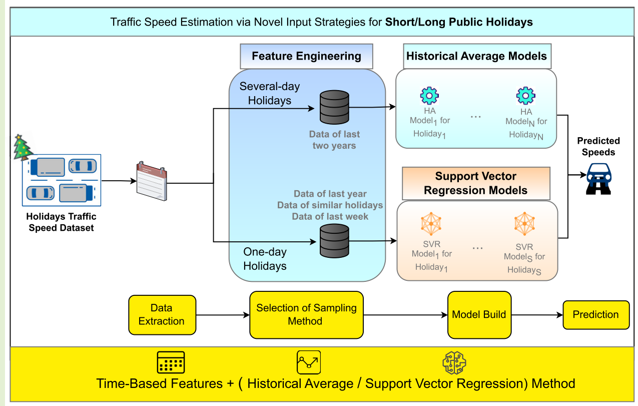
Traffic Characteristics of Short and Long Public Holidays: A Hybrid Holiday-Oriented Speed Prediction Approach via Feature Engineering

Irem Atilgan, H. Irem Turkmen, and M. Amac Guvensan, *Member, IEEE*

***Abstract*—Special holidays differ from regular week(end) days in terms of traffic characteristics due to spatio-temporal bursts within the city. This results in the time spent in traffic during holidays becoming too unpredictable for many people. In this study, it was revealed that public holidays contribute significantly to the overall traffic speed estimation problem and therefore should be handled separately. Contrary to other studies in the literature, in order to achieve successful results in all long and short holidays, a hybrid holiday-oriented approach that combines the support vector regression (SVR) algorithm and historical average (HA) method is proposed. We additionally feed the proposed model with three novel feature engineering strategies in order to make the system learn similar holidays' characteristics. To ensure the system's effectiveness across a broad geographic scope, training and testing were conducted on 441 road segments located in Istanbul, Turkey. The results demonstrate that the proposed method could achieve up to 29% improvement in terms of mean absolute percentage error (MAPE) values for holiday times over 441 different locations. Moreover, with the help of our novel approach, long-term speed estimation models exceed their limits and we could end up with an average minimization of 2% in terms of mean absolute error (MAE) and MAPE traffic speed prediction error over the year.**



***Index Terms*—Feature engineering, historical average (HA), long-short term memory (LSTM), short/long holiday characteristics, support vector regression (SVR), traffic speed prediction.**

I. Introduction

TODAY, people living in metropolitan cities and traveling without using rail systems, spend majority of their time stuck in traffic. Istanbul has the world's heaviest traffic with a congestion rate of 62%, which implies a commute that would take 30 min in the absence of traffic would take an additional 19 min. Thus, people face a great waste of their valuable resources such as time and money. Following this important problem, traffic characteristics of big cities have been analyzed thoroughly in order to perform a better management of these resources. However, while it is generally possible to predict the hours at which the traffic density will change, many unpredictable situations can lead to unexpected changes in traffic characteristics. Incidents, such as traffic accidents [1], road maintenance works, specially planned events, and holidays can create unexpected congestion in traffic and keep drivers at the vehicle for long hours.

In fact, public holidays, which are repeated every year, can create dramatic changes in the characteristics of the traffic, as they affect the current routine of workplaces, schools, and many other public institutions and their effects can span from a single day to several days. Throughout these days, the number of urban or intercity trips increases and creates a different pattern from traffic characteristics of a regular workday, which studies analyzing the effect of holidays [2], [3], [4], [5] have already proven. Municipalities, on the other hand, can barely provide quick solutions since they are not prepared for these abnormal volatile situations. As a result, it is concluded that holiday traffic should be handled differently than on regular days.

While there has been a significant amount of research on traffic speed prediction, only a limited number [6] of studies have delved into the impact of holidays and tried to estimate the range and severity of such effect. Studies that extract trend

Manuscript received 23 February 2023; revised 23 August 2023; accepted 23 August 2023. Date of publication 11 September 2023; date of current version 16 October 2023. This work was supported by the Turkish Scientific and Technological Research Council of Turkey (TUBITAK) under Grant TUBITAK1001-120E357. The associate editor coordinating the review of this article and approving it for publication was Dr. Ravibabu Mulaveesala. *(Corresponding author: M. Amac Guvensan.)*

The authors are with the Intelligent Systems Laboratory, Department of Computer Engineering, Yildiz Technical University, 34220 Istanbul, Turkey (e-mail: amac@yildiz.edu.tr).

Digital Object Identifier 10.1109/JSEN.2023.3312189

and seasonality information from traffic data using discrete Fourier transform (DFT) [7], construct models by grouping traffic into hourly intervals using the k-means algorithm [8], or aim to capture traffic variance through the fluctuation coefficient method [9] have been reported so far.

Upon reviewing the studies in the literature, three primary drawbacks become evident. First, systems that focused on public holidays tend to be restricted to smaller regions such as highways [7], [8], [9]. Another limitation is the concentration on a single particular holiday [8], [9]. These factors hinder the widespread usage of the developed systems. The third limitation involves the inability to generate high-resolution forecasts [7], which is crucial for the development of reliable prediction systems. In this study, we proposed a novel statistical and nonparametric prediction approach which combines support vector regression (SVR) and historical average (HA) methods considering the characteristics of different types of holidays for 441 segments of a metropolitan city. In this context, the contributions of this study are summarized below.

1) It is proven that handling the problem of traffic prediction on holidays as a separate issue increases even the performance of state-of-the-art traffic speed estimation methods such as SVR and long short-term memory (LSTM) that have already proven their successes in this domain.
2) To the best of our knowledge, this study is the first one that introduces separate traffic speed estimation models for short and long public holidays to improve prediction success. We present a hybrid model that can predict long-term traffic speed for short/long holidays by combining the SVR and HA methods which are nonparametric and statistical approaches, respectively.
3) We propose three novice feature engineering strategies using the information obtained from the similarity between holidays, last year's holiday characteristic, and last week's data, in order to feed the holiday-oriented prediction models.

The rest of this article is organized as follows. In Section II, the state-of-the-art algorithms for traffic speed estimation are presented and discussed regarding how these techniques are arranged for holidays and planned special events. Moreover, the differences between normal days and special days are mentioned. In Section III, it is explained to what extent holidays can affect traffic speed estimation and why estimating traffic speed on holidays should be considered as a separate problem. Section IV gives our methodology for traffic speed estimation at short/long public national holidays. Finally, in Sections V and VI, the results of the proposed hybrid method are demonstrated and the open issues are discussed for further studies.

II. Related Work

The literature review shows us that related studies fell under three headings: Studies on traffic speed estimation, studies on the impact of special days (holidays and planned special events) and studies on the estimation of traffic speed on special days.

