**SEYED HADI HOSSEINI**, M.Sc. (Ph.D. Candidate)
E-mail: hosseini@iauet.ac.ir
Department of Electrical Engineering
Science and Research Branch
Islamic Azad University
Pounak, Tehran, P.O. Box 14515-775, Iran
**BEHZAD MOSHIRI**, Ph.D.
E-mail: moshiri@ut.ac.ir
**ASHKAN RAHIMI-KIAN**, Ph.D.
E-mail: arkian@ut.ac.ir
**BABAK NADJAR ARAABI**, Ph.D.
E-mail: araabi@ut.ac.ir
Control and Intelligent Processing Center of Excellence
School of Electrical and Computer Engineering
University of Tehran
North Kargar, P.O. Box 14395-515, Tehran, Iran

# TRAFFIC FLOW PREDICTION USING MI ALGORITHM AND CONSIDERING NOISY AND DATA LOSS CONDITIONS: AN APPLICATION TO MINNESOTA TRAFFIC FLOW PREDICTION

*ABSTRACT*

*Traffic flow forecasting is useful for controlling traffic flow, traffic lights, and travel times. This study uses a multi-layer perceptron neural network and the mutual information (MI) technique to forecast traffic flow and compares the prediction results with conventional traffic flow forecasting methods. The MI method is used to calculate the interdependency of historical traffic data and future traffic flow. In numerical case studies, the proposed traffic flow forecasting method was tested against data loss, changes in weather conditions, traffic congestion, and accidents. The outcomes were highly acceptable for all cases and showed the robustness of the proposed flow forecasting method.*

*KEY WORDS*

*ITS; mutual information; traffic flow forecasting; noisy data; data loss;*

## 1. INTRODUCTION

### 1.1 Statement of the problem

In many countries, transportation problems and traffic jams are social challenges that increasingly demand national resources. Slow traffic on urban and rural highways increases air pollution and fuel consumption. An intelligent transportation system (ITS) is an advanced application that provides innovative services for different modes of traffic management. The term ITS was first introduced as an umbrella term to cover all technologies in information technology, communications, and control [1]. A solution to prevent traffic congestion using ITS is one that predicts traffic parameters such as traffic flow, speed, and density.

### 1.2 Literature review

Traffic flow is a macroscopic feature of traffic and is a real-time, completely non-linear, high-dimensional and non-stationary stochastic process [2]. Traffic flow forecasting is mainly classified into long-term and short-term prediction. In short-term prediction, traffic flow is forecast in the immediate future (typically 5 to 30 min.) on the basis of real-time online or historical data. Traffic conditions may vary from one moment to another in response to changing weather conditions, road accidents, cultural or political occasions and events, the types of vehicles, and driver characteristics [3]. In the present study, if traffic flow is not seriously hampered by these conditions and congestion has not occurred, traffic conditions are considered to be normal. Traffic flow on a normal day is similar to that on other days without sudden congestion. On an abnormal day, sudden congestion occurs as the result of an accident, heavy precipitation, or storms. In the past

few years, simple and complex methods have been proposed for predicting traffic flow under different traffic conditions. Experimental and data-driven prediction methods of traffic flow can be divided into parametric, non-parametric, and hybrid methods, each with its own advantages and disadvantages. Most research on traffic flow prediction has been implemented under normal traffic conditions and research has seldom addressed special traffic conditions such as climate, the existence of noise in the data, and disturbances on the highways [5].

*Parametric prediction methods*: Predictive parametric methods are modelled using time series of recorded data that predict the immediate future in steps. One advantage of these methods is their low prediction error. The major disadvantage is their frequently poor performance in the presence of noise and disturbance. Linear regression is a parametric prediction method that can predict the next variable online using real data. Sun et al. [5] have shown that this method depends on the rate at which data are recorded to predict the speed of traffic. Another parametric method is the historical mean average model. It shows poor performance in unpredictable circumstances of traffic because of its high dependence on recorded data [6]. The maximum likelihood (ML) model is robust for sensor failures and rapid change in conditions [7]. Despite the benefits of the exponential smoothing method for predicting traffic flow, it is very difficult to determine constant convergence for the model during major changes in traffic flow [8]. The simplicity and strong potential of time series models for online operation makes these models popular for most traffic predictions. One disadvantage of the time series model is determination of the degrees for the auto-regressive (AR) and moving average (MA) models for designing an accurate forecasting model. The second disadvantage is their high dependence on input data. Incomplete or inaccurate input data produce an inaccurate time series model, resulting in an incorrect prediction [8]. Time series models used to predict traffic flow include the auto-regressive integrated moving average (ARIMA) [9], seasonal ARIMA [10], vector ARIMA [11], and ARIMA with EXtra (ARIMAX) [12].

*Non-parametric prediction methods*: Non-parametric prediction methods predict traffic flow in proportion to road conditions and are enhanced using modern models rather than classic models. The complexity of these models and their strong dependence on large volumes of data are major disadvantages of these methods. The most popular non-parametric methods are neural networks, such as the multi-layer perceptron (MLP), radial basis function (RBF) and time delay neural network (TDNN) [13]. The biggest flaw in neural networks is the type of training required and their need to handle large volumes of

data to train the network weights. Fuzzy [14], k-nearest neighbour (KNN) [15], support vector machine [16-17], Bayesian networks [18] and wavelet [19] are other parametric methods used to predict traffic flow.

*Hybrid prediction methods*: Each of the above models responds well individually only under specific traffic conditions; when the conditions change, their efficiency decreases for forecasting traffic flow. In recent years, researchers have used linear and non-linear hybrids of parametric and non-parametric methods to increase traffic flow prediction accuracy. The accuracy of hybrid methods depends on the type of parametric and non-parametric methods used; however, the computational complexity and developmental costs of hybrid methods exceed the individual parametric and non-parametric methods [8]. Some of these methods are combinations of a neural network model and models such as the genetic algorithm (GA) [20], fuzzy [3], wavelet [4] and ARIMA [21] models. Although the hybrid methods have high accuracy for predicting traffic flow, their dependency on the type and volume of recorded data is a disadvantage. Their use requires the selection of good data from input data.

## 1.3 Contribution and structure of the paper

A general well-defined, robust, highly efficient and comprehensive method has yet to be developed to forecast traffic flow and deliver an accurate response for all aspects of traffic [8]. The high volume of data combined with inaccurate and noisy data is a major disadvantage of these three prediction methods. The present study proposes a new forecasting method that tolerates changes in traffic and road congestion. The proposed method can detect inaccurate, noisy and faulty input traffic flow data and uses a newly-developed data selection method to extract data so that only the most informative data are used.

