Int. J. Bio-Inorg. Hybr. Nanomater., 11(2): 97-103, Summer 2022

QSAR study for choosing biological and chemical parameters of water in the Anzali international wetland in growth Oligochaetes (L. Claparedeianus, L. Hoffmeisteri)

R. Rajabei Nezhad[1,], R. Sayyadi Kord Abadi[2], F. Nazarhaghighi[3], O. Alizadeh[2]*

[1] *Department of Fisheries, Bandar Anzali Branch, Islamic Azad University, Bandar Anzali, Iran*
[2] *Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran*
[3] *Administrative Office of Fisheries, Ardabil, Iran*

Received: 20 December 2021; Accepted: 22 February 2022

ABSTRACT: The Biological and chemical properties of water in the Anzali International Wetland were estimated using multiple linear regression (MLR), artificial neural network (ANN), and genetic algorithm(GA) and simulated annealing algorithm (SA) as optimization methods. The obtained results from MLR-MLR, SA-ANN and GA-ANN techniques were compared and a high predictive ability was observed for the GA-ANN model with the root-mean-sum-squared error (RMSE) of 0.0079 and 0.01535 in L. Claparedeianus and L. hoffmeisteri, respectively. The results obtained using the GA-ANN method indicated that abundance of L. Claparedeianus and L. hoffmeisteri in the Anzali International Wetland depends on different parameters, which include: that NH3 concentration, total nitrogen (TN), dissolved oxygen (DO), Sodium chloride (Sali), Nitrat (NO3), total phosphorus (TP), Biochemical Oxygen demand (BOD), Total dissolved solids (TDS) and electrical conductivity (EC) in water. In conclusion, the comparison of the quality of the ANN with different MLR methods showed that GA-ANN has a better predictive capability.

Keywords: *Genetic Algoritm, Limnodrilus claparedeianus, MLR, QSAR.*

INTRODUCTION

Quantitative structure-activity relationship (QSAR) methods are important for prediction of abundance of Oligochaetes based on mathematical and statistical relations. The data sets for QSAR modeling, which contain the Biological and chemical properties of water, are generated by experimental scientists and available in various data sources [1, 2]. There are several variable selection QSAR methods including multiple linear regression (MLR), genetic algorithm (GA), simulated Annealing algorithm (SA) and so on [3-9]. Benthic macro-invertebrate communities establish an important connection as food reservoir for fishes and other animals in aquatic food chain [10]. The Oligochaetes are benthic macro-invertebrate communities and abundant animal groups, which are strongly influenced by environmental conditions. Out of more than 5000 known species of class Oligochaeta, approximately 1100 species live in freshwater [11, 12]. Limnodrilus

(*) Corresponding Author - e-mail: reza_igl@yahoo.com

claparedeianus are considered as the most tolerant oligochaets to aquatic pollution. Therefore identification and biological characteristics of this taxon are of great help to evaluate the ecosystem of water bodies [13]. Limnodrilus hoffmeisteri, also known as red worm, is one of the most widespread and abundant oligochaetes in the world [14]. They live in fine, sandy and coarse sediments of watercourses, pond and lakes, hyporheic zone and groundwater, feeding on microorganisms and organic material [15, 16]. Due to the lack of gill, it breathes through skin. This worm is well recognized indicator of organic pollution and low dissolved oxygen [17]. It is increasingly being used to test the toxicity of sediment-associated contaminants and their bioaccumulation [18]. In present work, MLR and ANN modeling tools coupled with SA and GA optimization techniques were used to find the best set of bio-chemical parameters of the water (depth, temperature, dissolved oxygen, and pH) that correlate the abundance of Limnodrilus claparedeianus and Limnodrilus hoffmeisteri in the Anzali International Wetland.

Linear and Non-Linear Methods

Aquatic oligochaetes were collected twice a season from 13 sites in the Anzali Wetland from August 2012 to June 2013. The samples were taken with a bottom grab (0.04 m2) from the surface layer of bottom sediments among the submerged macrovegetation. During each sampling period, water temperature was measured with a thermometer with a sensitivity of 0.1°C, dissolved oxygen was measured with an oxygen meter WTW-OXI 330/SET, and pH was determined with a pH meter WTW pH 330/SET-1. Details of Sampling and sample processing of Limnodrilus claparedeianus and Limnodrilus hoffmeisteri were given in ref. [13].

