Journal of Mechatronics, Automation and Identification Technology
Vol. 4, No. 4. pp. 7 - 11

Feature Selection Using Combined Particle Swarm Optimization and Artificial Neural Network Approach

Sandra NEMET, Gordana OSTOJIĆ, Dragan KUKOLJ, Stevan STANKOVSKI, Dragan JOVANOVIĆ

RT-RK Institute for Computer Based Systems, Novi Sad, Serbia
Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
sandra.nemet@rt-rk.com, goca@uns.ac.rs, dragan.kukolj@rt-rk.uns.ac.rs, stevan@uns.ac.rs, draganj@uns.ac.rs
Corresponding Author: Sandra NEMET

***Abstract*— This paper deals with identification of the most influencing input attributes related to the accuracy of the prediction model. It is assumed that the prediction model may be represented by any machine learning-based models, including artificial neural networks, fuzzy models, etc. Selection of influencing attributes is based on particle swarm optimization (PSO) combined with neural networks. The role of neural networks is to estimate fitness of each particle during the search procedure implemented using a PSO algorithm. The presented feature selection method represents the first step in the prediction model design. The method is applied on a dataset characterized with weak correlation between the model's inputs and outputs. Selection of appropriate inputs improves the prediction model accuracy.**

***Keywords*— feature selection, variable subset selection, particle swarm optimization, neural network, prediction model**

I. INTRODUCTION

Feature selection is a process of taking a subset of features from the initial set without transformation. Generally, feature selection methods are classified in three groups of approaches: wrappers, filter approaches and their combination. The wrapper approach uses a machine learning algorithm, repeatedly called by a data re-sampling technique which is dependent on the search strategy. The filter approach looks for features which have stronger measure of association to the output variable. Different measures can be employed, such as similarity, information gain, distance, correlation or other statistical measures.

Growth in road traffic intensity as consequence has a rapid increase in road accidents. Identification of the most influencing factors related to traffic accidents represents a basic phase in analysis of accident causes and their consequences. Artificial neural networks (ANN) are the most frequent machine learning techniques employed in traffic accident severity prediction. Back-propagation neural networks were applied in [1, 2]. Ogrenci [3] predicted accident severity with Multilayer Perceptron (MLP) neural networks. Moreover, Support Vector Machines (SVM) were used for crash severity prediction [4], and traffic fatalities prediction was improved with an SVM model parameters optimization using hybrid particle swarm optimization [5]. In order to design more accurately, a robust and interpretable prediction model with reduction of input data by proper selection of key variables, should be performed [6].

This work describes the problem of finding fetures that are most influential as a problem of searching feature space. As a search technique, the Particle Swarm Optimization (PSO) algorithm was used [7]. The PSO algorithm uses a particle swarm made out of a vast number of particles in a solution candidate population that represent feature subsets. In each PSO iteration the particles „learn“ from one another and refresh their knowledge in relation to the current and best solution, as well as the best solution for the whole swarm. Every particle in the swarm is presented with its position and speed. Fitness of every particle in this PSO algorithm implementation presents an estimation of the crash severity estimated using an MLP type of ANN [8].

The paper is structured as following: In the second section, the attributes selection method is described using a hybrid PSO – ANN approach. The third section contains results of the attributes selection method applied on the traffic accident dataset. In the last section, concluding remarks are given.

II. FEATURE SELECTION USING PSO-ANN ALGORITHM

The feature selection algorithm presented here is based on a Particle Swarm Optimization (PSO) method, which is combined with an artificial neural network (hereinafter referred to as the PSO-ANN algorithm). Input of the ANN is a candidate subset of features, and it is trained to predict crash severity. The calculated prediction error is presented as fitness of the particle and it has an influence on the current position of the particle (feature subset), the direction of the particle, and preferably also on the entire swarm. Therefore, the used hybrid method of feature selection using the PSO-ANN algorithm is characterized with the following:

- Number of particles, i.e. population size, defines how many neural networks will be used during the search for the best feature combination.
- Every neural network has a different combination of input variables in one generation, but their number is constant.
- During the search of the solution space, synaptic weight coefficients of all neural networks are iteratively adapted with an aim to achieve a minimal error.
- The search method using the hybrid PSO-ANN algorithm has ended when the given number of generations is achieved.