Most studies compare the quality of a new prediction method with previous models such as ARIMA and a neural network model as parametric and non-parametric methods, respectively [1]. The historical mean average model is a time series model with fixed and equal weights, ARIMA is an accurate and useful time series model, and MLP as a popular neural network model; these models were selected for comparison with the proposed model. All of these models have been widely used in previous studies because of their accurate results.

Section two presents the MI algorithm and the best procedures for input feature selection. Section three discusses the data and the methods used for forecasting. Section four presents the results of the simulations and analyses of traffic data. Section five presents concluding remarks.

## 2. INPUT DATA SELECTION

An appropriate method for decreasing prediction error covariance is to decrease the dependence on observation error. Mutual information (MI) theory quantifies dependence on such error. It is similar to the cross-correlation method used to calculate the non-linear correlation between two quantities [22, 23]; MI calculates the non-linear correlation between two non-linear quantities. Input feature selection is an important aspect of data classification. The mutual information feature selection (MIFS) algorithm calculates the MI between the input data and the best data selected for prediction to decrease the computation load and increase computation accuracy. This makes the MI a very reasonable method for forecasting the traffic flow [24].

### 2.1 Estimation of MI using KNN method

In control systems, MI is used to measure non-linear interdependence of two random variables. The amount of MI determines the amount of information on random variable $X$ that was obtained from random variable $Y$ and is denoted as $I(X;Y)$. The calculation of MI helps to decrease the uncertainty of $X$ while $Y$ exists. MI is defined in Equation 1 in discrete form [25].

$$I(X;Y) = \sum_x \sum_y P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} \quad (1)$$

$P_{XY}(x,y)$ represents the probability density function for $X$ and $Y$ random variables. To calculate $I(X;Y)$, the $P_{XY}(x,y)$ function must be known; however, this function is unknown here and therefore must be estimated. Methods used to estimate $P_{XY}(x,y)$ include the Bayesian, ML, wavelet and KNN. Kraskov et al. [25] selected the KNN to design an estimator for the probability density function and estimate the amount of MI on the basis of the observed random data. MI estimation using the KNN method is based on Kraskov et al. as [25]:

$$\hat{I}(X;Y) = \Psi(K) - \frac{1}{K} - \frac{1}{n}\sum_{i=1}^{n}(\Psi(n_i^x) + \Psi(n_i^y)) + \Psi(n) \quad (2)$$

$$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du \quad, \quad \Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \frac{d}{dx}\ln\Gamma(x) \quad (3)$$

where $\Gamma(x)$ and $\Psi(x)$ represent the gamma function and the digamma function, respectively; $K$ is a fixed positive integer used to calculate the distance of the $K^{th}$ nearest neighbour; $n$ is the maximum number of pieces of data; and $n_i^x$ and $n_i^y$ are the numbers of sample points located in the neighbourhood of $X_i$ and $Y_i$, respectively.

The most important aspect of the KNN estimator is determination of the value of $K$. Because there are different values of $K$ and estimates for $I(X;Y)$, when $K$ is a small number, the estimate of $I(X;Y)$ will have low bias but large variance. When $K$ is a large number, it will have large bias but low variance [26]. For selecting the best value of $K$, a good rule of thumb is $K = \sqrt{n}$ [27]. As explained later for forecast Method E, the maximum amount of data for simulations is about $n = 38$. This means that the best value for $K$ is six.

### 2.2 MIFS algorithm

In real systems, a significant degree of uncertainty occurs in identification system output from the inadequacy of the initial data or suboptimal conditions in the system. In this study, inadequate initial data in the traffic systems were ruled out, thus, uncertainty can be attributed to the sensing system or noise pollution. Selecting appropriate data from a large pool of data is a working solution for this problem that can be accomplished using MI.

An algorithm must be used to select the best input data as an optimal subset of initial candidate input data. Battiti's MIFS algorithm is notable for the selection of efficient inputs. In the MIFS algorithm, the aim is to obtain a relationship between the inputs and the output to decrease the existing redundancy in the input data and at the same time select the data with the best relationship for the output. The goal is to select a subset of $m$ classes from an input set with $n$ data classes ($m < n$) having the highest level of relationship with the input set [28]. For a large value of $m$, computational complexity increases; for a small value of $m$, the accuracy decreases. A good rule of thumb for the number of $m$ outcomes is $m = n/4$. Here, the maximum amount of data in the simulations was $n = 38$, so $m = 10$ was considered [29].

Suppose that $T$ is the output set, $S$ is the empty set, and $l_i$ is the distinct input class that belongs to the $n$-member $L = \{l_i\}$ input set, $i = 1, 2, ..., n$. The aim is to obtain the MI amount for each input $l_i$ member of $L$ and output $T$; $I(l_i;T)$. The $l_j$ input for ($1 \leq j \leq n$ that maximizes $I(l_i;T)$) is selected and separated from set $L$ and added to set $S$ as the first input selection ($s_1 = l_j$). The $I(l_i;s_1)$ is computed for all pairs of variables ($l_i,s_1$) with an $l_i$ member of $L$ and ($i \neq j$). The $l_i$ input that maximizes the subsequent term is selected then separated from the set $L$ and added into set $S$ [28].

$$\max_i \{I(l_i;T) - \beta \times I(l_i;s_1)\} \quad (4)$$

This is repeated until all variables of $S$ are selected and $I(l_i;s_m)$ with $l_i$ member of $L$ and ($i \neq j$) and $s_m$ member of $S$ is computed to maximize the subsequent term; in each repetition, $m = 10$.

$$\max_i \left\{ I(l_i;T) - \beta \times \sum_m I(l_i;s_m) \right\} \quad (5)$$

The important parameter in the MIFS algorithm is $\beta$, which shows the augmentation between inputs. If $\beta = 0$, the algorithm looks for inputs with the greatest

correlation only with the output, making augmentation between the inputs redundant. As $\beta$ increases, the augmentation between inputs decreases gradually but the correlation between them and the output gradually lose effect. Neither a large nor a small value is appropriate for $\beta$. The main challenge to using the MIFS algorithm is finding the value of $\beta$ that is different for different non-linear systems. For data classification purposes, $0.5 \leq \beta \leq 1$ was the most appropriate. A series of numbers from 0.5 to 10, in increments of 0.5 was generated for $\beta$. The training and testing data (80% and 20% of input data, respectively) and the MIFS algorithm were used to predict the traffic flow using the averaging method. The best value ($\beta = 0.6$) was obtained for minimum prediction error with using the training and testing data.

