RESEARCH ARTICLE

Presentation of regression analysis, GP and GMDH models to predict the pedestrian density in various urban facilities

Iraj BARGEGOL[a], Seyed Mohsen HOSSEINIAN[b], Vahid NAJAFI MOGHADDAM GILANI[b*], Mohammad NIKOOKAR[a], Alireza OROUEI[c]

[a] School of Civil Engineering, University of Guilan, Rasht 41996-13776, Iran

[b] School of Civil Engineering, Iran University of Science and Technology (IUST), Tehran 13114-16846, Iran

[c] School of Civil Engineering, Islamic Azad University, Semnan 35131-37111, Iran

[*] Corresponding author. E-mail: vahid_najafi@alumni.iust.ac.ir

© Higher Education Press 2022

ABSTRACT   In this study, the relationship between space mean speed (*SMS*), flow rate and density of pedestrians was investigated in different pedestrian facilities, including 1 walkway, 2 sidewalks, 2 signalized crosswalks and 2 mid-block crosswalks. First, statistical analysis was performed to investigate the normality of data and correlation of variables. Regression analysis was then applied to determine the relationship between *SMS*, flow rate, and density of pedestrians. Finally, two prediction models of density were obtained using genetic programming (GP) and group method of data handling (GMDH) models, and k-fold and holdout cross-validation methods were used to evaluate the models. By the use of regression analysis, the mathematical relationships between variables in all facilities were calculated and plotted, and the best relationships were observed in flow rate-density diagrams. Results also indicated that GP had a higher R^2 than GMDH in the prediction of pedestrian density in terms of flow rate and *SMS*, suggesting that GP was better able to model *SMS* and pedestrian density. Moreover, the application of k-fold cross-validation method in the models led to better performances compared to the holdout cross-validation method, which shows that the prediction models using k-fold were more reliable. Finally, density relationships in all facilities were obtained in terms of *SMS* and flow rate.

KEYWORDS   pedestrian density, regression analysis, GP model, GMDH model

1   Introduction

Walking is the simplest way to move between the origin and destination. For many, it can be the only means of personal transportation and does not need any vehicles. Therefore, pedestrians constitute an inseparable part of transportation systems [1]. Some people are pedestrians during a day at some time. Safety should be considered [2], and traffic flow management is certainly an issue in which governments make significant investments in order to have an efficient transportation system [3].

Traffic accidents have significantly damaged transportation facilities in Iran. They reportedly caused 15932 fatalities in Iran in 2016, of which 78% were male and 22% were female. The estimated rate of the World Health Organization (WHO) for road traffic fatalities was 20.5 per 100000 population in 2016 [4]. They have also imposed substantial costs on governments and resulted in problems in other infrastructure and societal spheres, especially in developing countries, due to their low gross national income [5]. The occurrence of road traffic accidents is inevitable; however, the factors affecting road traffic accidents can be recognized and analyzed in order to establish and support prevention policies [6]. Road safety depends very much on the road user behavior. Therefore, identifying the behavior of pedestrians in the transportation network helps to manage and control traffic flow [7].

Since there are various walking habits, the behavior of pedestrians is variable because of features like the

number of people in a group, age, gender, clothing, and so on. Various studies have investigated the relationship between flow, speed and density of pedestrians. Hughes [8] proposed a modeling method based on a differential equation that analyzed the movement of a large group of pedestrians. This group of pedestrians moved continually and their density and speed were defined based on the three hypotheses that imply walking habits for each pedestrian type. Also, the independent and constant variables modeled the place and time. The relationships between speed, density and flow were specified in a fluid mechanics approach. Helbing et al. [9] presented a figure density method of pedestrians as simulations in a study on the movement of pilgrims in Mina/Makkah during the Hajj by video recording. The figure was based on a model of pedestrian movements under the social force condition, which was applied to compare the density in simulations. In a study performed by Liu et al. [10], the velocity-density and flow rate-density relationships were obtained using the pedestrian flow characteristics on stairways such as in train stations in China. Moreover, the vehicle flow and pedestrian flow were compared in the mentioned pedestrian distributing centers, such as the Shanghai Chifeng Road Light Orbit Station, Shanghai People's Square Subway Station and the Shanghai train station and so on. Results showed that the pedestrians' flow models on stairways were appropriate in some transfer stations. Chen et al. [11] examined longitudinal pedestrian flows in some passages such as two-direction (ascending and descending) stairways in a metro station in Shanghai. They used video recording to obtain pedestrian flow characteristics such as volume, density and speed in some high- and low-density conditions and analyzed the relationships obtained from the dataset. They finally applied statistical regression to present a model based on the density-volume and density-speed relationships. Plaue et al. [12] explored the pedestrian density by video recording in roofed buildings to gather data. They utilized some mathematical models to perform numeral simulations and eventually proposed a technique to calculate pedestrian density. Shafabakhsh et al. [13] conducted a simulation-based study to examine the effect of elderly populations on the movement flows of pedestrians in the sidewalks by the use of a micro-simulation method. By collecting the data of movement of pedestrians in 19 sidewalks in India and classifying the pedestrian facilities in terms of their width, Rastogi and Chandra [14] revealed that the relationships between speed and density follow the exponential model on the sidewalks of varying widths and the linear model on a non-exclusive facility. Moreover, the increase in width of facilities resulted in increased space available to a pedestrian, but reduced maximum flow rate and optimum density. In a research conducted by Bargegol et al. [15] by modelling the flow rate data of 8541 pedestrians in a walkable street and two

sidewalks of Rasht, results revealed that the capacity of the flow rate and the optimum density in all the mentioned facilities were 96 ($\text{ped} \cdot \text{m}^{-1} \cdot \text{min}^{-1}$) and 2.83 ($\text{ped} \cdot \text{m}^{-2}$), respectively. Pinna and Murrau [16] investigated the behavior of pedestrians and the simultaneous effect of age and speed of pedestrians on sidewalks in the Oristano downtown, Italia. Results indicated that there was not a linear relationship between age and speed and they stated that it is better to consider a polynomial relationship between the mean speed of walking, mean speed of individual pedestrian, and age class. Sun [17] provided a simulation of two-sided pedestrian flow with various walking speeds and presented factors to represent the walking preference and speeds of pedestrians. The simulation results exhibit transitions between three phases: freely flowing, lane formation, and fully jammed phases as a function of initial density of pedestrians. In the phase of lane formation, it was observed that faster pedestrians exceed the slower ones through a narrow walkway. Moreover, the relationships of density–flow and density–velocity were different from each other at different phases.

Various studies have been conducted on the speed, flow rate, and density of various facilities and the behavior of drivers, but few studies have been performed on the behavior of pedestrians in this regard [18]. Although there have been various studies on the behavior of pedestrians on issues such as sidewalks, design of walking systems, and safety of pedestrians, less attention has been paid to concepts including speed, density, flow rate for passing through and outside the crosswalk, and so on. Moreover, various researches have been performed by the use of group method of data handling (GMDH) and genetic programming (GP) methods in many engineering fields [19–24]. However, no studies have been done on the use of prediction models (for instance, GP and GMDH) to represent pedestrian density.

Pedestrian density and pedestrian behavior in general play important roles in the quality and design of pedestrian facilities. If the facility faces a density defect, it would lead to reduced safety in interactions between the pedestrians and vehicle flows caused by pedestrian crossings, as well as various violations of pedestrians, especially through crossings and intersections. Therefore, it is very important to investigate the speed, flow rate and density of pedestrians, and modeling this type of pedestrian behavior will increase their safety and reduce traffic accidents induced by pedestrian crossings. So, the objectives of this study were:

• obtain the *SMS*, flow rate and density values in various facilities, including 1 walkway, 2 sidewalks, 2 mid-block crosswalks and 2 signalized crosswalks,

• perform statistical analysis to investigate the normality of data and the correlation of variables,

• apply regression analysis to determine the relationship

between *SMS*, flow rate and density of pedestrians in various facilities,

• obtain prediction models of the pedestrian density using GMDH and GP models in various facilities,

• evaluate the model by the use of k-fold and holdout cross-validation methods,

• present pedestrian density relationships in all facilities using the superior model in terms of *SMS* and flow rate.

In this study, a field method was applied using video recording in order to collect data in various facilities, including 1 walkway, 2 sidewalks, 2 signalized crosswalks and 2 mid-block crosswalks. Next, the information of 8490 pedestrians was obtained. Moreover, statistical analysis was conducted to investigate the normality of data and correlation of variables. Regression analysis was then applied to determine the relationship between *SMS*, flow rate, and density of pedestrians in various facilities. Finally, the prediction models of the density of pedestrians were presented using GP and GMDH models in various facilities. Figure 1 indicates the process of this research.

2   Methods

2.1   Data collecting

In this study, the results of extracting pedestrian information from 7.5 hours of video recording from 2 sidewalks, 1 walkway, 2 mid-block crosswalks and 2 signalized crosswalks of Rasht metropolis were used during evening peak traffic, then flow rate (ped·min^{-1}), *SMS* (m·min^{-1}) and density (ped·m^{-2}) of pedestrians were determined using Tracker software. In these facilities, all 8490 pedestrian crossings have been occurred on two-sides. To explore the behavior of pedestrians, crosswalks at 2 busy signalized crosswalks, 2 mid-block crosswalks, 2 sidewalk areas, and 1 walkable street facilities in the most crowded area of Rasht were selected and used. Some pictures of the studied area for various facilities are also presented in Fig. 2.

2.2   Data description

In this section, the normality of data as well as the correlation coefficient of dependent and independent variables were explored.

2.2.1   Normality of dataset

Before modeling, to have fair comparisons, the normality of dataset was explored, the results of which are illustrated in Table 1. As can be seen, the normality of dataset was checked according to skewness and kurtosis parameters. Kurtosis illustrates a peak distribution measure and skewness represents an asymmetrical measure. Generally, if these two parameters are not in the range of $(-2, +2)$, the dataset is far from the normal distribution and needs corrections [25].