Some bio-chemical parameters of the water including mixture is sodium chloride and water (Salin), dissolved oxygen (DO), pH, TOM (total organic material), turbidity (turb), electrical conductivity (EC), total soluble solids (TSS), Total dissolved solids (TDS), total phosphorus (TP), NO_2, NO_3, NH_3, Viruses and Worms counting (Vw), total nitrogen (TN), Biochemical oxygen demand (BOD), chemical oxygen demand (COD), chlorophyll-a (chla), Photovoltaic (Pv), Non-invasive ventilation (Niv), Part By Weight (Pbw) , CrS, PbS, Pressure Sensor (Ps), NiS, Vs were measured at the sampling from 13 sites in the Anzali Wetland in 7 species of aquatic oligochaetes including L. claparedeianus and Limnodrilus hoffmeisteri [12, 13].

The physico-chemical parameters values in Linear and Nonlinear QSAR models were normalized using the equation (1) in Excel program.

$$X_{normalized} = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \quad (1)$$

SPSS [19] program was employed to select the best physico-chemical parameters. An ideal method is one that has low standard deviation, high correlation coefficient (R^2) and root mean of square error (RMSE) [20], where the RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_o)^2}{n}} \quad (2)$$

The correlation coefficient (R^2), root mean of square error (RMSE) were calculated by using the linear method (MLR) approach in the unscramble [20] program. The bio-chemical parameters chosen in the previous primary linear selection were implemented under further screening using QSAR methods including GA-ANN, SA-ANN, and MLR-GA as shown in Fig. 1. In each run via ANN, the 3-5 bio-chemical parameters selected by the optimization method (GA or SA) were used as the inputs and corresponding values of (abundance of L. claparedeianus) were utilized as the target values (Fig.1).

This study also used three neurons in the hidden layer of the ANN approaches 80%, 10% and 10% of data sets in these models were randomly chosen as training, validation and test sets, respectively. The networks were trained by using the TSET members with Levenberg-Marquart Algorithm [8], while Logarithmic sigmoid and linear transfer function were used as the hidden and output transfer function, respectively. Logarithmic sigmoid transfer function is defined as follows:

$$\log sigm = \frac{1}{1 + e^{-x}} \quad (3)$$

A genetic algorithm (GA) is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics bio-

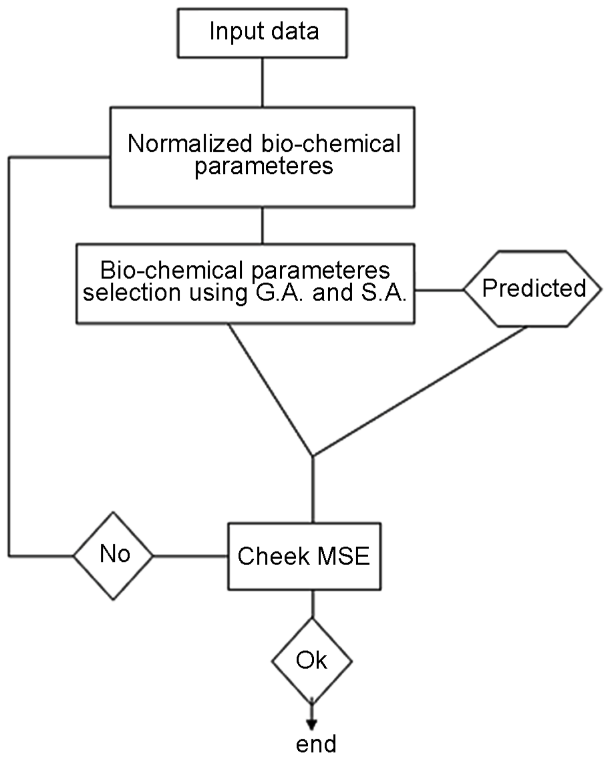

Fig. 1. The employed procedure for finding optimum descriptors of the ANN models.

logical evolution. The algorithm repeatedly modifies a population of individual solutions [21]. SA is a descent algorithm modified by random ascent moves in order to escape local minima which are not global minima. It has been used in statistical physics to choose sample states of a particle system model to efficiently estimate some physical quantities [6]. Matlab 2014a was used for modeling and optimization calculations.

