Kuwait J. Sci. **44 (3)** pp. 83-91, 2017

An approach to estimate occupational accidents using least-squares support vector machines

Hüseyin Ceylan[1], Şakir Parlakyildiz[2,*]

[1]*Kırıkkale Vocational School, Kırıkkale University, 71451, Kırıkkale, Turkey*

[2]*Bitlis Technical Sciences Vocational School, Bitlis Eren University, 13100, Bitlis, Turkey*

**Corresponding author: sp_yildiz@hotmail.com*

Abstract

Least-squares support vector machines represent an emerging technique that has been adopted to estimate accidents. In this study, occupational accident estimation models were developed using the least-squares support vector machine method for the Republic of Turkey. In addition, linear regression analysis, nonlinear regression analysis, and artificial neural network models were considered. During the development phase of the models, statistical data from 1970 to 2012 were used to consider parameters such as insured workers, workplaces, occupational accidents, deaths, and permanent incapacities. Using these models, the numbers of accidents, deaths, and permanent incapacities resulting from occupational accidents were estimated for three different scenarios in the Republic of Turkey through the end of 2025. The performances of the developed models were evaluated considering the mean absolute percent errors and the mean absolute errors. In addition, we compared the least-squares support vector machine, linear regression analysis, nonlinear regression analysis, and artificial neural network methods in terms of their estimation performances. Our simulation results demonstrate that the proposed least-squares support vector machine model outperforms other techniques in terms of accuracy and has a rapid convergence capability when estimating occupational accidents.

Keywords: Accident prediction models; artificial neural networks; estimation error; least squares support vector machine; occupational accident.

1. Introduction

Occupational accidents lead to serious problems in the Republic of Turkey, as in many other countries, where the most serious outcomes of occupational accidents are deaths, injuries, or disabilities. In addition, many other important work-related accidents may cause serious financial losses. The economy of Turkey experienced a major growth trend over the last decade. However, with the growth of the economy, both the number of workplaces and the number of employees significantly increased. Naturally, the number of employees affected by occupational accidents also increased at a significant rate. Many employees die or are seriously injured due to occupational accidents. Moreover, via evaluations of these undesirable incidents, it can be seen that the importance of occupational accidents is not fully grasped by society or the government. Each year, approximately 1,500 employees die and 2,100 employees become permanently incapacitated as a result of occupational accidents in the Republic of Turkey. Yet, this figure tends to increase linearly with the growth of the economy. In this study, the occupational accidents and the number of deaths and injuries resulting from these accidents are estimated through the end of 2025. We aim to reveal the true extent of this issue and attempt to form a perspective regarding workplace safety for decision makers (Ceylan, 2014).

Using various approaches and analyses, multiple prediction models have been developed over the last decade. In the study of prediction models, regression analysis techniques are simultaneously used with genetic algorithms and artificial neural networks (ANNs). Ceylan (2014) has developed models to predict the number of occupational accidents, deaths, and permanent incapacities resulting from occupational accidents in the Republic of Turkey through the end of 2025 using ANN technique. (Abdelwahab & Abdel-Aty, 2001) used ANN techniques to predict the severity of injuries to drivers in accidents occurring at signalized intersections. Similarly, Chiou (2006) used an ANN-based expert system for accidents resulting from the collision of two vehicles and discovered that the accident location and the use of alcohol were prevalent variables in the occurrence of an accident. (Akgüngör & Doğan, 2008) predicted the number of accidents, injuries, and deaths in the Republic of Turkey using ANN and nonlinear regression techniques.

2. Methods

2.1. Multiple nonlinear regression analysis.

We assume that there is a nonlinear relationship between the dependent and independent variables. This nonlinear relationship can be approximated by a polynomial equation. In this study, we used a surface-fitting tool with the help of the MATLAB Statistic Toolbox. The generalized equation form is provided as follows.

$$Y = F(x_i) = p_{00} + p_{10} * x_1 + p_{01} * x_2 + p_{20} * {x_1}^2 + p_{11} * x_1 * x_2 + p_{02} * {x_2}^2 \quad (1)$$

where the input variables are denoted by x_1 and x_2 and the response variable is denoted by Y. Here, x_1 represents the number of insured workers, x_2 represents the number of workplaces, and Y represents the number of occupational accidents in model 1, the number of deaths in model 2, and the number of permanent incapacities in model 3. P_{00}, P_{10}, P_{01}, P_{20}, P_{11}, and P_{02} are the constants obtained using the surface-fitting tool.