Particle swarm optimization algorithm was first described by J. Kennedy and R. C. Eberhart in 1995 [7]. The main idea of the method resulted from an association with social behaviour of an individual from a flock of birds or fishes. PSO is a stochastic search algorithm based on a population of adaptive solutions by simulating the behaviour of individuals in the swarm. The PSO approach has become popular because of the simplicity of the basic PSO algorithm implementation [9] and because it is characterized by low computational complexity [10].

Particles are moving through the hyper-dimensional space in search of new solutions. The position of a particle denoted by vector $\boldsymbol{x_i}$ (in iteration t) is changed by adding the speed vector $\boldsymbol{v_i}$ in the current position:

$$\boldsymbol{x_i}(t) = \boldsymbol{x_i}(t-1) + \boldsymbol{v_i}(t). \quad (1)$$

Therefore, equation (1) calculates the new particle position. Speed vector $\boldsymbol{v_i}$ defines where will the particle move in the next iteration, considering its previous best positions, as well as the „experience" of the most successful particle from the swarm. Speed vector $\boldsymbol{v_i}$ of the i-th particle is defined by the following equation:

$$\boldsymbol{v_i}(t) = W\boldsymbol{v_i}(t-1) + C_1 r_1 (x_{pbest_i} - \boldsymbol{x_i}(t)) + C_2 r_2 (\boldsymbol{x}_{gbest} - \boldsymbol{x_i}(t)), \quad (2)$$

where r_1 and r_2 are random values in the interval [0,1], W is the inertion coefficient, while C_1 and C_2 represent constants for controlling the influence of the best local solution for a particle $\boldsymbol{x}_{pbest}$ and global best solution for particle $\boldsymbol{x}_{gbest}$, respectively. In each iteration of the PSO algorithm, particles "learn" from each other and find new solutions on a better location in the solution space. Every particle memorizes the local best position (*pbest*) until the given moment, while the best solution for the entire swarm is memorized in the global best position of all particles (*gbest*). These two positions have an influence on every particle in the PSO set of solutions, besides the current moving direction of every particle.

The basic PSO algorithm pseudo-code is given in Table I. Stopping criteria for this algorithm is determined by at least one of the following conditions: maximum number of iterations, a defined number of iterations without achieving an improvement with the best solution, or the achievement of the given minimal error for the fitness function.

TABLE I PSEUDO-CODE OF THE BASIC PSO ALGORITHM VERSION

```
Input: Random initialization of position and speed of all particles:
        Xi (0), Vi (0)
Output: The approximate position of the global minima X*

while Stopping criteria not satisfied do
   for i = 1 to Particle_number do
      Calculate fitness function f
      Update local pbest and global gbest position of each particle
      Update particle speed using eq. (2)
      Update particle position using eq. (1)
   end for
end while
```

Feature selection is based on ranking features based on importance using the PSO-ANN method defined by a 100 particles (individual solutions) and 50 iterations. The complete algorithm is repeated 20 times, in order to reduce the stochastic nature of the search method. Inertia coefficient W =0.754 was adopted, while a value of C_1=C_2=1.5457 was adopted for the personal (C_1) and global (C_2) learning coefficient. Values of the best particles, i.e. particles with the best combination of input variables are used in forming the importance index of the following form:

$$\sum_i^m \sum_j^n \left(1 - \frac{c_{ij} - c_{min}}{c_{max} - c_{min}}\right), \quad (3)$$

with the meaning: m – number of repetitions; n – number of variables; c_{ij} – value of the solution for the i-th repetition in which the j-th variable participated; c_{max} – the weakest achieved result; and c_{min} – the best achieved result.

III. RESULTS OF FEATURE SELECTION OBTAINED WITH THE PSO-ANN ALGORITHM

The accident dataset used for this study contained traffic accident records from year 2008 to 2011 in the urban area of the city of Novi Sad [11]. The dataset comprised of the official statistics on traffic accidents, with a total of 42 features and 752 traffic accidents, from which 798 were pedestrian casualties. All factors involved in accidents are classified and labelled, including consequences which consist of four classes: fatalities, major injuries, minor injuries and without injuries.

A neural network with one hidden layer and fitness function in the form of the mean absolute error between the real and estimated output was embedded in each particle of the PSO+ANN algorithm. In case of five input variables, neural network has eleven neurons on the hidden layer and one output neuron with the meaning of variable *Consequences*.