RESULTS AND DISCUSSION

QSAR investigations in Biological and chemical properties of water in the L. claparedeianus and L. hoffmeisteri were performed using multiple linear regression (MLR) and artificial neural network (ANN) as modelling tools and simulated annealing (SA) and genetic algorithm (GA) as optimization methods. According to the types of variable section method and feature mapping techniques, these models were shown as MLR-MLR, SA-ANN, MLR-GA and GA-ANN. In MLR-MLR, MLR-PCR and MLR-PLS1 methods, the best physico-chemical parameters were selected using MLR procedure of SPSS [9] software .Then the selected physico-chemical parameters were used as input in unscramble software and statistical parameters were calculated using PCR and PLS1 methods (Tables 2, 4). The RMSE and the correlation coefficient (R^2) for predicted the abundance of L. claparedeianus and L. hoffmeisteri in MLR–MLR were found to be [0.1535-0.47015] and [0.01535-0.4701], respectively. Furthermore, the calculated parameters indicated that MLR–MLR method was better than the two other employed linear methods. The definition of the bio-chemical parameters in the MLR-MLR method is shown in Tables 1, 4. To establish the SA-ANN and MLR-GA and GA-ANN models, the 28 bio-chemical parameters were fed to the neural network to select the best bio -chemical parameters. Also we used 3 neurons in the hidden layer of the ANN models (Fig.1). The statistical parameters of all QSAR approaches are shown in Tables 2, 4. In GA-ANN, SA-ANN, MLR-GA methods, 80%, 10% and 10% of data sets were randomly chosen as training, validation and test sets, respectively. The results of the QSAR models proved that non-linear feature selection approaches were better than their linear models. The obtained results demonstrated that the GA-ANN method led to better

Table 1. The best selected bio-chemical parameters of water using QSAR Methods in L. Claparedeianus

MLR-MLR	SA-ANN	GA-MLR	GA-ANN
Sali	TN	COD	NH_3
Tss	NH_3	BOD	TN
COD	pH	Sali	DO
NH_3	Sali	TN	Sali
Chla	Chla	NH_3	BOD

Table 2. Statistical parameters of different linear QSAR models in L. Claparedeianus

QSAR Model	R^2	RMSE
MLR-PLS1	0.4691	0.1536
MLR-MLR	0.4701	0.1535
MLR-PCR	0.4684	0.1537
MLR-ANN	0.6543	0.0154
SA-ANN	0.7861	0.0095
GA-ANN	0.8292	0.0079
GA-MLR	0.8129	0.0084

Table 3. The best selected bio-chemical parameters of water using QSAR Methods in L. hoffmeisteri

MLR-MLR	SA-ANN	GA-MLR	GA-ANN
Sali	Ph	TSS	EC
Ph	TN	Sali	TDS
TSS	COD	NH_3	NH_3
NH_3	EC	Ph	TN

Table 4. Statistical parameters of different linear QSAR models in L. hoffmeisteri

QSAR Model	R^2	RMSE
MLR-PLS1	0.3039	0.1147
MLR-MLR	0.3555	0.1140
MLR-PCR	0.2523	0.1189
MLR-ANN	0.5932	0.0077
SA-ANN	0.7075	0.0055
GA-ANN	0.7668	0.0044
GA-MLR	0.6159	0.0073

results with good predictive ability than the other QSAR models. Therefore, the selected the best bio -chemical parameters using GA-ANN are discussed here (Tables 1, 3).