As is evident from Table 1, the skewness and kurtosis values of all variables in various facilities except *SMS* data in mid-block crosswalks (through pedestrian crossing) were between $(-2, 2)$, indicating that the dataset applied in this study, with that one exception, had normal distributions.

2.2.2   Correlation of variables

The Pearson correlation analysis was applied in this research because of the quantitative nature of the dataset. The Pearson's correlation coefficient calculates the linear correlation level of two variables. It is described as a value between +1 and −1, whereby −1 illustrates a full negative correlation, 0 shows no correlation and +1 illustrates a full positive correlation. The closer the absolute amount to 1, the stronger the correlation between the two variables [26]. The correlation results between the variables in various facilities used in this study are indicated in Table 2. The results of the Pearson correlation between flow rate, *SMS* and density variables indicated a high value for all facilities.
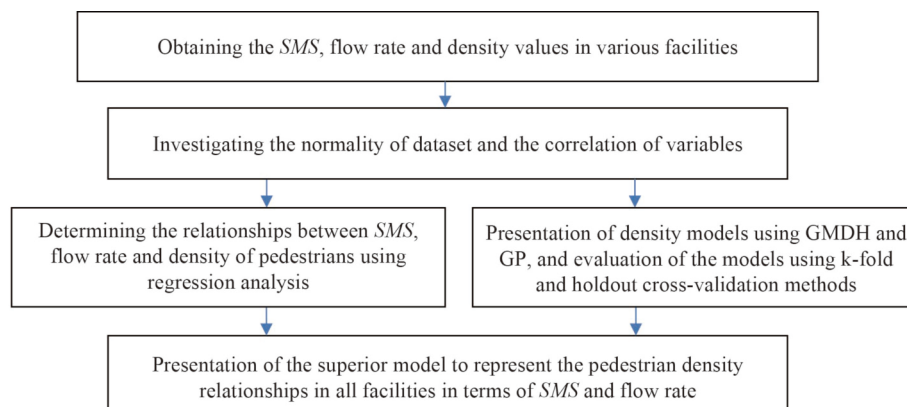

Fig. 1   The steps of the research method.

2.3 Modeling methods

In this study, regression analysis was first used to determine the relationship between *SMS*, flow rate, and density of pedestrians in various facilities. Finally, the prediction models of density in terms of flow rate and *SMS* were achieved using GMDH and GP.

2.3.1 Regression analysis

The relationships between response and independent variables in regressions are predicted using polynomials. To determine whether regressions are capable of estimating the response variable appropriately, the coefficient of

Fig. 2 Sample pictures of various facilities in the studied area.

determination (R^2) can be calculated. R^2 is a statistical measurement of dataset proximity to the fitted regression, showing the extent to which percentages of changes of response variable are explained by independent variables [27]. R^2 is always between 0% and 100%. 0% indicates that the model does not illustrate any variation of the independent data around its average amount and 100% shows that the model can illustrate all the variations. Higher R^2 values indicate that the observed values are closer to the fitted line in the regression model. The Pearson correlation coefficient was utilized due to the use of quantitative data in this research. The correlation coefficient (R) demonstrates the capability of linear relationships between two variables. R-value is between −1 and +1, which are applied for negative and positive linear correlations, respectively. R and R^2 are presented by Eqs. (1) and (2), respectively, where x and y are two parameters, y_{actual} is the actual density value, $y_{predicted}$ is the predicted density value, $\bar{y}_{actual}$ is the mean actual density value, $\bar{y}_{predicted}$ is the predicted mean density value, and N is the number of individuals [28].

$$R = \frac{N(\sum xy) - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}}, \quad (1)$$

$$R^2 = \frac{\left(\sum_{i=1}^{N}(y_{actual} - \bar{y}_{actual}) \times (y_{predicted} - \bar{y}_{predicted})\right)^2}{\sum_{i=1}^{N}(y_{actual} - \bar{y}_{actual})^2 \times \sum_{i=1}^{N}(y_{predicted} - \bar{y}_{predicted})^2}. \quad (2)$$

Table 1 The kurtosis and skewness results of the dataset

model	variables	kurtosis	skewness	standard deviation	mean	maximum statistic	minimum statistic
sidewalks	flow rate	5	87	37.85	25.77	0.62	−1.05
	SMS	55.16	63.51	61.01	2.31	−1.09	−0.09
	density	0.09	1.99	82.22	0.55	0.98	−0.31
walkable street	flow rate	10	60	37.6	9.17	−0.05	0.65
	SMS	53.8	60.86	57.74	1.79	−0.42	−0.63
	density	0.18	1.11	0.67	0.18	0.25	0.29
mid-block crosswalks (through pedestrian crossing)	flow rate	1	10	5.79	2.46	−0.17	−0.01
	SMS	55.92	103.85	65.19	9.35	2.65	8.77
	density	0.01	0.18	0.09	0.04	−0.22	−0.11
mid-block crosswalks (out of pedestrian crossing)	flow rate	3	31	12.69	6.24	0.98	1.44
	SMS	53.83	67.54	59.67	3.48	0.1	−0.37
	density	0.05	0.47	0.21	0.09	0.77	0.98
signalized crosswalks (through pedestrian crossing)	flow rate	1	31	8.78	6.47	1.03	0.51
	SMS	54.83	90	68.68	7.34	0.29	−0.62
	density	0.01	0.5	0.14	0.11	1.02	0.28
signalized crosswalks (out of pedestrian crossing)	flow rate	1	18	4.72	3.45	1.18	1.36
	SMS	47.41	86.84	66.89	9.63	0.02	−0.98
	density	0.01	0.32	0.08	0.07	1.24	1.23

Table 2 Investigation of the correlation between dependent and independent variables

model	variables	flow rate	*SMS*	density
sidewalks	flow rate	1	−0.962	0.981
	SMS	−0.962	1	0.983
	density	0.981	−0.983	1
walkable street	flow rate	1	−0.892	0.958
	SMS	−0.892	1	−0.937
	density	0.958	−0.937	1
mid-block crosswalks (through pedestrian crossing)	flow rate	1	−0.674	0.991
	SMS	−0.674	1	−0.719
	density	0.991	−0.719	1
mid-block crosswalks (out of pedestrian crossing)	flow rate	1	0.902	0.997
	SMS	0.902	1	0.887
	density	0.997	0.887	1
signalized crosswalks (through pedestrian crossing)	flow rate	1	−0.789	0.997
	SMS	−0.789	1	−0.808
	density	0.997	−0.808	1
signalized crosswalks (out of pedestrian crossing)	flow rate	1	−0.839	0.994
	SMS	−0.839	1	−0.854
	density	0.994	−0.854	1

2.3.2 Group method of data handling

GMDH neural network uses a class of polynomial to describe relationships between independent and dependent parameters and is constructed in feedforward and multi-layered networks and contains a set of neurons that are taken from several input pairs using second-order polynomials. Ivakhenko devised a technique in 1966 that could be applied in complex systems and create a self-organizing model that could be used to predict and identify system problems. To model complex systems containing a dataset with 1 output and many inputs, Kolmogorov-Gabor polynomials are applied as Eq. (3), in which y is the model output, $x = (x_1, x_2, \ldots, x_n)$ is the input vector and $a = (a_0, a_1, \ldots, a_k)$ is the vector of coefficients [29]:

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_i a_j a_k x_i x_j x_k + \cdots. \quad (3)$$

In this network, each layer contains one or more processor units, each of which has one output and two inputs. These units act as the components of the model and are assumed in a second-order polynomial form through Eq. (4), the coefficients of which are obtained by the use of regression models:

$$\hat{y}_n = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_1^2 + a_5 x_2^2. \quad (4)$$

The least-square errors should be minimized to compute the output value $\hat{y}_i$ as Eq. (5) [30]:

$$e = \sum_{i+1}^{n} (\hat{y}_i - y_i)^2. \quad (5)$$

To find the minimum errors, the partial derivative of Eq. (5) is applied. A matrix equation ($Aa = y$) can be presented by substituting from Eq. (4) into this partial derivative, in which $Y = \{y_1, \ldots, y_m\}^{\mathrm{T}}$, $a = \{a_0, a_1, a_2, a_3, a_4, a_5\}$ and matrix A is as Eq. (6):

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1p} & x_{1p}^2 & x_{1p}^2 & x_{1p} & x_{1p} \\ 1 & x_{2p} & x_{2q} & x_{2p}^2 & x_{2q}^2 & x_{2p} & x_{2q} \\ 1 & x_{np} & x_{nq} & x_{np}^2 & x_{nq}^2 & x_{np} & x_{nq} \end{bmatrix}. \quad (6)$$

The technique of singular value decomposition is utilized as a solution to the aforementioned matrix equation, through which the unknown a is calculated using Eq. (7), in which A^{T} is the transposition of the matrix:

$$a = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} y. \quad (7)$$

2.3.3 Genetic programming

GP was applied to propose the pedestrian density model in the study, which was first presented by Koza [31], and demonstrates an extendable approach for an evolutionary algorithm based on the evolution concept of Darwin. In general, this approach seeks to solve problem space instead of dataset space. The GP method has been presented as a development of genetic algorithms where programs are represented as tree graphs and illustrated in a practical programming language. In this approach, after the specification of the generation numbers and presenting the input and output parameter amounts, a random population of trees in each generation is constructed to obtain the greatest population diversity. Each member of a random population is a computer structure composed of functions and numbers, which are selected from two separate sets called the set of functions and the set of terminals. Functions are all arithmetic operators or Boolean logic functions that help to solve the problem better and faster. Terminals are also independent variables. In short, GP solves problems in either a mutation or a combination [32]. This evolutionary basis in GP is a well-regarded concept in many other swarm-based approaches such as particle swarm optimizer (PSO) [33], ant colony optimizer (ACO) [34], grey wolf optimizer (GWO) [35,36], Harris hawks optimizer (HHO) [37], whale optimizer (WOA) [38], and teaching-learning-based optimizer (TLBO) [39], differential evolution (DE)

[40], differential search (DS) [41], and other hybrid and enhanced techniques [42–44]. Such methods and recent evolutionary optimization algorithms have been applied in various complex space and practical problems [45–47]. In fact, GP is an optimization technique that applies the concept of natural selection to generate solutions for problems. Using these solutions, solving the problems or estimating the ultimate answers, without specific program-ming requirements, can be achieved. GP has been shown to be very useful for the discovery of the solutions which best fit the data, without presupposing the model struc-ture. These solutions are finally represented in a mathematical relationship form. Moreover, the created relationships can be easily manipulated practically [48].