2.2. Least-squares support vector machines.

Least-squares support vector machines (LS-SVM) are a modification of classical support vector machines (SVM), which are a set of related supervised learning methods that analyze data and recognize patterns. This technique is used for optimal control, classification, and regression analysis (Suykens & Vandewalle, 1998). In this sophisticated version, the solution is found by solving a set of linear equations instead of a convex quadratic programming problem, as in standard SVM. LS-SVM proposed by Suykens & Vandewalle (1999) and are a class of kernel-based learning methods. The theoretical basis of LS-SVM is created with a two-class classification, as in standard support vector classifiers. In the model construction, linear equation series are preferred.

$$\frac{1}{2}||W||^2 + \frac{C}{2}\sum_{i=1}^{n} {\xi_i}^2 \quad (2)$$

$$y_i[(w.x_i) + w_0] = 1 - \xi_i \;\; i = 1,2,\ldots.,n \quad (3)$$

The formula given in Equation (2) needs to be minimized depending on the condition expressed in Equation (3). As in a standard support vector machine classifier, the primary equation given by Equation (2) is converted into a dual problem, as seen in Equation (4), for LS-SVM (Suykens & Vandewalle, 1999). The optimal solution needs to be obtained in the presence of this dual problem.

$$Q(w,b,a,\xi) = \frac{1}{2}||W||^2 + \frac{C}{2}\sum_{i=1}^{n} {\xi_i}^2 - \sum_{i=1}^{n} a_i\{y_i\,[(wx_i) + w_0] - 1 + \xi_i\} \quad (4)$$

The formula expressed in Equation (4) and the formula used in the standard support vector machine classifier, are similar. The only difference between the two is that, while the “a” coefficient must be positive in the standard support vector machine classifier, it may be positive or negative in LS-SVM (Tsujinishi & Abe, 2003). This formulaic difference provides benefits, such as lower training time, to LS-SVM compared with standard SVM.

2.3. Artificial neural networks.

ANN is a data processing system resulting from the mathematical modeling of learning processes inspired by humans. It is based on simple processors that imitate the basic abilities of human brains, such as learning, remembering, forming new knowledge, making generalizations, and exploring.

The basic processor forming an ANN is an artificial neuron (Figure 1), which consists of inputs (x_i), weights (w_i), a summing function (Σ), an activation function (f), and an output (y).

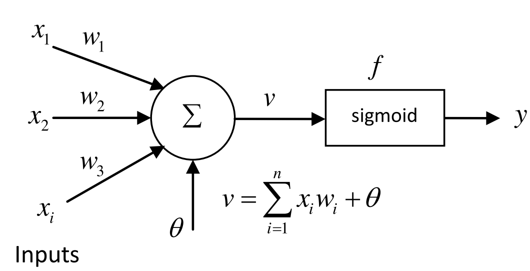

Fig. 1. Artificial neuron

In the basic ANN cell seen in Figure 1, the values are cell inputs, which are multiplied by certain weights and summed with threshold values varying from −1 to +1, which then form the net input variable using the activation function over the net input, and the cell output can then be transformed into the desired intervals. By comparing this output value with the known output for a system, an error ratio can be determined. According to this error ratio, the ANN cell updates the weight values of the inputs. Therefore, a loop is made to gather more accurate results. The learning of the network occurs by updating these weights, as follows.

$$y = f(\sum_{i=1}^{n} w_i x_i + \theta) \quad (5)$$

In this study, many different network architectures are considered, and it is seen that the 2-5-1 network architecture, provided in Figure 2, is the most appropriate scenario for this particular accident prediction application. For the activation function, at the first and second layer, the "tangent sigmoid" and "linear" functions are used, respectively. For the network type "feed-forwarding back-propagation algorithm" and for the learning function "trainlm," the performance functions MAPE (mean absolute percent errors) and MAE (mean absolute errors) are preferred, respectively. As seen in Figure 2, for the input parameter, only the numbers of workplaces and insured workers are used. The numbers of accidents, deaths, and permanent incapacities resulting from occupational accidents are the focus of our prediction model.

