Short-Term Traffic Flow Prediction Based on the Efficient Hinging Hyperplanes Neural Network

Tao, Qinghua; Li, Zhen; Xu, Jun; Lin, Shu; De Schutter, Bart; Suykens, Johan A.K.

DOI
10.1109/TITS.2022.3142728

Publication date
2022

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Intelligent Transportation Systems

Short-Term Traffic Flow Prediction Based on the Efficient Hinging Hyperplanes Neural Network

Qinghua Tao, Zhen Li, Jun Xu, *Member, IEEE*, Shu Lin, *Member, IEEE*, Bart De Schutter, *Fellow, IEEE*, and Johan A. K. Suykens, *Fellow, IEEE*

***Abstract*— Traffic flow (TF) prediction is an important and yet a challenging task in transportation systems, since the TF involves high nonlinearities and is affected by many elements. Recently, neural networks have attracted much attention for TF prediction, but they are commonly black boxes with complex architectures and difficult to be interpreted, e.g., the contributions of specific traffic elements are not explicit, hardly providing informative guidance. In this paper, we aim at addressing more interpretable short-term TF prediction with joint consideration to high accuracy, and thus introduces a pragmatic method by applying the efficient hinging hyperplanes neural network (EHHNN) simply built upon sparse neuron connections. In the proposed method, different traffic factors are incorporated into the inputs, including their spatial-temporal information. Besides the pursuit of accuracy, we further extend the ANOVA decomposition of EHHNNs to the interpretation analysis with specifications to traffic data, in which the contributions concerning specific traffic variables are detected quantitatively. As such, the proposed method firstly applies the EHHNN to filter out more important traffic variables for dimensionality reduction while maintaining accurate prediction. Then, variable interpretation analysis is performed from different perspectives, e.g. to quantitatively investigate the influence of traffic factors and also their spatial-temporal impacts. Therefore, a predictor and an analyzing tool can both be attained for the TF by exerting the flexibility and extending the interpretability of EHHNNs, which is promising to provide informative guidance to future traffic control. Numerical experiments verify the effectiveness and potential of the proposed method in TF prediction and analysis.**

***Index Terms*— Traffic flow prediction, piecewise linear neural networks, interpretation, variables analysis.**

Manuscript received January 14, 2021; revised July 30, 2021, November 2, 2021, and December 20, 2021; accepted December 31, 2021. This work was supported in part by the European Research Council (ERC) Advanced Grant E-DUALITY (787960); in part by the Research Council under Grant KUL C14/18/068; in part by the Flemish Government The Research Foundation-Flanders (FWO) Project GOA4917; in part by the Ford–KU Leuven Research Alliance Project KUL0076; in part by the EU H2020 ICT-48 Network TAILOR (Foundations of Trustworthy AI-Integrating Reasoning, Learning and Optimization); in part by the National Natural Science Foundation of China under Grant U1813224, Grant 61673366, Grant 62173113, and Grant 61903304; and in part by the Science and Technology Innovation Committee of Shenzhen Municipality under Grant JCYJ20200109113412326 and Grant GXWD20201230155427003-20200821173613001. The Associate Editor for this article was S. Selpi. *(Corresponding author: Jun Xu.)*

Qinghua Tao and Johan A. K. Suykens are with the Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, 3000 Leuven, Belgium (e-mail: qinghua.tao@esat.kuleuven.be; johan.suykens@esat.kuleuven.be).

Zhen Li is with the School of Mechanical Engineering and Automation, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China (e-mail: zueslee.hitsz@foxmail.com).

Jun Xu is with the School of Mechanical Engineering and Automation, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China, and also with the Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai 200240, China (e-mail: xujunqgy@hit.edu.cn).

Shu Lin is with the School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: slin@ucas.ac.cn).

Bart De Schutter is with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: b.deschutter@tudelft.nl).

Digital Object Identifier 10.1109/TITS.2022.3142728

I. Introduction

With the booming of economy, modern transportation systems have been developing rapidly, along with which traffic congestion and incidents are increasing. In the past a few decades, researchers have been working on the development of advanced intelligent transportation systems, in which the prediction of traffic flow (TF) appears as a crucial task [1]. The quality of system service depends to a large extent on TF prediction, since accurate prediction is very useful to provide information for proactive dynamic traffic control, so as to alleviate traffic congestion and to improve the operation efficiency [2].

TF prediction utilizes historical traffic data, normally the volume of vehicles (TF) in the predicted location, as the inputs to predict the TF in a location at future time steps. The existing methods for TF prediction can be mainly categorized as the types based on traffic simulation, non-parametric modelling and parametric modelling [3]. In [4], the LWR model was proposed by taking the TF as a fluid to predict. Daganzo constructed the cell transmission model as a simple approximation to the relationship of the density [5]. In [6], the stochastic features in urban traffic dynamics are also discussed with demand uncertainties. There is much literature of the theoretical basis of those simulation based methods. However, the accuracy remains unsatisfactory and the computational burden is high. Non-parametric methods developed rapidly and gained wide attention, since they have been found to provide stronger representability and flexibility to capture the nonlinearities of TF data, including support vector regression (SVR) [7], [8], k-nearest neighbor methods [9], etc. Parametric models include Kalman filtering models [10], historical average (HA), and time-series models relating to the autoregressive integrated moving average (ARIMA) [11]–[14], etc.

Particularly, the methods based on deep neural networks (NN) are getting increasing popularity and have brought great improvements in performance. Deep NNs use multilayer architectures to extract features of TF data from the original input variables. In [15], deep belief networks were applied to predict TF with multitask learning by layerwise training [3]. Later, convolutional neural networks (CNN) were used [16]–[19] and combined with different techniques, such as the

adversarial learning [19], and graph CNNs have also been found effective [20]–[22]. Moreover, the attention mechanism has been successfully applied for enhancing TF prediction accuracy [23], [24]. Recently in [25], different advanced techniques in deep learning were integrated for further enhancing the accuracy of short-term TF prediction, namely the attention based spatial-temporal graph convolution network (ASTGCN), which utilized spatial-temporal information of traffic data and showed state-of-the-art performance. In addition to the single-model prediction, hybrid deep learning models were also equipped by combining different models [26]–[29].

Although the aforementioned methods based on NNs have achieved fairly good accuracy, they require sophisticated techniques and varied tunings, and commonly have complex architectures. Moreover, the spatial-temporal correlations are inherently considered in the input variables, where the extracted features are high-level and used for enhancing the accuracy, so that it remains a black-box model and is difficult to analyze. In these predictors based on typical NNs, the contributions of different traffic variables to the prediction are not explicit, such as the influence of specific traffic variables (e.g., locations and historical time steps), and thus can hardly provide potentially informative analysis to facilitate future traffic control. As indicated in [30]–[35], more interpretable TF prediction is of great significance for future traffic control, but remains a challenge by far. Thus, rather than simply having an accurate predictor, it is desirable to perform interpretation analysis of the TF data and to investigate the modelling mechanism behind the predictors. Recently, a novel NN was proposed with a good trade-off between model flexibility and interpretability, namely the efficient hinging hyperplanes neural network (EHHNN) [36]. The EHHNN developed from the piecewise linear (PWL) representation of adaptive hinging hyperplanes (AHH), which involves a generic tree topology and thus is easier to interpret [37]. The EHHNN has shown great flexibility in dynamic system identification, and in particular, is decomposable via the analysis of variance (ANOVA) [36], [38], [39] owing to its sparse architecture, which greatly facilitates variable analysis.

In this paper, we revisit the problem of TF prediction and address joint considerations w.r.t. both accuracy and interpretation analysis. To this end, a pragmatic TF prediction and analysis method is established based on the EHHNN [36]. Compared to the existing methods using typical NNs, the TF prediction accuracy is guaranteed by the model flexibility of EHHNN, and yet the network architecture of EHHNN is much simpler. More importantly, we further extend the ANOVA decomposition of EHHNNs to varied interpretation analysis towards TF prediction, where careful considerations are taken with specifications to traffic variables and their physical explanations. In this way, it become viable to identify the particular TF variables contributing to the prediction output, whether they participate additively as univariate units or jointly in interaction with multiple traffic variables and how important they are concerning spatial or temporal aspects. Therefore, aside from the accurate predictor, our proposed method also leverages the interpretability of EHHNNs to detect the effects and correlations between different traffic variables, which yields a promising interpretable TF prediction. The main contributions of this study are summarized as follows:

- A programmatic method is introduced for accurate short-term TF prediction by applying the EHHNN, which is an accurate predictor yet with much simpler architectures than typical NNs. Different traffic factors are incorporated as the potential candidates to provide more possibilities not only in boosting accurate prediction but also in enhancing TF analysis afterwards. Those traffic factors include the TF, the average vehicle speed (AVS), and the road occupancy (RO), containing their spatial and temporal information.
- We further extend the ANOVA decomposition of EHHNNs to varied interpretation analysis, which is specified to traffic variables. In this way, we decouple traffic variables differently regarding their explicit physical meanings, so that their corresponding contributions to the prediction output can be detected quantitatively. Hence, 4 perspectives are introduced therein: variables selection, analysis of traffic factors (TF, AVS, and RO), spatial analysis and temporal analysis.
- An efficient training algorithm is developed for EHHNNs, so as to adapt to the practical application of real-world datasets. Numerical experiments demonstrate that the proposed method based on the EHHNN can achieve advantageous prediction accuracy and meanwhile effectively conduct varied analysis.

The rest of the paper is organized as follows. Section II introduces the preliminaries. Section III presents the details of the proposed method by extending the EHHNN and its ANOVA decomposition to TF prediction as well as interpretation analysis. Section IV gives the numerical experiments, and a brief conclusion is provided in Section VI.