2.3.4   Cross-validation

When the parameters consist of high ranges, using subsets of data that completely cover data trends is more obvious due to the increment of the performance of models. Today, many cross-validation methods are used to overcome this problem, one of which is k-fold cross-validation. Due to the dynamic nature of the k-fold method, it is able to cover all data trends in both training and testing samples. In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples, of which a single subsample is retained as the validation data for testing the model, and the remaining $k - 1$ subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results can then be averaged to produce a single estimation. In science, in general, a k value of 10 is commonly used, and this was applied in this research [49,50].

  In order to select the training and testing data for evaluating algorithms, the holdout method is adopted. In the holdout method, data points are randomly assigned to two sets, usually called the training set and the testing set. The size of each of the sets is arbitrary although the testing set is typically smaller than the training set. In typical cross-validation, results of multiple runs of model testing are averaged together; in contrast, the holdout method, in isolation, involves a single run. It should be used with caution because without such averaging of multiple runs, one may achieve highly misleading results [51,52].

3   Results

In this section, the relationship between *SMS*, flow rate, and density of pedestrians was determined and plotted using regression analysis. Then using the independent variables of flow rate and *SMS*, the prediction models of

pedestrian density were obtained by GMDH and GP models. GMDH and GP models using k-fold cross-validation method were applied to predict pedestrian density and results were compared with those of the holdout method (i.e., 70% of data was used for training and the rest was applied for testing the dataset). Finally, the dependent variable of density was predicted in various facilities using the superior model.

3.1   Regression analysis

The main goal of this research was to determine the relationship between *SMS*, flow rate (*V*) and density (*D*) of pedestrians. To specify these relationships, the number of pedestrians crossing each facility was first obtained by video recording. The time for all pedestrians crossing (8490 pedestrians) was then achieved using a stopwatch and *SMS* of movement of individuals was provided by dividing the traveled distance by the crossing time. Then, using $V = SMS \times D$ proposed by the HCM [53], the density for each minute was calculated. Moreover, the flow rate-density, *SMS*-flow rate and *SMS*-density curves were plotted by the dataset achieved in various facilities. Eventually, the mathematical relationship between vari-ables was formulated by the use of regression analysis, in which the formula is represented on the curves. The relationships between flow rate-density, *SMS*-flow rate and *SMS*-density of sidewalks are illustrated in Fig. 3. Also, these curves in other facilities such as walkable street (in Fig. 4), mid-block crosswalks (through pede-strian crossing) (in Fig. 5), mid-block crosswalks (out of pedestrian crossing) (in Fig. 6), signalized crosswalks (through pedestrian crossing) (in Fig. 7) and signalized crosswalks (out of pedestrian crossing) (in Fig. 8) are indicated separately.

  As shown in Figs. 3–8, the highest relationships between the variables were observed in flow rate-density curves. To observe the relationship between flow rate, *SMS* and density of pedestrians, the equations for all facilities were drawn. Results of the pedestrians crossing the sidewalks were achieved using Eqs. (8)–(10):
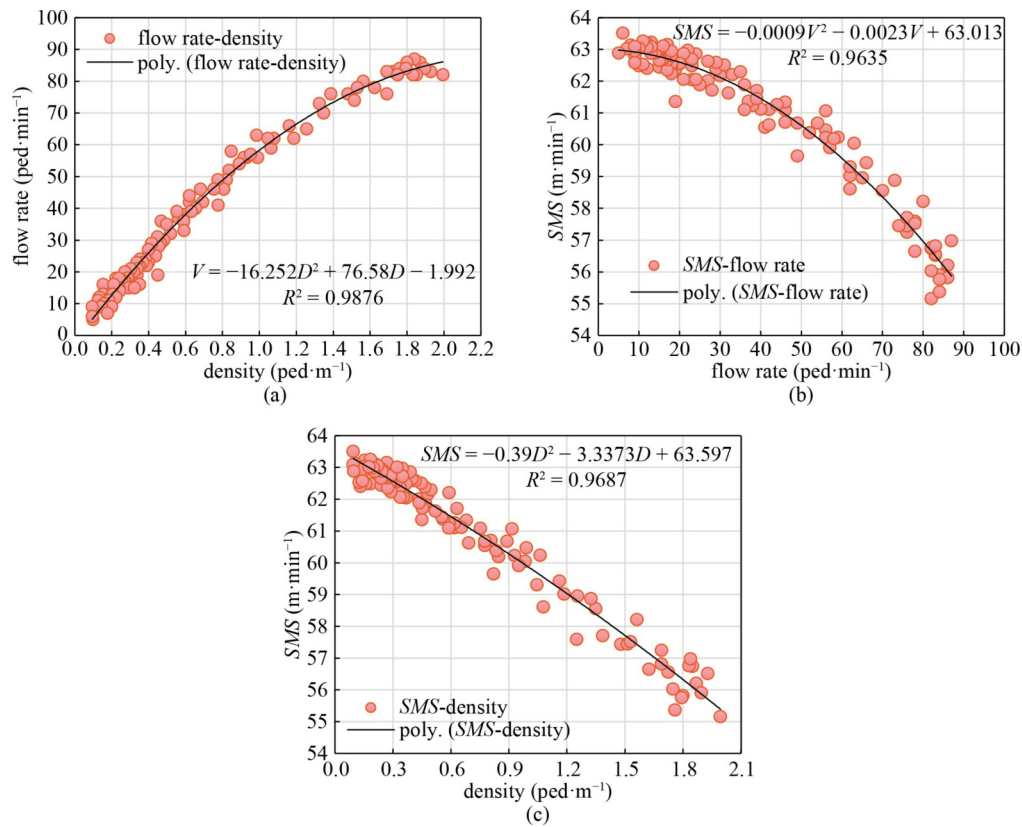
$$V = -16.252D^2 + 76.58D - 1.992, \tag{8}$$
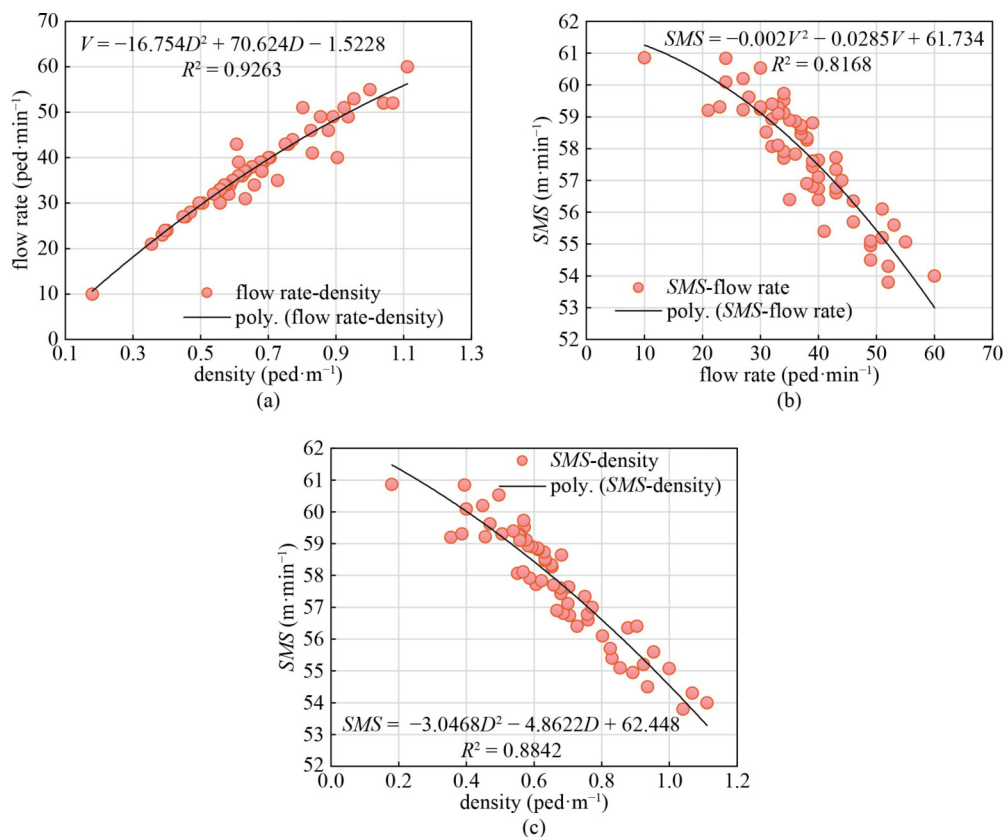
$$SMS = -0.0009V^2 - 0.0023V + 63.013, \tag{9}$$