A. Traffic Speed Estimation

Traffic speed estimation is an ongoing problem which hosts different approaches and cases. Considering the time frame which the solution covers, literature works mainly on short-term traffic speed estimation [10], [11], [12], [13], [14], [15] and recently on long-term predictions [16], [17]. On the other hand, considering the methodologies infers that studies on the prediction of traffic speed have followed two main approaches over the years [18], [19]: the model-based (analytical) approach and the data-driven approach. In the years when big data and computer power were lacking, model-based approach, including traffic-velocity model, cell transmission model, and store/forward model was frequently utilized [20], [21], [22]. However, these methods have difficulty in explaining a multidirectional, high variation, and abruptly changing area such as traffic. Additionally, these methods can provide a solution only for a limited region and can be easily affected by missing data. Therefore, as stated in [18], the data-driven approach has gained importance in recent years. The data-driven approach, which deals with the statistical variation trends of historical data, can offer more flexible and universal solutions than the model-based approach. Today, since it is easier to access high-dimensional, large numbers of data, this approach is on trend among traffic speed estimation methodologies. Models based on this approach are divided into two as *parametric* and *nonparametric* models [7], [18].

Parametric models predict the traffic flow by manual calibration on the parameters of the regression functions. Since it does not require much computational power, these models including autoregressive integrated moving average (ARIMA) [23] and its variations such as Kohonen ARIMA [24] and Seasonal ARIMA [25] are utilized especially for short-term predictions. Despite the benefits, their accuracy can be limited to a point where they do not adequately handle the nonlinearity and uncertain nature of traffic data. Hence, nonparametric models have been proposed to overcome this disadvantage of parametric models. Nonparametric models including artificial neural networks (ANNs) [12], [13], k-nearest neighbor (k-NN) algorithm [14], [15], decision trees (DT) [26], and SVR [7], [27] have recently emerged to improve the speed prediction accuracy. For example, in [7], it is stated that the presented SVR and ANN model could give approximately 5% better accuracy when compared to ARIMA's prediction accuracy.

The structure of deep learning, i.e., one of the nonparametric models, enabled nonlinear problems to be modeled as hierarchical and distributed features, which made deep learning models such as LSTM [11], [28] and convolutional neural network (CNN) [10], [29] preferable for solving data that contains many nonlinearities such as traffic data. For instance, utilizing LSTM helped researchers to comprehend nonlinear patterns in traffic flow, since this model is more resistant to peaks, and is successful with sequential data. Hsueh and Yang [11], exploit LSTM for short-term traffic speed estimation on a highway in Taiwan, achieving much lower mean absolute error (MAE) and mean absolute percentage error (MAPE) values compared to ARIMA, SVR, and RNN models. In another study [28], making predictions for more unusual traffic conditions, the authors developed a short-term traffic speed prediction model using LSTM for evacuation moments based on Hurricane Irma in Florida, USA. The study points out that other nonparametric approaches, such as k-NN and analytical neural network, can respond well to regular traffic demands, but, unlike LSTM, can perform poorly at recognizing sharp patterns in sudden changes.

In recent studies, with the increase in the richness of the data obtained, spatio-temporal information has also started to be used in nonparametric models [28], [29], [30]. This information is usually given as input to LSTM, CNN, and graph-based models and traffic flow predictions can be made accordingly. In [29], daily and weekly traffic speed features in the region are extracted using the CNN model. Then, the study merged the feature vector obtained from other spatio-temporal features and transforms the features to a final vector by employing LSTM. On the other hand, the authors take into account the R^2 score, which is a coefficient of determination in statistics, as a performance metric, and with a value of 0.651 for a 15-min estimation, it outperformed methods such as SVR, multilayered perceptron (MLP), and random forest (RF).

Studies in this section generally focused on increasing the overall success and tried to estimate the traffic speeds of rush hours on regular days, when the traffic was very congested and the unpredictability was increased. Although this strategy is beneficial, it ignores the effect of events such as concerts, sports competitions, and holidays, which causes the success of these studies to be limited at some point.

B. Traffic Speed Estimation for Special Days (Holidays and Planned Special Events)

The characteristics of the traffic can be critically affected by some events, such as weather conditions, roadworks, traffic accidents, rush hours, and special days. These factors are often referred to "bursts" in [7] and [18], as they can abruptly change the traffic speed pattern. Because these bursts could usually occur randomly and their degree of effect can vary widely, they pose a great challenge for prediction models. Although special days have been known previously, they include structural and nonstructural features such as the degree of impact, direction, time, and duration of impact which are not known in detail. Therefore, it is regarded as a burst.

Since planned special events including sporting events, concerts, festivals, and conventions occur at permanent multiuse venues (e.g., arenas, stadiums, racetracks, fairgrounds, amphitheaters, convention centers, etc.), they can also cause traffic jams as well as holidays. In [5], it was mentioned that drivers waited in traffic for about five hours for Rihanna's music concert in Johannesburg, South Africa. Even though special events and holidays are planned, their degree of impact is uncertain and the duration of the impact may vary. Special events can affect traffic for a few hours and holidays can affect for days. Hence, it can be said that traffic speed estimations on holidays are counted as long-term estimations, when they are compared to planned special events, which in case, require short-term estimations. Last but not least, the size of the area they affect can vary; Holidays can affect traffic across the city or country, while special events often burst in a limited venue and on road segments close to itself.

Holidays are a celebration of governmental, religious, or personal events and important memories for a country. They follow the calendar year and can be annually scheduled dates or a differently scheduled dates throughout the calendar. People usually make vacation plans, visit their relatives, meet their friends, have a sightseeing tour at holidays. Thus, the traffic forms different flow characteristics compared to ordinary week(end) days. Jun [2], Liu and Sharma [3], Datla et al. [4], Kwoczek et al. [5], and Luo et al. [7] analyzed the impact of holidays on traffic flow. When the effect of Thanksgiving on the traffic of the United States of America was examined in [7], it was found that the traffic speeds of the two days before and after the holiday were significantly different from normal weekday speeds. In another study conducted in Jiangsu, China [5], it has been proven that the first and the last day of the holiday show a different sequence from the general pattern of the traffic.

There are some underlying reasons for the unsuccessful results in estimating traffic speed both on holidays and special events.

1) It is difficult to catch/detect the beginning and the end of the anomaly experienced as a result of the events. Irregularities [31] may occur before and after the event, and the time and duration of these irregularities vary according to the characteristics of the special day.
2) In the datasets provided on traffic, special days take place at a very low rate. Chen et al. [32] states that the traffic burst data constitutes 2% of the total data. The scarcity of it creates the misconception that it has little effect on the overall error of the prediction model. Therefore, the model may ignore burst data when inferring patterns.

