University of Tehran Press

Pollution

https://jpoll.ut.ac.ir/

Print ISSN: 2383-451X
Online ISSN: 2383-4501

A Novel Deep Learning-based Prediction Approach for Groundwater Salinity Assessment of Urban Areas

Pouyan Abbasimaedeh[1]✉ | **Nasim Ferdosian**[2]

1. PhD, Geotechnical and Geoenvironmental Scientist, Perth, Australia
2. PhD, Research Fellow, Curtin University, Perth, Australia

Article Info

Article type:
Research Article

Article history:
Received: 02 Oct 2022
Revised: 16 Nov 2022
Accepted: 12 Jan 2023

Keywords:
Geoenvironment
Electrical conductivity
Groundwater
Deep learning
Prediction
Keras
Tensor Flow

ABSTRACT

The high amount of Electrical Conductivity (EC) in the groundwater is one of the major negative Geo-environmental problems which has a considerable effect on the quality of drinking water. To address this challenging problem we proposed an intelligent Machine Learning (ML) based approach to predict EC in urban areas. We applied the deep learning technique as one of the most applicable ML techniques with high capabilities for intelligent predictions. Five different deep neural networks (Net 1 to Net 5) were developed in this study and their reliability to predict EC with an emphasis on different settings of inputs, features, functions, and the number of hidden layers was evaluated. The achieved results showed that deep neural networks can predict EC parameters using minimum and economic input parameters. Results showed parameters Cl and SO_4 with a high range of correlation and pH with a low range of Pearson correlation properties are influential parameters to be used as the input of neural networks. Activation function Relu, optimization function Adam with a learning rate of 0.0005 and loss function Mean Squared Error with the minimum of two hidden dense layers from Keras laboratory of Tensor Flow developed an efficient and fast network to predict the EC parameter in urban areas. Maximum epochs for developed networks were defined up to 2000 iterations while epochs are reducible up to 200 to drive minimum loss function outcome. The maximum training and testing R2 for developed networks was 0.99 in both the training and testing parts.

Cite this article: Abbasimaedeh, P., & Ferdosian, N. (2023). A Novel Deep Learning-based Prediction Approach for Groundwater Salinity Assessment of Urban Areas. *Pollution*, 9 (2), 712-725.
http://doi.org/10.22059/POLL.2022.348405.1645

DOI: http://doi.org/10.22059/POLL.2022.348405.1645

INTRODUCTION

Communities' growth and overconsumption of clean water lead to a permanent demand and search for new water sources. Groundwater is an extensive source of drinking water supply in different urban areas around the world. Hence, water quality stability has always been of considerable geoenvironmental interest in most developed and developing countries (Maghrebi et al., 2022; Noori et al., 2022; Nasrabadi et al., 2010; Nasrabadi and Abbasi Maedeh, 2014a and 2014b). A remarkable number of academic studies have evaluated different attitudes toward groundwater such as reservoir issues, hydrogeology, quality, vulnerability, sustainability, and so on (Pandey and Kazama, 2011; 2012; Pandey et al., 2011; 2012; Chapagain et al., 2010). Although seas cover almost 75% of the earth's surface and finding new technologies to desalinate seawater takes a vast area of scientific research but it is not easy. Hence, quality monitoring to

*Corresponding Author Email: *Pouyan.Abbasimaedeh@gmail.com*

help for saving drinking water is located at the top of resource research nowadays in sustainable environments (Sun et. al. 2019)

A variety of geological, geotechnical, environmental, and human-made factors contribute to variations in groundwater quality (Nasrabadi and Abbasi Maedeh, 2014a and 2014b). For instance, Total Dissolved Solids (TDS) and Nitrate are among the most significant and common factors that may adversely affect the groundwater quality of an aquifer and both Anthropogenic and Geopogenic sources are known to greatly govern the extent of such contamination (Pazand et al., 2012). The lack of efficient and widespread sewer systems results in the direct and indirect discharge of municipal wastewater to surface and groundwater bodies. Leached fertilizers and pesticides can be introduced as another geoenvironmental threat to agriculture; the amount beyond the limit taken up by plants will be leached downwards to aquifers (Nolan 2001). Anthropogenic sources such as acid rain and sulphate fertilizers as well as geopogenic ones such as pyrite oxidation in a limestone aquifer can both contribute to groundwater quality degradation (Stigter et al.,1998; Chan 2000; Vizintin et al., 2009).