$$SMS = -0.39D^2 - 3.3373D + 63.597. \tag{10}$$

  In which *SMS* is space mean speed (m·min^{-1}), *V* represents flow rate (ped·min^{-1}) and *D* (ped·m^{-2}) is density. The results of the pedestrians crossing the walkable street are as Eqs. (11)–(13):
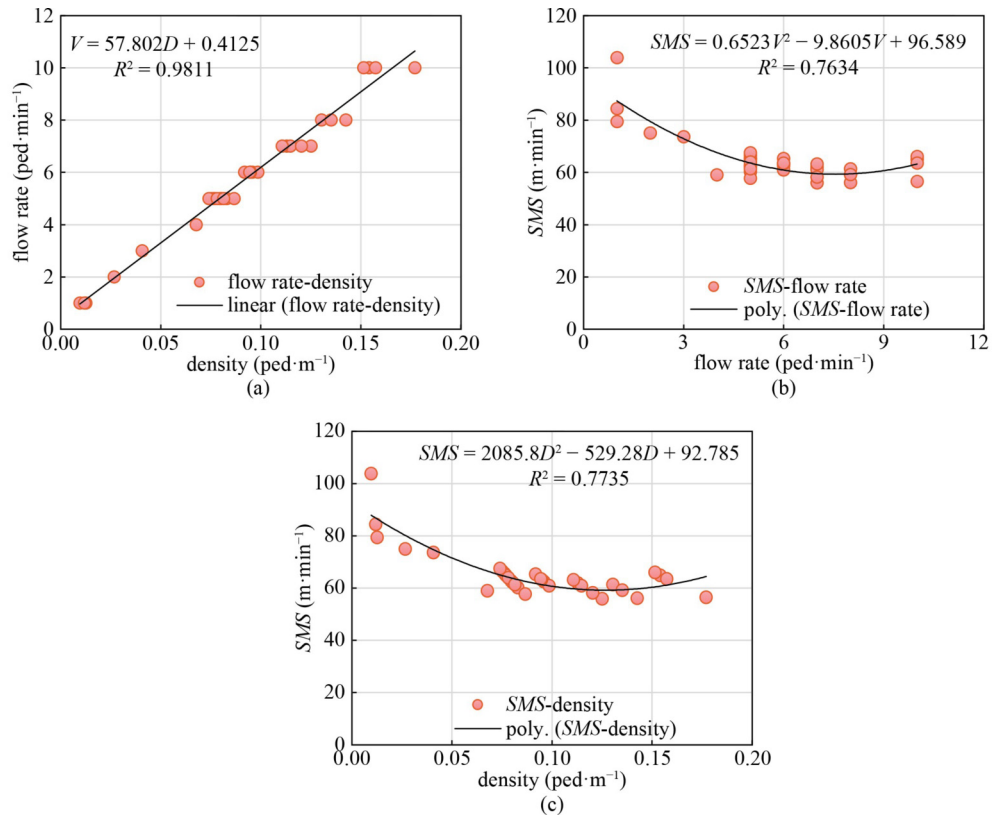
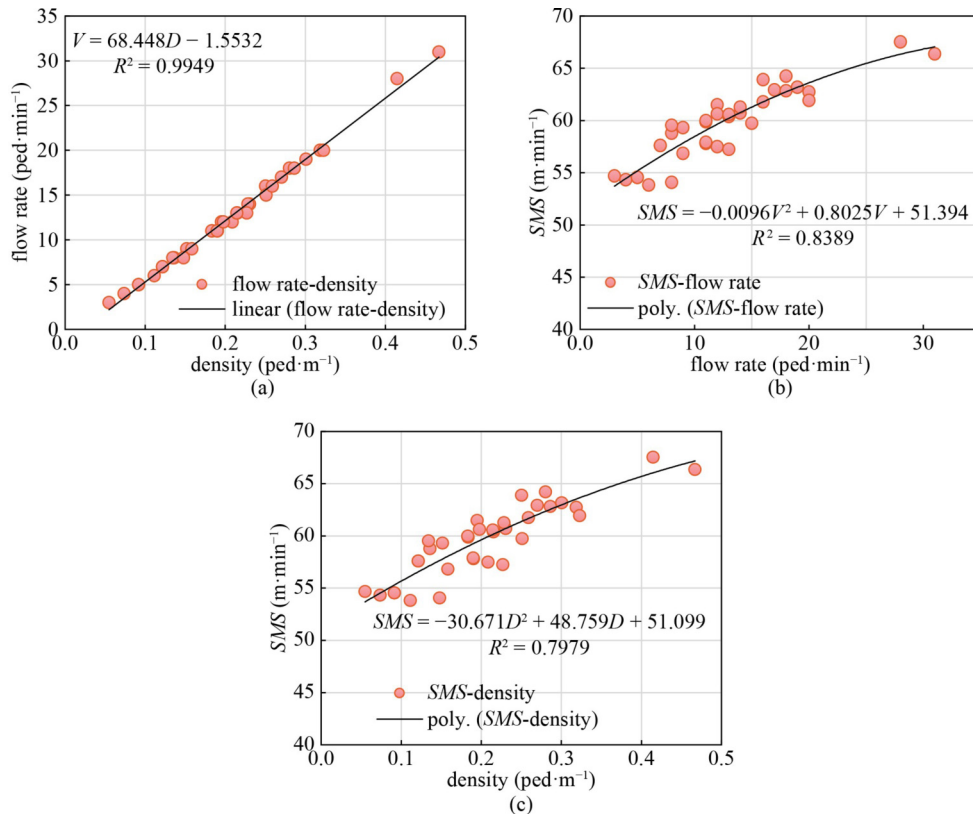$$V = -16.754D^2 + 70.624D - 1.5228, \tag{11}$$

Fig. 3 The relationship between variables in the sidewalks in terms of: (a) flow rate-density; (b) *SMS*-flow rate; (c) *SMS*-density.

Fig. 4 The relationship between variables in the walkable in terms of: (a) flow rate-density; (b) *SMS*-flow rate; (c) *SMS*-density.
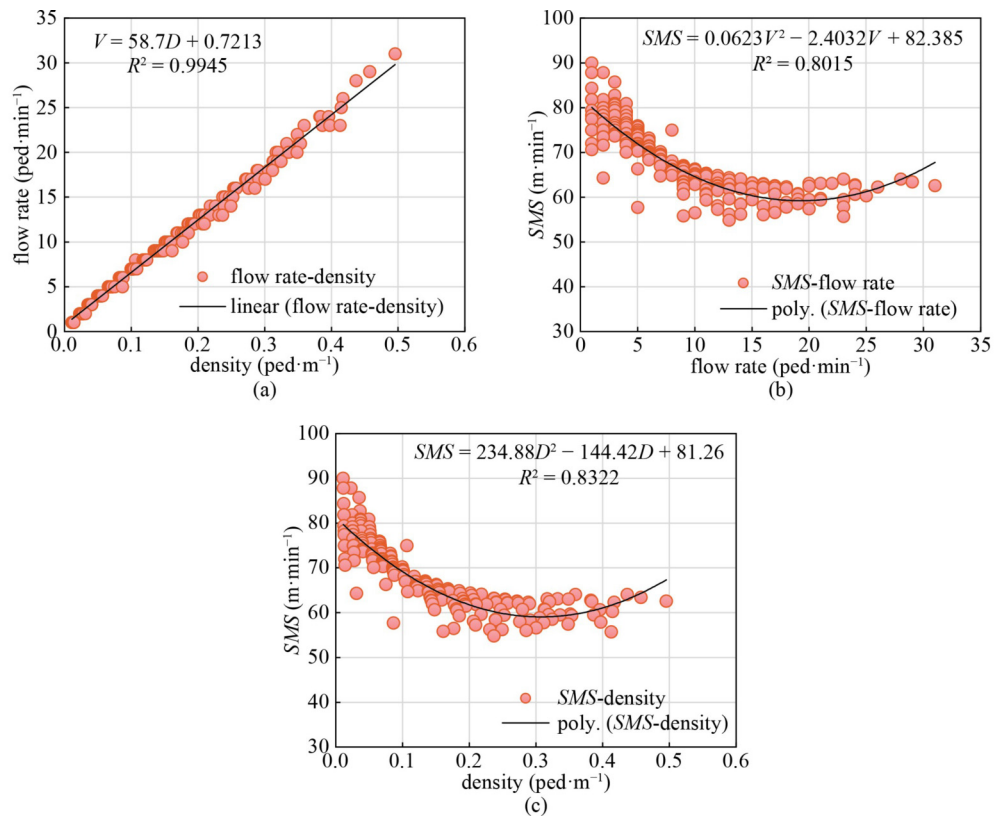
Fig. 5 The relationship between variables in the mid-block crosswalks (through pedestrian crossing) in terms of (a) flow rate-density; (b) *SMS*-flow rate; (c) *SMS*-density.
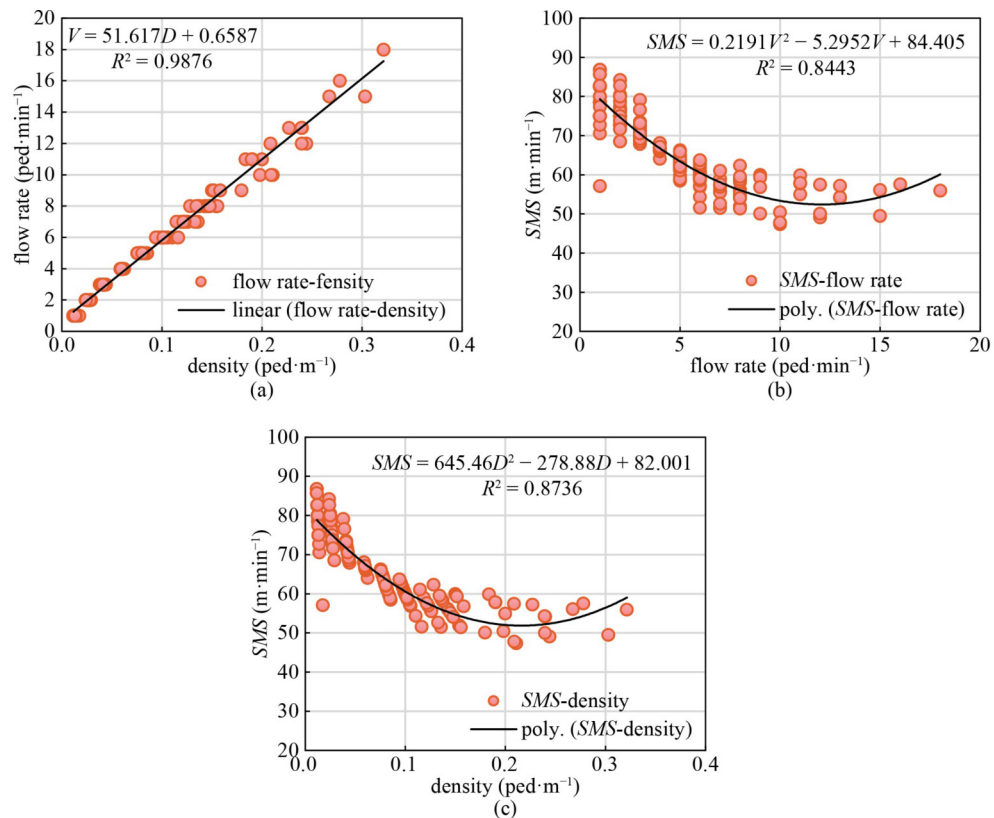

Fig. 6 The relationship between variables in the mid-block crosswalks (out of pedestrian crossing) in terms of: (a) flow rate-density; (b) *SMS*-flow rate; (c) *SMS*-density.

Fig. 7 The relationship between variables in the signalized crosswalks (through pedestrian crossing) in terms of: (a) flow rate-density; (b) *SMS*-flow rate; (c) *SMS*-density.

Fig. 8 The relationship between variables in the signalized crosswalks (out of pedestrian crossing) in terms of: (a) flow rate-density; (b) *SMS*-flow rate; (c) *SMS*-density.

$$SMS = -0.002V^2 - 0.0285V + 61.734, \quad (12)$$

$$SMS = -3.0468D^2 - 4.8622D + 62.448. \quad (13)$$

The results of the pedestrians crossing the mid-block crosswalks (through pedestrian crossing) are illustrated as Eqs. (14)–(16):

$$V = 57.802D + 0.4125, \quad (14)$$

$$SMS = 0.6523V^2 - 9.8605V + 96.589, \quad (15)$$

$$SMS = 2085.8D^2 - 529.28D + 92.785. \quad (16)$$

The results corresponding to the pedestrians crossing the mid-block crosswalks (out of pedestrian crossing) are presented as Eqs. (17)–(19):

$$V = 68.448D - 1.5532, \quad (17)$$

$$SMS = -0.0096V^2 + 0.8025V + 51.394, \quad (18)$$

$$SMS = -30.671D^2 + 48.759D + 51.099. \quad (19)$$

Moreover, the flow rate and *SMS* results of the pedestrians crossing the signalized crosswalks (through pedestrian crossing) are represented as Eqs. (20)–(22):

$$V = 58.7D + 0.7213, \quad (20)$$

$$SMS = 0.0623V^2 - 2.4032V + 82.385, \quad (21)$$

$$SMS = 234.88D^2 - 144.42D + 81.26. \quad (22)$$

Finally, the results of the pedestrians crossing the signalized crosswalks (out of pedestrian crossing) are represented as Eqs. (23)–(25):

$$V = 51.617D + 0.6587, \quad (23)$$

$$SMS = 0.2191V^2 - 5.2952V + 84.405, \quad (24)$$

$$SMS = 645.46D^2 - 278.88D + 82.001. \quad (25)$$

3.2 GMDH model

The density prediction model results obtained by the GMDH technique are illustrated in Table 3 using two cross-validation methods (k-fold and holdout). Moreover, due to the dynamic nature of k-fold method, the model was run k times and finally, the average results for R^2 and root-mean-square error (*RMSE*) were introduced. Table 3 indicates that the application of the k-fold method in the GMDH model led to better performances than the holdout method. Thus, the R^2 values of k-fold method for density model in all facilities, including sidewalks, walkable street, mid-block crosswalks (through pedestrian crossing), mid-block crosswalks (out of pedestrian crossing), signalized crosswalks (through pedestrian crossing) and signalized crosswalks (out of pedestrian crossing) were obtained as 0.9724, 0.9564, 0.9726, 0.9787, 0.9813, and 0.9851, respectively. This indicates that the model in the k-fold cross-validation method was able to represent the actual amounts appropriately. However, the main point of using k-fold returns to the proper utilization of data for training-testing processes and hence increasing the reliability of the prediction model [49].

The correlation of the predicted and actual amounts achieved by the density model using holdout and k-fold cross-validation methods is presented for all facilities mentioned, as shown in Fig. 9.

Figure 10 lists the outputs of the density model for various facilities aforementioned, and the comparison between the predicted and actual amounts. For example, Fig. 10(a) illustrates that the training process for the sidewalks was stopped after 120 repetitions. This point shows that the solutions will no longer improve from this point on, and this point with the *RMSE* of 0.0018 in the k-fold cross-validation method indicates the best point for the end of the calculations and creation of the model for the mentioned facility.

3.3 GP model

In this section, by the use of the independent variables of

Table 3   Density model results presented by GMDH method

number	model	holdout		k-fold	
		R^2	*RMSE*	R^2	*RMSE*
a	sidewalks	0.9691	0.1145	0.9724	0.0976
b	walkable street	0.9437	0.0444	0.9564	0.0383
c	mid-block crosswalks (through pedestrian crossing)	0.9554	0.0138	0.9726	0.0085
d	mid-block crosswalks (out of pedestrian crossing)	0.9755	0.0153	0.9787	0.0142
e	signalized crosswalks (through pedestrian crossing)	0.9783	0.0301	0.9813	0.0194
f	signalized crosswalks (out of pedestrian crossing)	0.9750	0.0232	0.9851	0.0086
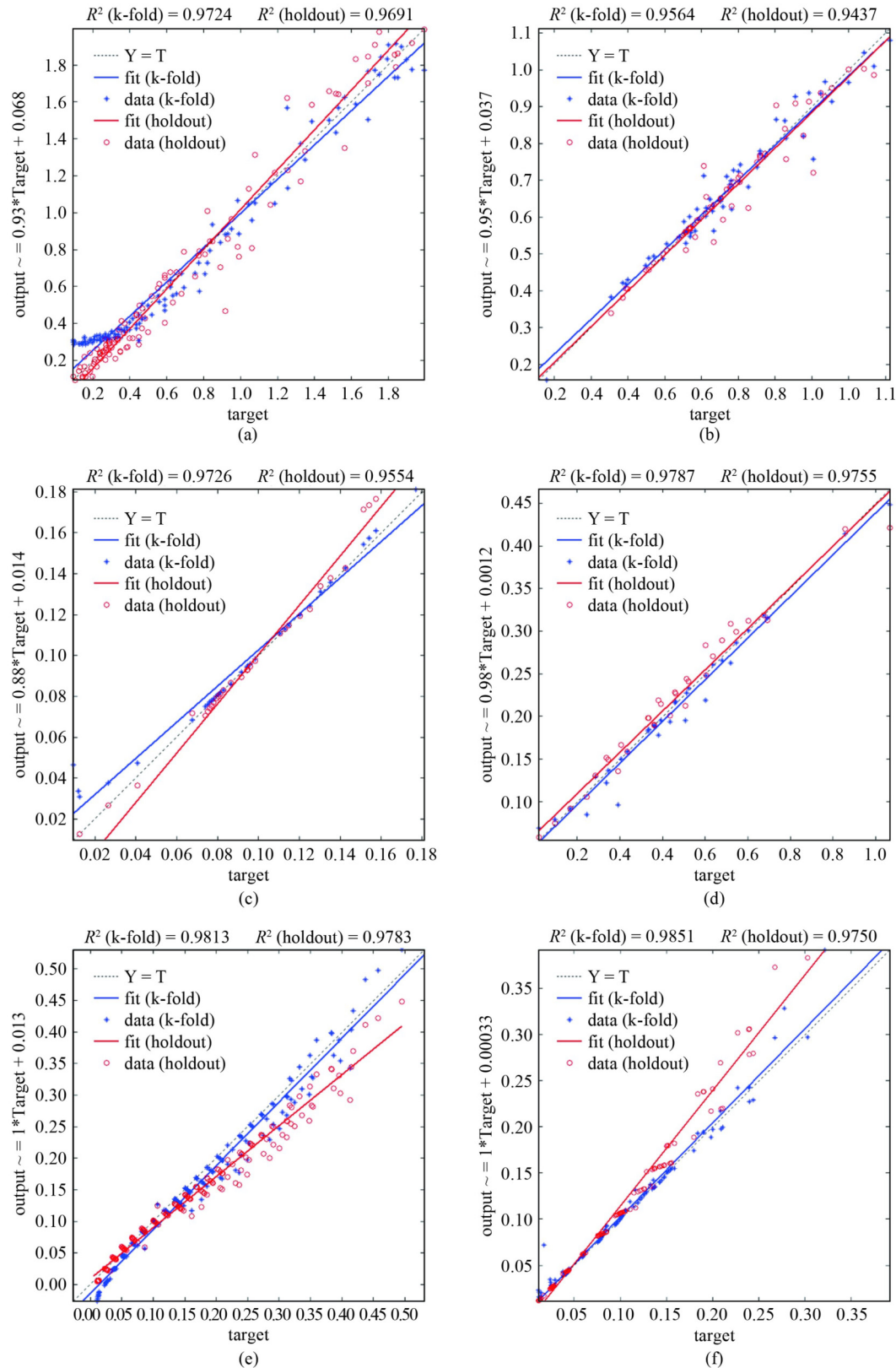
Fig. 9 The correlation of the predicted and actual amounts in GMDH model for: (a) sidewalks; (b) walkable street; (c) mid-block crosswalks (through pedestrian crossing); (d) mid-block crosswalks (out of pedestrian crossing); (e) signalized crosswalks (through pedestrian crossing); (f) signalized crosswalks (out of pedestrian crossing).
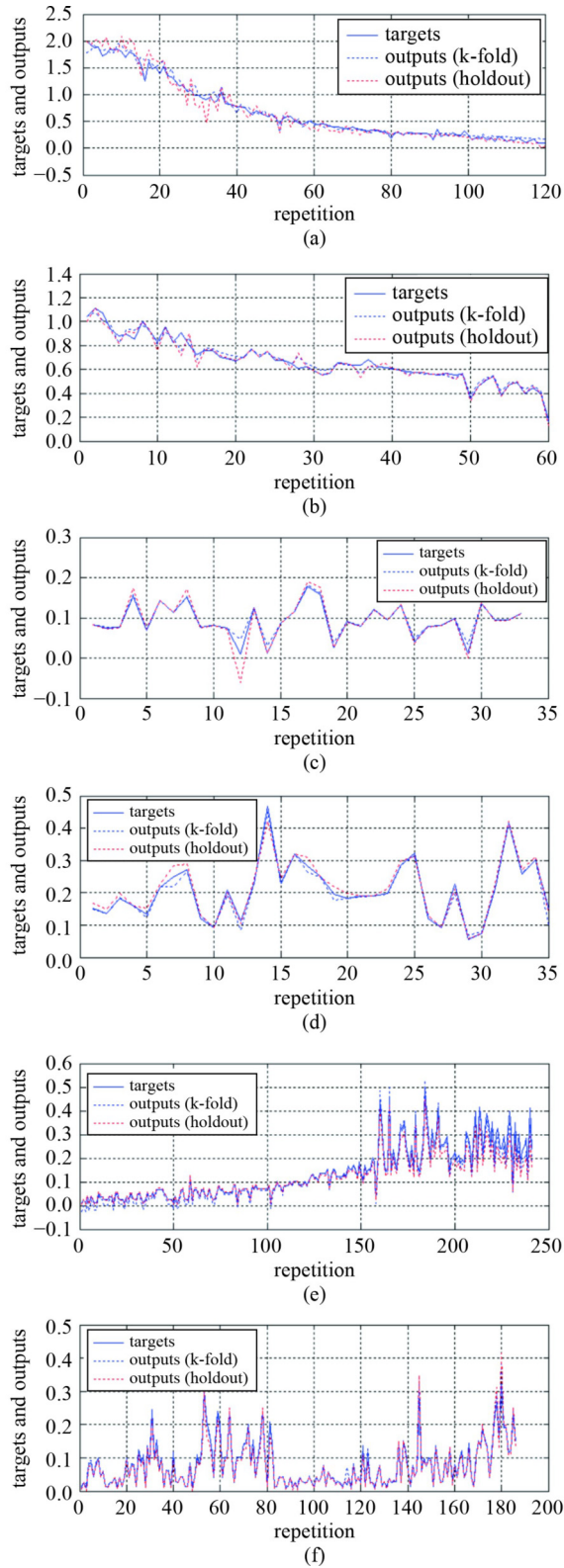
Fig. 10 The comparison chart of the predicted and actual amounts in GMDH method for: (a) sidewalks; (b) walkable street; (c) mid-block crosswalks (through pedestrian crossing); (d) mid-block crosswalks (out of pedestrian crossing); (e) signalized crosswalks (through pedestrian crossing); (f) signalized crosswalks (out of pedestrian crossing).