In general, studies on traffic speed estimation follow one of three ways regarding special days: ignoring/extracting anomalies from data, developing adaptive models that can adjust itself during these time periods [18], [33], [34], or developing appropriate methods by targeting special days only. Ren et al. [18] formulated the traffic estimation problem using the Markov decision process (MDP) and present traffic burst sensitive model (TBSM). By using the short-term trend information obtained from the data and the observed speed, the degree of effects was weighted, and the balancing and generalization of these weights was provided by the deep deterministic policy gradient (DDPG) network containing LSTM. The model created as a result of the study can make a short-term traffic speed estimation of 2–6 min in real-time. Thus, a flexible system for bursts has been developed.

In another study that makes traffic speed estimation for the planned event [5], a method which is named as specific inbound prediction approach (CIPA) was created using k-NN algorithm with traffic speed data and event type. The study noted that one event produces two ripple effects; The first wave was defined as 2 h before the event and the second wave defined as 2 h after the event. The presented model estimates the traffic flow for the second wave using the data of the first wave and since it requires first wave data, the solution space is limited. The training and testing of the model were provided using seven months of data from Cologne, Germany.

There are also some studies that exploit the components that make up the speed of traffic and develop separate methods for each. For instance, [9] divides traffic speed into two sides ordinary and fluctuant. Ordinary flow is estimated by the LSTM model, and fluctuant flow is estimated by the "fluctuation coefficient method," which is also used in some studies. The authors believe that this method can overcome the constraints of historical data often encountered in holiday data. They also state that the proposed model outperforms an ordinary LSTM model. Likewise, one of the studies conducted in recent years, [7] presents a hybrid methodology for traffic speed estimation on holidays using DFT and SVR. First, traffic flow data were decomposed using DFT and trend/residual information were extracted. Then, trend estimation was made using extreme extrapolation. Residual estimation, on the other hand, is made with the help of the SVR model. The final estimation is obtained by combining the estimated trend and residual values. When the method was tested for Tomb-Sweeping Day and National Day celebrated in China, MAPE values were calculated as 12.11% and 9.62%, respectively, and much more successful results were obtained than ARIMA and empirical mode decomposition (EMD)-SVR methods. In addition, the authors found that the method predicted 15.04% more accurate on long holidays than on short holidays, indicating that the method works better with stochastic traffic flow data.

The aforementioned studies generally take special days on a day-by-day basis, and accordingly make predictions for the whole day. However, the hours of the day can also show different speed patterns. Following this fact, [8] presented the segment estimation algorithm (SPA) which groups velocity data using k-means algorithm and builds hourly linear regression models according to these groups. While evaluating the results of the predictions made for Columbus Day in USA, this study also preferred LSTM as the baseline method. When the results were examined, the proposed SPA method proved to be more successful with a MAPE value of 1.14% than the LSTM model with a MAPE value of 1.52%.

Although the magnitude of the effect of these dates has been proven many times in several studies [2], [3], [4], [5], it can be said that special days (planned special events and holidays) are generally ignored, and the high rate of estimation failures made on these dates was barely considered for traffic estimation models. Moreover, the scope of these works is limited to small regions (highway, several road segments) or a single holiday. All these takeaways show that the subject has not been dealt with comprehensively yet. Based on the foregoing, this study proposes a methodology for estimating the traffic speed of long and short-term holidays in Turkey, with the aim of analyzing the effect of holidays on traffic speeds, minimizing the MAE and MAPE values for traffic flow estimation on holidays, and contributing to future research in this field.

TABLE I
HOLIDAYS CELEBRATED IN TURKEY

Holiday Name	Date	Holiday Type
New Year's Day	January 1st	One-Day
National Sovereignty and Children's Day	April 23rd	One-Day
Labor Day	May 1st	One-Day
Atatürk Commemoration, Youth and Sports Day	May 19th	One-Day
Democracy and National Unity Day	July 15th	One-Day
Victory Day	August 30th	One-Day
Republic Day	October 29th	One-Day
Eid al-Fitr	Varying	Several-Day
Eid-al-Adha	Varying	Several-Day

III. CHARACTERISTICS OF NATIONAL HOLIDAYS

Official holidays in Turkey can be divided into two: one-day and several-day holidays. The names, the dates, and the types (one-day, several-day) of these national holidays are given in Table I. However, Eid-al-Fitr and Eid-al-Adha are celebrated 11–12 days earlier each year, as they shift according to the Hijri calendar, so it is assigned as "varying" in the Date column. Eid-al-Fitr lasts for a total of four days while Eid-al-Adha lasts for five days.

As emphasized in the Introduction, traffic on public holidays has different characteristics than regular days. In order to reveal the importance of the difference, first, traffic speeds at holidays obtained from 441 road segments in Istanbul were compared with the average of the speeds four weeks before these holidays. The results for Victory Day, Republic Day, first day of Eid al-Fitr, and second day of Eid al-Adha are shown in Fig. 1. The charts placed in Fig. 1 reveals that the pattern of traffic speeds on holidays and normal days differs significantly. Speeds that are similar to each other only during night hours generally begin to diverge from 07:00 in the morning, and this gap can only narrow after about 21:00 in the evening. One of the reasons for this difference is that there are often known "rush hours," which reduce the traffic speed considerably on regular days. As there are no specific rush hours during the holidays, there is no major drop in traffic speed in the evening hours.

MAPE is one of the most common performance metric to measure forecast accuracy. It ensures that relative errors do not depend on the scale of the dependent variable. Thus, it allows comparison of forecast accuracy of differently scaled time series data. When the speed prediction was made for the holidays using the HA [35] method (which will be explained in Section IV-A) exploiting the average speed of four previous weeks, MAPE was calculated as 18.38% which can be interpreted as quite high compared to the results obtained in the literature. In addition, an LSTM model, which was trained using the same data source and estimated traffic speed for all days of the year, was also examined, and compared with the actual speed data of holidays in 441 road segments. The estimation results obtained from the LSTM model for Victory Day, Republic Day, first day of Eid al-Fitr, and second day of Eid al-Adha are shown in Fig. 2. Analyzing these graphs shows that the LSTM model which is trained with all the data throughout the year generally makes erroneous predictions for holidays. The most important reason for such inaccurate results is that the number of holidays is less by far compared to regular days. As a consequence, LSTM produces a generalized model that mainly considers regular days. The greatest effect