The Electrical Conductivity (EC) parameter constitutes the next fundamental parameter regards drinking, industrial and agricultural water. EC is directly linked to water salinity, sodium absorption coefficient, and drinking water quality (Asghari Moghaddam et al., 2006; Mehrdadi et al., 2012). Electrical Conductivity is a scale measuring the water's salinity and has been measured from 0 to 50,000 uS/cm. Electrical Conductivity is measured in micro Siemens per centimetre (uS/cm). Freshwater is usually between 0 and 1,500 uS/cm and typical seawater has a conductivity value of about 50,000 uS/cm. A narrow range of salinity could be found naturally in waterways which is significant for plants and animals growing. High levels of salinity in freshwater can lead to aquatic ecosystem problems as well as convoluted drinking uses. A linear relationship has been detected between EC and TDS which has been described in the literature (Maedeh et. Al., 2013). From a geoenvironmental point of view, EC fluctuation might be affected by an anthropogenic source or a high amount of groundwater usage. Subsidence, cracks, aquifer salinity, and death of ground in the high EC area might be a major probable crisis (Nasrabadi and Abbasi Maedeh, 2014a and 2014b).

To measure all the participated quality factors of drinking water, especially groundwater sources, laboratory tests are necessary, while most of these lab tests take time and are not cost-efficient during the time. In addition, for a fast reaction to finding where is the source of the pollution in some critical situations like biological pollution leakage, radioactive detection, oil leakage or so on, and what is the effect of measured parameters on other quality factors, laboratory tests are not so efficient and most of the time prepared fast response kits or probs will be more efficient. Nowadays, different Machine Learning (ML) approaches are used to define nonlinear and smart relationships between parameters and calculate the effective weight of each parameter on the targeted parameters to predict the specific parameters and their relationship with other factors (Abbasi Maedeh et. Al., 2013).

Machine Learning is applied in systems like Geotechnical and Environmental Health Monitoring, Terrific, sea and air navigation, and medical system. The significance of using ML stems from cost reduction, human footprint and effort, and reliability. Moreover, ML is capable to simulate systems, even if mathematical models are complex and time-consuming. This is the case for geoenvironmental and groundwater monitoring systems. Neural Networks which are the main part of ML can model the dynamic behaviour of a non-linear process only through training. For instance, a General Regression Neural Network (GRNN) and a single-objective optimization model are suggested to predict quality parameters in groundwater in previous research (Taulli, 2019).

A Neural Network with more than one simple hidden layer has been called Deep Neural Network, also known as Deep Learning (Sutton and Barto., 2018). Deep Learning methodology has been considered a convenient substitute for regressions and empirical models to predict the behaviour of data series, owing to their time reliability and adaptability to unpredicted changes (Zheng and Sabuncu, 2018). The Deep Learning method could be applied not only for qualitative predictions of water but also for predicting groundwater situation and volume (Sun et.al 2019). Strong programming languages and highly advanced programming libraries such as TensorFlow, and Keras help scientists to design, develop and use different Deep Learning approaches to make predictions. Such advanced libraries are widely being used in different industries and cutting-edge technology interdisciplinary research (Aggarwal, 2015; Sun et.al 2019).

A variety of previous research has used deep learning to predict targeted parameters in Geotechnical and environmental engineering (Coulibaly et al., 2001; Kumar et al., 2002; Coppola et al., 2003; Hosaini et al., 2007; Khalili et al., 2008; Dehghani et al., 2009; Aggarwal, 2015; Zhang & Sabuncu, 2018; Sun et.al 2019). For instance, previous research on Geoenvironmental engineering showed strong perdition and precise outcome of the TDS parameter of groundwater in urban areas (Mehrdadi, et al., 2012; Abbasimaedeh, et.al., 2013).