flow rate and *SMS*, the prediction model of pedestrian density using two cross-validation methods (k-fold and holdout) was obtained using the GP model. The results of the density prediction model in GP technique are illustrated in Table 4, which indicates the superiority of the k-fold method over the holdout method in the GP model. The R^2 values of the k-fold method for density model in all facilities, including sidewalks, walkable street, mid-block crosswalks (through pedestrian crossing), mid-block crosswalks (out of pedestrian crossing), signalized crosswalks (through pedestrian crossing) and signalized crosswalks (out of pedestrian crossing) were obtained as 0.9967, 0.9785, 0.9952, 0.9975, 0.9981, and 0.9999, respectively, indicating that the GP model in the k-fold cross-validation method for all facilities was also able to illustrate the observed amounts well. Moreover, the population size, number of generations, tournament size, maximum tree depth and maximum genes for all facilities were 200, 100, 4, 8, and 7, respectively, and the best fitness values achieved for the aforementioned facilities were 0.00075663, 0.020988, 0.00037482, 0.00081453, 0.00065317, and 0.04674, respectively.

According to Fig. 11, the correlation of the predicted and actual amounts by GP model for all facilities mentioned is presented.

The proposed GP model using the k-fold method (as the superior method) for all facilities, including sidewalks, walkable street, mid-block crosswalks (through pedestrian crossing), mid-block crosswalks (out of pedestrian crossing), signalized crosswalks (through pedestrian crossing) and signalized crosswalks (out of pedestrian crossing) is presented in Eqs. (26)–(31), in which the parameters are already defined.

$$
\begin{aligned}
D = {} & 0.1015 \times \sin V - 0.0162 \times \cos(\sin SMS) \\
& \times e^{V + \tanh(\cos(\cos SMS))} - 0.0125 \times \cos(\cos SMS) \\
& \times \cos(V + SMS \times V) \times SMS^2 - 1.692 \times e^{\tanh(\cos SMS)} \\
& + \cos(\tanh(\cos^2 V + V^3)) + 1.548,
\end{aligned} \tag{26}
$$

$$
\begin{aligned}
D = {} & 0.0257 \times \cos(SMS + \sin(\cos SMS^2) + 6.802) \\
& - 0.0825 \times V - 1.201 \times e^{\cos SMS^3} + 0.0387 \\
& \times \sin(2 \times V + SMS - 0.8248) - SMS \times e^{\sin SMS + 0.8248} \\
& + 0.1509 \times V \times \tanh V - 0.087 \\
& \times (SMS - \sin(2 \times V + SMS - 0.8248)) + 6.765,
\end{aligned} \tag{27}
$$

$$
\begin{aligned}
D = {} & 0.0439 \times \sin(e^{SMS} \times (\tanh(V - SMS) + \cos SMS^3 + \\
& 0.3868 \times e^V + 0.408)) - 0.0009 \times SMS + 0.0021 \\
& \times \sin(V^2 \times \sin(0.3713 \times e^{SMS}) + 0.1514 \times \sin(2 \times \tanh V \\
& - 2.409 \times SMS + 0.9081) + \sin(V^3 \times SMS) \\
& + 1.408 \times \cos V + 0.0246,
\end{aligned} \tag{28}
$$

Table 4 Density model results illustrated in GP technique

number	model	holdout		k-fold	
		R^2	*RMSE*	R^2	*RMSE*
a	sidewalks	0.9895	0.0579	0.9967	0.0018
b	walkable street	0.9676	0.0331	0.9785	0.0298
c	mid-block crosswalks (through pedestrian crossing)	0.9913	0.0029	0.9952	0.0021
d	mid-block crosswalks (out of pedestrian crossing)	0.9962	0.0042	0.9975	0.0014
e	signalized crosswalks (through pedestrian crossing)	0.9966	0.0064	0.9981	0.0012
f	signalized crosswalks (out of pedestrian crossing)	0.9998	0.0009	0.9999	0.0004

$$
\begin{aligned}
D = {} & 0.4347 - 0.0561 \times SMS - 1.692 \\
& \times (V - 1.8524 \times \cos SMS - 1.52 \times \sin^2 V) \\
& - 1.113 \times SMS^2 \times \tanh(SMS \times V^3) - 0.001 \\
& \times e^{2 \times V + \sin SMS - \cos V} - 0.0327 \times SMS \times e^{\sin V^2} \\
& + V \times \sin(V^2 + 3 \times SMS - 0.8248) \\
& + \tanh(\cos SMS) - 1.0413,
\end{aligned} \tag{29}
$$

$$
\begin{aligned}
D = {} & 0.0629 \times \cos(e^{SMS} \times (V - 9.28)) \\
& - 0.2231 \times \tanh(SMS + \sin(SMS \times V^3)) \\
& - 0.3817 \times \sin(e^{V \times (SMS^2 - 9.358)^3 \times (\cos(SMS - 2 \times V) - 1.457)} \\
& - e^{V \times (3 \times SMS^2 - 9.153)} + 0.0134 \times \sin(\sin(SMS \times (V - 9.28))) \\
& \times \sin(\cos V) \times (SMS - \tanh V^2) \\
& \times (\tanh(2 \times SMS) - 9.28) + 0.6887,
\end{aligned} \tag{30}
$$

$$
\begin{aligned}
D = {} & 1.816 \times SMS + 0.0265 \times \tanh(SMS - V^2) \\
& + 0.6829 \times \tanh(\cos SMS) + 0.2109 \times \tanh(e^{3.7146 \times V^3}) \\
& + 0.0041 \times \sin(\cos V^2) \times \cos^2 SMS \\
& - \tanh(SMS^2 \times (V + 0.6731)) - 0.0741 \\
& \times \tanh(\cos SMS) \times e^{SMS \times \tanh V} + 2.2405 \\
& \times \tanh(\tanh V^3) \times e^{\cos SMS^2 + 0.051} \times SMS \\
& + \sin(\tanh(\cos(0.1437 \times SMS^2))) \\
& \times (3.545 \times V + e^{\sin SMS}) + 2.275.
\end{aligned} \tag{31}
$$

4   Conclusion and future directions

In this research, the relationship between *SMS*, flow rate, and density of pedestrians was investigated in order to explore the behavior of pedestrians in various facilities. First, the normality of data and the correlation of variables were examined. Then regression analysis was used to determine the relationship between these variables in the facilities. Eventually, density prediction models were obtained using GP and GMDH models in two cross-validation methods (k-fold and holdout) for different facilities. The most significant results of the research are as follows.