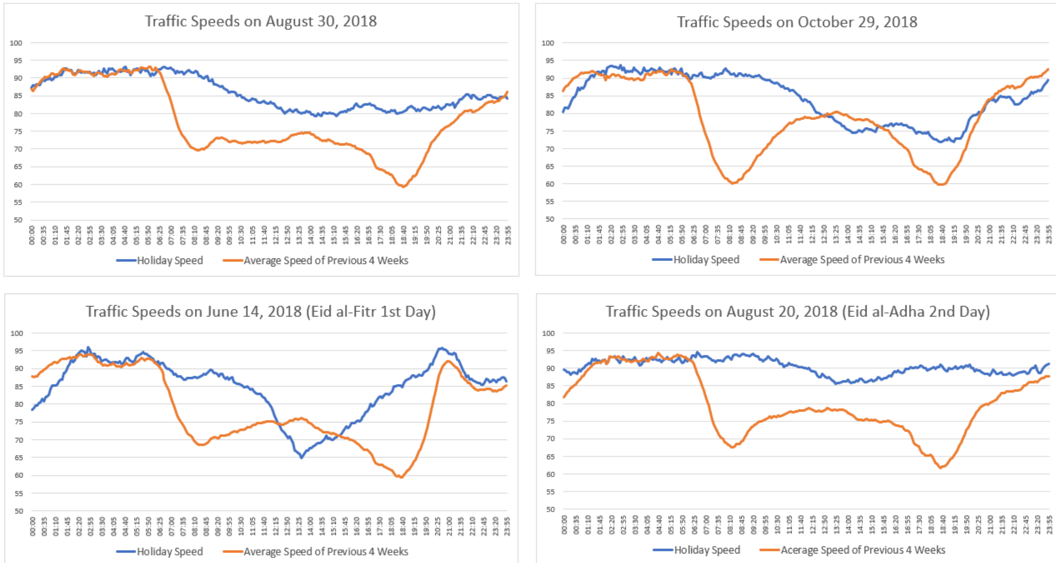


Fig. 1. Comparison of average traffic speeds between holidays and previous four weeks.

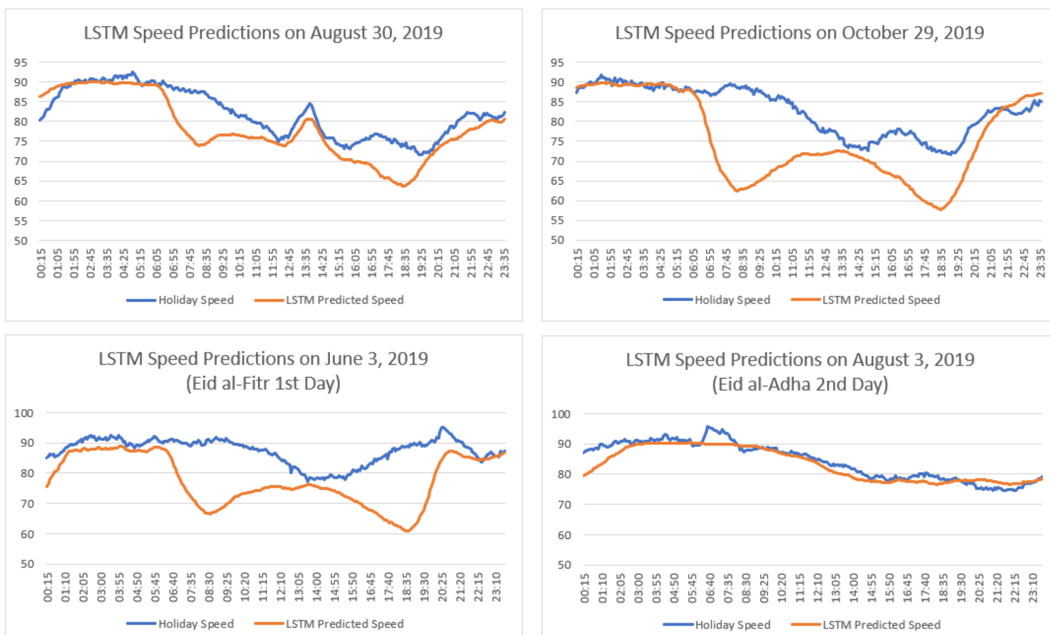


Fig. 2. Comparison of average traffic speeds between holidays and LSTM model predictions.

of this generalization could be seen in rush hours. The model usually predicts low traffic speeds during these hours since the data often follows such a pattern. However, the model only predicts second day of Eid al-Adha correctly, which can be explained by third of August is a weekend day where traffic speeds might show a similar characteristic to a given holiday.

Consequently, our analysis showed us that, holidays differ from nonholiday (regular) days in their characteristics and should therefore be treated separately. Moreover, models created for general traffic speed prediction on the other hand may be insufficient in predicting speed during holidays as they principally generalize according to the majority of data. On the contrary, Figs. 1 and 2 show that holidays, even though they occur on different months or days of the year, exhibit similar patterns of traffic speeds. It should also be noted that holidays in 2018 and 2019 shows consistent patterns in terms of speed. As a result, both of these findings encouraged us to explore and utilize the speed characteristics of holidays as well as the parallels between them for the same and/or previous year.

IV. Methodology

Parametric/nonparametric methods are frequently used in traffic speed estimation. Models created by parametric methods are directly affected by certain parameters whereas nonparametric models are composed of the relationships between the variables and the inference of the parameters. In this study, the HA method as the parametric method, the SVR and LSTM methods of the nonparametric methods will be mentioned and their performances will be discussed to prove the success of our hybrid approach.

A. HA Method

The HA method is a parametric method that can give efficient results as well as being quite simple to apply in long-term speed forecasting. Keskin and Güvensan [35] proposed this method as "mean filtering estimation" while using the speed values of five previous weeks of predicted time. At the end of the study, the authors determined the results to be promising for long-term predictions and highlighted the method's strength in its simplicity. When calculating the average traffic speed s for the date-time T of the year D, (1) is applied where Y indicates the number of years prior to the current date

$$\bar{s}_{D,T} = \frac{\sum_{i=1}^{Y} s_{D-i,T}}{Y}. \tag{1}$$

In this study, this method was preferred when estimating the traffic speed for Eid al-Fitr and Eid al-Adha, which are holidays that last for several days. When these holidays are examined, it is observed that, unlike other holidays, they reveal a very similar pattern for three years. Therefore, their stability enables them to be predictable and makes the HA method a suitable method for the problem.

B. Support Vector Regression

SVM used for regression problems are known as SVR models. SVR provides flexibility in defining how much error is acceptable in the model and finds a suitable line or hyperplane to fit the data in that direction. The most important advantages of SVR are that it can work in high-dimensional spaces and can show success when the number of dimensions is more than the number of samples.