In this study, we applied deep learning prediction theory to predict the EC parameters of the groundwater reservoir of very busy and populated urban areas. The updated database of urban groundwater quality factors for Tehran city, the capital city of Iran, has been imported into our model as the input. We aimed to predict the salinity and Electrical Conductivity of the urban groundwater which is the main resource of drinking water in Tehran province. Authors have previous research on the groundwater quality of similar areas and in this article, they apply different deep learning methods to predict future groundwater quality based on limited and economic input data, location, and seasonal properties. The innovation of this study could help to reduce the bug of traditional data interpolation to estimate parameters and we have developed different deep neural networks to predict local and global EC parameters of groundwater sources in urban areas.

MATERIALS AND METHODS

Tehran is situated in northern Iran and is identified as the most populated city in Iran. Tehran has an approximate population of more than 15 million people and is prominent among the Middle East capitals (Nasrabadi and Abbasi Maedeh, 2014a and 2014b). Although a massive rise rate of the population during the last 70 years, its density is relatively constant at around 120 persons per hectare (Asadpour and Nasrabadi 2011). Such a case confirms an overwhelming urban enlargement and consequently an expanded anthropogenic pollution discharge. The locations of Tehran Province and Tehran County are illustrated in Figure 1.

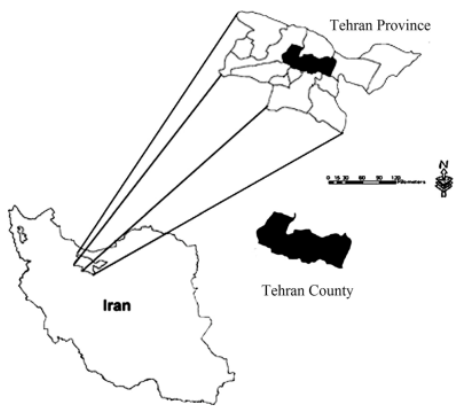

Fig. 1. Location of Tehran Province and Tehran county in Iran

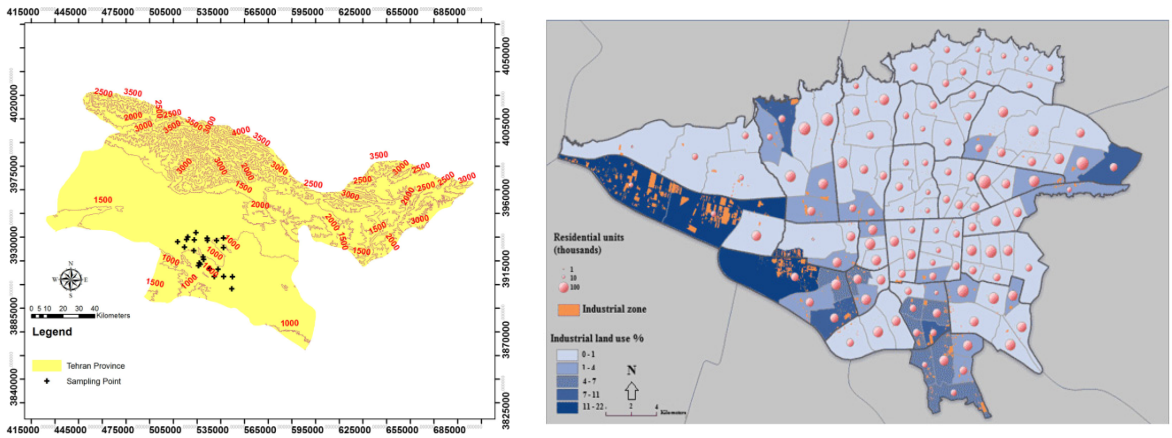

Fig. 2. Elevation distribution of Tehran Province (left); Residential and industrial areas distribution (right)

To identify the groundwater flow direction within the study area and a better understanding of industrial areas, the general topography, and industrial and residential spatial distribution of Tehran province are depicted in Figure 2. Regarding the fast socio-economic and industrial growth, land use has been playing a crucial role in groundwater quality degradation in Tehran province. Agriculture is the dominant land use in the southern parts of the province while industrial and residential parts are concentrated in the western and northern parts. The density of industrial zones is higher in the southwestern and southern parts of the city, while a relatively even distribution is observed regarding residential land use (Noori et al. 2022). In general, the groundwater flows southwards. Tehran city and its land use are considered the main groundwater pollution source in the southern parts of the province, which are located at relatively low elevations (Nasrabadi and Abbasi Maedeh, 2014a and 2014b).