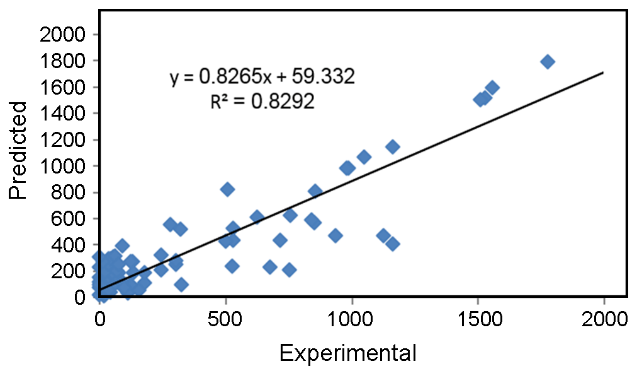

Fig. 2. Correlation between experimental and predicted abundance values calculated by using GA-ANN method L. Claparedeianus.

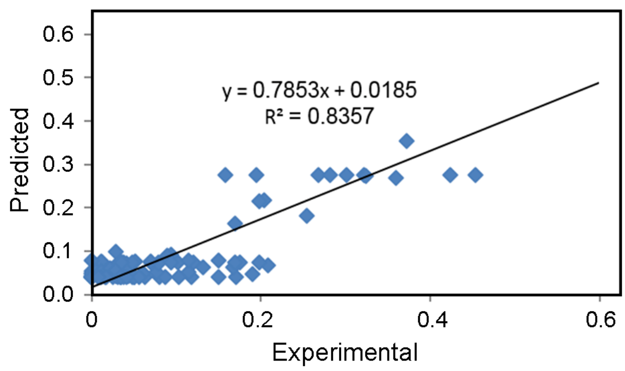

Fig. 3. . Correlation between experimental and predicted abundance values calculated by using GA-ANN method L. hoffmeisteri.

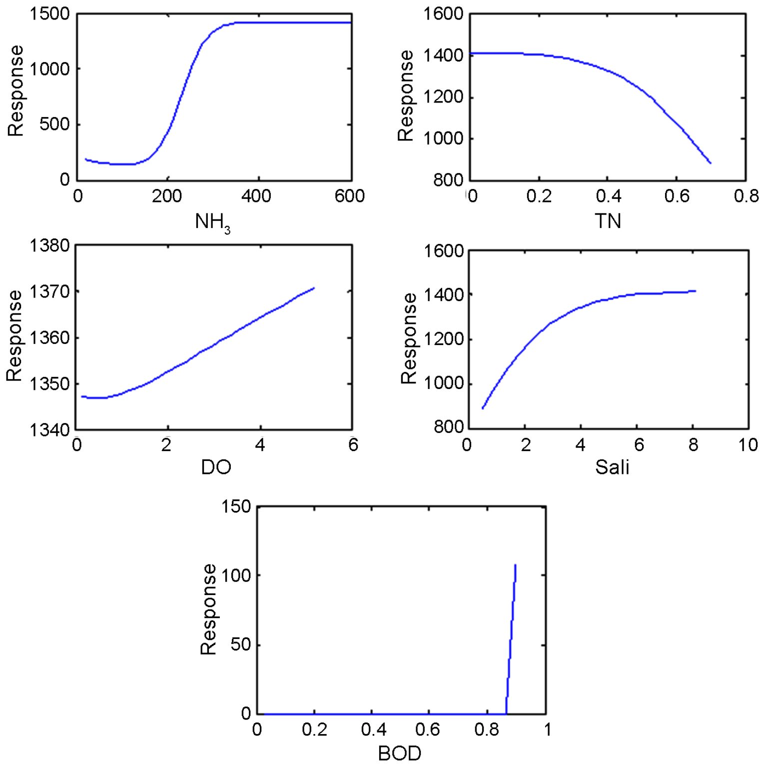

Fig. 4. Experimental values of abundance versus NH_3, TN, DO, Sali, BOD Values in GA-ANN method in L. Claparedeianus.