For a given training set $K = \{(x_i, y_i) \mid 1 \leq i \leq m, x_i \in \mathbb{R}^n, y_i \in \mathbb{R}^n\}$, where x_i is an input variable, y_i is the corresponding output value, and m is the size of the training data. With the nonlinear $\phi(.)$ function, x_i is mapped to the high-dimensional space. The resulting feature vector is expressed by the following equation:

$$f(x) = w^T \phi(x) + b, w \in \mathbb{R}^n, \quad b \in \mathbb{R} \tag{2}$$

where w represents the weight vector and b stands for bias. The SVR tries to minimize the following objective function, represented by the following equation:

$$\min \left[\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i) \right]$$
$$\text{s.t.} \begin{cases} (wx_i + b) - y_i \leq \varepsilon + \xi_i \\ y_i - (wx_i + b) \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \ \ \xi_i^* \geq 0, \ \ \varepsilon \geq 0 \end{cases} \tag{3}$$

where $1/2\left\|w^2\right\|$ represents the complexity of the decision function, C the penalty coefficient, ξ_i the upper training error, ξ_i^* the lower training error, and ε the insensitive factor.

Examining the traffic speed data of the last three years in Istanbul showed that the pattern of one-day holidays does not change much over the years, but can be affected by factors such as the day of the week, the days before and after it. For this reason, a flexible model that can learn the pattern is needed. SVR was preferred considering its high profile for traffic speed estimation problems and the advantages mentioned above [7], [27].

C. Long Short-Term Memory

LSTM is a type of recurrent neural networks (RNNs) that is frequently used in deep learning and is effective in learning order dependencies for sequence prediction. It is more resistant to time delays than RNNs and solves the existing exploding gradient and vanishing gradient problem in RNN. In recent years, it has become quite popular because it can draw very successful results from natural language processing (machine translation, speech recognition, etc.) problem and data given as time series. Due to the proven success of LSTM with nonlinear traffic speed characteristics [11], [16], [29], an LSTM model was also established for general long-term traffic speed estimation and compared with the hybrid approach.

D. Hybrid Approach

By trying to take the strengths of the above-mentioned HA and SVR methods on different holidays, a hybrid approach, shown in Fig. 3, was followed and the two proposed models were combined. For a given holiday whose traffic speed is to be predicted, the system follows one of two different strategies, depending on the holiday type. Traffic speeds on several-day (long) holidays are predicted by using HA models, whereas SVR models are used for one-day holidays. It is important to note that different HA models are generated for each road segment and holiday, whereas separate SVR models are generated for each holiday only, since the data space in which holidays occur is too limited to create segment-specific models.

While predicting the traffic flow using SVR and HA methods, determining the data samples to feed the relevant models has a great effect on the results. Therefore, in this study, three novel feature engineering strategies are proposed for short and long holidays.

1) *Last Year's Data:* Traffic speed data of a year before the holiday.
2) *Similar Holidays:* Traffic speed data of the holiday which is most similar to the current holiday in the previous year.
3) *Last Week's Data:* Traffic speed data of one week before the current holiday.

These strategies are applied while performing estimation by SVR and HA methods separately and the most effective ones are favored in the proposed hybrid approach. While detecting similar holidays, the average speeds of each holiday for all road segments are vectorized, and then the Euclidean distance to the vectors of other holidays is calculated. Finally, the holiday with minimum distance is selected for training stage. The closest holiday to the holiday vector h_j is found by (4), where l is the number of holidays and the vectors h are the

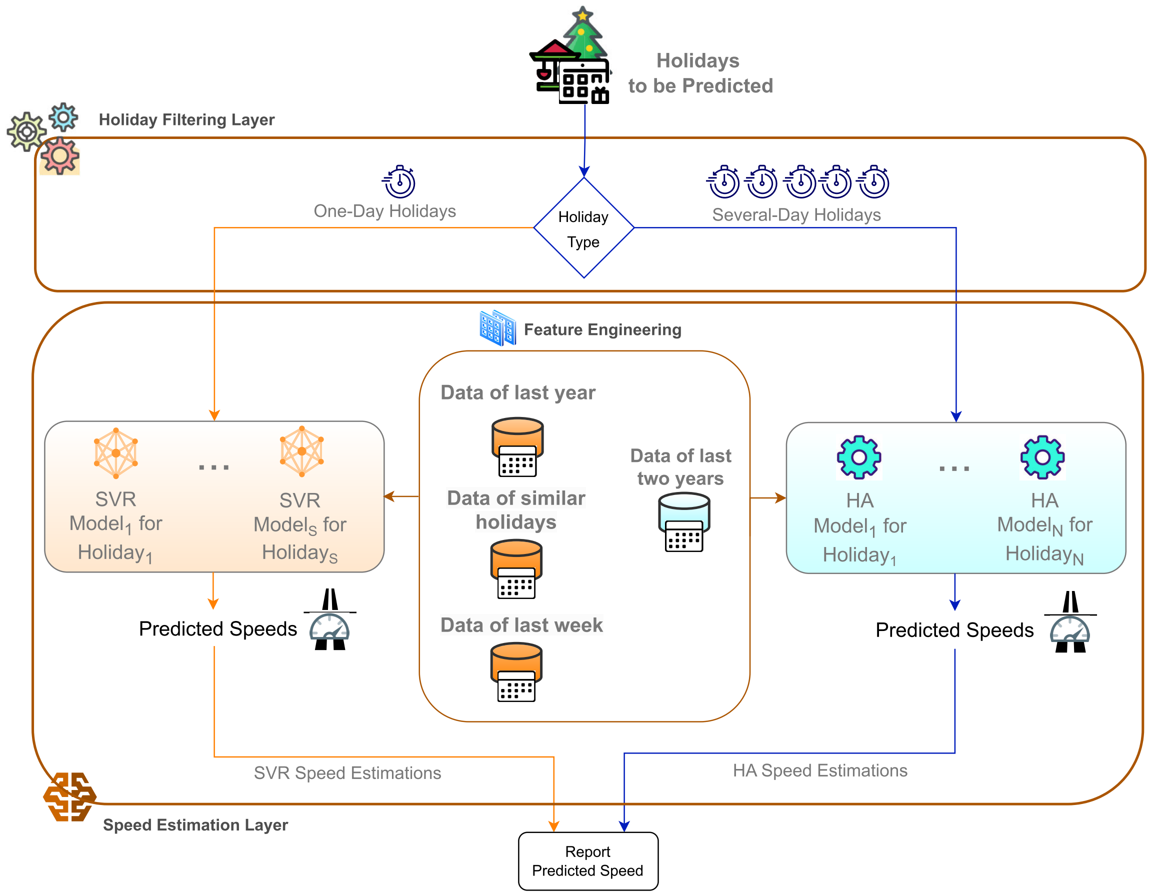


Fig. 3. Hybrid holiday-oriented speed prediction model for short/long public holidays.