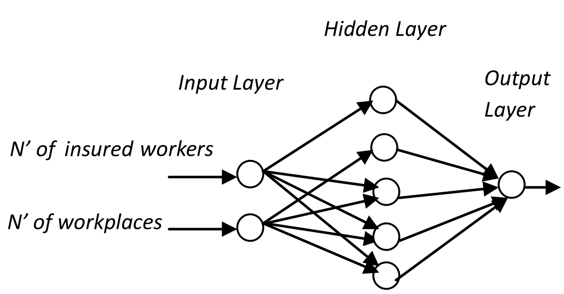

Fig. 2. Structure of artificial neural network

3. LS-SVM regression model for occupational accidents

Occupational accidents occur due to the interaction of multiple factors. The desired model is intended to be simple and reliable. Therefore, the numbers of workplaces and insured workers are considered to be directly related to the accidents and are applied as the model parameter (Table 1) in the accident prediction models developed for the Republic of Turkey. The data used in this study include the numbers of workplaces, insured workers, occupational accidents, deaths, and permanent incapacities resulting from occupational accidents between 1970 and 2012, which were obtained for statistical operations from data provided by the Social Security Institution (SSI) (SSI, 2012).

Table 1. Input and output data used in the LS-SVM regression analysis.

Variable name	Variable meaning
Input 1	Number of insured workers
Input 2	Number of workplaces
Output 1	Number of occupational accidents
Output 2	Number of deaths resulting from occupational accidents
Output 3	Number of permanent incapacities resulting from occupational accidents

3.1. LS-SVM regression model to predictthe number of occupational accidents.

In this model, X_1 represents the number of insured workers, X_2 represents the number of workplaces, and the number of occupational accidents is denoted by Y. The estimated function for occupational accidents using the LS-SVM Toolbox in MATLAB is provided in Figure 3.

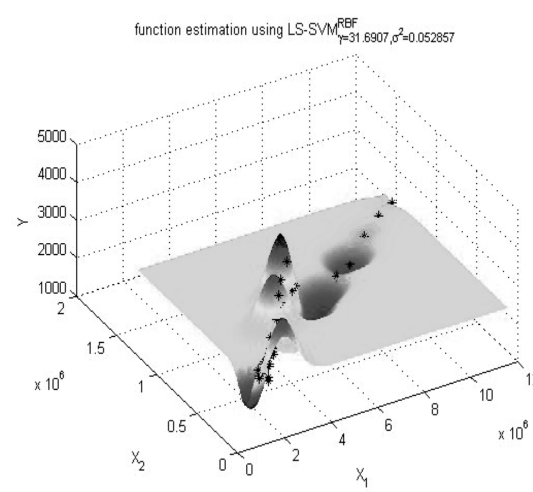

Fig. 3. Estimated function model for occupational accidents

3.2. LS-SVM regression model to predict the number of deaths resulting from occupational accidents.

In this model, the number of insured workers is denoted by X_1, while X_2 represents the number of workplaces and Y represents the number of deaths resulting from occupational accidents. The estimated function for the number of deaths resulting from occupational accidents using the LS-SVM Toolbox in MATLAB is provided in Figure 4.

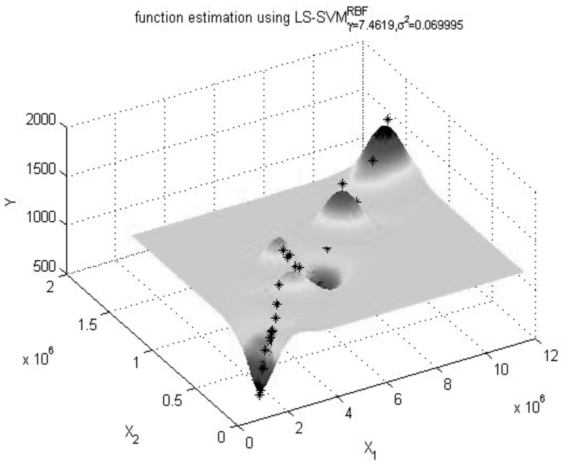

Fig. 4. Estimated function model for number of deaths resulting

3.3. LS-SVM regression model to predict the number of permanent incapacities resulting from occupational accidents.

In this model, X_1 is the number of insured workers, X_2 is the number of workplaces, and Y is the number of permanent incapacities resulting from occupational accidents. The estimated function for the number of permanent incapacities resulting from occupational accidents using the LS-SVM Toolbox in MATLAB is provided in Figure 5.