Table 1 shows that NH_3 concentration, total nitrogen (TN), Dissolved oxygen (DO), Sodium chloride (Sali), Biochemical Oxygen demand (BOD) are the best selected Biological and chemical properties of water in L. Claparedeianus in GA-ANN method. In Table 3 the electrical conductivity (EC), Total dissolved solids (TDS), NH_3, total nitrogen (TN) are the best selected biological and chemical properties of water in L. hoffmeisteri in GA-ANN method. The plot showing the variation of observed versus predicted abundance values in L. Claparedeianus and L. hoffmeisteri are shown in Figs. 2, 3. A good correlation between the calculated and empirical values of abundance can be observed in these Figures that approves the appropriateness of the developed model. The graphs of the found most effective Biological and chemical properties of water in L. Claparedeianus (NH_3,TN,DO,Sali,BOD) and L. hoffmeisteri (EC,TDS, NH_3, TN) versus the abun-

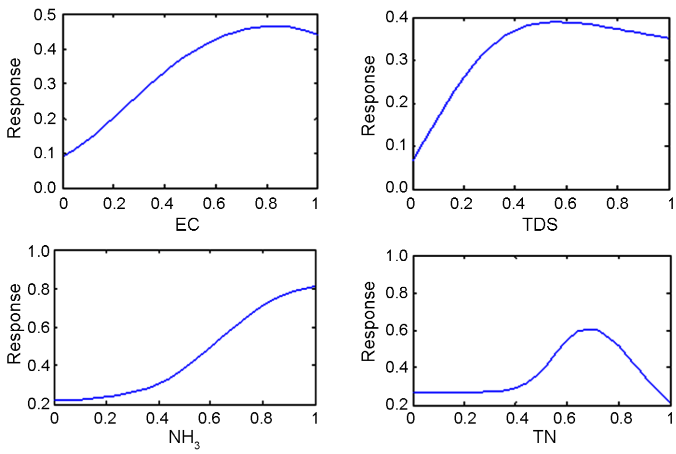

Fig. 5. Experimental values of abundance versus EC, TDS, NH_3 and TN Values in GA-ANN method in L. hoffmeisteri.

dance are plotted in Figs. 4, 5.

The charts showed that the abundance value in the L. Claparedeianus increases with increasing NH_3 concentration, Dissolved oxygen (DO), Sodium chloride in water (Sali). Also the total nitrogen (TN) increased the abundance value decreased. As the Biochemical Oxygen demand (BOD) increased to 0.8, no change in abundance was observed. Thus during this period, a bar was seen in the abundance. The charts in Fig. 5 showed that with increase in EC, TDS and NH_3 concentration. The abundance values in the L. hoffmeisteri were increased. By increasing the TN the response was increased and then reduced. The results in this work proved Nitrogen fixation was also enhanced in soils enriched with L. Claparedeianus and L. hoffmeisteri.

CONCLUSION

The obtained results from QAR models showed that GA-ANN combination were better than the other models used and also proved that NH_3 concentration, total nitrogen (TN), Dissolved oxygen (DO), Sodium chloride (Sali), Biochemical Oxygen demand (BOD), Total dissolved solids (TDS) and electrical conductivity (EC) were more significant than other Biological and chemical properties of water and predicting abundance values in L. Claparedeianus and L. hoffmeisteri substitution patterns. Thus NH_3 concentration, Dissolved oxygen (DO), Sodium chloride (Sali) and total dissolved solids (TDS) and electrical conductivity (EC) and NH3 concentration increase the population of L. Claparedeianus and L. hoffmeisteri and the soil becomes fertile. It can be concluded that simultaneous use of GA-ANN methods give deeper and more comprehensive knowledge of the effect of Biological and chemical properties of water on the abundance values in L. Claparedeianus and L. hoffmeisteri worms.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support provided by the Islamic Azad University of Rasht and Administrative Office of Ardabil Fisheries.

REFERENCES

[1] Liao, S.Y., Chen, J.C., Qian, L., Shen, Y., Zheng, K.C. (2008). QSAR studies and molecular Design of Phenanthrene-based Tylophorine derivatives with anticancer a activity. QSAR Comb. Sci., 27, 280–288.

[2] Wang, X., Yang, W., Xu, X., Zhang, H., Wang, Y. (2010). Studies of benzothiadiazine derivatives as Hepatitis C Virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics. Curr. Med. Chem., 17, 2788–2803.

[3] Putta, S., Eksterowicz, J., Lemmen, C., Stanton, R. (2003). A Novel Subshape Molecular Descriptor. J. Chem. Inf. Compute. Sci., 43, 1623.