To appropriately cover the study area, more than 450 boreholes were selected for groundwater sampling in this study, and the lab result of each borehole for at least 15 years, different sessions, and months were collected in the database. The overlapped aerial photo of proposed borehole locations and distributed map of land elevation in the study area have been shown in Figure 3. It is observed that most of the boreholes were concentrated at the elevation of 1000 to 2000 m and limited numbers were located in the higher levels, especially in the eastern part.

The northern zone of Tehran Province includes parts of the Alborz mountains, while southern Tehran areas are located in the central Iran plain. Tehran flat generally has a plain-like physiographic feature, in which hilly land structure forms a harsh morphology with south-eastern Tehran heights, southern Karaj hills, and southern Kahrizak hilly lands. It is reliable that Tehran Province belongs to Alborz and central Iran structural zones. These areas lie on the northern Tehran fault.

The oldest stratigraphic units of Tehran Province including a late Precambrian to mid-Triassic platform sequence comprise several geologic formations and alluvial unconformities (Nasrabadi and Abbasi Maedeh, 2014a and 2014b). In the southern half of the Tehran plain, against the northern half, Neogene units were formed mainly of marl, red sandstones, and conglomerates with red hilly land features. In Tehran Province, several epeirogenic and orogenic phenomena have been caused in the recent morphotectonics area. Geological structures (e.g., faults and folds) have an NW–SE general trend in the western parts turning to an NE trend in the east (Nasrabadi and Abbasi Maedeh, 2014a and 2014b).

Parameters such as temperature, pH, Electrical Conductivity, and Dissolved oxygen have been evaluated using portable instruments at the sampling site and the other parameters have been analysed in the laboratory. Nitrate, nitrite, and sulphate have been measured by the HACH

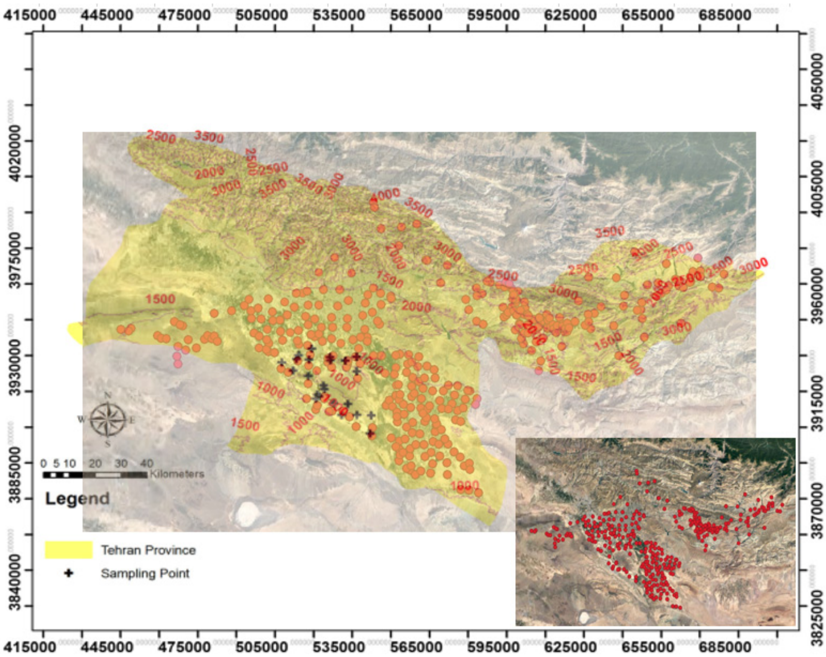

Fig. 3. Overlapped picture of study area elevation versus boreholes location

instrument and US-EPA methods 8039, 8507, 8051, and 8029 have been employed respectively. All cations have been measured via the EPA-3005 method and through the inductive flame atomic absorption spectrometer (FAAS). Standardized method number 4500 has been used to measure carbonate and bicarbonate anions and chloride measurement has been carried out applying the argentometric method (Coppola et al., 2003; Biswas, 2005; Abbasimaedeh and Mehrdadi, 2012).