average speeds of road segments, aggregated into one day

$$\underset{i,j\in\{1,2,\ldots,l\},i\neq j}{\operatorname{argmin}} \left\|h_j - h_i\right\|^2 . \tag{4}$$

V. Experimental Setup

The traffic speed dataset obtained from the Traffic Branch Directorate of Istanbul Metropolitan Municipality includes traffic flow speed data collected from 441 different points in Istanbul for the years 2017, 2018, and 2019 with a measurement frequency of 1 min. For each measurement, the data involves *segment id*, *direction of the segment*, and *time* information. Data frequency is decreased to 5-min in order to reduce data gaps(missing values) within the records. To demonstrate the effectiveness of the proposed hybrid method, the SVR, HA, and LSTM methods were all exploited to estimate traffic speeds both at single-day holidays and several-day holidays. We first evaluated the performance of our similar holiday strategy. We utilized the data of 2017 and 2018 for determining the holiday with similar traffic characteristics. Then, the data of 2018 is used to train SVR models whereas holidays in 2019 were considered for test results. Setting up the SVR model, RBF (radial basis function) was chosen as the kernel function. The kernel coefficient gamma value, the tolerance value, the epsilon value, the regularization parameter C value was determined as 0.01, 1e-4, 0.1, and 12, respectively. For the HA method, holiday traffic speed data for 2017 and 2018 were utilized to estimate the traffic speed in 2019. While estimating the several-day holidays with the HA method, the average of the traffic speeds of the previous two years is taken into consideration as given in (1) and assigned as the traffic speed to the current date.

Two performance metrics, MAPE and MAE, were utilized in order to show the success ratio of the proposed methods. Both metrics were calculated by (5) and (6), respectively where n represents the sample size while y_i and $\bar{y}_i$ symbolize the actual and predicted values

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \bar{y}_i|}{y_i} \tag{5}$$

$$\text{MAE} \;= \frac{1}{n}\sum |y_i - \bar{y}_i| . \tag{6}$$

A. Experimental Results

In this section, the test results of several experiments regarding the proposed methods and data strategies introduced in Section IV were thoroughly discussed including nine public holidays celebrated in Turkey. In the first step, three different feature engineering strategies (Last Year, Similar Holidays, Last Week) were tested on the SVR models created for each holiday. The success rates of the traffic speed predictions for 2019 were given in Table II. The letter O in parentheses in the table denotes one-day holidays, whereas the letter

TABLE II

MAE and MAPE Results for Varying Length Holidays, Different Feature Engineering and Traffic Estimation Methods

Prediction Method	Method and Feature Engineering Strategy	MAE	MAPE	MAE (S)	MAPE (S)	MAE (O)	MAPE (O)
SVR	Last Week	10.90	17.35%	11.56	18.71%	9.90	15.26%
	Similar	9.77	15.82%	10.03	16.30%	9.48	15.26%
	Last Year	8.52	13.42%	8.63	13.23%	8.35	13.71%
	Last Year & Similar	8.18	12.90%	8.32	12.74%	7.98	13.14%
	Last Year & Last Week	8.22	12.80%	8.56	13.05%	7.69	12.42%
	Last Year & Similar & Last Week	**7.99**	**12.49%**	8.33	12.70%	**7.48**	**12.16%**
HA	Similar 2 Holidays	9.62	15.48%	10.17	16.55%	8.88	14.08%
	Last 2 Years & Similar 2 Holidays	7.84	12.51%	7.77	12.37%	7.94	12.69%
	Last 2 Years	**7.93**	**12.04%**	**7.43**	**11.05%**	8.58	13.34%
Hybrid	SVR(Last Year & Similar & Last Week) and HA(Last 2 Years)	**7.46**	**11.60%**	**7.43**	**11.05%**	7.51	12.34%

S represents several-day holidays. Analyzing those results shows that all three sampling techniques partially contributed to the model success. Especially traffic speed data of last year has a high impact on model performance. However, as expected, using last week's traffic speed data alone yielded the most unsuccessful results. The inadequacy of utilizing the speed data of the previous week proved that traffic characteristics at holidays are quite different from regular days.

The HA method was also evaluated with the same sampling strategies. Table II shows the results of the averaging method for different holidays. It is revealed that averaging the speeds of the last two years and averaging the speeds of the two most similar holidays yields the best result. The use of data for only two years is especially effective for several-day holidays, while the use of similar holiday data has been observed to have a positive impact on one-day holidays. The comparison between SVR models using all sampling techniques and the results obtained from the HA method shows us that both methods have advantages for certain cases. SVR models reveal lower MAE and MAPE values for one-day holidays compared to the HA method, whereas HA models achieve much more successful results for several-day holidays. As a result, we propose a hybrid approach to combine these methods where SVR models will be used to make predictions for one-day holidays, whereas HA models will be providing estimation for several-day holidays. The results of the proposed hybrid approach are given in the last row of Table II. Table II clearly shows that the hybrid approach has the lowest MAE and MAPE values among other approaches. Exploiting the strengths of the HA and SVR methods yielded the best results for both one-day and several-day holidays.

Additionally, the performance of the proposed method on holidays according to the general traffic speed models was also investigated. For this comparison, the LSTM model, which has been frequently used [29], [30], [36], [37] for traffic speed prediction in recent years, was selected and trained with traffic speed data of 441 road segments in Istanbul. The results obtained from the LSTM method for predicting traffic flow on holidays and the results of the hybrid method are given in Table III. Overall, MAE and MAPE values which are obtained

TABLE III

Comparison of MAE and MAPE Values of the Holiday-Oriented Hybrid Approach Against Traditional Whole-Year Oriented LSTM Model

Method Name	MAE	MAPE	MAE (S)	MAPE (S)	MAE (O)	MAPE (O)
Hybrid	7.46	11.60%	7.43	11.05%	7.51	12.34%
LSTM	11.42	16.39%	12.52	19.02%	10.05	13.43%

TABLE IV

Performance Comparison of the Proposed Methods in Terms of MAPE

Method	Baseline Method	Baseline Method and Hybrid Approach	Rate of Change (%)
HA	22.03	21.76	-1.23
SVR	14.27	14.04	-1.6
LSTM	13.36	13.13	-1.71

predicting both one-day and several-day holidays are given in the first two columns.

It is clear that the proposed hybrid method outperforms the LSTM method by far for holidays. Exploiting the proposed holiday-oriented approach could provide a 34.6% decrease in the MAE value and a 29.2% decrease in the MAPE value for holidays over 441 segments. These results also show that traditionally trained LSTM performed quite poorly, especially during the several-day holidays. One of the most important reasons for this is that the model generalizes to regular days and does not take the traffic characteristics at holidays into account. This fact is clearly shown in Fig. 4 where the predictions of our hybrid model can adapt to the actual traffic speed, while the LSTM make wrong estimations especially at "rush hour"s due to the characteristic on regular days.

In order to express the efficiency of the proposed hybrid method, the prediction results of the whole year which are obtained by using the HA, SVR, and LSTM methods are compared with the results obtained by using the hybrid method for holidays and traditional methods for the remaining days. Tables IV and V show the success of HA, SVR, and LSTM models and how much this success has increased after combining them with the proposed hybrid method. The results showed that utilizing hybrid method for short/long holidays helps reduce the total MAE and MAPE values for all possible prediction models. Considering that the holidays in Turkey last only a total of 16 days, the effect rate of the improvement on

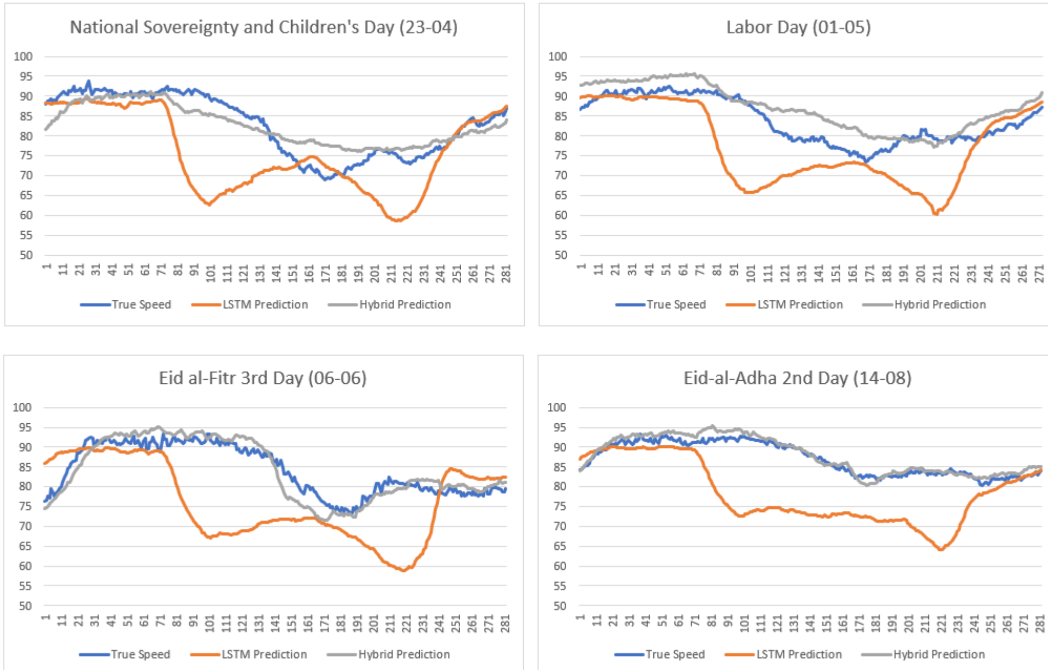


Fig. 4. Comparison of average speeds between true speeds, LSTM predictions and proposed model (Hybrid) predictions on different holidays in 2019.

TABLE V
PERFORMANCE COMPARISON OF THE PROPOSED METHODS IN TERMS OF MAE (%)

Method	Baseline Method	Baseline Method and Hybrid Approach	Rate of Change (%)
HA	11.90	11.72	-1.56
SVR	7.39	7.21	-2.36
LSTM	7.23	7.06	-2.32

these days is dramatically high. It is proven once again that the issue of traffic speed estimation during holidays should be handled separately.

Following the test results, we can safely state that applying a hybrid holiday-oriented approach on one-day holidays and several-day holidays using SVR and HA methods has proven to be an effective solution. The SVR's ability to work well with nonlinear historical data, its flexibility, and its robustness against outliers; combined with the HA method's intelligibility and its ability to capture simple relationships, contributed to the overall success of the approach.

VI. Conclusion and Future Work

The unique traffic characteristic of public holidays encouraged us to design a new methodology for such special days. They are categorized into two groups, one-day and several-day holidays, where the mobility behavior of people is strongly affected by the duration of the public holiday. Following this observation, we designed two different strategies and combined them in one solution to minimize the prediction errors for special holidays. This article introduces a novel holiday-oriented hybrid approach regarding the traffic characteristics of short and long period public holidays by combining the strengths of HA and SVR algorithms. We also integrated three novice feature engineering methods to feed the relevant models. Test results showed that the proposed holiday-oriented method could outperform the traditional whole-year oriented LSTM model up to 29% and 35%, respectively, in terms of MAPE and MAE for public holidays. Additionally, integrating this proposed strategy into baseline methodology helped us to improve the overall performance of 441 segments for seven short holidays and two long holidays in 2018 by up to 1.71% and 2.36% in terms of MAPE and MAE. We also observed that the overall improvement for short holidays are considerably better compared to long holidays, whereas the combination of the information obtained from the similarity between holidays, last year's holiday characteristic, and last week's data gives the most successful MAPE and MAE values.

As future work, we plan to include environmental factors such as weather conditions and road maintenance to improve the overall accuracy. On the other hand, since people could bind public holidays with weekend and/or prolong their vacations by taking annual leave, we believe that days before and after short/long public holidays should be elaborated as well as public holidays.

References

[1] M. Humayun, M. F. Almufareh, and N. Z. Jhanjhi, "Autonomous traffic system for emergency vehicles," *Electronics*, vol. 11, no. 4, p. 510, Feb. 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/4/510

[2] J. Jun, "Understanding the variability of speed distributions under mixed traffic conditions caused by holiday traffic," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 4, pp. 599–610, Aug. 2010.

[3] Z. Liu and S. Sharma, "Statistical investigations of statutory holiday effects on traffic volumes," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1945, no. 1, pp. 40–48, Jan. 2006.

[4] Z. Liu, S. Sharma, and S. Datla, "The necessity of understanding temporal and spatial variations of holiday traffic," *ITE J. Web*, vol. 78, no. 11, pp. 69–74, 2008.

[5] S. Kwoczek, S. D. Martino, and W. Nejdl, "Predicting and visualizing traffic congestion in the presence of planned special events," *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 973–980, Dec. 2014.

[6] W. Zhang, R. Yao, X. Du, and J. Ye, "Hybrid deep spatio-temporal models for traffic flow prediction on holidays and under adverse weather," *IEEE Access*, vol. 9, pp. 157165–157181, 2021.

[7] X. Luo, D. Li, and S. Zhang, "Traffic flow prediction during the holidays based on DFT and SVR," *J. Sensors*, vol. 2019, Jan. 2019, Art. no. 6461450.