[4] Gupta, S., Singh, M., Madan, A.K. (1999). Superpendentic index: a novel topological descriptor for predicting biological activity. J. Chem. Inf. Compute. Sci., 39, 272.

[5] Consonni, V., Todeschini, R., Pavan, M. (2002). Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies. J. Chem. Inf. Compute. Sci., 42, 693-705.

[6] Kirkpatrick, S., Gelatt, Jr. C.D., Vecchi, M.P. (1983). Optimization by Simulated Annealing. Science, 220, 671–680.

[7] Cerný, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. J. Optimum Theory Appl., 45, 41–51.

[8] Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. Q. Appl. Math., 2, 164–168.

[9] SPSS, Version 19, (2010). available at http://www.spssscience.com.

[10] Blay, J, Jr., Dongdem, F. (1996). Preliminary observations on the benthic macrofauna of a polluted coastal lagoon in Ghana (West Africa). Tropical Ecology, 37, 127 -133

[11] Suriani, A.L., Franca, R.S., Pamplin, P.A.Z., Marchese, M., Lucca, J.V., Rocha, O. (2007). Species richness and distribution of oligochaetes in six reservoirs on Middle and Low Tiete river (SP, Brazil). Acta. Limnologica. Brasilia, 19, 415-426.

[12] Nazarhaghighi, F., Timm, T., Shabanipourd, N., Nadoushan, R.M. (2014). Oligochaetes (Annelida, Clitellata) in the Anzali International Wetland, north-western Iran. Estonian. J. Ecol., 63, 130–144.

[13] Nazarhaghighi, F., Mousavi Nadoushan, R., Shabanipour, N., Fatemi, M., Mashinchian Moradi, A. (2015). First record of Limnodrilus claparedeianus Ratzel, 1868 (Annelida: Oligochaeta: Tubificidae) in Anzali Wetland, Guilan province, Iran Caspian J Environ. Sci., 13(4), 407-416.

[14] Kennedy, C.R. (1965). The distribution and habitat of limnodrilus claparède (oligochaeta: tubificidae). Oikos, 16, 26–38.

[15] Lafont, M., Vivier, A. (2006). Oligochaete assemblages in the hyporheic zone and coarse surface sediments: their importance for understanding of ecological functioning of watercourses. Hydrobiologia, 564, 171–181.

[16] Vivien, R., Tixier, G., Lafont, M. (2014). Use of oligochaete communities for assessing the quality of sediments in watercourses of the Geneva area (Switzerland) and Artois-Picardie basin (France): proposition of heavy metal toxicity thresholds. Ecohydrol. Hydrobiol., 14, 142–151.

[17] Lauritsen, D., Mozley, S.C., White, D. (1985). Distribution of oligochaetes in Lake Michigan and comments on their use as indices of pollution. J Great Lake Res., 11, 67–76.

[18] Lotufo, G.R., Fleeger, J.W. (1996). Toxicity of sediment-associated pyrene and phenanthrene to (oligochaeta: Tubificidae). Environ. Toxic Chem., 15, 1508–1516.

[19] Jalali-Heravi, M., Parastar, M.F. (2000). Use of artificial neural networks in a QSAR study of anti-HIV activity for a large group of HEPT derivatives. J. Chem. Inf. Compute. Sci., 40, 147.

[20] Unscrambler v.9.7, www.camo.com.

[21] Mauri, A., Consonni, V., Pavan, M., Todeschini, R. (2006). Dragon Software: an Easy Approach to Molecular Descriptor Calculations. MATCH Commun. Math. Comput. Chem., 56, 237-248.

AUTHOR (S) BIOSKETCHES

Reza Rajabei Nezhad, Assistant Professor, Department of Fisheries, Bandar Anzali Branch, Islamic Azad University, Bandar Anzali , Iran, *Email: reza_igl@yahoo.com*

Rrobabeh Sayyadi Kord Abadi, Assistant Professor, Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran

Fatemeh Nazarhaghighi, Assistant Professor, Administrative Office of Fisheries, Ardabil, Iran

Omid Alizadeh, Assistant Professor, Department of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran