a larger state cannot be directly employed for training the model in other states. Such a situation makes using the “extended” approach impossible, while using the pretrained model and transfer learning approach is still doable.

5.3. Discussion. The first set of experiments suggest that utilizing information from a previous dataset can improve traffic volume estimates in a target location—even if the locations and time periods differ. However, it is very difficult to maintain a model’s accuracy when applied directly to a

different location and time period than it was trained. Directly applying a model trained in one location (e.g., Florida) to a new target location (e.g., New Hampshire) yields reasonable “typical” performance—implying that, in general, the model captures the relation between input variables and traffic volumes, but suffers from severe outliers. Instead, a more promising approach appears to be fine-tuning the model originally trained in one location using data from the target location—an approach that incorporates patterns in the target location, and thus helps to avoid many outliers. Finally, if available, incorporating all data for training purposes (e.g., adding all Florida data for training a model to estimate New Hampshire volumes) yields the best results.

Given the findings from the first set of experiments—namely, that a previous dataset can be used to improve estimates in a new location, the second set of experiments seek to understand how much data are needed at the target location to achieve acceptable accuracy. Focusing,

in particular, on the “transfer learning” with fine-tuning and “extended” approaches, it finds that given a sufficiently large alternate dataset to leverage (i.e., Florida), both approaches can achieve reasonable results with as little as one week of data in the target (i.e., New Hampshire) location.

Overall, it appears that the machine learning-based method of volume estimation requires large datasets to learn complex spatial and temporal patterns between the input variables and traffic volumes. Given financial limitations and the fact that some states and jurisdictions are small and have limited count stations on various road classes, it can be challenging to collect sufficient data to train a reliable estimation model. However, the approaches highlighted in this paper show that input features and corresponding traffic counts can be leveraged from other locations, and used in conjunction with small amounts of data from the target location to develop better overall models. The results indicate that if possible, the best approach is the “extended” one, which uses all previous data in the training process, and does the best job of eliminating outliers. However, in cases where this is not possible (perhaps due to licensing of previous data), the “transfer learning” with fine-tuning approach can be used. This approach only requires access to a previously trained model (not the raw data), whose hyperparameters are subsequently tuned while training the model based on data from the target location.

6. Conclusion

This paper explores the transferability of volume estimation machine learning models, seeking to understand the extent to which a model trained to predict traffic volumes in one location can be used directly or modified slightly to do so elsewhere. The implications of this research question are significant, as it is expensive to acquire GPS trajectory data—a key model input— and time consuming to preprocess data sources. If existing large datasets can be utilized for training purposes, smaller, less expensive datasets can be purchased in new locations and used to build accurate models—potentially saving transportation agencies time and money.

The experimental results suggest that a key component of model performance is having sufficient data for training. If access to a larger, existing dataset is available, the optimal approach is to train the volume estimation model on all available data—using both locations in the target geography and previously collected data. However, even if the raw dataset is not available from a previous location (perhaps due to licensing restrictions), an existing model can be trained and fine-tuned using a small amount of data from the target locations. Interestingly, the larger dataset used to improve performance appears to be useful even if it comes from different places and time periods, and when the target dataset is as small as one week.

Note that the results provided in this study are limited to the hourly volume estimation model’s transferability at the state level. The states whose data were available to test the proposed approach encompass both urban and rural areas with different land-use characteristics. Additionally, all roads used for the analyses are FRC 1 or 2 due to data

accessibility limitations. Therefore, overgeneralizing the results to various geographical and road levels should be avoided. However, given the availability of the required data for different locations, it is possible to test the proposed approach in other areas to investigate the extent to which it can be generalized.

A key future direction for research includes investigating how much data need to be collected from continuous count stations in order to constitute a satisfactorily large dataset for transfer learning, and whether there are certain temporal or spatial characteristics that are necessary for the data to be transferable. Additionally, it would be beneficial to understand how often the models should be fine-tuned or retrained in order to optimize performance. Based on the promising results presented in this paper, these research questions will be explored in more detail in future modeling efforts.

Appendix

A. Model Selection

This appendix describes the experiments performed to verify whether the solution selected in [4] (a fully connected artificial neural network) is still the most suitable for the presented task. Similar to Section 5, two experiments were carried out.

First experiment—carried out on the entire dataset. The primary purpose of this experiment was to analyze alternative models and compare them with artificial neural networks. This experiment corresponds to the experiment described in Section 5.1. *Experiment 1: model comparison.*

Second experiment—carried out on reduced datasets. It was conducted to check if the conclusions based on the results from the first experiment are valid for much smaller datasets. This experiment corresponds to the experiment described in Section 5.2. *Experiment 2: impact of the dataset size.*

A.1. Analyzed Models. In the first step of the presented analysis, the authors analyzed different models and selected some of them for further tests. The following models were selected for future analysis:

Fully connected neural network (dense): The model used in [4]. This model is a neural network built of three fully connected (dense) hidden layers, 256 neurons in each layer. The hidden layer neurons use ELU activation function.

Random forest (RF): In our experiments, this model performs only slightly worse than the dense neural network. RF models tend to overfit, so we applied regularization; namely, we limited the minimum samples per leaf (values 5, 10, 15, 95, and 100 were tested). The model with the smallest median EMFR was selected. This model was employed for both experiments.

Linear regression (LR): A basic linear regression model was used as a baseline for our experiments.

Bayesian ridge regression (BR): an approach to linear regression that tries to solve the problem of poorly distributed data by using probability distributions rather than point estimates.

Polynomial regression (PR): A variation of linear regression that uses polynomial features. Technically, the model is linear, but additional features are provided to ensure the polynomial output. For example, if the original input features were X_1 and X_2 , and we are interested in the second degree polynomial, the following set of features shall be provided as a model input: $X_1, X_2, X_1 * X_2, X_1^2, X_2^2$.

The following models were not considered in this comparison:

Support vector machines (SVM): SVM is a very powerful solution, but it does not scale well to large datasets. Our training dataset contained over 830,000 samples, thus it was impossible to train an SVM model with a radial basis function (RBF) kernel that ensures the best performance of this approach. In fact, it turned out that with the Sci-kit Learn library, it was also impossible to train an SVM model using polynomial or linear kernels that are expected to generate much worse results than an RBF kernel. Therefore, we did not include this class of models in our analysis.

Long-short-term memory neural networks (LSTM)—LSTM networks seem to be the most suitable architecture for the presented task, as they can generate estimates using also the previously observed features as inputs, and simultaneously, not excessively increase the complexity of the networks. During previous (unpublished) research, we used LSTM networks and achieved slightly better results than with dense neural networks. However, there are two main reasons that prevented us from using these models in the presented research.

Depending on the size of the train dataset, it takes 6–48 hours to train the model. Overall, we trained more than 500 models; thus, it would not be possible to repeat the entire procedure with LSTM models in a reasonable timespan.

Although LSTM models are slightly better than dense neural networks, we also discovered that for the given type of data they are very susceptible to small changes in the hyperparameters. The gist of our research was to discover how the models behave in given different data scenarios. Had we used an LSTM model and got worse results for some scenario, we could not be sure whether this would be the problem with the scenario or the model hyperparameters. One can argue that we could fine-tune the hyperparameters for each scenario, but given the aforementioned training time, it would take far too long.

A.2. Datasets. All the models were trained and tested with the same datasets. During the first experiment, the data were divided into two subsets:

The train dataset comprising of the entire Florida dataset (172 continuous count stations, 691,486 data points) and a part of the New Hampshire dataset (25 Continuous Count Stations, 140,306 data points).

The test dataset comprising of 17 continuous count stations (35,098 data points) located in New Hampshire.

Both train and test datasets contained the features described in Section 3.2.

In the second experiment, only the New Hampshire dataset (25 continuous count stations, 175,404 data points) was selected. To perform this experiment, the authors followed the procedure described in Section 5.2. *Experiment 2: impact of dataset size*, namely, the size of the dataset was gradually reduced from twelve weeks to one week. Each time, the entire crossvalidation procedure was carried out to determine the models' accuracy.

A.3 Results. The final results of the first experiment are presented in Table 4. For random forest, the authors trained the models for different minimum samples per leaf values and chose the best model (the model with the smallest median EMFR value).

A fully connected (dense) neural network turned out to be the best approach, although the results generated by the random forest model were very alike. This is consistent with the results presented in [4]. The results generated by the polynomial regression model are surprising. The medians of the metrics are not unexpected, although the means are very poor. Overall, the model behaved well, but it completely misestimated the traffic for one CCS, in one direction. The authors believe that this behavior is an overfitting problem, which was caused by fitting the model with too many features. To generate a second degree polynomial model as presented in the table, the authors used 3,240 features instead of the 79 used in all other models. Due to the observed overfitting, the authors did not test higher degree polynomial models (for example, to generate a third degree polynomial model, 88,560 features are needed, and a fourth degree polynomial model requires 1,837,620 features).

The second experiment's results overview is presented in Figure 7. Additionally, Table 5 consists of the detailed results for the smallest (one week long) dataset. First of all, both linear and polynomial regression could not deal with outliers. Similar to polynomial regression in the first experiment, they completely misestimated the traffic for one CCS in one direction, which resulted in terrible mean results. The median results presented in Figure 7 show that these models cannot handle the datasets smaller than 9 weeks of data. Bayesian ridge regression results were stable and did not deteriorate heavily with the train dataset size reduction. Moreover, Bayesian Ridge Regression managed to deal with

outliers, regardless of the size of the dataset. However, these results were significantly worse than the ANN-based results. Random forest turned out to be the best of the alternative models. For a two-week-long training set, RF results were similar to the results obtained with a “base” artificial neural network, and for a one-week-long dataset, they were even better. However, regardless of the size of the dataset, RF results were worse than the results of both approaches that leveraged the Florida dataset, namely, extended and transfer learning-based models.

The first three rows of the table present results obtained with dense neural networks. *Base* stands for the model trained on New Hampshire data only, *TL* stands for transfer learning (the model was trained on the Florida dataset, and then fine-tuned with NH dataset), and *Extended* stands for the extended dataset (both Florida and New Hampshire datasets were included into training set).

Data Availability

The data used to support the findings of this study are not available due to third-party rights.

Disclosure

An early version of this manuscript has been published at the 99th Annual Meeting of the Transportation Research Board. An earlier version of this paper was presented at the TRB annual meeting.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Acknowledgments

The authors would like to thank the I-95 Corridor Coalition for funding this work through the Volume and Turning Movement Project. The authors would also like to thank the Florida DOT and New Hampshire DOT for providing their CCS data. This support is gratefully acknowledged, but it implies no endorsement of the findings. This work was supported by the I-95 Corridor Coalition through the Volume and Turning Movement (VTM) project.

References

- [1] H.-h. Chang and S.-h. Cheon, “The potential use of big vehicle GPS data for estimations of annual average daily traffic for unmeasured road segments,” *Transportation*, vol. 46, no. 3, pp. 1011–1032, 2019, <https://link.springer.com/10.1007/s11116-018-9903-6>.
- [2] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin, “Deep learning on traffic prediction: methods, analysis and future directions,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–16, 2021.
- [3] J. Li, F. Guo, A. Sivakumar, Y. Dong, and R. Krishnan, “Transferability improvement in short-term traffic prediction using stacked LSTM network,” *Transportation Research Part C: Emerging Technologies*, vol. 124, Article ID 102977, 2021.
- [4] P. Sekuła, N. Marković, Z. Vander Laan, and K. Farokhi Sadabadi, “Estimating historical hourly traffic volumes via machine learning and vehicle probe data: a Maryland case study,” 2017.
- [5] N. R. E. L. CATT, “Volume & turning mvmt project,” 2018, <https://i95coalition.org/wp-content/uploads/2015/02/I-95CC-VTM-SCMtg-02-13-2018-final-combined.pdf?x70560>.
- [6] N. Slimani, I. Slimani, N. Sbiti, and M. Amghar, “Traffic forecasting in Morocco using artificial neural networks,” *Procedia Computer Science*, vol. 151, pp. 471–476, 2019, <https://www.sciencedirect.com/science/article/pii/S1877050919305265>.
- [7] S. Deng, S. Jia, and J. Chen, “Exploring spatial-temporal relations via deep convolutional neural networks for traffic flow prediction with incomplete data,” *Applied Soft Computing*, vol. 78, pp. 712–721, 2019, <https://www.sciencedirect.com/science/article/pii/S1568494618306082>.
- [8] P. Shang, X. Li, and S. Kamae, “Chaotic analysis of traffic time series,” *Chaos, Solitons & Fractals*, vol. 25, no. 1, pp. 121–128, 2005, <https://www.sciencedirect.com/science/article/pii/S0960077904006472>.
- [9] M. Castro-Neto, Y. Jeong, M. K. Jeong, and L. D. Han, “AADT prediction using support vector regression with data-dependent parameters,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 2979–2986, 2009.
- [10] S. Islam, “Estimation of annual average daily traffic (AADT) and missing hourly volume using artificial intelligence,” Ph.D. thesis, Clemson University, Clemson, SC, USA, 2016.
- [11] Y. Zhang and Y. Liu, “Traffic forecasting using least squares support vector machines,” *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009 b.
- [12] Y. Xu, Q.-J. Kong, and Y. Liu, “Short-term traffic volume prediction using classification and regression trees,” *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 493–498, 2013.
- [13] W. Alajali, W. Zhou, S. Wen, and Y. Wang, “Intersection traffic prediction using decision tree models,” *Symmetry*, vol. 10, no. 9, p. 386, 2018.
- [14] B. Abdulhai, H. Porwal, and W. Recker, “Short-term traffic flow prediction using neuro-genetic algorithms,” *Intelligent Transportation Systems Journal*, vol. 7, no. 1, pp. 3–41, 2002.
- [15] S. Sun, C. Zhang, and G. Yu, “A bayesian network approach to traffic flow forecasting,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–132, 2006.
- [16] S. Yang and G. A. Davis, “Bayesian estimation of classified mean daily traffic,” *Transportation Research Part A: Policy and Practice*, vol. 36, no. 4, pp. 365–382, 2002, <https://www.sciencedirect.com/science/article/pii/S0965856401000088>.
- [17] E. Castillo, J. M. Menéndez, and S. Sánchez-Cambronero, “Predicting traffic flow using bayesian networks,” *Transportation Research Part B: Methodological*, vol. 42, no. 5, pp. 482–509, 2008, <https://www.sciencedirect.com/science/article/pii/S0191261507001300>.
- [18] A. Pascale and M. Nicoli, “Adaptive bayesian network for traffic flow prediction,” in *Proceedings of the Statistical Signal Processing Workshop*, pp. 177–180, Nice, France, June 2011.
- [19] B. L. Smith, B. M. Williams, and R. Keith Oswald, “Comparison of parametric and nonparametric models for traffic flow forecasting,” *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002, <https://www.sciencedirect.com/science/article/pii/S0968090X02000098>.
- [20] K. Jha, N. Sinha, S. S. Arkatkar, and A. K. Sarkar, “A comparative study on application of time series analysis for traffic

- forecasting in India: prospects and limitations,” *Current Science*, vol. 110, no. 3, p. 373, 2016.
- [21] Y. Zhang and Y. Liu, “Comparison of parametric and non-parametric techniques for non-peak traffic forecasting,” *International Journal of Mathematics and Computer Science*, vol. 3, no. 3, pp. 172–178, 2009 a, <https://publications.waset.org/vol/27>.
- [22] J. Salotti, S. Fenet, R. Billot, N.-E. El Faouzi, and C. Solnon, “Comparison of traffic forecasting methods in urban and suburban context,” in *Proceedings of the 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 846–853, Volos, Greece, November 2018.
- [23] X. Luo, L. Niu, and S. Zhang, “An algorithm for traffic flow prediction based on improved sarima and ga,” *KSCE Journal of Civil Engineering*, vol. 22, no. 10, pp. 4107–4115, 2018.
- [24] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, “A hybrid deep learning based traffic flow prediction method and its understanding,” *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 166–180, 2018.
- [25] L. Li, B. Ran, J. Zhu, and B. Du, “Coupled application of deep learning model and quantile regression for travel time and its interval estimation using data in different dimensions,” *Applied Soft Computing*, vol. 93, Article ID 106387, 2020.
- [26] L. Li, X. Qu, J. Zhang, Y. Wang, and B. Ran, “Traffic speed prediction for intelligent transportation system based on a deep feature fusion model,” *Journal of Intelligent Transportation Systems*, vol. 23, no. 6, pp. 605–616, 2019, <https://www.tandfonline.com/doi/full/10.1080/15472450.2019.1583965>.
- [27] S. C. Sharma, P. Lingras, F. Xu, and G. X. Liu, “Neural networks as alternative to traditional factor approach of annual average daily traffic estimation from traffic counts,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1660, no. 1, pp. 24–31, 1999.
- [28] M. G. Karlaftis and E. I. Vlahogianni, “Statistical methods versus neural networks in transportation research: differences, similarities and some insights,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 3, pp. 387–399, 2011.
- [29] W. Chen, J. An, L. Renfa et al., “A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial-temporal data features,” *Future Generation Computer Systems*, vol. 89, 2018.
- [30] Z. Khan, S. M. Khan, K. Dey, and M. Chowdhury, “Development and evaluation of recurrent neural network-based models for hourly traffic volume and annual average daily traffic prediction,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 7, pp. 489–503, 2019.
- [31] M. Gastaldi, G. Gecchele, and R. Rossi, “Estimation of annual average daily traffic from one-week traffic counts. a combined ann-fuzzy approach,” *Transportation Research Part C: Emerging Technologies*, vol. 47, no. 1, pp. 86–99, 2014.
- [32] P. Kanestørn, “Traffic flow forecasting with deep learning. m.s. thesis ntnu, 2017,” 2017, https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2445851/14739_FULLTEXT.pdf Master thesis.
- [33] S. Zahedian, P. Sekuła, A. Nohekhan, and Z. Vander Laan, “Estimating hourly traffic volumes using artificial neural network with additional inputs from automatic traffic recorders,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 3, pp. 272–282, 2020.
- [34] A. Nohekhan, S. Zahedian, and A. Haghani, “A deep learning model for off-ramp hourly traffic volume estimation,” *Transportation Research Record*, vol. 2675, no. 7, 2021.
- [35] X. Cao, X. Zhu, Z. Tian, J. Chen, D. Wu, and W. Du, “A knowledge-transfer-based learning framework for airspace operation complexity evaluation,” *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 61–81, 2018, <https://www.sciencedirect.com/science/article/pii/S0968090X18309860>.
- [36] H. Dong, J. Man, L. Jia, X. Wang, Y. Qin, and K. Liu, “Traffic speed estimation using mobile phone location data based on longest common subsequence,” in *Proceedings of the IEEE Conference on Intelligent Transportation Systems, ITSC*, pp. 2819–2824, Maui, HI, USA, November 2018.
- [37] Z. Zhang, M. Li, X. Lin, and Y. Wang, “Network-wide traffic flow estimation with insufficient volume detection and crowdsourcing data,” *Transportation Research Part C: Emerging Technologies*, vol. 121, p. 102870, 2020.
- [38] Y. Zhao, H. Zhang, L. An, and Q. Liu, “Improving the approaches of traffic demand forecasting in the big data era,” *Cities*, vol. 82, pp. 19–26, 2018, <https://www.sciencedirect.com/science/article/pii/S0264275117315081>.
- [39] K. Andrade, S. Kagaya, K. Uchida, A. Dantas, and A. Nicholson, “Investigating the temporal transferability of transport modal choice models: an approach based on gis data base,” *Proceedings of the Eastern Asia Society for Transportation Studies*, vol. 7, p. 80, 2007.
- [40] S. Sikder, A. R. Pinjari, S. Srinivasan, and R. Nowrouzian, “Spatial transferability of travel forecasting models: a review and synthesis,” *International Journal of Advances in Engineering Sciences and Applied Mathematics*, vol. 5, no. 2-3, pp. 104–128, 2013.
- [41] H. Gunn, “Spatial and temporal transferability of relationships between travel demand, trip cost and travel time,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 37, no. 2-3, pp. 163–189, 2001.
- [42] J. Fox and S. Hess, “Review of evidence for temporal transferability of mode-destination models,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2175, no. 1, pp. 74–83, 2010, <https://journals.sagepub.com/doi/10.3141/2175-09>.
- [43] J. Fox, A. Daly, S. Hess, and E. Miller, “Temporal transferability of models of mode-destination choice for the Greater Toronto and Hamilton Area,” *Journal of Transport and Land Use*, vol. 7, no. 2, pp. 41–62, 2014.
- [44] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” 2018, <https://arxiv.org/abs/1808.01974>.
- [45] N. T. Ratrouf and U. Gazder, “Factors affecting performance of parametric and non-parametric models for daily traffic forecasting,” *Procedia Computer Science*, vol. 32, pp. 285–292, 2014, <https://www.sciencedirect.com/science/article/pii/S1877050914006267>.
- [46] T. Ma, Z. Zhou, and C. Antoniou, “Dynamic factor model for network traffic state forecast,” *Transportation Research Part B: Methodological*, vol. 118, no. C, pp. 281–317, 2018, <https://ideas.repec.org/a/eee/transb/v118y2018icp281-317.html>.
- [47] RITIS, “The regional integrated transportation information system (RITIS),” <https://www.cattlab.umd.edu/?portfolio=ritis>, 2017.
- [48] Iowa State University, 2018, <https://mesonet.agron.iastate.edu/>.
- [49] D. Schrank, B. Eisele, T. Lomax, and J. Bak, “Appendix A: methodology for the 2015 Urban Mobility Scorecard,” Texas Transportation Institute, Texas A&M University, 2015, <https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-scorecard-2015-appx-a.pdf> Technical Report.

- [50] HPMS, 2017, <https://www.fhwa.dot.gov/policyinformation/hpms.cfm>.
- [51] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," <http://arxiv.org/abs/1412.6980>, 2014.
- [52] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, <https://arxiv.org/abs/1207.0580>.