## 3. TRAFFIC FLOW FORECASTING DATA AND METHODS

The aim of forecasting is to survey the performance of the MIFS algorithm in the presence of different data types, especially normal data, incomplete data (failure detectors) and noisy data (occurrence of accidents, rain or snow and heavy traffic). Since current data is inexact, using the MIFS algorithm allows the detection or omission of the false data and will increase the forecasting accuracy. Simulations using MATLAB software were carried out to determine the performance of the MIFS algorithm when choosing pre-eminent data to decrease the amount of data and increase the accuracy of predictions.

### 3.1 Data source

The data were collected from sensors installed on the highway network of the metro area in Minnesota, USA, and were provided by the Transportation Data Research Laboratory [30]. The traffic data were taken from three detectors at station 286 located on Highway I-394, as shown in *Figure 1*. The traffic flow data used for forecasting were collected during the first six months of 2012 at 15 min. intervals in real time. About 96 pieces of data were collected every day. Data for holidays and working days were separated and data for non-holiday days were used for simulations and forecasting.

No data were saved on 2 January (Monday) and on 19 April (Thursday) because of technical failure and malfunctioning detectors; all data from those days equalled zero. On 29 February (Wednesday) heavy snow resulted in heavy traffic. On 1 June (Friday), 3 April (Tuesday), 7 March (Wednesday) and 16 March (Friday) heavy traffic and congestion occurred; congestion on the last two days was likely the result of an accident on the road. This analysis is based on the data specified in *Table 1*.

*Table 1 - Traffic congestion on specific days in first six months of 2012 [30, 31]*

| Day | Date | Problem Description |
|-----|------|---------------------|
| Monday | 2nd January | Detector failure (incomplete data) |
| Wednesday | 29th February | Heavy traffic (heavy snow) |
| Wednesday | 7th March | Heavy traffic (congestion) |
| Friday | 16th March | Heavy traffic (congestion) |
| Tuesday | 3rd April | Heavy traffic (probably an accident) |
| Thursday | 19th April | Detector failure (incomplete data) |
| Friday | 1st June | Heavy traffic (probably an accident) |

### 3.2 Prediction models

*Mean average model*: The first model is simply the average of all input data used to predict the next step ahead (subsequent 15 min. interval).

*ARIMA model*: This is a common regression model intended for obtaining the relationship between past and future data. Traffic flow has an erratic variance and is a non-static process that can be modelled using the ARIMA time series. The mathematical model of ARIMA(p,d,q) consists of polynomials AR and MA as shown in Equation 6 [32]. ARIMA(2,2,0) uses data from the past and the time series of traffic, as shown in Equation 8:

$$\phi(B)\nabla^d X_t = \theta(B)\varepsilon_t \tag{6}$$

$$\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \ldots - \phi_p B^p,$$
$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \ldots - \theta_q B^q \tag{7}$$

$$\hat{X}_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \phi_4 X_{t-4} + \varepsilon_t,$$
$$\varepsilon_t \sim N(0,1) \tag{8}$$

where $B$ is the delay factor and is defined as $BX_t = X_{t-1}$; $\nabla$ is the difference and is defined as $\nabla = (1 - B)$; $d$ is the degree of difference and is defined as $\nabla^d = (1 - B)^d$; $\varepsilon_t$ is the error rate at time $t$ and is considered to be white noise. Also, $\phi(B)$ is a polynomial of AR; $\theta(B)$ is a polynomial expression of MA; and $p$ and $q$ are the degrees of the polynomials, respectively.

*MLP model*: This neural network predictive model is composed of input, intermediate and output layers. The size of the data set selected by the MIFS algorithm chosen was ten, thus, the number of input layer neurons is assumed to be ten. The number of hidden layer neurons of the neural network equals the number in the input layer (ten), and the number of output layer neurons for simplicity is set at one. The training algorithm used for the neural network is the Levenberg-
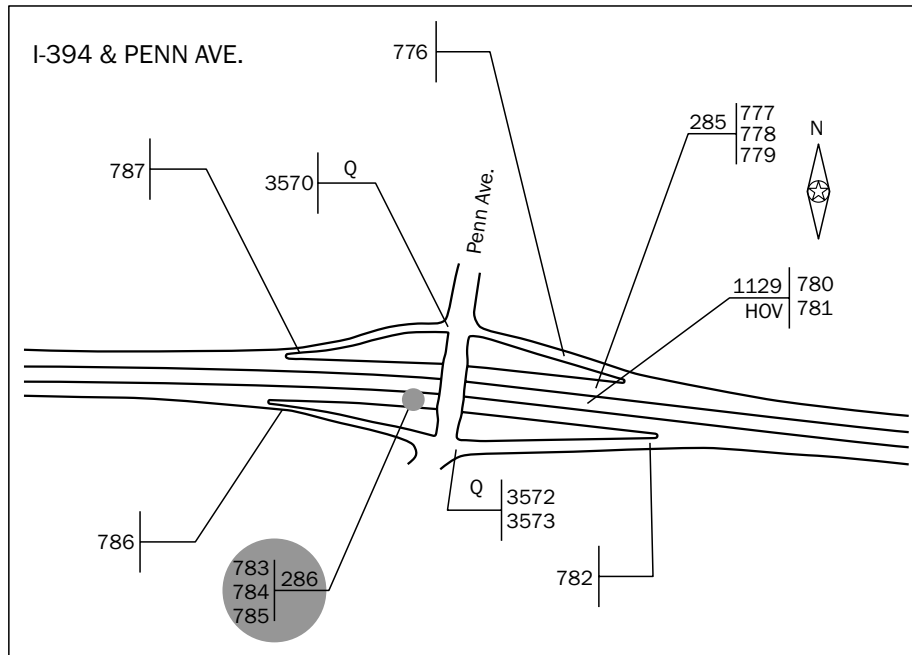
*Figure 1 - Map of station 286 and detectors d783, d784, d785 on Highway EB, I-394 [30]*

Marquardt method, the fastest back propagation algorithm in the MATLAB NN-toolbox. MLP neural network is first modelled separately using the training data and then forecasts using the test data. About 85% of data selected by the MIFS algorithm are used to train the network and the remaining 15% are used to test each simulation. The MLP predictive model is shown in *Figure 2*.