Different cases for a variable set of input variables were considered. When considering the choice of the best five features, the PSO-ANN algorithm has selected variables shown on Figure 1. The initial set with 25 input variables corresponds to the variable list given on the ordinate axis in Fig. 1. Bars shown above the number of the corresponding variable represent a variable's importance index (3). Based on the importance index from equation (3), a subset of variables [9, 11, 20, 24, 25] can be selected, i.e. [*Gender*, *Behaviour*, *Age groups*, *Pedestrian crossing*, *Road crossing*]. Interestingly, a combination of features [11, 17, 20, 24, 25] was achieved using the PSO-ANN algorithm, i.e. [*Behaviour*, *Gender groups*, *Age groups*, *Pedestrian crossing*, *Road crossing*], with a minimal value of 0.026 for the objective function.

As an illustration of PSO-ANN algorithm convergence, Fig. 2 shows 50 iterations of one PSO algorithm execution (randomly selected a single run of the algorithm execution). The algorithm converged towards the best combination of features [9, 11, 20, 24, 25] in a first few iterations, i.e. [*Gender*, *Behaviour*, *Age groups*, *Pedestrian crossing*, *Road crossing*], where after the seventh iteration it still minimized the fitness function, but with the same feature combination. This solution represents one of the best solutions achieved.