An ANN is a computation technique to accelerate the learning process and tries to trace the input space (input layer) and a desirable environment (output layer) by using functions called neurons by identifying inherent relationships between data (Esmaeili Vakili et al., 2004; Taherion, 2006). Hidden layers receive data from the input layer, process it, and send it to the output layer. Each neural network receives training through instances.

Network learning is carried out when the linkage weight between the layers changes in a way that the difference between the predicted and the measured values are within permissible limits. Prosperity in this status fulfils the learning proceeding. This declares the weight of memory and neural network know-how. Trained neural networks can be used to predict outputs corresponding to actual new data (Abbasimaedeh et al., 2013). Considering the structure of neural networks, features such as speed processing, learning capability by a schema design method, conception of proficiency after learning, flexibility against surprising defeats, and lack of significant disruption on the part of the connection could be the result of weight distribution.

Deep neural networks with more than two layers that have been designed to predict EC were developed in this study. A maximum of 4 hidden layers were used for preliminary networks and layers were reduced to two layers in the optimized design. The major difference between the current study networks with normal ANNs is to use of very deep networks and a high amount of inner features to increase prediction accuracy. Dense layers for multilayer neural networks were selected for network structures (Fig.4). The inner features for hidden layers were selected between 1000 to 500 for primary hidden layers and were reduced up to 10 in the final hidden layers on deep neural networks. The Relu activation function has been used for all of the hidden layers except the final layers which have no activation function.

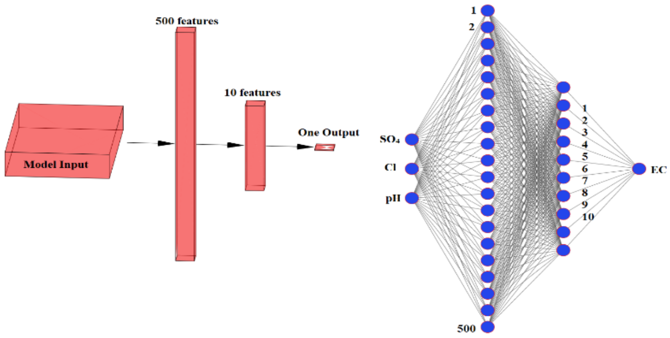

Fig. 4. The structure of the deep neural network (NET 5)

Table 1. Primary data analysis of the dataset

	pH	TDS (ppm)	EC (μS/Cm)	Cations (meq/L)				Anions (meq/L)				NO3 (meq/L)
				K	Na	Mg	Ca	SO_4	Cl	HCO_3	CO_3	
Count	3287	3287	3287	3287	3287	3287	3287	3287	3287	3287	3287	3287
Mean	7.83	1171.08	1906.23	0.08	8.52	3.85	6.88	7.12	8.25	3.80	0.02	27.36
Std. Dev	0.36	1622.76	2585.27	0.07	13.63	5.15	9.92	10.75	16.58	1.72	0.10	17.62
Min	6.42	109.00	191.00	0.00	0.07	0.08	0.70	0.18	0.10	0.25	0.00	6.20
25%	7.60	326.50	573.00	0.03	1.60	1.20	2.56	1.45	1.00	2.58	0.00	18.60
50%	7.82	557.10	936.90	0.08	3.25	2.00	3.76	2.90	2.40	3.38	0.00	27.36
75%	8.10	1157.50	1887.00	0.10	8.98	3.96	6.58	7.22	6.30	4.63	0.00	31.00
Max	9.78	16100.00	25000	0.90	136	42	109	108	171	15.48	1.06	148.80

To achieve the best prediction outcome with the highest amount of accuracy the Adam optimizer function with a learning rate of 0.0005 for updating neuron weights has been applied to the deep networks. All the designed deep networks in this study were updated with their weights with 2000 to 1000 epochs. Furthermore, 70% and 30% of the total data have been considered for the training and testing stages respectively. In addition, 20% of training data have been chosen for the validation of network output in proposed deep networks. The mean squared error (MSE) and R-squared amount were used in proposed deep networks to estimate the errors of predictions. The relationship for loss function as MSE errors is shown in Equation 1.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2 \tag{1}$$