1) The results of normality of dataset based on kurtosis and skewness showed that *SMS*, flow rate and density in various facilities (except *SMS* data in mid-block crosswalks (through pedestrian crossing)) had a normal distribution.

2) The Pearson correlation analysis results indicated a high Pearson correlation between flow rate, *SMS*, and density of pedestrians for all facilities.

3) Regression analysis illustrated the mathematical relationships between flow rate-density, *SMS*-flow rate and *SMS*-density in all facilities and showed that the highest relationships between the variables were in flow rate versus density curves.

4) Results indicated that the application of k-fold method in GMDH and GP models led to better performances than the holdout method, meaning that using k-fold as the cross-validation method increased the model reliability.

5) The model presented by GMDH technique showed that the R^2 values obtained for the k-fold method for all facilities, including sidewalks, walkable street, mid-block crosswalks (through pedestrian crossing), mid-block crosswalks (out of pedestrian crossing), signalized crosswalks (through pedestrian crossing) and signalized crosswalks (out of pedestrian crossing) were 97.24%, 95.64%, 97.26%, 97.87%, 98.13%, and 98.51%, respectively.

6) The GP model predicted the density of pedestrians in terms of flow rate and *SMS* with greater accuracy and less error compared to GMDH, so that the R^2 values of GP model in k-fold cross-validation method for all aforementioned facilities were 99.67%, 97.85%, 99.52%, 99.75%, 99.81%, and 99.99%, respectively.

7) For future research work, other machine learning techniques [54–57], deep learning models [58–62], and optimization algorithms [63–65] can be incorporated into the proposed approaches to obtain more accurate results in the prediction of pedestrian density. Moreover, this research can be easily extended to other case studies. The effects of various variables, such as environmental factors and travel type (recreational, educational, etc.), on the *SMS*, flow rate, and density of pedestrians and the relationship between them can also be examined. Furthermore, calibration of the simulation software can be performed with the results from the present study.
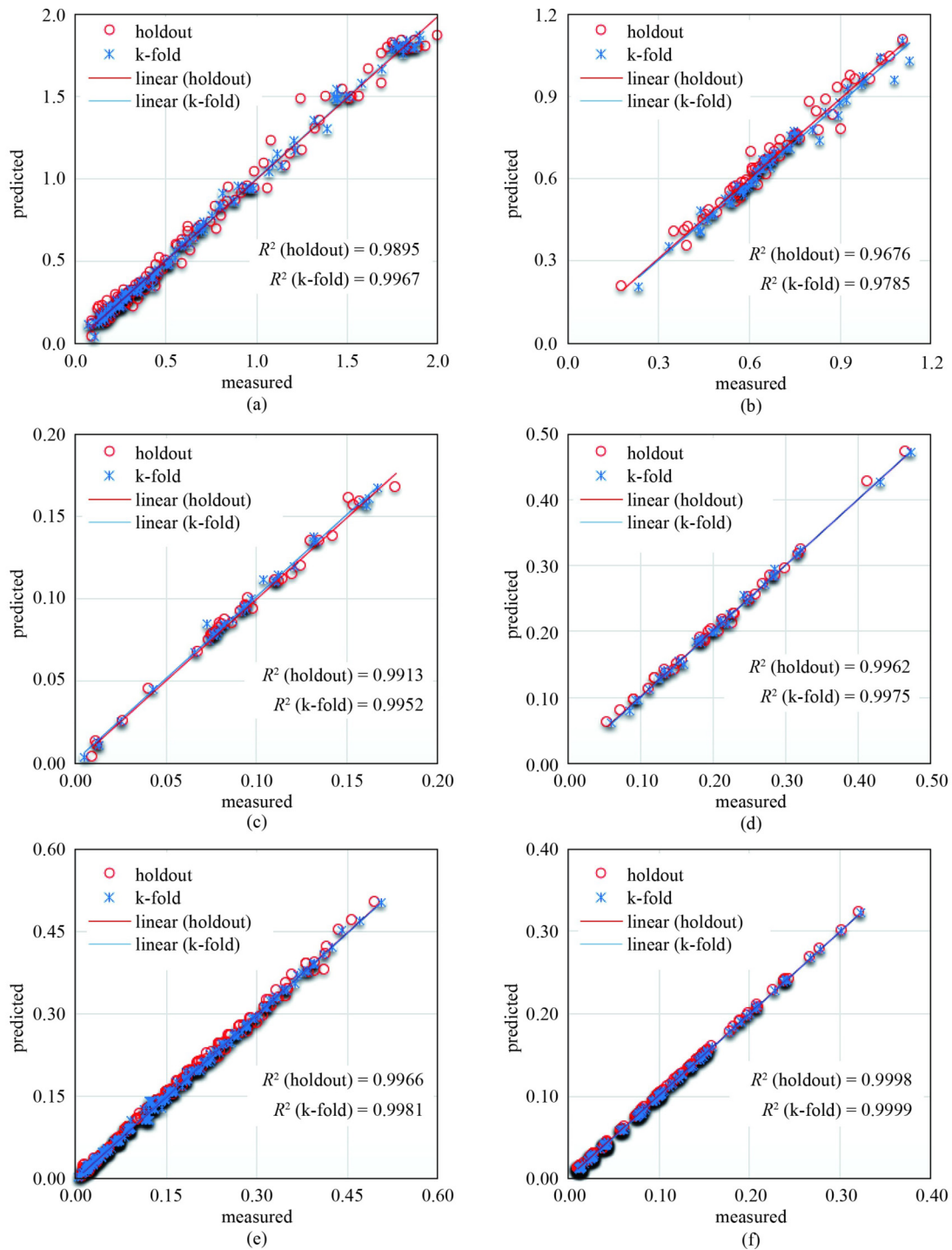
Fig. 11　The correlation chart of the predicted and actual amounts in GP model for: (a) sidewalks; (b) walkable street; (c) mid-block crosswalks (through pedestrian crossing); (d) mid-block crosswalks (out of pedestrian crossing); (e) signalized crosswalks (through pedestrian crossing); (f) signalized crosswalks (out of pedestrian crossing).

References

1. Haghighi M, Bakhtari F, Sadeghi-Bazargani H, Nadrian H. Strategies to promote pedestrian safety from the viewpoints of traffic and transport stakeholders in a developing country: A mixed-method study. Journal of Transport & Health, 2021, 22: 101125

2. Chen S, Fu L, Fang J, Yang P. The effect of obstacle layouts on pedestrian flow in corridors: An experimental study. Physica A, 2019, 534: 122333

3. Guidoni D L, Maia G, Souza F S H, Villas L A, Loureiro A A F. Vehicular traffic management based on traffic engineering for vehicular ad hoc networks. IEEE Access: Practical Innovations, Open Solutions, 2020, 8: 45167−45183

4. World Health Organization. Global Status Report on Road Safety 2018. 2018

5. Mohammadi A, Yousefi M, Taghipour A, Ebrahimipour H, Varmaghani M. Burden of disease caused by road traffic accidents in the city of mashhad. Health Scope, 2020, 9(4): 101657

6. Hadaye R S, Rathod S, Shastri S. A cross-sectional study of epidemiological factors related to road traffic accidents in a metropolitan city. Journal of Family Medicine and Primary Care, 2020, 9(1): 168−172

7. McIlroy R C, Kokwaro G O, Wu J, Jikyong U, Nam V H, Hoque M S, Preston J M, Plant K L, Stanton N A. How do fatalistic beliefs affect the attitudes and pedestrian behaviours of road users in different countries? A cross-cultural study. Accident Analysis and Prevention, 2020, 139: 105491

8. Hughes R L. A continuum theory for the flow of pedestrians. Transportation Research Part B: Methodological, 2002, 36(6): 507−535

9. Helbing D, Johansson A, Al-Abideen H Z. Dynamics of crowd disasters: An empirical study. Physical Review. E, 2007, 75(4): 046109

10. Liu W, Zhou H, He Q. Modeling pedestrians flow on stairways in Shanghai metro transfer station. In: 2008 International Conference on Intelligent Computation Technology and Automation (ICICTA). Changsha: IEEE, 2008

11. Chen X, Ye J, Jian N. Relationships and characteristics of pedestrian traffic flow in confined passageways. Transportation Research Record: Journal of the Transportation Research Board, 2010, 2198(1): 32−40

12. Plaue M, Chen M, Bärwolff G, Schwandt H. Trajectory extraction and density analysis of intersecting pedestrian flows from video recordings. In: ISPRS Conference on Photogrammetric Image Analysis. Berlin: Springer, 2011

13. Shafabakhsh G, Mohammadi M, Mirzanamadi R. Analysis of pedestrians' walking speed in Iran's sidewalks (considering the elderly). Journal of Basic and Applied Scientific Research, 2013, 3(3): 172−182

14. Rastogi R, Chandra S. Pedestrian flow characteristics for different pedestrian facilities and situations. European Transport, 2013, 53: 1−21

15. Bargegol I, Gilani V, Jamshidpour F. Modeling pedestrian flow at central business district. Jurnal UMP Social Sciences and Technology Management, 2015, 3(3): 217−222

16. Pinna F, Murrau R. Age factor and pedestrian speed on sidewalks. Sustainability, 2018, 10(11): 4084

17. Sun Y. Kinetic Monte Carlo simulations of bi-direction pedestrian flow with different walk speeds. Physica A, 2020, 549: 124295