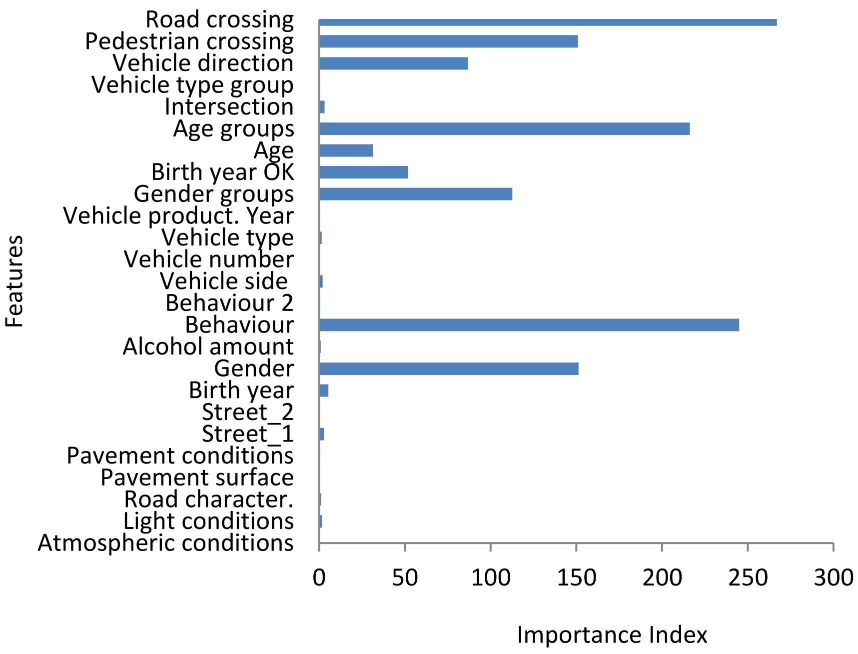


Fig. 1 Selection of five features using the PSO-ANN algorithm

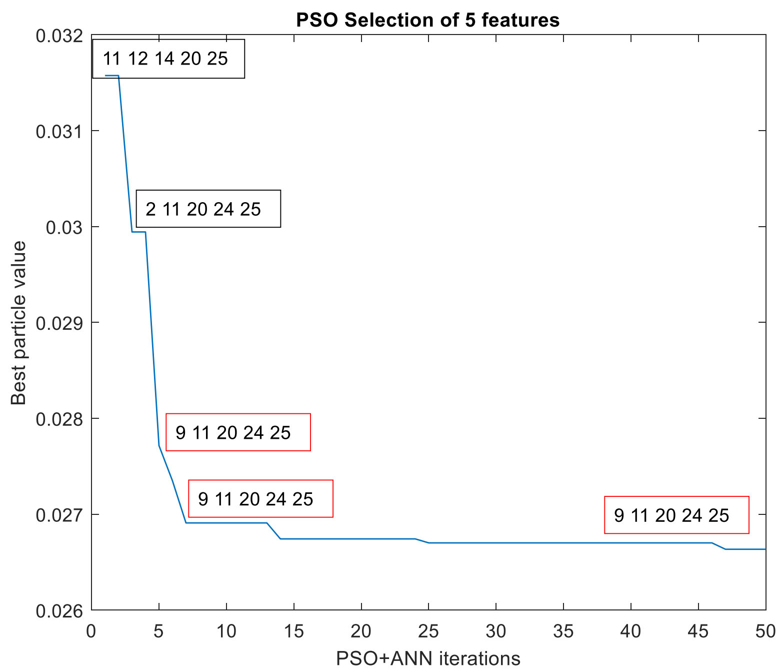


Fig. 2 Convergence of the PSO-ANN algorithm during selection of five features

The following text describes the results of the PSO-ANN algorithm execution during the variable's selection for an interval from two to seven input variables. The number of iterations was 100, while the global learning coefficient of PSO algorithm (2) was changed from C_2=1.5457 to C_2=0.85. The entire method was repeated 20 times. Fig. 3 shows importance indexes for all variables in the selected subsets from two to seven variables. Dominant variables in the selected subsets are *Gender*, *Behaviour 2*, *Age groups*, *Pedestrian's age*, *Vehicle type*, *Vehicle direction*, and *Road crossing*.

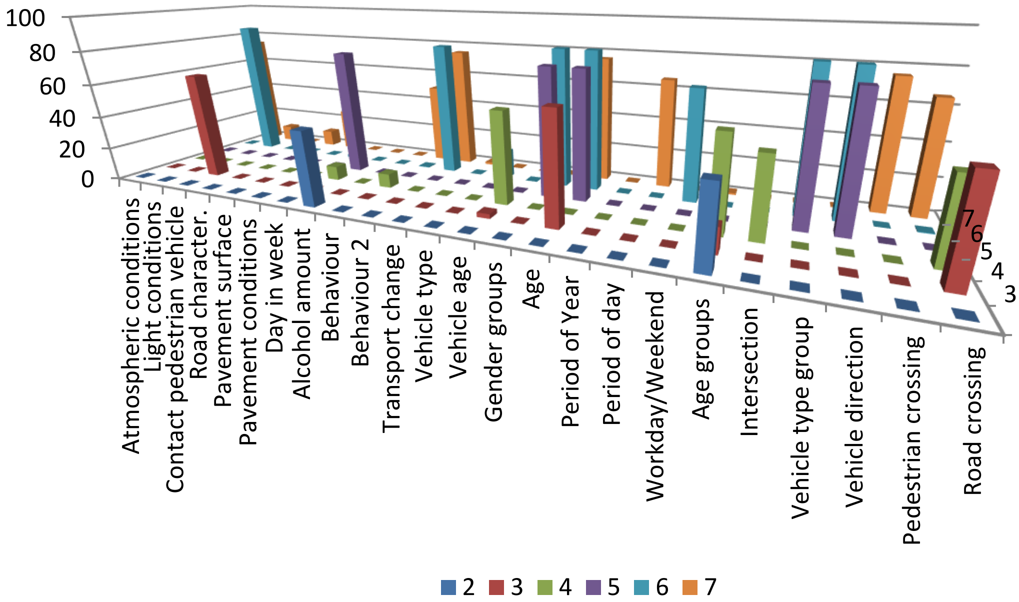
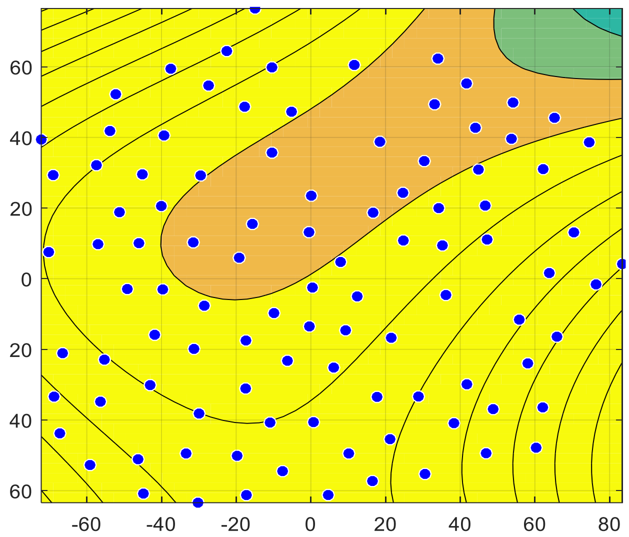
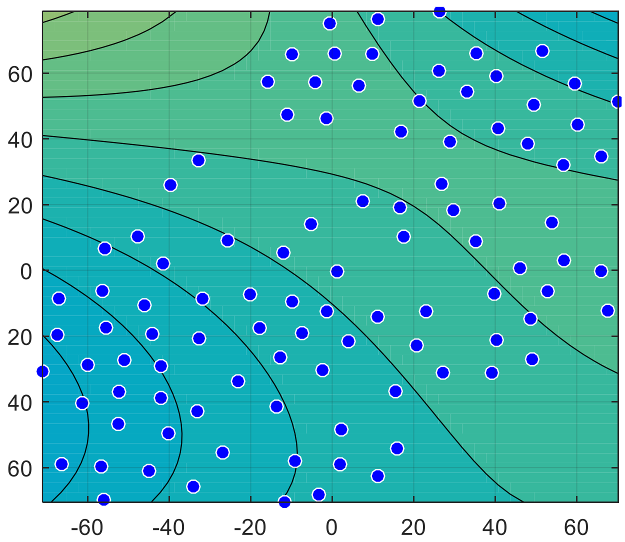