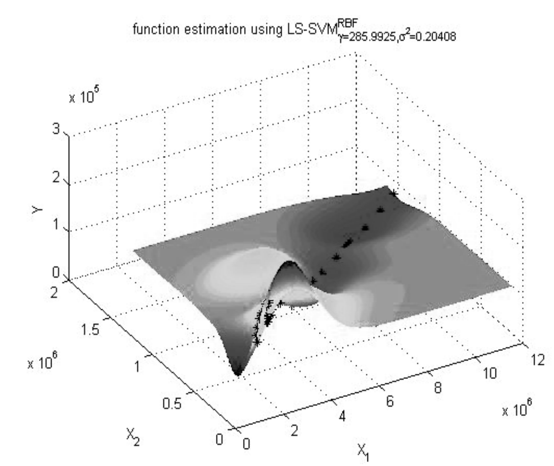

Fig. 5. Estimated function model for number of permanent incapacities

4. Results of the application of the LS-SVM regression models for Turkey

To predict the values of the numbers of occupational accidents, deaths, and permanent incapacities resulting from occupational accidents, three different LS-SVM models were implemented.

First, to determine the efficiency of the models, the values predicted by the models for the numbers of occupational accidents, deaths, and permanent incapacities resulting from occupational accidents between 2000 and 2012 were compared with the actual values.

In Figure 6, the number of occupational accidents predicted by the LS-SVM model and the actual number of occupational accidents from 2000 to 2012 are given for the Republic of Turkey.

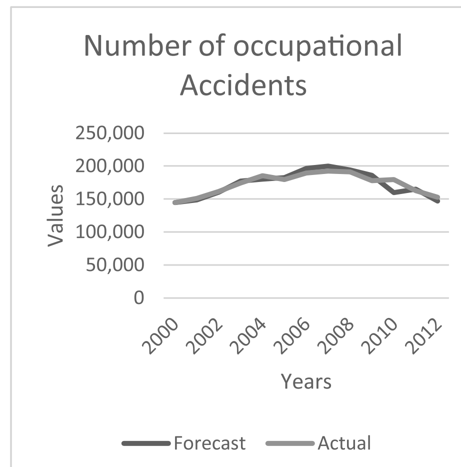

Fig. 6. Actual number of occupational accidents from 2000 to 2012

In Figure 7, the number of deaths resulting from occupational accidents predicted by the LS-SVM model and the actual number of deaths resulting from occupational accidents from 2000 to 2012 are given for the Republic of Turkey.

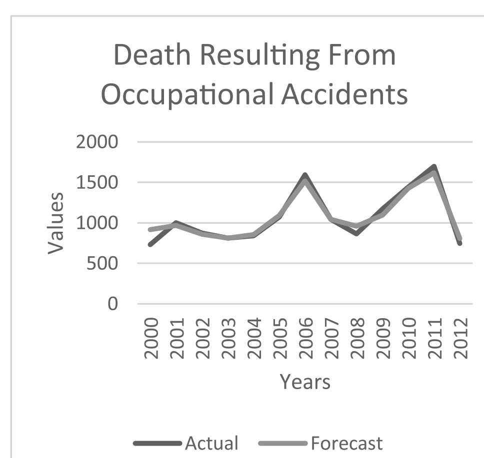

Fig. 7. Actual number of death resulting from 2000 to 2012

In Figure 8, the numbers of permanent incapacities resulting from occupational accidents predicted by the LS-SVM model and the actual permanent incapacities resulting from occupational accidents from 2000 to 2012 are given for the Republic of Turkey.