MATERIALS AND METHODS

The quantitative analysis of groundwater quality parameters can be reflected with an emphasis on different geological, geotechnical, and environmental indexes which have been assessed based on geoenvironmental engineering legislations. The maxima, minima, averages, and standard deviations of the physicochemical parameters were calculated and have been shown as the primary data analysis in Table 1. Statistical analysis has been done on 3287 results of each set for Anions, Cations, pH, No3, TDS, and EC. The result showed only 25 percent of data had EC amounts lower than 573 mS/cm and 75% of EC amounts were detected for less than

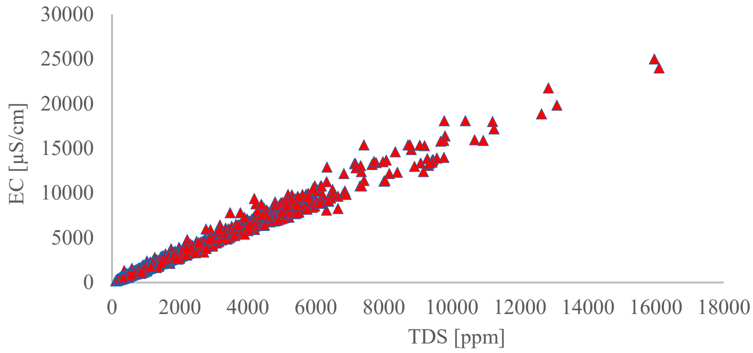

Fig. 5. Scatter graph of measured EC and TDS

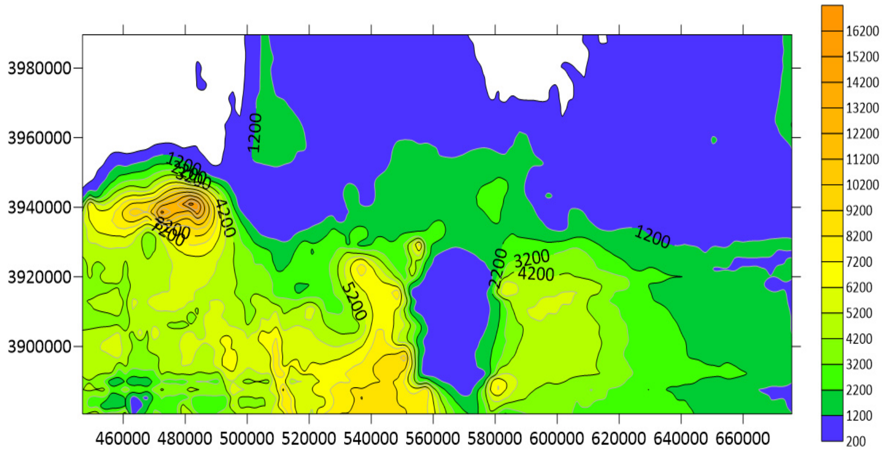

Fig. 6. Distributed map of EC (mS/cm) based on Kriging interpolation

1887 mS/cm. It is concluded the study area is going to be the salinized area. According to WHO standards, the EC value should not exceed 400 µS/cm. The current investigation indicated that the EC value was 179.3–20 µS/cm with an average value of 192.14 µS/cm.

A linear relationship between EC and TDS has been developed based on the current study's collected dataset. The proposed relationship could estimate the TDS parameter with the highest degree of accuracy and shows R-squared 0.99 in its TDS development. The proposed relationship to estimate TDS based on EC in our study of the urban area is presented in Eq. 2.

$$EC_{(\mu S/cm)} = 1.58\ (TDS_{\ (ppm)}) + 53.36 \quad (2)$$

The kriging interpolation method has been used to estimate the EC amount in the areas between actual boreholes for current study boundaries. Kriging is an interpolation method that makes predictions at unsampled locations using a linear combination of observations at nearby sampled locations (DiBiase et al., 2006). The result of EC prediction based on the Kriging method has been depicted in Figure 6. The result shows the northern part has been filled with blue or null. The null is per the lack of measured data in this area which has mountain geomorphology. The distribution of EC in other areas shows by Equipotential Line with the range of 1000 mS/cm difference. The maximum amount of EC shows in the north-western southwestern western part of the Tehran province. This interpolation significantly has confirmed the authors' previous studies in this area.