*W* and *Z* are the input and intermediate weighting matrices, respectively, and function *f( . )* is the neuron transfer function.

### 3.3 Forecasting methods

Five forecasting methods were employed in each simulation to compare the predictive power of the models and demonstrate the performance of the MIFS algorithm.

*Method A: Mean average model and all data from previous days*: All data and the average of all data from previous days are used to predict traffic flow of the last day of each simulation. For example, average data traffic flow at about 6 p.m. for the last 22 non-holiday days is used to predict traffic flow at 6 p.m. of the last day of the month. There are usually 22 × 96 pieces of data, but this may vary in different simulations.

*Method B: ARIMA model and data from the previous four hours*: Only the data from the previous four hours is used to predict traffic flow for the last day in each simulation. About 80% of data is used to identify the regression relations using ARIMA(2,2,0) and 20% is used to predict traffic flow. To predict traffic flow on the last day of the month, at least four days of data are required. Using four pieces of data for each one
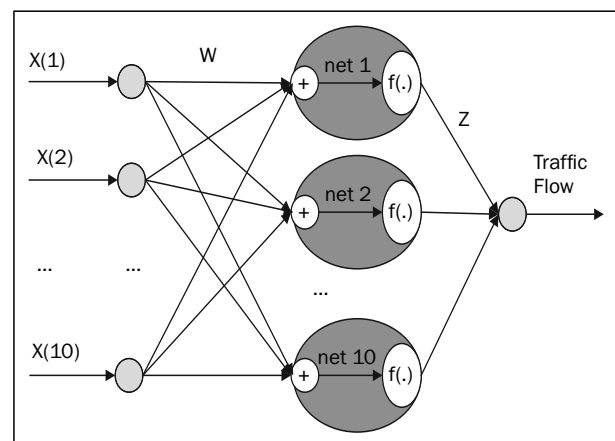


*Figure 2 - Three-layer MLP predictive model*

hour means that only the last 16 pieces of data (one per 15-minute data interval) are used to predict traffic flow at *t*. This totals about 16×96 pieces of data for all simulations. For example, to predict traffic flow at 6 p.m. on the last day of the month using ARIMA(2,2,0), all traffic flow data related to the previous four hours (2 p.m. to 6 p.m.) are required.

*Method C: Mean average model and data from the same days of the week*: This method is used to forecast traffic flow on the last day of each simulation with the selected data. It uses the mean of all data from the same day of each week. For example, traffic flow is predicted for 6 p.m. of the last Wednesday of the month by averaging traffic flow data at 6 p.m. of the four previous Wednesdays. There are 4×96 pieces of data used in this model.

*Method D: Mean average model using MIFS algorithm and all data from previous days with the previ-*

*Table 2 - Traffic flow prediction error for Scenario 1*

| | Predictive Method | Method A | Method B | Method C | Method D | Method E |
|---|---|---|---|---|---|---|
| Scenario 1 | MAE | 57.0293 | 45.3600 | 33.2734 | 24.6948 | **17.9686** |
| | MAPE (%) | 8.5632 | 10.3316 | 6.3387 | 5.4330 | 6.8657 |
| | VAPE (%) | 0.3753 | 1.6927 | 0.4762 | 0.4036 | 1.7509 |

*ous four hours*: Data from Methods A and B are used to forecast traffic flow on the last day of each simulation. In offline mode using the MIFS algorithm, a set of ten optimal pieces of data from among all input data is averaged. For example, to predict traffic flow at 6 p.m. on the last day of the month, the input traffic flow data for the previous four hours plus traffic flow data at 6 p.m. for the previous 22 days is used. Using the MIFS algorithm, the mean of ten optimized pieces of data extracted from all input data is calculated. There are about 38×96 pieces of data (22×96 for Method A and 16×96 for Method B), and these are different for each simulation.

*Method E: MLP using MIFS algorithm and all data from previous days with the previous four hours*: This method is similar to Method D but uses the MLP neural network instead of a mean averaging model. MIFS algorithm outputs require that a set of ten optimal pieces of data be selected for the input layer of the MLP model and traffic flow is predicted in the output layer. The volume of training and testing data is about 85% and 15% of the 10×96 pieces of data selected.

### 3.4 Numerical comparison

To evaluate the efficiency and accuracy of the forecasting methods and to numerically evaluate the MIFS algorithm, the prediction error is obtained using mean absolute error (MAE), mean absolute percentage error (MAPE) and variance absolute percentage error (VAPE). Mathematical models for all three error criteria are shown below.

MAE: $\frac{1}{N}\left(\sum |Real\ Value - Predicition\ Value|\right)$ (9)

MAPE: $\frac{1}{N}\left(\sum \frac{|Real\ Value - Prediction\ Value|}{Real\ Value}\right) \times 100\%$ (10)

VAPE: $Var\left(\frac{|Real\ Value - Prediction\ Value|}{Real\ Value}\right) \times 100\%$ (11)

### 4. SIMULATIONS AND RESULTS

Five simulations were used to show the effect of the MIFS algorithm in decreasing computational complexity and increasing the accuracy of ITS and traffic flow forecasting. The proposed method with noise and lost sensor data was also analysed under different traffic conditions.

### 4.1 Scenario 1: Forecasting a normal day using incomplete data

Traffic flow was predicted for a normal day using a data set containing incomplete data (failure of detectors) using the five methods and MIFS algorithm. The simulated data for January 2012 are used to predict traffic flow on 31 January (Tuesday) using data from the most recent non-holiday days of the month (21 days). On 2 January, detector failure occurred and no data were available (zero information). The simulation results are shown in *Table 2* and *Figure 3*(a).

Method A contains incomplete data. When compared with Method C, Method A shows obvious bias and the impact of detector failure in its prediction of traffic flow. Method C adequately removed this effect; however, if the detector failure had occurred on the day of the week that provides data for Method C, it would have worsened the bias instead of improving it. The ARIMA model did not experience incomplete data, but used lag data for prediction, which means it did not perform well. Generally, when data are selected in Methods C, D and E using whole data, the results are good. Methods D and E produce more accurate predictions than Method C because the MIFS algorithm selects data with stronger logic. As expected, the MIFS algorithm decreased the amount of input data and increased the accuracy of forecasting. Method E most accurately tracked and decreased the prediction error, which indicates the high power of neural network in forecasting.