Fig. 3 Representation of features for reduced subsets in the range from 2 to 7 selected features

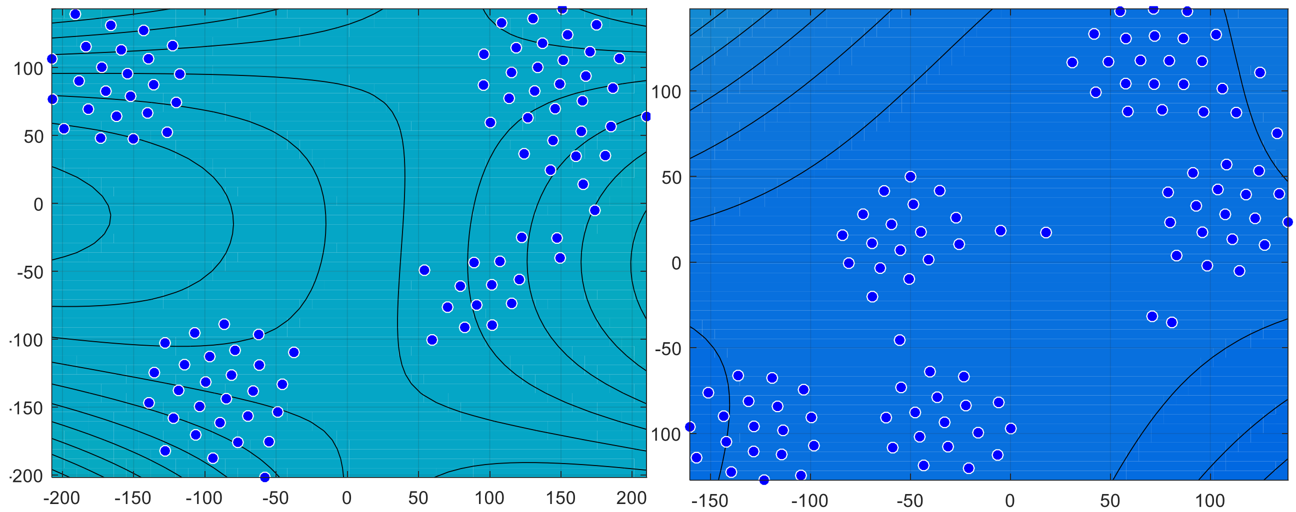
Fig. 4 shows solution positions of 100 particles after the first, fifth, twentieth, and fiftieth iteration of the PSO+ANN algorithm. The selection algorithm searched for five best features. Combinations of particle's features quintet were transformed into a 2D representation using the tSNE algorithm [12] in order for the solution layout to be displayed. Illustration on Fig. 4.a) shows an expected higher dispersion after the first iteration, i.e. solution diversity, while in later iterations it is noted that the solution's dispersion was gradually decreasing. More precisely, the result was that in later iterations, the solution clusters in the lower-value regions of the PSO+ANN algorithm's fitness function. The lower values are indicated in blue, while the highest values of the fitness functions are displayed in yellow.



a) Positions of particles after first iteration



b) Positions of particles after 5-th iteration



c) Positions of particles after 20-th iteration

d) Positions of particles after 50-th iteration

Fig. 4 Positions of particles representing subsets of selected variables during selection of five features using the PSO+ANN algorithm

IV. Conclusions

The feature selection method in the form of a specific realization of the hybrid algorithm for optimizing the swarm of solutions with artificial neural networks (PSO-ANN) has shown that it enables choosing independent input variables necessary for generating the prediction model. Specificity of the PSO-ANN algorithm realization is defined by: (a) the method is composed of an environment which is based on the PSO search rules and generates a large number of variable subsets as a list of candidates; (b) it uses neural networks for fitness functions estimation; and (c) the entire method is repeatedly executed in order to reduce the stohastic characteristics in the PSO-ANN algorithm and easily identify the most frequently selected variables. The analysis of the feature selection PSO-ANN algorithm's results has shown that the most suitable subset consists of the following five variables: *Gender*, *Behavior*, *Age groups*, *Pedestrian crossing*, and *Road crossing*.

Acknowledgment

This work was partially supported by the Provincial Secretariat for Higher Education and Scientific Research through the project: "Implementation of IoT tracking system for fresh food produces from Vojvodina" by a grant 142-451-2484/2017-01/01, and Ministry of Education, Science and Technology Development of the Republic of Serbia under the Grant TR32034.

References

[1] X. Meng, Q. Hou and Y. Shi, "Research on Accident Prediction Models for Freeways in Mountainous and Rolling Areas," in 7th Int. Conf. on Measuring Techn. & Mechatronics Automation, 2015, pp. 838-838.

[2] S. Li, and D. Zhao, „Prediction of Road Traffic Accidents Loss Using Improved Wavelet Neural Network," in Proceedings of IEEE TENCON'02, 2002, pp. 1526-1529.

[3] Ogrenci, A.S., "Complex Event Post Processing for Traffic Accidents," in 13th IEEE International Symposium on Computational Intelligence and Informatics - CINTI, 20–22 Nov., Budapest, 2012, pp. 341-345.

[4] A. Iranitalaba and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," Accident Analysis & Prevention, vol. 108, pp. 27-36, 2017, doi:10.1016/j.aap.2017.08.008.

[5] X. Gu, T. Li, Y. Wang, L. Zhang, Y. Wang and J. Yao, "Traffic fatalities prediction using support vector machine with hybrid particle swarm optimization," Journal of Algorithms & Computational Technology, vol. 12(1), pp. 20-29, 2017, doi:10.1177/1748301817729953.

[6] S. Nemet, D. Kukolj, G. Ostojić, S. Stankovski, D. Jovanović, "Aggregation Framework for TSK Fuzzy and Association Rules: Interpretability Improvement on a Traffic Accidents Case," Applied Intelligence, 49 (11), pp. 3909 - 3922, Nov. 2019, DOI: https://doi.org /10.1007/s10489-019-01485-6.

[7] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in Proceedings of the IEEE International Conference on Neural Networks, Piscataway, New Jersey, 1995, pp. 1942–1948, doi: http://dx.doi.org /10.1109/ICNN.1995.488968.

[8] S. Nemet, D. Kukolj, D. Jovanović, G. Ostojić, S. Stankovski, "Identification of influencing attributes for the traffic accident severity prediction model," Proc. 7th Int. Conference ''TOWARDS A HUMANE CITY'', pp. 389-394, Novi Sad, Serbia, 6-7 Dec. 2019.

[9] M. Reyes-Sierra and C. A. Coello, "Multi-objective particle swarm optimizers: A survey of the state-of-the-art," International Journal of Computational Intelligence Research (IJCIR), vol. 2, pp. 287–308, 2006.

[10] J. Kennedy and R. C. Eberhart, Swarm Intelligence, Morgan Kaufmann Publishers, San Francisco, California, 2001.

[11] A. Bulajić, D. Jovanović, B. Matović, S.D. Bačkalić, "Identification of High-Density Locations with Homogenous Attributes of Pedestrian Accident in the Urban Area of Novi Sad," in Proc. of XII Int. Symposium, Road Accidents Preventions, 2014, pp. 89-98.

[12] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of Machine Learning Research, vol. 9, pp. 2579-2605, Nov. 2008.