[8] Z. Gao, X. Yang, J. Zhang, H. Lu, R. Xu, and W. Diao, "Redundancy-reducing and holiday speed prediction based on highway traffic speed data," *IEEE Access*, vol. 7, pp. 31535–31546, 2019.

[9] G. Lu, J. Li, J. Chen, A. Chen, J. Gu, and R. Pang, "A long-term highway traffic flow prediction method for holiday," in *Advanced Multimedia and Ubiquitous Engineering*. Singapore: Springer, 2019, pp. 153–159.

[10] M. Chen, X. Yu, and Y. Liu, "PCNN: Deep convolutional networks for short-term traffic congestion prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3550–3559, Nov. 2018.

[11] Y.-L. Hsueh and Y.-R. Yang, "A short-term traffic speed prediction model based on LSTM networks," *Int. J. Intell. Transp. Syst. Res.*, vol. 19, no. 3, pp. 510–524, Sep. 2021.

[12] Z. Ma, G. Luo, and D. Huang, "Short term traffic flow prediction based on on-line sequential extreme learning machine," in *Proc. 8th Int. Conf. Adv. Comput. Intell. (ICACI)*, Feb. 2016, pp. 143–149.

[13] H.-F. Yang, T. S. Dillon, and Y. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2371–2381, Oct. 2017.

[14] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 61–78, May 2016.

[15] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "k-nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *J. Transp. Eng.*, vol. 142, no. 6, Jun. 2016, Art. no. 04016018.

[16] M. Méndez, M. G. Merayo, and M. Núñez, "Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model," *Eng. Appl. Artif. Intell.*, vol. 121, May 2023, Art. no. 106041.

[17] S. Reza, M. C. Ferreira, J. J. M. Machado, and J. M. R. S. Tavares, "A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks," *Expert Syst. Appl.*, vol. 202, Sep. 2022, Art. no. 117275.

[18] Y. Ren, H. Jiang, N. Ji, and H. Yu, "TBSM: A traffic burst-sensitive model for short-term prediction under special events," *Knowl.-Based Syst.*, vol. 240, Mar. 2022, Art. no. 108120.

[19] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[20] Q. Wang, L. Li, J. Hu, and B. Zou, "Traffic velocity distributions for different spacings," *J. Tsinghua Univ. Sci. Technol.*, vol. 51, no. 3, pp. 309–312, Mar. 2011.

[21] P. Wei, Y. Cao, and D. Sun, "Total unimodularity and decomposition method for large-scale air traffic cell transmission model," *Transp. Res. B, Methodol.*, vol. 53, pp. 1–16, Jul. 2013.

[22] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward based methods for the signal control problem in large-scale congested urban road networks," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 2, pp. 163–174, Apr. 2009.

[23] C. Han, S. Song, and C. H. Wang, "A real-time short-term traffic flow adaptive forecasting method based on ARIMA model," *Acta Simulata Systematica Sinica*, vol. 16, no. 7, pp. 1530–1535, 2004.

[24] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, Oct. 1996.

[25] X. Luo, L. Niu, and S. Zhang, "An algorithm for traffic flow prediction based on improved SARIMA and GA," *KSCE J. Civil Eng.*, vol. 22, pp. 4107–4115, May 2018.

[26] S. R. K. S. Prasad, "An efficient traffic forecasting system based on spatial data and decision trees," *Int. Arab J. Inf. Technol. (IAJIT)*, vol. 11, no. 2, pp. 186–194, 2014.

[27] J. Ahn, E. Ko, and E. Y. Kim, "Highway traffic flow prediction using support vector regression and Bayesian classifier," in *Proc. Int. Conf. Big Data Smart Comput. (BigComp)*, 2016, pp. 239–244.

[28] R. Rahman and S. Hasan, "Short-term traffic speed prediction for freeways during hurricane evacuation: A deep learning approach," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1291–1296.

[29] M. Cao, V. O. K. Li, and V. W. S. Chan, "A CNN-LSTM model for traffic speed prediction," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.

[30] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, Mar. 2017.

[31] P. Kim, J. Park, S. Lee, and J. Kim, "Similarity-based historical input selection to predict irregular holiday traffics in real-time," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2022, pp. 2290–2296.

[32] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, "The retrieval of intra-day trend and its influence on traffic prediction," *Transp. Res. C, Emerg. Technol.*, vol. 22, pp. 103–118, Jun. 2012.

[33] H. Li, X. Li, L. Su, D. Jin, J. Huang, and D. Huang, "Deep spatio-temporal adaptive 3D convolutional neural networks for traffic flow prediction," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 2, pp. 1–21, Apr. 2022.

[34] M. Z. Mehdi, H. M. Kammoun, N. G. Benayed, D. Sellami, and A. D. Masmoudi, "Entropy-based traffic flow labeling for CNN-based traffic congestion prediction from meta-parameters," *IEEE Access*, vol. 10, pp. 16123–16133, 2022.

[35] C. Keskin and M. A. Guvensan, "Traffic hyperparameters for long-term traffic forecasting," in *Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA)*, Aug. 2021, pp. 1–6.

[36] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, and S. Dong, "A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction," *Phys. A, Stat. Mech. Appl.*, vol. 583, Dec. 2021, Art. no. 126293.

[37] D. Kang, Y. Lv, and Y.-Y. Chen, "Short-term traffic flow prediction with LSTM recurrent neural network," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.

Irem Atilgan received the degree in computer engineering and industrial engineering from Yildiz Technical University, Istanbul, Turkey, in 2023.

She is currently working as a Technology Consultant. Her project "Outfit Recommendation Application on Smartphones" was supported by the Turkish Scientific and Technological Research Council of Turkey (TUBITAK), in 2020. Her research interests include data analysis, machine learning, and cloud technologies.

H. Irem Turkmen received the B.Sc., M.Sc., and Ph.D. degrees in computer engineering from Yildiz Technical University, Istanbul, Turkey, in 2005, 2008, and 2013, respectively.

She is currently an Assistant Professor with the Department of Computer Engineering, Yildiz Technical University. Her research interests include medical image processing, pattern recognition, speech processing, and machine learning.

M. Amac Guvensan (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in computer engineering from Yildiz Technical University, Istanbul, Turkey, in 2002, 2006, and 2011, respectively.

From 2009 to 2010, he visited the Wireless Networks and Embedded Systems Laboratory, University at Buffalo, The State University of New York, Buffalo, NY, USA, for a period of six months. He is currently an Associate Professor with the Department of Computer Engineering, Yildiz Technical University. His research interests include pervasive and ubiquitous computing, mobile technologies and applications, intelligent transportation systems, machine learning, and the Internet of Things.