The meaningful relationship between Anions, Cations, pH, TDS, and EC in the statistical

Table 2. Pearson correlation for anion, cation, EC, and pH

	K	Na	Mg	Ca	So_4	Cl	Hco_3	Co_3	No_3	pH	TDS	EC
K	1											
Na	0.59	1										
Mg	0.60	0.77	1									
Ca	0.60	0.74	0.77	1								
So_4	0.65	0.91	0.86	0.81	1							
Cl	0.61	0.90	0.82	0.91	0.84	1						
Hco_3	0.18	0.21	0.32	0.10	0.24	0.07	1					
Co_3	-0.04	-0.02	-0.05	-0.07	-0.04	-0.05	-0.08	1				
No_3	0.18	0.23	0.17	0.26	0.22	0.22	0.20	-0.04	1			
pH	-0.17	-0.11	-0.15	-0.20	-0.14	-0.14	-0.25	0.51	-0.07	1		
TDS	0.64	0.94	0.87	0.89	0.94	0.96	0.21	-0.05	0.25	-0.16	1	
EC	0.65	0.94	0.87	0.91	**0.94**	**0.97**	0.21	-0.05	0.25	**-0.16**	0.99	1

correlation analysis has been assessed in this study. Preliminary statistical analysis showed a semi-linear relationship among parameters which could be a strong reason to select the Pearson Correlation analysis for the current study. Pearson correlation coefficients measure the linear relationship between the variables (Abbasimaedeh et al., 2013). This method is a relationship in which the variables tend to move in the same/opposite direction but not necessarily at a constant rate whereas the rate is constant in a linear relationship. According to the rule of correlation coefficients, the strongest correlation is considered when the value is closest to +1 or -1. A positive correlation coefficient implies that the variables are affiliated directly.

The result of the Pearson correlation showed the strongest relationship between EC, K, Na, Mg, Ca, SO_4, and Cl. The result of correlation analysis showed a limited or negligible negative correlation between EC and the remaining parameters in this study database. Results showed a negative relationship between pH, $Co_{3,}$ and EC while other relationships are positive. The achieved relationship between EC and pH is not enough strong in linear correlation but based on the lowest price and easy accessibility of the pH test, this parameter will be considered as an input in the proposed neural networks of this study. In addition, the achieved strong linear and meaningful correlation between EC and TDS has confirmed the principle of developed relationship 1. The result of the Pearson correlation has been illustrated in Table 2.

Keras and Tensor Flow libraries were used for developing deep-learning neural network structures. One of the main advantages of these libraries in addition to their development is that the models which are made based on these libraries can run with both GPU and CPU. Due to the constant number of input parameters for our proposed models, a constant amount of learning rate was considered for the model optimizer function. To reduce the weight saturation phenomenon in proposed deep networks, the activation function Relu was considered in our proposed deep networks and Tangent Hyperbolic and Sigmoid functions have been neglected.

The optimized number of hidden layers and features has been considered in this study based on try and error. Different amounts for learning rates were applied to the model for the Adam optimizer function, while based on try and error it is concluded the amount 0.0005 was highly efficient in achieving acceptable accuracy of prediction. The results of the accuracy of prediction in terms of the R-squared values related to the EC parameter predicted by all the proposed deep neural networks have been reported in Table 3.

Five different deep networks with different numbers of hidden layers, features, and epochs were designed to predict EC parameters in the groundwater of the Tehran urban area. The input