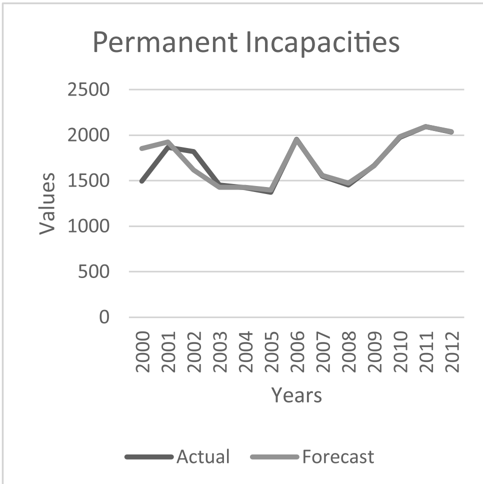

Fig. 8. Actual number of permanent incapacities from 2000 to 2012

4.1. Performance comparison of the developed LS-SVM prediction models with the NRA, linear regression analysis (LRA), and ANN models.

To compare the performance of the developed model with respect to other models, a performance comparison was conducted between the NRA, LRA, ANN (Ceylan, 2014), and LS-SVM models. Therefore, the same estimations were made using the NRA, LRA, and ANN models. The performances of all the developed models were evaluated via MAPE and MAE, where the results are provided in Table 2.

$$MAPE = \frac{1}{n}\sum[\left|\frac{o_i - t_i}{o_i}\right| * 100] \quad (6)$$

$$MAE = \tfrac{1}{n}\sum(|o_i - t_i|) \quad (7)$$

Table 2. Comparison of the estimation error values for the NRA, LRA, ANN, and LS-SVM models.

Applied model	Error code	Prediction of occupational accident	Prediction of death	Prediction of permanent incapacity
NRA	MAPE	1,13	4,41	2,61
	MAE	292	0,37	1,84
LRA	MAPE	12,99	17,89	16,91
	MAE	14716,31	175,26	357,406
ANN	MAPE	3,24	6,21	7,55
	MAE	2169	65	132
LS-SVM	MAPE	0,1575	0,013	0,0072
	MAE	0	0	0

The results provided in Table 2 indicate that the proposed LS-SVM model has the best performance and provides an accurate and quick accident estimation technique.

5. Accident scenarios for Turkey

Predictions through to the end of 2025 using LS-SVM models, whose estimation error values are given in Table 2, were made and the predictions were estimated using three different scenarios. In Scenario I, 80% of the average annual increase (calculated between the years 1970 and 2012) in the numbers of insured workers and workplaces was assumed. In Scenarios II and III, 100% and 120% of the average annual increase in the numbers of insured workers and workplaces were assumed, respectively.

The predicted increases were determined to be 202,402 for insured workers and 27,212 for workplaces in Scenario I; 253,003 for insured workers and 34,015 for workplaces in Scenario II; and 303,604 for insured workers and 40,818 for workplaces in Scenario III.

The results for Scenario I using the LS-SVM model are presented in Table 3. According to this table, it can be seen that, in Turkey, there will be increases in the numbers of occupational accidents, deaths, and permanent incapacities in the near future. It is predicted that the number of permanent incapacities will reach 2,219 by 2025, with an increase of 8.98% compared with 2012. The number of accidents will rise to 102,955 by 2025. The number of deaths due to occupational accidents is estimated to reach 1,094 by 2025 with an increase of 47% compared with 2012.

Table 3. Predictions of the numbers of occupational accidents, deaths, and permanent incapacities produced by the LS-SVM models for Scenario I in Turkey.

Year	Number of insured workers	Number of workplaces	Number of occupational accidents	Deaths resulting from occupational accidents	Permanent incapacities resulting from occupational accidents
2013	12.142.022	1.565.218	78.219	744	2.064
2014	12.344.425	1.592.430	82.019	783	2.106
2015	12.546.827	1.619.642	85.968	866	2.150
2016	12.749.230	1.646.854	89.755	953	2.185
2017	12.951.632	1.674.066	93.140	1.020	2.205
2018	13.154.034	1.701.278	95.982	1.060	2.214
2019	13.356.437	1.728.490	98.234	1.081	2.218
2020	13.558.839	1.755.702	99.923	1.089	2.219
2021	13.761.242	1.782.914	101.124	1.092	2.219
2022	13.963.644	1.810.126	101.935	1.093	2.219
2023	14.166.046	1.837.338	102.455	1.094	2.219
2024	14.368.449	1.864.550	102.772	1.094	2.219
2025	14.570.851	1.891.762	102.955	1.094	2.219

The simulation results for the study of Scenario II using the LS-SVM model are provided in Table 4. Table 4 indicates that there will be a steep increase in the number of permanent incapacities in the near future in the Republic of Turkey. It is predicted that the number of permanent incapacities will reach 2,317 by the end of 2025, with an increase of 13.8% compared with 2012. The number of accidents will increase to 102,811 by the end of 2025. The number of deaths is estimated to reach 1,093 by 2025 with an increase of 46.9% compared with 2012.

Table 4. Predictions of the numbers of occupational accidents, deaths, and permanent incapacities produced by the LS-SVM models for Scenario II in Turkey.

Year	Number of insured workers	Number of workplaces	Number of occupational accidents	Deaths resulting from occupational accidents	Permanent incapacities resulting from occupational accidents
2013	12.192.623	1.572.021	79.086	791	2.072
2014	12.445.626	1.606.036	83.885	864	2.127
2015	12.698.629	1.640.051	88.662	972	2.175
2016	12.951.632	1.674.066	92.905	1.043	2.202
2017	13.204.635	1.708.081	96.320	1.077	2.213
2018	13.457.638	1.742.096	98.834	1.089	2.216
2019	13.710.641	1.776.111	100.536	1.092	2.317
2020	13.963.644	1.810.126	101.600	1.093	2.317
2021	14.216.647	1.844.141	102.215	1.093	2.317
2022	14.469.650	1.878.156	102.543	1.093	2.317
2023	14.722.653	1.912.171	102.706	1.093	2.317
2024	14.975.656	1.946.186	102.780	1.093	2.317
2025	15.228.659	1.980.201	102.811	1.093	2.317

The results for the study made in accordance with Scenario III using the LS-SVM regression model are presented in Table 5. This table predicts that the number of permanent incapacities will reach 2,223 by 2025, with an increase of 9.18% compared with 2012. It can be seen that the values of the estimated occupational accidents tend to increase. The number of deaths due to occupational accidents is estimated to reach 1,094 by 2025 with an increase of 47% compared with 2012.

Table 5. Predictions of the numbers of occupational accidents, deaths, and permanent incapacities produced by the LS-SVM models for Scenario III in Turkey.

Year	Number of insured workers	Number of workplaces	Number of occupational accidents	Deaths resulting from occupational accidents	Permanent incapacities resulting from occupational accidents
2013	12.243.224	1.578.824	79.881	765	2.084
2014	12.546.827	1.619.642	85.539	876	2.153
2015	12.850.431	1.660.460	90.853	997	2.200
2016	13.154.034	1.701.278	95.148	1.064	2.218
2017	13.457.638	1.742.096	98.201	1.088	2.222
2018	13.761.242	1.782.914	100.133	1.093	2.223
2019	14.064.845	1.823.732	101.231	1.094	2.223
2020	14.368.449	1.864.550	101.793	1.094	2.223
2021	14.672.052	1.905.368	102.052	1.094	2.223
2022	14.975.656	1.946.186	102.160	1.094	2.223
2023	15.279.260	1.987.004	102.200	1.094	2.223
2024	15.582.863	2.027.822	102.214	1.094	2.223
2025	15.886.467	2.068.640	102.218	1.094	2.223

6. Conclusion

In this study, three different models were developed using LS-SVM to predict the numbers of occupational accidents, deaths, and permanent incapacities resulting from occupational accidents in the Republic of Turkey. To develop each accident model, the numbers of insured workers, workplaces, occupational accidents, deaths, and permanent incapacities were considered as the optimization variables. MAPE and MAE values were calculated for the LS-SVM, NRA, LRA, and ANN models in Table 3 to measure and compare the performances of the models with respect to each other. As a result of this performance comparison, it was demonstrated that the LS-SVM model performed the best. Therefore, it was implemented as the accident prediction model. Three scenarios were constructed for the Republic of Turkey through the end of 2025, and, as is inferred from Tables 3, 4, and 5, there will be an increase in the number of occupational accidents ending in death or permanent incapacity. Therefore, to prevent occupational accidents and their undesirable results, occupational safety plans and policies need to be extensively revised.

It is required that all accidents occurring in the Republic of Turkey be reported to the SSI within three working days. In this study, official occupational accident data belonging to the SSI were used. Unfortunately, failing to report incidents is a significant issue in the Republic of Turkey, as well as in the rest of the world. Therefore, a large number of the occupational accidents occurring in Turkey may not be recorded. Accordingly, it is inevitable that the actual situation in Turkey, in terms of occupational accidents, deaths, and permanent incapacities resulting from occupational accidents, is worse than the numbers used for our study imply (Ceylan, 2014; Ceylan, 2012).

References

Abdelwahab, H.T. & Abdel-Aty, M.A. (2001). Development of artificial neural network models to predict driver injury severity in traffic accident at signalized intersection. Transportation Research Record, **17**(46): 6-13

Akgungor, A.P. & Dogan, E. (2008). Estimating road accidents of turkey based on regression analysis and artificial neural network approach. Advances In Transportation Studies, **2008**(16):11-22

Ceylan, H. (2014). An artificial neural networks approach to estimate occupational accident: a national perspective for Turkey. Mathematical Problems In Engineering, Article ID 756326, http://dx.doi.org/10.1155/2014/756326, 10 pages.

Ceylan, H. (2012). Analysis of occupational accidents according to the sectors in Turkey. Gazi University Journal of Science, **25**(4):909-918

Chiou, Y.C. (2006). An artificial network-based expert system for appraisal of two-car crash accidents. Accident Analysis and Prevention, **38**(4):777-785

S.S.I., Statistical Yearbook. (2012). Ankara, 1970-2012 [in Turkish]. http://www.sgk.gov.tr/wps/portal/tr/kurumsal/istatistikler

Suykens, J.A.K. & Vandewalle, J. (1998). Nonlinear modeling: advanced black-box techniques, Boston: Kluwer Academic Publishers, 55-85

Suykens, J.A.K. & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural Processing Letters, **9**(3):293-300

Tsujinishi, D. & Abe, S. (2003). Fuzzy least squares support vector machines for multi-class problems. Neural Networks Field, **16**(5):785-792

Submitted **:** 26/08/2015
Revised **:** 20/10/2016
Accepted **:** 08/02/2017

نهج لتقدير الحوادث المهنية باستخدام آلات متجهات دعم المربعات الصغرى

[1]حسين سيلان، [2,*]شاكير بارلاكيلدز

[1]مدرسة كيريكال المهنية، جامعة كيريكال، 71451، كيريكال، تركيا

[2]مدرسة بيتليس المهنية للعلوم التقنية، جامعة بيتليس ايرين، 13100، بيتليس، تركيا

[*]المؤلف المراسل: sp_yildiz@hotmail.com

ملخص

تعتبر آلات متجهات دعم المربعات الصغرى تقنية ناشئة تم اعتمادها لتقدير الحوادث. في هذه الدراسة، تم تطوير نماذج لتقدير الحوادث المهنية باستخدام طريقة آلات متجهات دعم المربعات الصغرى في جمهورية تركيا. بالإضافة إلى ذلك، تم الوضع في الاعتبار تحليل الانحدار الخطي، وتحليل الانحدار اللاخطي، ونماذج الشبكات العصبية الاصطناعية. خلال مرحلة تطوير النماذج، استُخدِمت بيانات إحصائية من عام 1970 إلى عام 2012 للنظر في معلمات، مثل: العمال المؤمن عليهم، وأماكن العمل، والحوادث المهنية، والوفيات، والعجز الدائم. وباستخدام هذه النماذج، تم تقدير عدد الحوادث والوفيات والعجز الدائم الناجم عن الحوادث المهنية بثلاثة سيناريوهات مختلفة في جمهورية تركيا حتى نهاية عام 2025. تم تقييم أداء النماذج المُطَوره بالنظر إلى متوسط أخطاء النسبة المطلقة ومتوسط الأخطاء المطلقة. بالإضافة إلى ذلك، قارنا آلة متجهات دعم المربعات الصغرى، وتحليل الانحدار الخطي، وتحليل الانحدار غير الخطي، وأساليب الشبكات العصبية الاصطناعية من حيث تقدير أدائها. وتبين نتائج المحاكاة لدينا أن نموذج آلة متجهات دعم المربعات الصغرى المُقترح يتفوق على التقنيات الأخرى من حيث الدقة ولديه خاصية التقارب السريع عند تقدير الحوادث المهنية.