### 4.2 Scenario 2: Forecasting a normal day using incomplete and noisy data

Traffic flow for a normal day was predicted using an incomplete data set with noisy data (congestion because of an accident). The data were for April 2012 and the goal was to predict the traffic flow on 26 April (Thursday) using data from 18 non-holiday days from the month of April. On 19 April, detector failure occurred and no data were available (zero information). Heavy traffic congestion (accident) occurred on 3 April. The simulation results are shown in *Table 3* and *Figure 3*(b).

As seen, the results of Scenario 2 are similar to those of Scenario 1; however, incomplete data were included in Method C, so its forecast was worse than that for Method A. Unlike Methods A, B and C, Methods

*Table 3 - Traffic flow prediction error for Scenario 2*

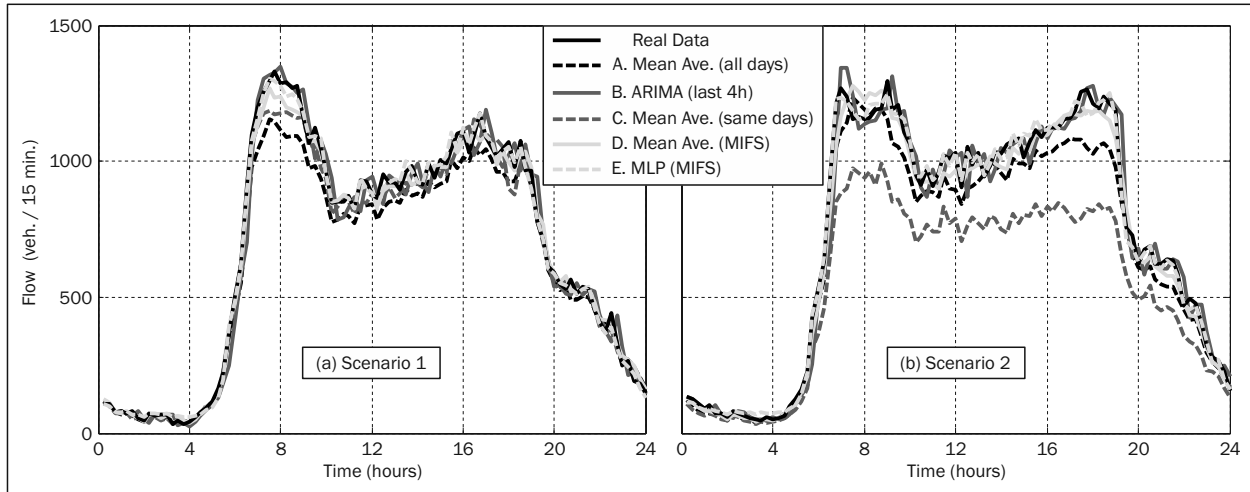| | Predictive Method | Method A | Method B | Method C | Method D | Method E |
|---|---|---|---|---|---|---|
| Scenario 2 | MAE | 65.2976 | 47.6813 | 191.0208 | 37.7969 | **19.6291** |
| | MAPE (%) | 9.8626 | 8.9111 | 25.6343 | 6.5645 | 6.2734 |
| | VAPE (%) | 0.3278 | 0.8568 | 0.3184 | 0.2215 | 0.9829 |



*Figure 3 - Traffic flow prediction for (a) Scenario 1, (b) Scenario 2*

E and D easily detected noisy data and set them aside with the use of the MIFS algorithm and even selected the ten best pieces of data. The MIFS algorithm clearly identified noisy, incomplete and lag data (four previous hours). For this simulation, the best results were achieved by Method E.

## 4.3 Scenario 3: Forecasting a normal day using a variety of noisy data

Forecasting normal traffic flow for one day was done using a data set containing a variety of noisy data to test the performance of the five methods and the MIFS algorithm under emergency conditions. Data collection occurred during the month from 28 February (Tuesday) to 28 March (Wednesday). The goal was to predict traffic flow on 28 March (Wednesday) using data from the previous month (21 non-holiday days). On 29 February (Wednesday), traffic was very heavy as a result of heavy precipitation and the data were similar to that for a holiday. Severe congestion was also recorded on 7 March (Wednesday) and 16 March (Friday). The simulation results are shown in *Table 4* and *Figure 4*(a).

The results of this scenario are similar to Scenario 2. Methods A and C were affected by the noisy data and were biased and did not provide accurate predictions. Method B used lag data (four previous hours) and should have performed well under normal conditions, however, under abnormal conditions; ARIMA could not predict traffic flow. MIFS-based Methods D and E identified and removed the erroneous delay-ridden and noisy information. This simulation shows that the MIFS algorithm clearly identified and tracked the real data, despite noisy, defective and abnormal data. In this simulation, the best prediction was made again by Method E.

## 4.4 Scenario 4: Forecasting a normal day using a large volume of different data types

The aim of this scenario was to predict traffic flow for a normal day using a large volume of different data types that included noisy and incomplete data caused by precipitation, accidents, and detector failure. The difference between this scenario and the previous scenarios are the large volume of data and

*Table 4 - Traffic flow prediction error for Scenario 3*

| | Predictive Method | Method A | Method B | Method C | Method D | Method E |
|---|---|---|---|---|---|---|
| Scenario 3 | MAE | 39.8646 | 48.3648 | 61.4844 | 32.9479 | **23.8240** |
| | MAPE (%) | 7.0389 | 10.3569 | 9.1038 | 6.2382 | 7.4289 |
| | VAPE (%) | 0.4098 | 1.4083 | 0.4701 | 0.3503 | 1.5903 |