II. Preliminaries

In this section, the prerequisites of this manuscript are presented, including some basic descriptions for the problem of TF prediction and a brief introduction to the EHHNN applied in our method.

A. Basic Descriptions for TF Prediction Problem

Given a sequence $X = \{X_1, \ldots, X_n\}$ of the observed traffic data from the historical n time steps, the task is to predict the TF y^{t+T} at a future time step $t + T$ (T time steps ahead) in the target location [3]. In TF prediction, a model $f(\cdot)$ is employed as the predictor, which aims to shrink the gap between the prediction results $\hat{Y} = f(X)$ and the observed data Y. Generally, the prediction model $f(\cdot)$ is estimated by minimizing the ℓ_2 norm loss, i.e.,

$$\min_{\boldsymbol{\theta}} ||Y - f(\boldsymbol{\theta}|X, Y)||_2^2, \tag{1}$$

where $\boldsymbol{\theta}$ is the vector of parameters. The problem of TF prediction normally consists of two stages, i.e., the determination of the relevant input variables and the prediction of the targeted TF [15]. The process of determining the input variables can be abstracted as determining a mapping $g(\cdot)$,

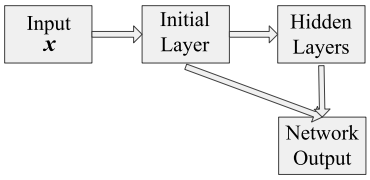


Fig. 1. Structure framework of the EHHNN.

which selects the most representative variables X from the given candidates F, i.e., $X = g(F)$. Most existing work conducts the first stage by manual selection from the historical TF at the prediction location [31], or employing supervised models to extract high-level features and then feed them into the subsequent predicting models [40], [41]. In the second stage, different models can be used to perform the prediction by utilizing the input data or the extracted features.

B. The EHHNN

In our method, the model used for TF prediction is the EHHNN, see Section III, where the terminology "hinging hyperplanes (HH)" comes from the work of [42], [43]. The HH is a popular PWL model to approximate nonlinear systems, and consists of the so-called hinges. The model of AHH also employs the hinges, but is built on compositing multiple univariate hinges. The topology of AHH can be seen as a generic recursive regression tree [37]; thus AHH has better interpretability than other PWL and even general nonlinear models. The EHHNN is an efficient network variant of AHH and inherits the universal approximation ability from AHH.

1) Network Architecture: The architecture of EHHNNs consists of three parts, i.e., the initial layer, the hidden layers, and the output connections [36]. Different from other NNs, the EHHNN is built upon the compositions of "max" and "min" operators without connecting weights between hidden layers. In this way, the neurons in EHHNNs can be connected to each other in a skip-layer manner, constructing a flexible model output, i.e., Fig. 1. More importantly, this special architecture enables the network to be decomposable via ANOVA and to be easy to obtain the explicit neuron expressions concerning specific variables.

2) Neuron Formulation: In the initial layer, the "max" operator can be regarded as the univariate rectifier linear unit (ReLU) [44] restricted to a specific dimension. For instance, in neuron $z_{1,s} = \max\{0, x_i - \beta_{k_i}\}$, x_i is the i-th component of $\boldsymbol{x}$ and β_{k_i} is called the k-th splitting knot over x_i. Then, the "min" operator becomes the activation function, which utilizes the existing neurons to formulate more flexible PWL neurons in the subsequent hidden layers, i.e., $z_{k,s}(\boldsymbol{x}) = \min\{z_{k_1,s_1}(\boldsymbol{x}), z_{k_2,s_2}(\boldsymbol{x})\}$ denoting the s-th neuron in the k-th layer with $k = k_1 + k_2$, given by

$$z_{k,s}(\boldsymbol{x}) = \min_{\upsilon_{s_1},\ldots,\upsilon_{s_k} \in J_{k,s}} \{\max\{x_{\upsilon_{s_1}} - \beta_{s_1}, 0\}, \ldots, \max\{x_{\upsilon_{s_k}} - \beta_{s_k}, 0\}\}, \quad (2)$$

where $J_{k,s} = \{\upsilon_{s_1}, \ldots, \upsilon_{s_k}\}$ contains the indices of the interacting variables with $\{\upsilon_{s_1}, \ldots, \upsilon_{s_k}\} \subseteq \{1, \ldots, d\}$. The cardinality $|J_{k,s}|$ is k, i.e., k variables are interacting in neurons of the k-th layer. Thus, the layer index of EHHNNs is determined by the number of variables in the neurons.

The output layer of EHHNN is the weighted sum of all neurons by skipping layers:

$$f(\boldsymbol{x}) = w_0 + \sum_{s=1}^{n_1} w_{1,s} \max\{x_{\upsilon_{1,s}} - \beta_{1,s}, 0\} + \sum_{k=2}^{K} \sum_{s=1}^{n_k} w_{k,s} \min_{\upsilon \in J_{k,s}} \{\max\{x_\upsilon - \beta_{\upsilon_{k,s}}, 0\}\}, \quad (3)$$

where n_k is the number of neurons in the k-th layer, $w_{k,s} \in \mathbb{R}$ are the weights connecting the network output, and $w_0 \in \mathbb{R}$ is the constant bias. Skip-layer connections are also used in the highway NNs [45] and the residual NNs [46]. Differently, in EHHNNs, multiplication operations and connecting weights are not involved across neurons, and all variables are not necessarily coupled in each neuron, so that the network structure is much simpler and easier to train and analyze.

III. TF Prediction and Analysis Based on the EHHNN

In this section, the proposed method for TF prediction and analysis is introduced in detail. The flexible EHHNN is firstly applied to boost accurate TF prediction, where different traffic factors are considered in the inputs together with their spatial-temporal information. Then, interpretation analysis is conducted from varied perspectives with specifications to traffic variables via extending the ANOVA decomposition of the EHHNN predictor.

A. Prediction for TF Data

In this subsection, we aim at formulating the TF predictor by applying the EHHNN with the traffic data, and then corresponding interpretation studies will follow up in the next subsection.

The problem of TF prediction can be abstracted as follows. Let X_i^t denote the data vector at the t-th time step in the i-th observation segment of the selected transportation system. Given the current time step t and the data sequence $H = \{X_{o_i}^t | o_i \in \mathcal{I}_i,\ t \in \mathcal{T}\}$ in the selected observation segments of $\mathcal{I}_i$ at previous n time steps of $\mathcal{T} = \{t-1, t-2, \cdots, t-n\}$, the task is to predict the TF y_i^{t+T} at a future time step $t+T$ within the i-th observation segment. Then, the corresponding forecasting model is described as

$$\hat{y}_i^{t+T} = f(X_{o_i}^t), \quad o_i \in \mathcal{I}_i,\ t \in \mathcal{T}, \quad (4)$$

where $\hat{y}_i^{t+T}$ denotes the EHHNN prediction in our method. In many existing methods, the inputs $X_{o_i}^t$ are normally taken as the historical TF data in the prediction location, i.e., $\mathcal{I}_i = \{i\}$ [31] or the historical TF data in different locations equipped with observation detectors [15], [40].

In practice, various factors can influence the TF [2]. In this paper, besides the historical TF, the attributes of AVS and RO are also incorporated as the potential traffic factors affecting the TF, providing more possibilities to boost the prediction

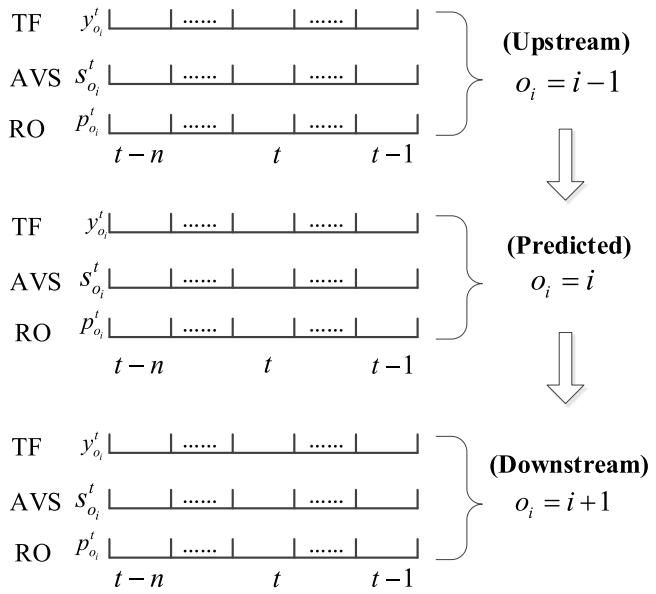


Fig. 2. An illustration on the descriptions of input variables.

accuracy by including different traffic variables. In addition, the corresponding data of those traffic factors from different adjacent observation segments are also incorporated into the input variables for spatial information. The historical data of these selected factors, i.e., TF, AVS and RO, are considered for temporal information. Those variables are seen as the candidates which potentially affect the TF, but their importance to the prediction can vary significantly.

For spatial information, we mainly focus on the upstream segment (the $(i-1)$-th observation) as well as the downstream segment (the $(i+1)$-th observation) herein. This idea can be easily extended to different spatial considerations, see Section IV-B.3 and Section IV-C. Therefore, the EHHNN for TF prediction is formulated as

$$\hat{y}_i^{t+T} = f_{\text{EHHNN}}(y_{o_i}^t, s_{o_i}^t, p_{o_i}^t), \quad o_i \in \mathcal{I}_i, \ t \in \mathcal{T}, \tag{5}$$

where $\mathcal{I}_i = \{i, i+1, i-1\}$ and $\mathcal{T} = \{t-1, \ldots, t-n\}$.

In the constructed EHHNN predictor (5), the spatial-temporal information is contained in the input data w.r.t. $\mathcal{I}$ and $\mathcal{T}$, where $y_{o_i}^t, s_{o_i}^t$, and $p_{o_i}^t$ denote the historical TF, AVS, and RO within the observation segment o_i at time step t, respectively. Therefore, the dimensionality of the input variables is $3 \cdot |\mathcal{T}| \cdot |\mathcal{I}|$, where $|\cdot|$ represents the cardinality of a set. Fig. 2 illustrates the descriptions of the input variables.

As described above, different traffic factors are incorporated together with spatial-temporal considerations in this paper. With such input data containing rich traffic information and the flexibility of EHHNNs, accurate TF prediction can be expected. Thus, the EHHNN is now successfully applied as a TF predictor.

However, when all the potential variables are considered, and the resulting prediction model can be complicated and computationally expensive regarding the redundant information. Therefore, it is necessary to filter individual variables and to select the ones that contribute influentially to the prediction. Then, a more concise predictor model can be obtained, while the prediction accuracy can also be guaranteed with those influential variables selected.

Based on the special network architecture of EHHNNs, it is amenable to perform interpretation analysis towards the TF prediction, while other NNs commonly extract high-level features from input data to enhance accuracy and are difficult to explicitly detect the contributions of specific traffic variables. Therefore, there are two major parts in the proposed method, i.e., an EHHNN model (5) is trained and performed with variables analysis and selections and then a more accurate EHHNN predictor with the selected traffic variables is attained, see the following subsection.

B. Interpretation Towards TF Data

In the neurons of EHHNNs, variables are not coupled in a linearly weighted combination, but by "min" and "max" operators, i.e., the network is decomposable w.r.t. input variables. It has been shown in [36] that the ANOVA decomposition can be used in EHHNNs to select influential variables. In fact, for practical applications, variables are assigned with explicit physical explanations, so that the analysis can be enriched by careful considerations specified on real-world data, rather than simply detecting influential individual variables in [36].

In this subsection, we further extend the ANOVA decomposition of EHHNNs to the application of more interpretable TF prediction, and thus present different analysis towards TF data. In this paper, the interpretation analysis is performed from 4 perspectives: variables selections similar to [36], and the detections on factor influence (TF, AVS, and RO), spatial influence and temporal influence.

1) Variable Decomposition: In the TF predictor based on EHHNNs, the explicit expressions of all neurons in each layer can be rearranged as

$$f_{\text{EHHNN}}(\boldsymbol{x}) = w_0 + \sum_i f_i(x_i) + \sum_{i,j} f_{ij}(x_i, x_j) + \cdots, \tag{6}$$

in which the first sum is over all neurons involving only a single variable, the second sum is over those neurons representing two-variable interactions, and so on. In the EHHNN predictor for TF, the input traffic variables can both additively and interactingly contribute to the prediction output, and their corresponding contributions can be decoupled.

More specifically, $f_i(x_i)$ is composed of the sum over all source neurons involving the particular input variable x_i, and $f_{ij}(x_i, x_j)$ indicates the sum over all neurons in the second hidden layer involving the particular pair of variables x_i and x_j. Then, we have the ANOVA functions, i.e.,

$$\begin{aligned} f_{r_1}(x_{r_1}) &= \sum_{(k,s):J_{k,s}=\{r_1\}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}}) \\ f_{r_1 r_2}(x_{r_1}, x_{r_2}) &= \sum_{(k,s):J_{k,s}=\{r_1,r_2\}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}}) \\ &\vdots, \end{aligned} \tag{7}$$

where the set $J_{k,s}$ contains the indices of variables $\tilde{\boldsymbol{x}}$ interacting in the neuron $z_{k,s}$, $\tilde{\boldsymbol{x}}$ is x_{r_1} in the first equation, $\tilde{\boldsymbol{x}}$ is

$[x_{r_1}, x_{r_2}]^T$ in the second, and so on. In this way, the EHHNN is decomposed into different network components characterized by individual traffic variables or their interactions.

To assist the following analysis, besides each individual traffic variable, we can decompose the network components from 3 aspects, i.e., the traffic factors ($\tilde{\boldsymbol{x}}^{\mathrm{TF}}$, $\tilde{\boldsymbol{x}}^{\mathrm{AVS}}$, and $\tilde{\boldsymbol{x}}^{\mathrm{RO}}$), the spatial information ($\tilde{\boldsymbol{x}}^{\mathrm{spa}(\hat{t})}$), and the temporal information ($\tilde{\boldsymbol{x}}^{\mathrm{tep}(\hat{i})}$). Accordingly, the EHHNN predictor can be decomposed into different components concerning

$$\begin{aligned}
\tilde{\boldsymbol{x}}^{\mathrm{TF}} &= [y_{o_{i-1}}^{t-1}, \ldots, y_{o_{i-1}}^{t-n}, y_{o_i}^{t-1}, \ldots, y_{o_i}^{t-n}, y_{o_{i+1}}^{t-1}, \ldots, y_{o_{i+1}}^{t-n}]^T, \\
\tilde{\boldsymbol{x}}^{\mathrm{AVS}} &= [s_{o_{i-1}}^{t-1}, \ldots, s_{o_{i-1}}^{t-n}, s_{o_i}^{t-1}, \ldots, s_{o_i}^{t-n}, s_{o_{i+1}}^{t-1}, \ldots, s_{o_{i+1}}^{t-n}]^T, \\
\tilde{\boldsymbol{x}}^{\mathrm{RO}} &= [p_{o_{i-1}}^{t-1}, \ldots, p_{o_{i-1}}^{t-n}, p_{o_i}^{t-1}, \ldots, p_{o_i}^{t-n}, p_{o_{i+1}}^{t-1}, \ldots, y_{o_{i+1}}^{t-n}], \\
\tilde{\boldsymbol{x}}^{\mathrm{spa}(\hat{t})} &= [y_{o_{i-1}}^{t-\hat{t}}, y_{o_i}^{t-\hat{t}}, y_{o_{i+1}}^{t-\hat{t}}, \ldots, s_{o_{i+1}}^{t-\hat{t}}, p_{o_{i-1}}^{t-\hat{t}}, p_{o_i}^{t-\hat{t}}, p_{o_{i+1}}^{t-\hat{t}}]^T \\
\tilde{\boldsymbol{x}}^{\mathrm{tep}(\hat{i})} &= [y_{o_{i-1}}^{t-1}, \ldots, y_{o_{i-1}}^{t-n}, y_{o_i}^{t-1}, \ldots, y_{o_i}^{t-n}, y_{o_{i+1}}^{t-1}, \ldots, y_{o_{i+1}}^{t-n}]^T.
\end{aligned} \tag{8}$$

In the EHHNN predictor, the index sets representing the indices of neurons which contain variables $\tilde{\boldsymbol{x}}^{\mathrm{TF}}$, $\tilde{\boldsymbol{x}}^{\mathrm{AVS}}$, $\tilde{\boldsymbol{x}}^{\mathrm{RO}}$, $\tilde{\boldsymbol{x}}^{\mathrm{spa}(\hat{t})}$ and $\tilde{\boldsymbol{x}}^{\mathrm{tep}(\hat{i})}$ are now denoted as J^{TF}, J^{AVS}, J^{RO}, $J^{\mathrm{spa}(\hat{t})}$ and $J^{\mathrm{tep}(\hat{i})}$, respectively.

Similar to the strategy in [36], [38], with the training data, the standard variance (the σ value) of the decomposed network components concerning specific variables, i.e., the ANOVA functions in (7), can be computed to reflect their relative contributions to the prediction output. Given an ANOVA function and the index set J of the variables $\tilde{\boldsymbol{x}}$ contained on it, we rewrite such an ANOVA function listed in (6) as $f_J(\cdot)$, where the variable ($\tilde{\boldsymbol{x}} = x_{r_1}$) is or the variables ($\tilde{\boldsymbol{x}} = [x_{r_1}, x_{r_2}, \cdots]^T$) are coupled in $f_J(\cdot)$. With the given training data, the standard variance (the σ value) of each $f_J(\cdot)$ in (7) is calculated by

$$\begin{aligned}
\sigma_J &= \sqrt{\mathrm{VAR}(f_J(\tilde{\boldsymbol{x}}))}, \\
f_J(\tilde{\boldsymbol{x}}) &= \sum_{(k,s):J_{k,s}=J} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}}),
\end{aligned} \tag{9}$$

where $\mathrm{VAR}(\cdot)$ denotes the corresponding variance of the prediction output related to the variables ($\tilde{\boldsymbol{x}}$) in set J. The larger σ_J is, a greater impact the corresponding network component has on the prediction output, i.e., the particular TF input variables $\tilde{\boldsymbol{x}}$ are more influential on the EHHNN prediction output, so that the relative importance of different network components is revealed explicitly.

2) Variable Analysis: By extending the ANOVA decomposition of EHHNNs, the TF prediction process becomes more interpretable, since it is now possible to detect the contributions of different individual traffic variables to the predictor or the interactions of them.

On the one hand, more influential input variables need to be selected for the predictor. On the other hand, interpretation analysis on the prediction outputs is also of great significance to have more in-depth understanding of the TF data, potentially providing guidance to future traffic control [30], [32]–[35]. With the variable decomposition in Section III-B.1, the method proposed for TF variable analysis is presented mainly from 4 aspects, i.e., variable selection, traffic factor analysis, spatial analysis, and temporal analysis. To this end, we apply ANOVA decomposition to the source nodes in a single-layered EHHNN, where the relative importance of all candidate input variables is revealed by the σ values in (9). We unify the vector of complete input variables as $\boldsymbol{z} = [y_{o_i}^t, s_{o_i}^t, p_{o_i}^t]^T$ with $o_i \in \mathcal{I}_i$ and $t \in \mathcal{T}$.

In analysis part of the proposed method, variable selection is firstly conducted by sorting the σ value regarding each variable z_k of $\boldsymbol{z}$ in descending order and selecting the top d ones for further prediction, which reduces model dimensionality and maintains prediction accuracy.

Next, the analysis of different traffic factors is cast by computing the accumulative σ values concerning the corresponding traffic variables to reveal their relative importance. For example, the importance of historical TF is evaluated by the σ value over

$$f_{\mathrm{TF}} = \sum_{(k,s):J_{k,s}=J^{\mathrm{TF}}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}}). \tag{10}$$

Similar analysis can be done for the factors of AVS and RO in (8) based on the index sets J^{AVS} and J^{RO}, respectively. Analogously, spatial analysis is performed by computing the corresponding σ value of

$$f_{\mathrm{spa}} = \sum_{(k,s):J_{k,s}=J^{\mathrm{spa}}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}}), \tag{11}$$

and then temporal analysis is done by calculating the σ value concerning

$$f_{\mathrm{tep}} = \sum_{(k,s):J_{k,s}=J^{\mathrm{tep}}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}}). \tag{12}$$

In addition, we can also perform extra analysis to detect the influence of different variable interactions in deeper layers of the EHHNN, showing the effects of different traffic variable interactings.

C. Method Summary

In this subsection, a more efficient training algorithm is proposed for the EHHNN, so as to better adapt to practical applications with real-world datasets. In EHHNNs, the splitting knots β_{k_i} are pre-allocated as the quantiles of each dimension, and the composites of subsequent hidden layers do not requires weights in between. Consequently, the weights connecting all neurons to the output are the only parameters to optimize. It is generally difficult to obtain the optimal connections among neurons, and thus the EHHNN employs a descent algorithm for searching the locally optimal network structure, and then the finally parameters are determined [36]. However, such descent algorithm traverses all neuron combinations by repeatedly updating the parameters. Given the total number of neurons P and the number of data points M, the computation costs in the existing training algorithm for EHHNN come from two aspect for each cycle of the descent searching, i.e., the optimization on the network structure and the ADMM solver for the network parameters [36]. Here, the adjacency matrix determines the network structure, and its optimization is based on the traversal scheme, leading to the worst-case

complexity as $O(P^4M)$ for 1-cycle optimization. The number of cycles in optimization is problem-dependent, and it was shown in [36] that 10 cycles could obtain good results for the tested problems, which is very time-consuming and even intractable for high dimensions and large data sizes.

To make the EHHNN more computationally applicable, an efficient training algorithm is hereby proposed based on stacking subnetworks, each of which is generated by random searches followed with a pruning procedure by ADMM. More specifically, the size of each subnetwork is tentatively set excessively larges for flexibility, and then an ℓ_1 norm regularization is used to trim the model to moderate size, i.e., we adopt the least absolute shrinkage and selection operator (LASSO). Analogously, the training of each EHHNN subnetwork is done through the optimization problem

$$\min_{w_{k,s}} \frac{1}{2}||Y - \sum_k \sum_s w_{k,s} z_{k,s}(X) - w_0||_2^2 + \lambda \|w_{k,s}\|_1, \quad (13)$$

where $\lambda > 0$ is the regularization coefficient. In this paper, we solve the resulting LASSO problem (13) with the alternating direction method of multipliers (ADMM) algorithm [47]. After this, the weights $w_{r,s}$ are determined and redundant neurons are removed, so that the network gets more condensed and maintains flexibility.

To alleviate the effect of the randomness in generating the EHHNN network structure, a stacking strategy is adopted to attain the final network, similar to the procedures in [48]. Specifically, multiple datasets are generated by dividing the original training dataset $\mathcal{V}$ into subsets $\mathcal{V}_1, \ldots, \mathcal{V}_L$, each of which is used to train an EHHNN subnetwork $f_{\text{EHHNN},j}(\boldsymbol{x})$. Then, the final network $f_{\text{EHHNN}}(\boldsymbol{x})$ is stacked by $f_{\text{EHHNN}}(\boldsymbol{x}) = \sum_{j=1}^{L} \gamma_j f_{\text{EHHNN},j}(\boldsymbol{x}, \boldsymbol{\theta}_j)$, i.e.,

$$\min_{\gamma_j} \sum_{(\boldsymbol{x}(k), y(k)) \in \mathcal{V}} (y(k) - \sum_{j=1}^{L} \gamma_j f_{\text{EHHNN},j}(\boldsymbol{x}(k), \boldsymbol{\theta}_j)^2. \quad (14)$$

The stacked EHHNN can be merged into a compact network by rearranging all neurons layer by layer in each subnetwork. Then, the merged EHHNN is more concise in expression and maintains the interpretability. Now each subnetwork is randomly generated and the parameters are then solved by the ADMM, so that the optimization of the adjacency matrix, i.e., the traversal procedure, is omitted for obtaining the network structure. It was pointed out in [36] that the computational burden of the subsequent ADMM solver for parameters was way less and even negligible, compared to the existing traversal scheme in optimizing the network structure. Besides, the proposed training algorithm only conducts a single cycle, and thus it is distinctively more efficient, even if L times of ADMM are performed for the parameters of subnetworks. In the next section, the resulting accuracy using this efficient training algorithm will be verified through the extensive numerical experiments in TF prediction.

In summary, the proposed method for TF prediction and analysis based on the EHHNN is presented in Algorithm 1, where the training of EHHNNs in Step 1/4 is performed with the algorithm mentioned above, i.e., (13) and (14).

Algorithm 1 TF Prediction and Analysis Based on EHHNN

Input: Input sequence $H = \{X_{o_i}^t(m)\}$ and output sequence $Y = \{y_i^t(m)\}$ with $o_i \in \mathcal{I}_i,\ t \in \mathcal{T},\ m = 1, \ldots, M$.

Output: The EHHNN predictor $f_{\text{EHHNN}}(\boldsymbol{x})$ and the corresponding analysis results.

1: Train a single-layered EHHNN with data H and Y;
2: Conduct ANOVA decomposition (6) on the EHHNN obtained by Step 1, i.e., calculating the σ values (9);
3: *Variable Selection*: sort the σ value concerning each variable z_k of $\boldsymbol{z}$ in descending order (Sort$\{\sigma_k\}$) and select the top d ones;
4: Train the TF prediction model using a multilayer EHHNN $f_{\text{EHHNN}}(\boldsymbol{x})$ with the variables selected by Step 3;
5: *Traffic Factor Analysis*: compute the accumulative σ value over $f_{\text{TF}} = \sum_{J_{k,s}^{\text{TF}}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}})$, $f_{\text{AVS}} = \sum_{J_{k,s}^{\text{AVS}}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}})$ and $f_{\text{RO}} = \sum_{J_{k,s}^{\text{RO}}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}})$;
6: *Spatial Analysis*: compute the accumulative σ values over $f_{\text{spa}(\hat{\imath})} = \sum_{J_{k,s}^{\text{spa}(\hat{\imath})}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}}), \forall \hat{\imath} \in \mathcal{T}$;
7: *Temporal Analysis*: compute the accumulative σ values over $f_{\text{tep}(\hat{\imath})} = \sum_{J_{k,s}^{\text{tep}(\hat{\imath})}} w_{k,s} z_{k,s}(\tilde{\boldsymbol{x}}), \forall \hat{\imath} \in \mathcal{I}$;
8: Additional Analysis: repeat Step 2 for the EHHNN predictor and calculate the σ values of variable interactions.

IV. Numerical Experiments

In this section, numerical experiments are performed to evaluate the proposed method for TF prediction and analysis by applying the EHHNN and extending its ANOVA decomposition, where real world datasets are used for TF prediction in different settings and various related methods are compared. Then, other real world examples are evaluated to further demonstrate the effectiveness of the proposed method and its flexibility to adapt to different traffic scenarios.

A. Data and Settings

The experimental data are collected from the Caltrans Performance Measurement System (PeMS) which can be accessed via https://pems.dot.ca.gov/, and are sampled every 5 minutes in each detector. In this paper, the traffic data from March 11th to April 7th of the year 2019 are used, where the data of the first 3 weeks are chosen for training and that of the remaining 1 week for testing. The EHHNN is implemented on a platform Matlab R2016b and a 3.20GHz Intel(R) Core(TM) i7-8700 CPU of 16.0 GB. The data are pre-processed to have the unit scale, i.e., $\boldsymbol{x} \in [0, 1]^d$. We first present a series of experiments on the selected data to do the TF prediction together with the comparison with other effective methods, and then perform the analysis on TF data from different perspectives.

To evaluate the prediction performance, similar to [49], [50], we use the mean absolute error (MAE), the root mean squared error (RMSE), and the R-squared index R^2 as companion criteria. The R^2 ranges from 0 to 1 and is used to measure the data similarity, where a larger value indicates a more accurate

approximation. The criteria are defined as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{M}\sum_{m=1}^{M}|y_m - \hat{y}_m| \\ \text{RMSE} &= \sqrt{\frac{1}{M}\sum_{i=1}^{M}(y_m - \hat{y}_m)^2} \\ \text{R}^2 &= 1 - \frac{\sum_{m=1}^{M}(y_i - \hat{y}_m)^2}{\sum_{m=1}^{M}(y_m - \bar{y})^2}, \end{aligned} \tag{15}$$

where M is the number of data, y_m is the observed TF with mean value $\bar{y}$, and $\hat{y}_m$ denotes the predicted TF.

For the EHHNN predictor in this paper, $L = 10$ subnetworks are generated using subsets of the sample data, where the penalty coefficient in the LASSO problems is selected from $\{0.01, 0.05, 0.1, 0.5, 1\}$. Considering that the training data are intrinsically time series, the training data sets for generating subnetworks are chosen as

$$\mathcal{V}_1 = \{\boldsymbol{x}(m), y(m)\}_{m=1}^{M-L}, \ldots, \mathcal{V}_{10} = \{\boldsymbol{x}(m), y(m)\}_{m=1}^{M-1}. \tag{16}$$

For the final stacked EHHNN (14), the complete sample dataset is used for training, i.e., $\mathcal{V} = \{\boldsymbol{x}(m), y(m)\}_{m=1}^{M}$. For the initial layer, the source neurons are set as $\max\{0, x_i - \beta_{k_i}\}$ with $i \in \{1, \ldots, d\}$ and $\beta_{k_i} \in \{0, 0.25, 0.5, 0.75\}$.

B. TF Prediction and Analysis on Road Segments

In this subsection, the road segment described in Section III-A is presented as a numerical study to comprehensively evaluate the proposed method in Algorithm 1, concerning both prediction accuracy and interpretation analysis. As shown in Fig. 2, the spatial-temporal data of TF, AVS and RO of detectors in the corridor US-50 with indices 311974, 313658, and 314559 are chosen.

In summary, the experiments mainly contain two parts, i.e.,

- Performance Evaluation and Comparisons: The candidate variables in Table I are considered, and the more important ones are selected based on the σ values in (9) to train the EHHNN. Then, the prediction performance of the EHHNN is compared with 5 related methods;
- Variable Analysis: To interpret the EHHNN predictor, variable analysis is conducted mainly concerning the aspects of network components, traffic factors, spatial influence and spatial influence based on the σ values (9).

1) Performance Evaluation and Comparisons: To evaluate the overall prediction performance of the proposed method, 5 related methods are presented for comparisons. In this experiment, we incorporate the NN, random walk (RW) [51], radial basis function (RBF), support vector regression (SVR), and the attention based spatial-temporal graph convolution network (ASTGCN) models, among which the RW model is regarded as a simple baseline that takes the current TF as the future outcome while the recently proposed ASTGCN has shown state-of-the-art performance in TF prediction. For the network models, the number of layer is 3 and the number of hidden units in each hidden layer is $[10, 20, 50, 100]$ and the AdamOptimizer is used with a batch size of 50 and a learning rate as 0.01. The ASTGCN is set with default settings described in [25]. Then, the final structure is determined as the best one evaluated by 5-fold cross validation, where the training dataset is randomly partitioned into 5 equal-sized subsets, one of which is used in turn as the validation data.

TABLE I
POTENTIAL INPUT VARIABLES

Historical time steps n / Factor / Segment	TF	AVS	RO
Upstream $o_i = i - 1$	10	10	10
Prediction $o_i = i$	10	10	10
Downstream $o_i = i + 1$	10	10	10

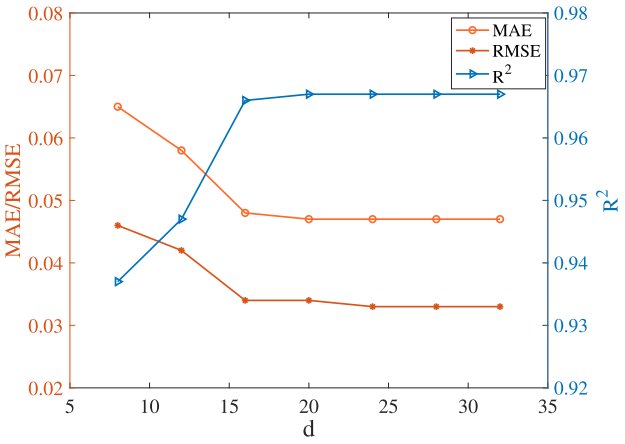


Fig. 3. An illustration of the prediction performance with d variables selected.

TABLE II
COMPARISON ON VARIABLE SELECTION

	MAE	RMSE	R^2
d=16	0.033	0.046	0.969
d=90	0.039	0.052	0.959

In our proposed method, the EHHNN is more than a predictor, and it is also an effective analyzing tool to identify the importance and effects of the input variables. Taking $T = 1$ for instance, we can select more influential variables according to their σ values, by removing the least influential ones. The corresponding prediction performance can be seen in Fig 3, where d is the number of selected variables.

From Fig 3, we can find that the prediction accuracy preliminarily improves with increasing number of variables (approximately when $d \leq 16$), and then stays almost unchanged, suggesting that approximately no less than 16 variables need to be selected as the final input X. Then, we employ a multilayer EHHNN to further verify the effect of variable selection and Table II presents the results. Table II demonstrates that less input variables than $d = 16$ can even yield higher prediction accuracy than considering all candidate variables ($d = 90$), which further verifies the effectiveness of variable selection with the EHHNN. The 74 removed variables are mainly

TABLE III
COMPARISON ON TF PREDICTION PERFORMANCE ON ROAD SEGMENTS OF DIFFERENT METHODS

Prediction	EHHNN			NN			RW		
Step	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
$T=1$	**0.033**	**0.047**	**0.969**	0.036	0.048	0.966	0.039	0.055	0.957
$T=3$	**0.041**	**0.058**	**0.949**	0.047	0.066	0.935	0.050	0.072	0.924
$T=6$	0.050	0.069	0.930	0.060	0.083	0.897	0.065	0.093	0.872
	SVR			ASTGCN			RBF		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
$T=1$	0.040	0.055	0.956	0.042	0.056	0.956	0.036	0.051	0.961
$T=3$	0.050	0.067	0.933	0.042	0.057	0.956	0.043	0.061	0.946
$T=6$	0.062	0.082	0.900	**0.046**	**0.063**	**0.941**	0.050	0.069	0.929

contained in the downstream segment, for which the corresponding spatial analysis will be introduced in Section IV-B.2.

As the experiments are for short-term prediction, we aim at predicting the TF within the next half hour. In this paper, the predicted future time step T takes 3 different values, i.e., $T = 1, 3, 6$, representing prediction horizons of 5, 15, 30 minutes, respectively. The comparison results with other TF prediction methods are shown in Table III, where the highest testing accuracy of each measure criterion is marked in bold.

Table III demonstrates that the EHHNN outperforms other methods for most of the cases, where the RBF model shows similar results with the EHHNN regarding MAE and RMSE when $T = 6$ and the ASTGCN shows to be slightly advantageous when $T = 6$. In particular, the ASTGCN employs many sophisticated deep learning techniques with complex network structures and shows state-of-the-art performance in TF prediction [25], and it is worth noting that the 3-layered EHHNN can achieve competitive and even advantageous performance compared to the ASTGCN, which integrates graphical CNNs and the attention mechanism. In our 3-layered EHHNN predictor, there are at most 3 variables interacting in a single neuron, and such an EHHNN can already achieve high prediction accuracy to well describe the TF data, while most existing methods based on NNs have all variables coupled in each single neuron. This reflects the concise structure and great flexibility of the EHHNN predictor for TF, and also indicates that the variables should not always be fully coupled with each other in TF prediction, providing possible interpretation analysis about interaction levels and relations among different traffic variables.

2) Variable Analysis: As shown in Algorithm 1, we then perform variable analysis from 4 perspectives: network component influence (variable selection), factor influence (TF, AVS, and RO), spatial influence, and temporal influence. Here, we set the prediction horizon as 5 minutes as a case study to illustrate the interpretability.

• *Network component influence:* When selecting more influential variables in the EHHNN in Fig 3, the relative importance of different input variables can be revealed. Then, when the more flexible EHHNN predictor is applied to TF prediction, the interactions of the selected variables (network components in hidden layers) can be detected. The number of hidden units in each hidden layer is [50, 50, 50], i.e., there are 50 neurons containing single variables, 50 neurons involving two interacting variables, and 50 neurons with three interacting variables. Those neurons all directly connect and contribute to the final network outputs.

TABLE IV
ANOVA DECOMPOSITION OF INDIVIDUAL VARIABLES

No.	Variable	σ
1	y_{i-1}^{t-1}	0.083
2	y_i^{t-2}	0.043
3	y_i^{t-1}	0.043
4	y_{i-1}^{t-3}	0.041
5	y_{i-1}^{t-4}	0.003

Firstly, we detect the influence of different individual traffic variables to the EHHNN predictor, based on the σ values in (9). Table IV shows the largest 5 σ values of the source neurons, reflecting the impacts of individual variables. It can be seen that the TF in the upstream segment at the previous time step (y_{i-1}^{t-1}) is the most influential element affecting the TF prediction and the TF in the prediction segment 2 time steps before was also quite important (y_i^{t-1} and y_i^{t-2}). The TF in the upstream segment close to (3 time steps before) the predicted time step (y_{i-1}^{t-3}) is also influential. Our analysis results tell that a lagging impact from the upstream segment plays a critical role in affecting the TF, which coincides with the practical assumptions in transportation systems, verifying the effectiveness of our analysis. Moreover, the concrete time steps of influential TF variables can also be revealed, which has potential to be utilized for proactive traffic control, such as restricting the TF in the upstream segment in advance to mitigate the traffic congestion in the prediction segment. In addition, the variables y_{i-1}^{t-3} and y_{i-1}^{t-4} are in the list while y_{i-1}^{t-2} does not appear. This may be due to the reason that the historical TF values within different time steps in the upstream segment are not independent, and the vehicles are not evenly distributed in the prediction location as time is passing by.

Then, we show the interactions among variables with the top 5 σ values, i.e., the network components in hidden layers, in Table V, where the interacting variables within the corresponding neurons are given in the second column.

Table V shows that the neurons containing the interactions between TF (y) and AVS (s), i.e., No. 1, 3, and 5, and also

TABLE V
ANOVA DECOMPOSITION OF VARIABLE INTERACTION

No.	Interacting Variables	σ
1	$y_i^{t-7}, y_i^{t-2}, s_i^{t-1}$	0.0084
2	$y_{i-1}^{t-2}, y_i^{t-2}, y_i^{t-3}$	0.0076
3	$y_{i-1}^{t-2}, s_i^{t-2}, y_{i-1}^{t-1}$	0.0069
4	$y_{i-1}^{t-1}, y_i^{t-3}, y_i^{t-4}$	0.0049
5	y_i^{t-7}, s_i^{t-1}	0.0049

TABLE VI
ATTRIBUTE CONTRIBUTION

	AVS	RO	TF
σ	0.019	0.032	0.328

the interactions between the TF from different historical time steps, i.e., No. 2 and 4, are more influential to the prediction output. For example, for the first row, the TF in the prediction segment in previous time steps $t-7$ and $t-2$, as well as with the AVS in the prediction segment in time step $t-1$, contribute jointly to the output, reflecting that the combinational effect of TF and AVS in the prediction segment is significant. For the second row, the TF in the prediction segment in previous time steps $t-3$ and $t-2$ has a joint impact together with the historical TF from the upstream segment, revealing that the TF in the upstream segment in previous time steps $t-2$ can possibly arrive in the prediction segment at the predicted time step, and affect the TF hereby. The other interactions can be explained similarly and their relative importance can also be revealed, which explicitly unfolds the constitution of the predictor and provides an interpretable predictor. Compared to Table V, the influence of single variables shown in Table IV is higher than those of variable interactions, indicating that for our problem, the variables mainly contribute additively to the prediction output, which reveals rather simple relations between those variables.

• *Traffic factor influence:* TF, AVS, and RO are basic traffic factors that reflect the characteristics of the TF. Through the ANOVA decomposition, the influence of these 3 traffic factors is shown by the σ values in Table VI.

We can see from Table VI that TF, AVS, and RO all influence the TF prediction and it is effective to deploy them in the input variables. Specifically, the historical TF is significantly more influential to the prediction output than the other 2 factors. At the same time, it demonstrates that the factors AVS and RO also have an impact on the prediction output to some extent, verifying the effectiveness of deploying multiple traffic factors, where AVS and RO seem to have comparable importance but less than that of the historical TF. Note that to calculate the variance σ, neuron outputs relating to that particular factor are all considered. For example, for the AVS, neuron outputs containing $s_{o_i}^t, o_i \in \{i-1, i, i+1\}, t \in \{t-1, \ldots, t-10\}$ are all summed to calculate the σ value, i.e., (8) and (10).

• *Spatial influence:* Besides the prediction segment, we also employ the upstream and the downstream segments for spatial information. Here, the relative contributions of these segments are shown in Table VII, reflecting the spatial influence. Notice that the σ value of the i-th segment is calculated by summing all neurons containing the corresponding TF, AVS, and RO.

TABLE VII
SEGMENT CONTRIBUTION

	Upstream($i+1$)	Prediction(i)	Downstream($i-1$)
σ	0.144	0.185	0.050

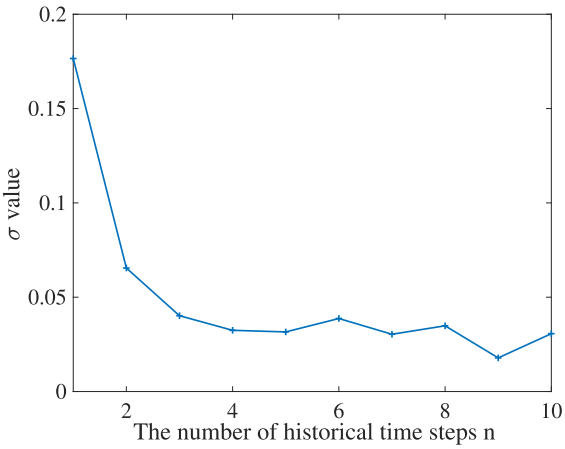


Fig. 4. Temporal influence based on the σ value.

Table VII demonstrates that the variables from the upstream segment show a comparable level of importance as the ones from the prediction segment, while the contribution of the downstream segment is shown to be much lower. This analysis result can be considered as an indication that most of the TF in the predicted future time step comes from the upstream segment, and the vehicles in the downstream segment have significantly less impact under the given scenario.

• *Temporal influence:* In the prediction formula (5), we can find that the number of historical time steps n is a tuning parameter that reflects the temporal influence. A large value of n means that the historical data far away from the current predicted time step are also taken into account. Fig. 4 depicts the influence of n, where $n = 10$ represents the furthest time step and $n = 1$ denotes the closest one from the current prediction.

We can see from Fig. 4 that the influence decreases when the historical time steps are selected farther away from the current prediction time step, which is in accordance with practical scenarios, i.e., the vehicles passing through a long time ago are unlikely to pose significantly impacts when predicting the TF at future time steps. It can also be noticed that the temporaral effect becomes lower with increasing n. Thus, it can be regarded as an indication for determining the number of previous time steps in the historical data. These experiments show that considering the historical data from the previous $n = 10$ time steps is sufficient to take the temporal information into account.

3) Extended Scenarios: We enrich the modelling in (5) by deploying multiple detectors in both upstream and downstream

TABLE VIII
SEGMENT CONTRIBUTION BASED ON THE σ VALUE

T	Upstream			Pred	Downstream		
	$i+3$	$i+2$	$i+1$	i	$i+1$	$i+2$	$i+3$
1	0.071	0.115	0.113	0.207	0.066	0.048	0.035
3	0.108	0.116	0.113	0.209	0.133	0.136	0.113
6	0.156	0.156	0.155	0.190	0.051	0.266	0.182

observation segments, each of which incorporates the information from 3 individual detectors, i.e., the prediction model $f_{\text{EHHNN}}(y^t_{o_i}, s^t_{o_i}, p^t_{o_i})$ in (5) is modified into

$$o_i \in \mathcal{I}_i, \quad \mathcal{I}_i = \{i+3, i+2, i+1, i, i-1, i-2, i-3\}, \tag{17}$$

where a larger value of $|o_i - i|$ indicates a longer distance of the corresponding detector from the prediction segment. The indices of the selected detectors are 311974, 318282, 314270, 314559, 319631, 313852, 313658 in PeMS. Then, we conduct the experiment with the same settings as Section IV-B.2, and now the total number of candidate variables is $3 \cdot 7 \cdot 10 = 210$. Considering the performance of the compared methods, we then simply report the prediction accuracy with ASTGCN, i.e., the RMSE results of EHHNN/ASTGCN are **0.032**/0.058 ($T = 1$), **0.055**/0.060 ($T = 3$) and **0.057**/0.058 ($T = 6$), further verifying the effectiveness of the proposed method in prediction accuracy. Similarly, with multiple upstream and downstream segments, spatial analysis can be conducted. Table VIII shows the relative importance of the selected segments, where "pred" is abbreviated for the prediction segment.

Table VIII shows a similar tendency to that in Table VII under the same prediction step $T = 1$, where the detectors in the upstream segments are more influential to the prediction output and those in the downstream segments have lower impact. When the prediction step T increases, meaning that we aim to predict the TF a bit farther away from the current time step ($T = 6$), the influence of faraway upstream segments, i.e., $i+3$ and $i+2$, shows to be higher, since it takes time for the vehicles in faraway upstream segments ($i+3$ and $i+2$) to arrive in the prediction segment. Moreover, vehicles normally move from upstream to downstream, so that the vehicles in the prediction segment are expected to be the ones that also passed the upstream segments in previous time steps, whereas this extended scenario in Table VIII shows that the information from downstream segments can also affect the TF at a bit faraway future time step, such as $T = 6$. One inference could be expected that there are likely traffic congestions in the downstream segments and a backpropagation of delay on the TF may occur when performing the prediction at later future time steps. This experiment provides versatile information of different road segments relating to TF, AVS, and RO that may cause traffic congestions in the prediction location under different road configurations in traffic systems, which is promising to facilitating the design of proper strategies to control the TF in the influential segments at proper time steps ahead.

Besides the spatial influence, we can also explore a similar interpretation from the temporal aspect. Fig. 5 shows the

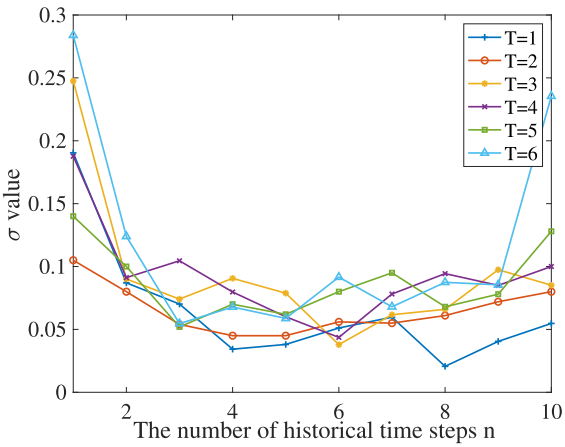


Fig. 5. Temporal influence based on the σ value with different prediction horizons.

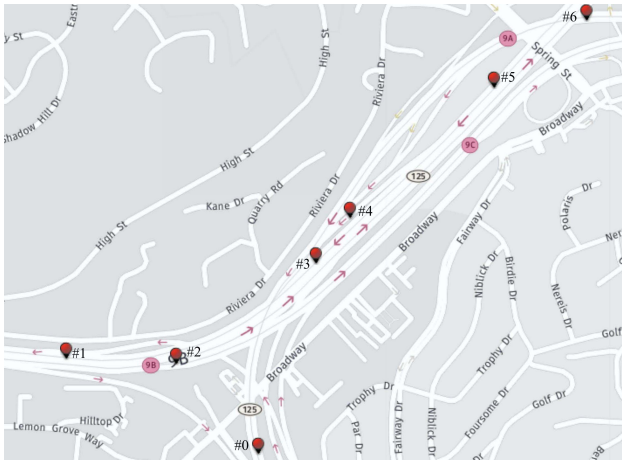


Fig. 6. Distribution of the selected detectors, where the information from detectors in detectors #0, . . . , #6 is incorporated into the TF prediction in #0.

results. We can see that the influence of those variables with larger lagged values becomes more distinctive when predicting the TF at a relatively faraway time step (larger T). Combined with the above Table VIII, we can see from the analysis that more spatial-temporaral information could be needed if a longer prediction horizon (larger T) is set, and the proposed interpretation analysis based on EHHNNs can provide a suggestion to detect such information for variable selection.

C. TF Prediction on Transportation Intersections

The proposed method can be easily adapted to different scenarios to perform TF prediction and analysis. In this subsection, we introduce another practical scenario in transportation systems as a case study to further demonstrate the effectiveness and extendability of the proposed method. Instead of the road segment, we conduct the TF prediction on a mainline nearby a transportation intersection, which is connected with multiple roads. The data are also collected from PeMS, and the distribution of the selected detectors is shown in Fig. 6.

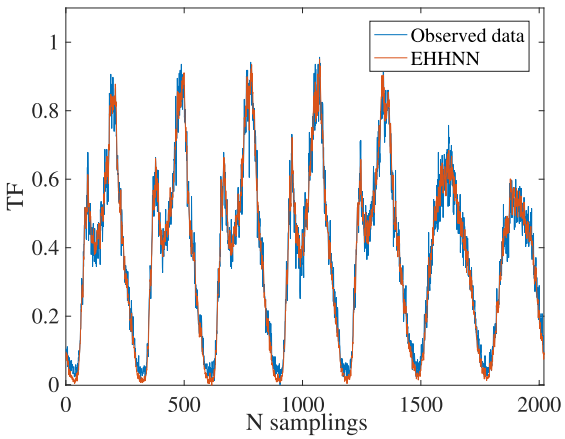


Fig. 7. An illustration of the prediction result on detector #0 with the EHHNN, where the testing MAE, RMSE and R^2 are 0.029, 0.036 and 0.977, respectively.

In Fig. 6, the indices of detectors regarding detectors #0, . . . , #6 are 1113138, 1113173, 1113181, 1113126, 1113297, 1113301, and 112021, respectively, in which the data are also sampled from March 11th to April 7th of the year 2019. Similar with the settings in Section IV-B, we employ the EHHNN to predict the TF, and Fig. 7 gives an illustration on the prediction results with prediction step $T = 1$.

Fig. 7 shows that the used EHHNN can well predict the TF nearby a traffic intersection with high volume and more complicated components influencing the TF, which further indicates the effectiveness of the proposed method. We also briefly compare the results with ASTGCN, i.e., the RMSE of EHHNN/ASTGCN is **0.036**/0.054 ($T = 1$), **0.048**/0.055 ($T = 3$), and 0.069/**0.057** ($T = 6$).

Similarly, the relative importance of those detectors can also be investigated, and the results are given in Table IX. It can be seen that the information from detector #5 shows to be more influential than other selected detectors, including the prediction detector, indicating that a large portion of the TF in #0 can be possibly coming from the roads monitored by detector #5. Moreover, the importance of those selected detectors varies when prediction time step T changes, and this can be due to the fact that the distances between these selected detectors and the prediction detector are different, and the corresponding TF from them can arrive at the prediction area in different future time steps. Such analysis results can be potentially adopted to prevent future traffic congestions in the prediction area, such as offering informative guidance to drivers heading towards the prediction area and providing useful information to the traffic controllers nearby. For instance, we could limit the TF in the more influential areas ahead of time, according to our analysis, such as limiting the TF in the areas nearby detectors #5, #2 and #1 to support the optimization of traffic control.

In the case study of Fig. 6, temporal analysis can also be conducted analogously to demonstrate the influence of variables from different historical time steps, in which a similar tendency is obtained with Fig. 5. This analysis can facilitate to determine the number of historical time steps n, when different prediction steps T are required. More specifically, the most influential variables are presented in Table X with different prediction time steps T.

TABLE IX

DETECTOR CONTRIBUTION BASED ON THE σ VALUE

T	#0	#1	#2	#3	#4	#5	#6
1	0.066	0.025	0.017	0.020	0.021	0.173	0.008
3	0.075	0.090	0.067	0.041	0.045	0.165	0.025
6	0.132	0.147	0.130	0.141	0.107	0.209	0.049

TABLE X

VARIABLE INFLUENCE BASED ON THE σ VALUE WITH DIFFERENT PREDICTION TIME STEP T

$T = 1$		$T = 3$		$T = 6$	
Variable	σ	Variable	σ	Variable	σ
y_5^{t-1}	0.132	y_5^{t-1}	0.063	y_5^{t-1}	0.059
y_0^{t-1}	0.025	y_0^{t-1}	0.033	y_1^{t-1}	0.033
y_5^{t-2}	0.019	y_5^{t-2}	0.026	y_1^{t-2}	0.028
y_4^{t-1}	0.015	y_4^{t-1}	0.023	y_5^{t-2}	0.023
y_0^{t-2}	0.013	y_1^{t-1}	0.018	y_0^{t-1}	0.016

The analysis from our method indicates that the TF in detector #5 is the most influential for each $T = 1, 3, 6$, given the current considerations. When T is increasing, the impact of historical TF in the prediction observation detector decreases and the more influential variables are mainly from other detectors and different historical time steps. This analysis can be helpful to explore how the TF in the prediction detector is formulated and how to adjust the TF in different roads when dealing with the corresponding traffic congestion or cooperative control. Overall, to control the future TF in #0, one can adjust the TF of the roads nearby #5 in advance. When predicting the TF after 5 minutes ($T = 1$) nearby #0, the TF around #4 can also be taken into account, while it has less influence when predicting the TF after 30 minutes ($T = 6$). Besides, for route planning guidance to drivers, such analysis result is also useful. For instance, when the drivers arrive close to #5 and encounter intensive congestion here at the current moment, they can be suggested not to drive towards #0, which may also suffer a high volume of TF at the upcoming time steps. Although the aforementioned analysis is not always guaranteed to correspond exactly to the ground truth under different real-world scenarios, it can be seen as a promising attempt to perform accurate and more interpretable TF prediction.

V. CONCLUSION

In this paper, short-term TF prediction is addressed together with interpretation analysis, and thus a pragmatic method is introduced based on the recently proposed EHHNN, which is a universal approximator and yet has quite simple architectures of sparse neuron connections. In the proposed method, the EHHNN is firstly applied for accurate TF prediction by utilizing its model flexibility. Compared with black-box models, the EHHNN predictor can be performed with ANOVA decomposition to quantitatively detect the contributions of different variables. On such basis, we further extend the ANOVA

decomposition of EHHNNs with specifications to traffic data, and then perform variable selection and varied interpretation analysis covering traffic factors, road segments (spatial) as well as historical time steps (temporal) of specific traffic variables, presenting a more interpretable TF prediction process. Thus, the resulting EHHNN is not only an accurate predictor for TF herein, but also an effective tool to perform interpretation analysis to approach more in-depth understanding, which is promising to provide versatile information to future traffic control. Numerical experiments confirm the effectiveness and potentials of the proposed method under different settings.

For future work, it would be interesting to investigate the impacts of other traffic factors and possibly their assistance to provocative traffic control. More analysis based on the interpretability of EHHNNs is also worthy of further investigations.

ACKNOWLEDGMENT

This paper reflects only the authors' views and the union is not liable for any use that may be made of the contained information.

REFERENCES

[1] L. A. Klein, M. K. Mills, D. Gibson, and L. A. Klein, "Traffic detector handbook," Federal Highway Admin., MaLean, VA, USA, Tech. Rep. FHWA-HRT-06-108, 2006, vol. 2.

[2] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016.

[3] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Dec. 2015.

[4] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proc. Roy. Soc. London A, Math. Phys. Sci.*, vol. 229, no. 1178, pp. 317–345, 1955.

[5] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transp. Res. B, Methodol.*, vol. 28, no. 4, pp. 269–287, Aug. 1994.

[6] S. Lin, T. Pan, W. Lam, R. Zhong, and B. D. Schutter, "Stochastic link flow model for signalized traffic networks with uncertainty in demand," in *Proc. 15th IFAC Symp. Control Transp. Syst.*, 2018, vol. 51, no. 9, pp. 458–463.

[7] H. Su, L. Zhang, and S. Yu, "Short-term traffic flow prediction based on incremental support vector regression," in *Proc. 3rd Int. Conf. Natural Comput. (ICNC)*, vol. 1, Aug. 2007, pp. 640–645.

[8] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6164–6173, 2009.

[9] X. Lv *et al.*, "Short-term power load forecasting based on balanced KNN," in *Proc. IOP Conf., Mater. Sci. Eng.*, 2018, vol. 322, no. 7, Art. no. 072058.

[10] Y. Gong and Y. Zhang, "Research of short-term traffic volume prediction based on Kalman filtering," in *Proc. 6th Int. Conf. Intell. Netw. Intell. Syst. (ICINIS)*, Nov. 2013, pp. 99–102.

[11] T. Thomas, W. Weijermars, and E. Van Berkum, "Predictions of urban volumes in single time series," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 71–80, Mar. 2010.

[12] 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, 1996.

[13] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transp. Res. Rec.*, vol. 1678, no. 1, pp. 179–188, 1999.

[14] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.

[15] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[16] B. Du *et al.*, "Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 972–985, Mar. 2020.

[17] Y. Liang *et al.*, "Revisiting convolutional neural networks for citywide crowd flow analytics," in *Proc. ECML-PKDD*, 2020, pp. 578–594.

[18] B. Du, X. Hu, L. Sun, J. Liu, Y. Qiao, and W. Lv, "Traffic demand prediction based on dynamic transition convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1237–1247, Feb. 2021.

[19] Y. Zhang, S. Wang, B. Chen, J. Cao, and Z. Huang, "TrafficGAN: Network-scale deep traffic prediction with generative adversarial nets," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 219–230, Jan. 2021.

[20] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3634–3640.

[21] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "GeoMAN: Multi-level attention networks for geo-sensory time series prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3428–3434.

[22] L. Zhao, Y. Song, C. Zhang, and Y. Liu, "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[23] L. N. N. Do, H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung, "An effective spatial-temporal attention based neural network for traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 108, pp. 12–28, Nov. 2019.

[24] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 1234–1241.

[25] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 922–929.

[26] S. Du, T. Li, X. Gong, Y. Yang, and S. J. Horng, "Traffic flow forecasting based on hybrid deep learning framework," in *Proc. 12th Int. Conf. Intell. Syst. Knowl. Eng. (ISKE)*, Nov. 2017, pp. 1–6.

[27] L. Chen, L. Zheng, J. Yang, D. Xia, and W. Liu, "Short-term traffic flow prediction: From the perspective of traffic flow decomposition," *Neurocomputing*, vol. 413, pp. 444–456, Nov. 2020.

[28] T. Ma, C. Antoniou, and T. Toledo, "Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast," *Transp. Res. C, Emerg. Technol.*, vol. 111, pp. 352–372, Feb. 2020.

[29] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 6910–6920, Nov. 2021.

[30] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 4, pp. 606–616, 2011.

[31] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.

[32] Y.-A. Daraghmi, C.-W. Yi, and T.-C. Chiang, "Negative binomial additive models for short-term traffic flow forecasting in urban areas," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 784–793, Apr. 2014.

[33] 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, no. 1, pp. 3–19, 2014.

[34] Y. Li and C. Shahabi, "A brief overview of machine learning methods for short-term traffic forecasting and future directions," *SIGSPATIAL Special*, vol. 10, no. 1, pp. 3–9, Jun. 2018.

[35] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 53–72, May 2009.

[36] J. Xu, Q. Tao, Z. Li, X. Xi, J. A. K. Suykens, and S. Wang, "Efficient hinging hyperplanes neural network and its application in nonlinear system identification," *Automatica*, vol. 116, Jun. 2020, Art. no. 108906.

[37] J. Xu, X. Huang, and S. Wang, "Adaptive hinging hyperplanes and its applications in dynamic system identification," *Automatica*, vol. 45, no. 10, pp. 2325–2332, Oct. 2009.

[38] J. H. Friedman, "Multivariate adaptive regression splines," *Annu. Statist.*, vol. 19, no. 1, pp. 1–67, Mar. 1991.

[39] I. Lind and L. Ljung, "Regressor and structure selection in narx models using a structured anova approach," *Automatica*, vol. 44, no. 2, pp. 383–395, 2008.

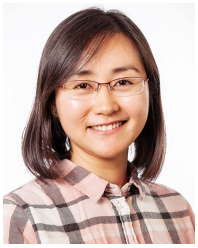
[40] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with Conv-LSTM," in *Proc. 9th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2017, pp. 1–6.

[41] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.
[42] L. Breiman, "Hinging hyperplanes for regression, classification, and function approximation," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 999–1013, May 1993.
[43] S. Wang and X. Sun, "Generalization of hinging hyperplanes," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4425–4431, Dec. 2005.
[44] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[45] R. Kumar Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," 2015, *arXiv:1505.00387*.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[47] R. S. Kar, Z. Miao, M. Zhang, and L. Fan, "ADMM for nonconvex AC optimal power flow," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2017, pp. 1–6.
[48] L. Breiman, "Stacked regressions," *Mach. Learn.*, vol. 24, no. 1, pp. 49–64, Jul. 1996.
[49] W. Feng, H. Chen, and Z. Zhang, "Short-term traffic flow prediction based on wavelet function and extreme learning machine," in *Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Nov. 2017, pp. 531–535.
[50] K.-L. Li, C.-J. Zhai, and J.-M. Xu, "Short-term traffic flow prediction using a methodology based on ARIMA and RBF-ANN," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 2804–2807.
[51] X. Liu, M. Zhang, and G. Xu, "Construction of distributed LDoS attack based on one-dimensional random walk algorithm," in *Proc. IEEE 2nd Int. Conf. Cloud Comput. Intell. Syst.*, vol. 2, Oct./Nov. 2012, pp. 685–689.

Qinghua Tao received the B.S. degree from Central South University, China, in 2014, and the Ph.D. degree from Tsinghua University, China, in 2020. She is currently a Post-Doctoral Researcher with ESAT-STADIUS, KU Leuven, Belgium. Her research interests include the analysis and applications of piecewise linear neural networks, machine learning, dynamic systems, and optimization.

Zhen Li received the B.S. degree in automation from Changzhou University, Changzhou, China, in 2017, and the M.S. degree in control science and engineering from the Harbin Institute of Technology (Shenzhen), Shenzhen, China. His current research interests include ensemble and boosting machine learning methods, piecewise linear neural networks, and their applications.

Jun Xu (Member, IEEE) received the B.S. degree in control science and engineering from the Harbin Institute of Technology, China, in 2005, and the Ph.D. degree in control science and engineering from Tsinghua University, China, in 2010. She is currently an Associate Professor with the School of Mechanical Engineering and Automation, Harbin Institute of Technology (Shenzhen), Shenzhen, China. Her research interests include piecewise linear functions and their applications in machine learning, as well as nonlinear systems identification and control.



Shu Lin (Member, IEEE) received the M.Sc. degree in control science and engineering from Shandong University, Jinan, China, in 2006, and the Ph.D. degree from the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands, in 2011. From 2011 to 2013, she was a Post-Doctoral Researcher with the Department of Automation, Shanghai Jiao Tong University, Shanghai, China. Currently, she is an Associate Professor at the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China. Her research interests mainly include modeling, optimization, and control for intelligent traffic networks.

Bart De Schutter (Fellow, IEEE) is a Full Professor at the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. His current research interests include intelligent transportation and infrastructure systems, hybrid systems, and multi-level control. He is a Senior Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Johan A. K. Suykens (Fellow, IEEE) was born in Willebroek, Belgium, in May 1966. He received the M.S. degree in electro-mechanical engineering and the Ph.D. degree in applied sciences from Katholieke Universiteit Leuven, Leuven, Belgium, in 1989 and 1995, respectively.

In 1996, he was a Visiting Post-Doctoral Researcher with the University of California at Berkeley, Berkeley, CA, USA. He has been a Post-Doctoral Researcher with the Fund for Scientific Research FWO Flanders, Belgium. He is currently a Professor (Hoogleraar) with KU Leuven, where he is also serving as the Program Director for master AI. He is the author of the books: *Artificial Neural Networks for Modelling and Control of Non-linear Systems* (Kluwer Academic Publishers) and *Least Squares Support Vector Machines* (World Scientific), a coauthor of the book: *Cellular Neural Networks, Multi-Scroll Chaos and Synchronization* (World Scientific), and an Editor of the books: *Nonlinear Modeling: Advanced Black-Box Techniques* (Kluwer Academic Publishers) and *Advances in Learning Theory: Methods, Models, and Applications* (IOS Press). He has been elevated as an IEEE Fellow in 2015 for developing least squares support vector machines. He has been awarded the ERC Advanced Grant 2011 and 2017. He was a recipient of the International Neural Networks Society INNS 2000 Young Investigator Award for significant contributions in the field of neural networks. He received the IEEE Signal Processing Society 1999 Best Paper (Senior) Award and several best paper awards at international conferences. In 1998, he organized an International Workshop on Nonlinear Modeling with Time-Series Prediction Competition. He has served as the Director and an Organizer for the NATO Advanced Study Institute on Learning Theory and Practice (Leuven, in 2002), a Program Co-Chair for the 2004 International Joint Conference on Neural Networks and the 2005 International Symposium on Nonlinear Theory and its Applications, an Organizer for the 2007 International Symposium on Synchronization in Complex Networks, a Co-Organizer for the NIPS 2010 Workshop on Tensors, Kernels and Machine Learning, and the Chair for the International Workshop on Advances in Regularization, Optimization, Kernel Methods, and Support Vector Machines: Theory and Applications (ROKS) 2013. He has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1997 to 1999 and 2004 to 2007 and the IEEE TRANSACTIONS ON NEURAL NETWORKS from 1998 to 2009.