18. Bargegol I, Gilani V N M, Jamshidpour F. Relationship between pedestrians' speed, density and flow rate of crossings through urban intersections (case study: Rasht metropolis). International Journal of Engineering, 2017, 30(12): 1814−1821

19. Najafzadeh M, Barani G A, Azamathulla H M. Prediction of pipeline scour depth in clear-water and live-bed conditions using group method of data handling. Neural Computing & Applications, 2014, 24(3): 629−635

20. Guido G, Haghshenas S S, Haghshenas S S, Vitale A, Gallelli V, Astarita V. Development of a binary classification model to assess safety in transportation systems using GMDH-type neural network algorithm. Sustainability, 2020, 12(17): 6735

21. Koopialipoor M, Nikouei S S, Marto A, Fahimifar A, Jahed Armaghani D, Mohamad E T. Predicting tunnel boring machine performance through a new model based on the group method of data handling. Bulletin of Engineering Geology and the Environment, 2019, 78(5): 3799−3813

22. López J R, Gonzalez L C, Wahlstrom J, Montes y Gomez M, Trujillo L, Ramirez-Alonso G. A genetic programming approach for driving score calculation in the context of intelligent transportation systems. IEEE Sensors Journal, 2018, 18(17): 7183−7192

23. Das S, Raju N, Maurya A K, Arkatkar S. Evaluating lateral interactions of motorized two-wheelers using multi-gene symbolic genetic programming. Transportation Research Record: Journal of the Transportation Research Board, 2020, 2674(9): 1120−1135

24. Pattanaik M L, Choudhary R, Kumar B. Prediction of frictional characteristics of bituminous mixes using group method of data handling and multigene symbolic genetic programming. Engineering with Computers, 2020, 36(4): 1875−1888

25. Blanca M J, Arnau J, López-Montiel D, Bono R, Bendayan R. Skewness and kurtosis in real data samples. Methodology, 2013, 9(2): 78−84

26. Liu Y, Mu Y, Chen K, Li Y, Guo J. Daily activity feature selection in smart homes based on pearson correlation coefficient. Neural Processing Letters, 2020, 51(2): 1−17

27. Xu H, Wu T, Liu Q, Li J. Research on the cut-throwing performance of chopper of sugarcane harvester. Computational Research Progress in Applied Science & Engineering (CRPASE), 2019, 05(03): 85−91

28. Xu B, Lin B. Investigating drivers of CO_2 emission in China's heavy industry: A quantile regression analysis. Energy, 2020, 206: 118159

29. Ivakhnenko A G, Lapa V G. Cybernetic Predicting Devices. New York: CCM Information Corp., 1996

30. Ahmadi M H, Mohseni-Gharyehsafa B, Ghazvini M, Goodarzi M, Jilte R D, Kumar R. Comparing various machine learning approaches in modeling the dynamic viscosity of CuO/water nanofluid. Journal of Thermal Analysis and Calorimetry, 2020, 139(4): 2585−2599

31. Koza J R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. Vol. 1. Cambridge, MA: MIT press, 1992

32. Shahin M, Yun L, Chin C M M, Gao L, Wang C, Niu X, Goyal A, Garg A. An application of genetic programming for lithium-ion battery pack enclosure design: Modelling of mass, minimum natural frequency and maximum deformation case. IOP Conference Series: Earth and Environmental Science, 2019, 268(1): 012065

33. Bai B, Guo Z, Zhou C, Zhang W, Zhang J. Application of adaptive reliability importance sampling-based extended domain PSO on single mode failure in reliability engineering. Information Sciences, 2021, 546: 42−59

34. Zhao D, Liu L, Yu F, Heidari A A, Wang M, Liang G, Muhammad K, Chen H. Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy. Knowledge-Based Systems, 2021, 216: 106510

35. Hu J, Chen H, Heidari A A, Wang M, Zhang X, Chen Y, Pan Z.

Orthogonal learning covariance matrix for defects of grey wolf optimizer: Insights, balance, diversity, and feature selection. Knowledge-Based Systems, 2021, 213: 106684

36. Zhao X, Zhang X, Cai Z, Tian X, Wang X, Huang Y, Chen H, Hu L. Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients. Computational Biology and Chemistry, 2019, 78: 481−490

37. Chen H, Heidari A A, Chen H, Wang M, Pan Z, Gandomi A H. Multi-population differential evolution-assisted Harris hawks optimization: Framework and case studies. Future Generation Computer Systems, 2020, 111: 175−198

38. Cao Y, Li Y, Zhang G, Jermsittiparsert K, Nasseri M. An efficient terminal voltage control for PEMFC based on an improved version of whale optimization algorithm. Energy Reports, 2020, 6: 530−542

39. Gao N, Luo D, Cheng B, Hou H. Teaching-learning-based optimization of a composite metastructure in the 0–10 kHz broadband sound absorption range. Journal of the Acoustical Society of America, 2020, 148(2): EL125−EL129

40. Sun G, Li C, Deng L. An adaptive regeneration framework based on search space adjustment for differential evolution. Neural Computing & Applications, 2021, 33(15): 1−17

41. Liu J, Wu C, Wu G, Wang X. A novel differential search algorithm and applications for structure design. Applied Mathematics and Computation, 2015, 268: 246−269

42. Zhang Y, Liu R, Heidari A A, Wang X, Chen Y, Wang M, Chen H. Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis. Neurocomputing, 2021, 430: 185−212

43. Wang M, Chen H. Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. Applied Soft Computing, 2020, 88: 105946

44. Xu X, Chen H. Adaptive computational chemotaxis based on field in bacterial foraging optimization. Soft Computing, 2014, 18(4): 797−807

45. Xu Y, Chen H, Luo J, Zhang Q, Jiao S, Zhang X. Enhanced Moth-flame optimizer with mutation strategy for global optimization. Information Sciences, 2019, 492: 181−203

46. Li C, Hou L, Sharma B Y, Li H, Chen C S, Li Y, Zhao X, Huang H, Cai Z, Chen H. Developing a new intelligent system for the diagnosis of tuberculous pleural effusion. Computer Methods and Programs in Biomedicine, 2018, 153: 211−225

47. Xia J, Chen H, Li Q, Zhou M, Chen L, Cai Z, Fang Y, Zhou H. Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach. Computer Methods and Programs in Biomedicine, 2017, 147: 37−49

48. Nguyen S, Mei Y, Zhang M. Genetic programming for production scheduling: A survey with a unified framework. Complex & Intelligent Systems, 2017, 3(1): 41−66

49. Esmaeelzadeh R, Borhani Dariane A. Long-term streamflow forecasting by adaptive neuro-fuzzy inference system using k-fold cross-validation (Case study: Taleghan Basin, Iran). Journal of Water Sciences Research, 2014, 6(1): 71−83

50. Wong T T, Yeh P Y. Reliable accuracy estimates from k-fold cross validation. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(8): 1586−1594

51. Nikbakhsh N, Dehghani G, Zamani F. Comparing classification algorithms of data mining in diagnosis of diabetes and assessing the effectiveness of k-fold cross validation in the accuracy of the constructed model. In: International Conference on Engineering and Computer Science. Najafabad: Islamic Azad University Najafabad Branch, 2016

52. Yadav S, Shukla S. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC). Bhimavaram: IEEE, 2016

53. Elefteriadou L A. The highway capacity manual 6th edition: A guide for multimodal mobility analysis. Ite journal, 2016, 86(4): 14−18

54. Zhang Y, Liu R, Wang X, Chen H, Li C. Boosted binary Harris hawks optimizer and feature selection. Engineering with Computers, 2021, 37(4): 3741−3770

55. Tu J, Chen H, Liu J, Heidari A A, Zhang X, Wang M, Ruby R, Pham Q V. Evolutionary biogeography-based whale optimization methods with communication structure: Towards measuring the balance. Knowledge-Based Systems, 2021, 212: 106642

56. Anitescu C, Atroshchenko E, Alajlan N, Rabczuk T. Artificial neural network methods for the solution of second order boundary value problems. Computers, Materials and Continua, 2019, 59(1): 345−359

57. Gharahbash J, Jazani N. An intelligent method for understanding consumers' perception of luxury hotel brands using convolutional neural networks. Computational Research Progress in Applied Science & Engineering (CRPASE), 2020, 06(01): 9−14

58. Chen H, Chen A, Xu L, Xie H, Qiao H, Lin Q, Cai K. A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources. Agricultural Water Management, 2020, 240: 106303

59. Guo H, Zhuang X, Rabczuk T. A deep collocation method for the bending analysis of Kirchhoff plate. Computers, Materials and Continua, 2019, 59(2): 433−456

60. Samaniego E, Anitescu C, Goswami S, Nguyen-Thanh V M, Guo H, Hamdia K, Zhuang X, Rabczuk T. An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications. Computer Methods in Applied Mechanics and Engineering, 2020, 362: 112790

61. Zhuang X, Guo H, Alajlan N, Zhu H, Rabczuk T. Deep autoencoder based energy method for the bending, vibration, and buckling analysis of Kirchhoff plates with transfer learning. European Journal of Mechanics. A, Solids, 2021, 87: 104225

62. Addeh A, Iri M. Brain tumor type classification using deep features of MRI images and optimized RBFNN. ENG Transactions, 2021, 2(1): 1−7

63. Zhang X, Wang J, Wang T, Jiang R, Xu J, Zhao L. Robust feature learning for adversarial defense via hierarchical feature alignment. Information Sciences, 2021, 560: 256−270

64. Golilarz N A, Gao H, Addeh A, Pirasteh S. ORCA optimization algorithm: A new meta-heuristic tool for complex optimization problems. In: 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). Chengdu: UESTC press, 2020: 198−204

65. Addeh A, Hemmati A, Lari A, Munir H. A hybrid diagnostic system to detect COVID-19 Based on selected deep features of chest CT images and SVM. ENG Transactions, 2021, 2(2): 1−18