*Table 5 - Traffic flow prediction error for Scenario 4*

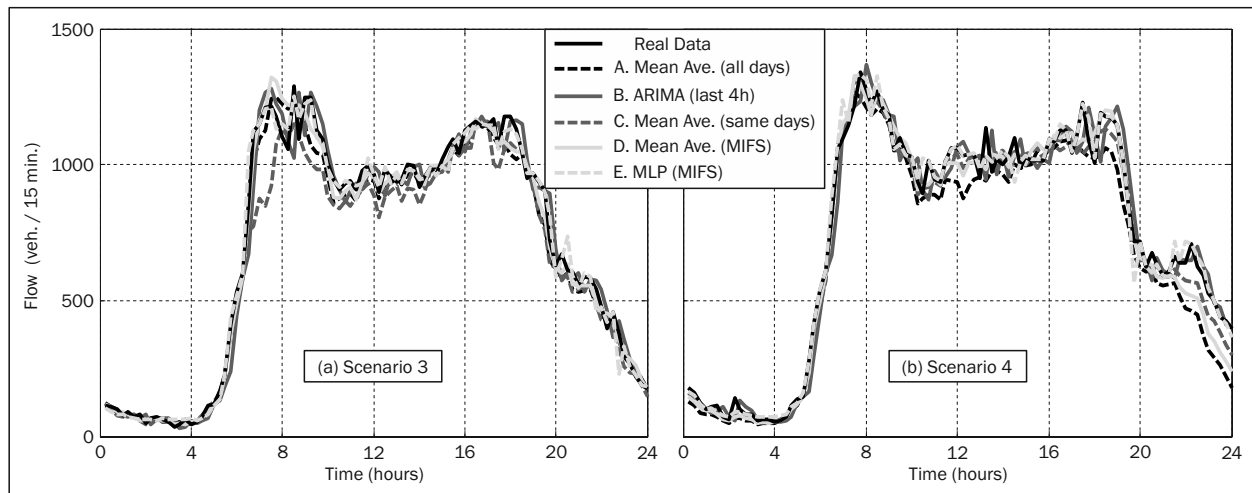| | Predictive Method | Method A | Method B | Method C | Method D | Method E |
|---|---|---|---|---|---|---|
| Scenario 4 | MAE | 68.1694 | 51.8848 | 37.2362 | 44.4250 | **21.0289** |
| | MAPE (%) | 13.1962 | 9.6257 | 6.5807 | 8.8018 | 4.8880 |
| | VAPE (%) | 1.6546 | 1.2467 | 0.3899 | 0.9404 | 0.5912 |



*Figure 4 - Traffic flow prediction for (a) Scenario 3, (b) Scenario 4*

different types of uncertainty. Data collected during the first six months of 2012 were used to analyze the performance of the methods and the MIFS algorithm. Traffic flow for 29 June (Friday) was predicted using data from workdays of the previous six months (128 days). The simulation results are shown in *Table 5* and *Figure 4*(b).

As seen, the results are similar to the results of Scenario 3. The large amount of input data should have been sufficient for most methods to predict the traffic flow. The amount of data for Method A was 128×96 data, for Method B was 16×96 data, for Method C was 25×96 data, and for Methods D and E was 10×96. The most accurate data should have been that for Method A, which had the largest amount input data; however, the results show that Methods C, D and E selected data more accurately. Methods D and E used the MIFS algorithm with ten pieces of data and succeeded in increasing forecasting speed and decreasing computational complexity with considerable accuracy. This simulation demonstrates that the MIFS algorithm is sensitive to and can identify noisy, incomplete and lag data. The MLP neural network performed better than the mean average and ARIMA models. Method E again provided the most accurate results.

## 4.5  Scenario 5: Forecasting an abnormal day using a large volume of different data types

This scenario forecasted traffic flow of an abnormal day using a variety of data. The aim was to test forecasting performance of the models, forecasting methods and the MIFS algorithm under adverse conditions. The traffic flow was forecast for 1 June (Friday) using normal, incomplete and noisy data. The initial forecast was based on data from the previous month (1 May to 31 May; 22 days). The simulation was then repeated using data for the three previous months (1 March to 31 May; 79 days). Further evaluation was made using data from the five previous months (1 January to 31 May; 108 days). The results of these simulations are shown in *Table 6* and *Figure 5*.

The increase in volume of input data allowed all methods to predict traffic flow well. The 16×96 lag data used to forecast traffic flow for 1 June (Friday) in Method B produced the same result for all three scenarios. It can be seen in *Table 6* that Method E used the MIFS algorithm and MLP model in all three scenarios to achieve the best performance. Increasing the amount of input data had no effect on prediction. It was expected that increasing the accurate data would increase the system precision; however, Method E with the MIFS algorithm and ten pieces of data produced the same result as did Method A using a greater amount of data. This confirms that the MIFS algorithm manifests excellent accuracy with decreased complexity that can identify noisy, incomplete and different data. Method E provided the best predictions for all three versions of this scenario using the power of the combined MIFS-MLP.

*Table 6 - Traffic flow prediction error for Scenario 5*

| Predictive Method | | Method A | Method B | Method C | Method D | Method E |
|---|---|---|---|---|---|---|
| Scenario 5.1 | MAE | 74.6373 | 72.4838 | 65.4766 | 62.9448 | **50.5098** |
| | MAPE (%) | 13.3450 | 13.7725 | 11.2670 | 11.0766 | 15.8618 |
| | VAPE (%) | 3.8248 | 5.0654 | 4.3353 | 3.4967 | 8.7116 |
| Scenario 5.2 | MAE | 82.1697 | 72.4838 | 63.9928 | 68.4167 | **50.0121** |
| | MAPE (%) | 15.1209 | 13.7725 | 11.0448 | 12.4220 | 11.9449 |
| | VAPE (%) | 3.8040 | 5.0654 | 3.9913 | 3.8626 | 5.0868 |
| Scenario 5.3 | MAE | 90.2490 | 72.4838 | 64.1954 | 83.3177 | **45.3508** |
| | MAPE (%) | 16.4285 | 13.7725 | 10.6859 | 15.0473 | 12.5537 |
| | VAPE (%) | 3.5539 | 5.0654 | 3.5651 | 3.7647 | 7.6106 |
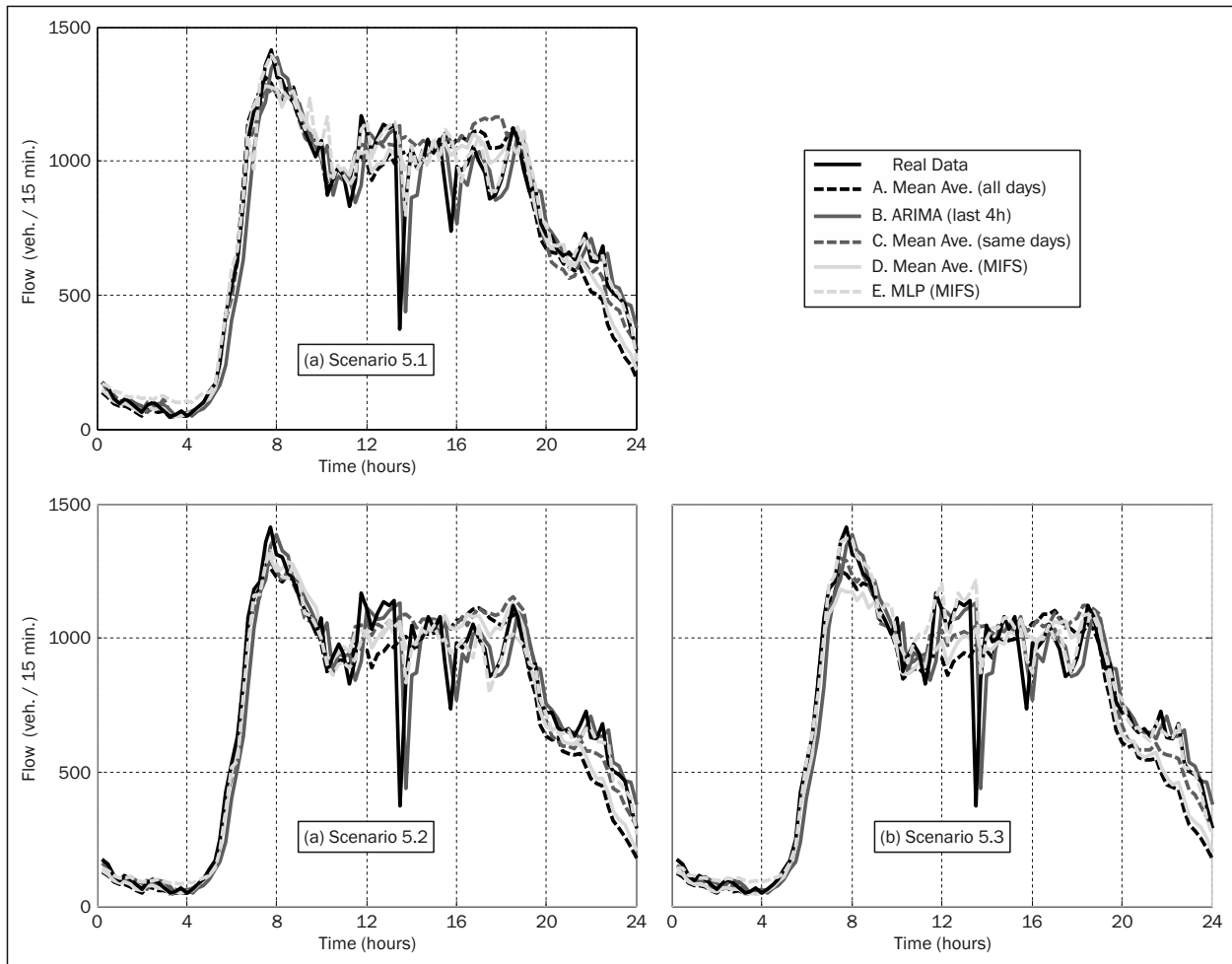


*Figure 5 - Traffic flow prediction for (a) Scenario 5.1, (b) Scenario 5.2, (c) Scenario 5.3*

## 5. CONCLUSION

The present study describes the calculation of non-linear dependence of recorded traffic flow data determining inaccurate and missing data. Then the best data were selected from the large amount of data provided to decrease the volume of calculations and improve the forecasting accuracy.

Numerical case studies from traffic flow data for Minnesota interstate Highway I-394 were used. The data contained incomplete data (failure of detectors), noisy data (rain, snow, congestion or accidents) and lag data (from the previous four hours) to test the forecasting methods under different traffic conditions.

The simulations results showed good robustness for the combined MIFS-MLP neural network model. The numerical results also indicated that the proposed MIFS-MLP flow predictor decreased forecasting error over the results of the mean average and ARIMA models that used traffic data from previous time periods to

ignore the interdependency of the input data. The proposed algorithm can detect and track incorrect data efficiently.

**سید هادی حسینی**

دانشجوی دکتری رشته مهندسی برق، گرایش کنترل، دانشکده مهندسی برق، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

**بهزاد مشیری**

دکتری رشته مهندسی برق، گرایش کنترل، استاد دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران، تهران، ایران

**اشکان رحیمی کیان**

دکتری رشته مهندسی برق، گرایش کنترل، دانشیار دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران، تهران، ایران

**بابک نجار اعرابی**

دکتری رشته مهندسی برق، گرایش کنترل، دانشیار دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران، تهران، ایران

**چکیده**

**روش ترکیبی جدید برای پیش‌بینی هوشمند جریان ترافیک براساس کنترل‌کننده فازی در حضور انواع مختلف داده‌های ترافیکی**

پیش‌بینی جریان ترافیک یکی از راه‌حل‌هایی است که برای جلوگیری از بروز ازدحام در بزرگراه‌ها ارائه می‌شود. در این مقاله یک روش ترکیبی جدید براساس ساختار کنترل‌کننده فازی جهت پیش‌بینی کوتاه مدت جریان ترافیک ارائه شده است و با انواع مختلف و پرکاربرد روش‌های پیش‌بینی سابق مورد مقایسه قرار گرفته است که نتایج شبیه‌سازی‌ها بیان‌کننده دقت بالای پیش‌بینی برای مدل جدید می‌باشد. با توجه به متفاوت بودن الگوهای رفتاری جریان ترافیک در واقعیت، انواع حالت‌های مختلف اغتشاشی همچون امکان از دست دادن داده‌ها به‌دلیل خرابی حسگرها و یا معیوب شدن آنها تحت شرایط مختلف آب و هوایی نظیر بارش باران یا برف و شرایط مختلف ترافیکی مانند وقوع ازدحام و تصادف در جاده، را در نظر گرفته و عملکرد مدل پیش‌بینی جدید ارائه شده را در حضور این داده‌ها مورد ارزیابی قرار داده‌ایم که حاکی از مقاوم بودن این مدل در حضور انواع مختلف داده‌های اغتشاشی می‌باشد. برای نشان دادن عملکرد این مدل جدید شبیه‌سازی‌هایی را براساس داده‌های واقعی در نرم افزار MATLAB انجام داده‌ایم.

**کلمات کلیدی**

پیش‌بینی جریان ترافیک، شبکه‌های هوشمند ترکیبی، کنترل فازی، سری‌های زمانی، داده‌های اغتشاشی.

## REFERENCES

[1] **Castro-Neto M**, **Jeong YS**, **Jeong MK**, **Han LD**. *On-line-SVR for short-term traffic flow prediction under typical and atypical traffic conditions*. Expert Sys with Appl. 2009 Apr;36(3):6164-6173.

[2] **Zhang Y**. *How to provide accurate and robust traffic forecasts practically?*. Intel Transportation Sys. 2012 Mar;8:189-206.

[3] **Stathopoulos A**, **Dimitriou L**, **Tsekeris T**. *Fuzzy modeling approach for combined forecasting of urban traffic flow*. Comp Aid Civil and Infra Eng. 2008 Oct;23(7):521-535.

[4] **Jiang X**, **Adeli H**. *Dynamic wavelet neural network model for traffic flow forecasting*. J Transportation Eng. 2005 Oct;131(10):771-779.

[5] **Sun H**, **Liu HX**, **Xiao H**, **Ran B**. *Short-term traffic forecasting using the local linear regression model*. Center for Traffic Simulation Studies [Internet]. UC Irvine: Center for Traffic Simulation Studies. 2002; Available from: http://escholarship.org/uc/item/540301xx

[6] **Smith BL**, **Williams BM**, **Oswald RK**. *Comparison of parametric and nonparametric models for traffic flow forecasting*. J Transportation Res Part C. 2002 Aug;10(4):303-321.

[7] **Ramezani A**, **Moshiri B**, **Rahimi-Kian A**, **Araabi BN**, **Abdulhai B**. *Distributed maximum likelihood estimation for flow and speed density prediction in distributed traffic detectors with Gaussian mixture model assumption*. IET Intel Transportation Sys. 2012 Jun;6(2):215-222.

[8] **Zhang Y**, **Ye Z**. *Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods*. J Intel Transportation Sys. 2008 Jul;12(3):102-112.

[9] **Chang H**, **Lee Y**, **Yoon B**, **Baek S**. *Dynamic near-term traffic flow prediction: System-oriented approach based on past experiences*. IET Intel Transportation Sys. 2012 Sep;6(3):292-305.

[10] **Prochazka A**. *Neural networks and seasonal time-series prediction*. 5th International Conference on Artificial Neural Networks. 7-9 July 1997, Cambridge, UK; 1997. p. 36-41.

[11] **Chandra SR**, **Al-Deek H**. *Predictions of freeway traffic speeds and volumes using vector autoregressive models*. J Intel Transportation Sys. 2009 Apr;13(2):53-72.

[12] **Williams BM**. *Multivariate vehicular traffic flow prediction: An evaluation of ARIMAX modeling*. Transportation Res Board. 2002 Mar; 00824511:194-200.

[13] **Ledoux C**. *An urban traffic control system integrating neural networks*. 8th International Conference on Road Traffic Monitoring and Control. 23-25 April 1996, London, UK; 1996. p. 197-201.

[14] **Zhang Y**, **Ye Z**. *Short-term traffic flow forecasting using fuzzy logic system methods*. J Intel Transportation Sys. 2008 Aug;12(3):102-112.

[15] **Lim S**, **Lee C**. *Data fusion algorithm improves travel time predictions*. IET Intel Transportation Sys. 2011 Dec;5(4):302-309.

[16] **Yu D**, **An S**, **Hu Q**. *Fuzzy mutual information based min-redundancy and max-relevance heterogeneous feature selection*. Int J Comp Intel Sys. 2011 Mar;4(4):619-633.

[17] **Wei D**, **Liu H**. *An adaptive-margin support vector regression for short-term traffic flow forecast*. J Intel Transportation Sys. 2013 Nov;17(4):317-327.

[18] **Ghosh B**, **Basu B**, **O'Mahony M**. *Bayesian time-series model for short-term traffic flow forecasting*. J Transportation Eng. 2007 Mar;133(3):180-189.

[19] **Xie Y**, **Zhang Y**. *A wavelet network model for short-term traffic volume forecasting*. J Intel Transportation Sys. 2007 Jan;10(3):141-150.

[20] **Abdulhai B**, **Porwal H**, **Recker W**. *Short-term traffic flow prediction using neuro-genetic algorithms*. J Intel Transportation Sys. 2002 Jan;7:3-41.

[21] **Zeng D**, **Xu J**, **Gu J**, **Liu L**, **Xu G**. *Short term traffic flow prediction using hybrid ARIMA and ANN models*. Workshop on Power Electronics and Intel Transportation Sys. 2-3 August 2008, Guangzhou, China; 2008. p. 621-625.

[22] **Messai N**, **Thomas P**, **Lefebvre D**, **El-Moudni A**. *A Neural Network Approach for Freeway Traffic Flow Prediction*. International Conference on Control App. 18-20 September 2002, Glasgow, UK; 2002. p. 984-989.

[23] **Yuan J**, **Mills K**. *A Cross-Correlation Based Method for Spatial-Temporal Traffic Analysis*. J Perform Eval. 2005 Jul;61(2-3):163-180.

[24] **Williams JW**, **Li Y**. *Estimation of mutual information: A survey*. 4th International Conference on Rough Sets and Knowledge Technology. 14-16 July 2009, Gold Coast, Australia; 2009. p. 389-396.

[25] **Kraskov A**, **Stögbauer H**, **Grassberger P**. Estimating Mutual Information. Physical Rev E. 2004 Jun;69(066138):1-16.

[26] **Stögbauer H**, **Kraskov A**, **Astakhov SA**, **Grassberger P**. *Least Dependent Component Analysis Based on Mutual Information*. Physical Rev E. 2004 Sep;70(066123):1-18.

[27] **Dasarathy BV**. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. Mc Graw-Hill: Computer Science Series, IEEE Computer Society Press: Las Alamitos, California; 1991.

[28] **Battiti R**. *Using mutual information for selecting features in supervised neural net learning*. IEEE Trans Neural Net. 1994 Jul;5(4):537-550.

[29] **Deng K**. *OMEGA: On-line Memory-based General Purpose System Classifier* [dissertation]. Carnegie Mellon University; Pittsburgh (PA): University of Pittsburgh; 1998.

[30] [Cited 2013 Jun 30]. Available from: http://www.d.umn.edu/tdrl/traffic/.

[31] [Cited 2014 Jul 10]. Available from: http://weatherspark.com/history/30956/2012/Minneapolis-Minnesota-United-States.

[32] **Tan MC**, **Wong SC**, **Xu JM**, **Guan ZR**, **Zhang P**. *An Aggregation Approach to Short-Term Traffic Flow Prediction*. IEEE Trans Intel Transportation Sys. 2009 Mar;10(1):60-69.