Table 3. Structure of developed deep neural networks and their outcomes

Network structure	Net 1	Net 2	Net 3	Net 4 (Tehran)	Net 5 (Global)
Input	UTMy, UTMx, K, Na, Mg, Ca, So4, Cl, Hco3, Co3, No3, pH, Month, Year	UTMy, UTMx, K, Na, Mg, Ca, So4, Cl, pH, Month, Year	UTMy, UTMx, K, Na, Mg, Ca, So4, Cl, pH, Month, Year	UTMy, UTMx, So4, Cl, pH, Month, Year	So4, Cl, pH
Layer 1	Layer type= Dense Feature=1000 Function*= Relu	Layer type= Dense Feature=1000 Function= Relu	Layer type= Dense Feature=500 Function= Relu	Layer type= Dense Feature=500 Function= Relu	Layer type= Dense Feature=500 Function= Relu
Layer 2	Layer type= Dense Feature=500 Function= Relu	Layer type= Dense Feature=500 Function= Relu	Layer type= Dense Feature=500 Function= Relu	Layer type= Dense Feature=10 Function= Relu	Layer type= Dense Feature=10 Function= Relu
Layer 3	Layer type= Dense Feature=200 Function= Relu	Layer type= Dense Feature=200 Function= Relu	Layer type= Dense Feature=200 Function= Relu		
Layer 4	Layer type= Dense Feature=10 Function= Relu	Layer type= Dense Feature=10 Function= Relu	Layer type= Dense Feature=10 Function= Relu		
Out	Layer type= Dense, Feature=1	Layer type= Dense, Feature=1	Layer type= Dense, Feature=1	Layer type= Dense, Feature=1	Layer type= Dense, Feature=1
Optimizer	Adam (LR** 0.0005)	Adam (LR 0.0005)	Adam (LR 0.0005)	Adam (LR 0.0005)	Adam (LR 0.0005)
Val.Split* **	20%	20%	20%	20%	20%
Epochs	2000	1000	2000	1000	1000
R-squared Train	0.99	0.99	0.99	0.99	0.99
R-squared Test	0.99	0.95	0.95	0.99	0.99

*Activation function ** Learning Rate *** Validation split

parameters of each proposed network have been shown in Table 3. The aim of deep layers neural network design is to reduce the inputs with an emphasis on the rate of correlation between EC and other available parameters, accessibility, ease to measure, and the price of the lab test. We considered two different deep learning models as targeted deep neural networks which have different targets. The first network with an emphasis on predicting based on the location and time of the sampling in addition to principal quality parameters (NET 4) and the second one was only designed based on principal quality parameters (NET 5).

Cl and SO_4 are very important and common parameters that normally will measure in different technical statements like soil mechanics reports. These parameters are highly influential on human health, industrial equipment health and durability, and construction durability. These parameters are easy to access and measure in both soil and water environment, while laboratory tests to measure these two parameters are fairly common and cost-efficient. On the other hand, developed probs which are very common and are beneficial to show pH in real time are so cheap and common in environmental monitoring.

Proposed NET 4 was designed to predict EC in the study area and it can consider the time of the sampling (year and month) as inputs, while proposed NET 5 has been designed for a global perspective to predict EC from three simple parameters. All the information for the current study proposed deep neural networks have been explained in Table 3.

RESULT AND DISCUSSION

Five different proposed deep neural networks with an emphasis on the explained input parameters and described configuration were developed under the Linux operating system and Python version 3.9. To compare the accuracy of developed prediction models (Net 1 to Net 5), we have used the R-squared values related to the EC parameter predicted by the models. The results of the accuracy of prediction in terms of the R-squared values related to the EC parameter predicted by all the proposed deep neural networks for both train and test datasets were reported in Table 3. We found that all five methods provide high R-squared values when we train them. For instance, Net 1, Net 4, and Net 5 provide the highest amount of accuracy with an R-squared value equal to 0.99.

The actual amount and predicted amount of EC parameter (mS/cm) for each developed deep neural network in the stages of training and test were depicted in Figure 7. These results showed the minimum accuracy was measured in the test steps of developed Net 2 and Net 3, while other networks achieved the highest amount of accuracy for the actual and predicted amount of EC parameter (mS/cm). The lack of prediction accuracy is detected where the amount of actual EC exceeds 10000 mS/cm in proposed NET 2 and NET 3, hence the R-squared error for these two proposed networks showed a lower amount of resolution in a comparison with the amount achieved by other proposed networks. Outcomes showed Net 4 and Net 5 could predict the amount of EC (mS/cm) with the highest accuracy of 0.99 and using a minimum range of input parameters.

Considering only the R-Squared parameter is not sufficient in comparing all investigated

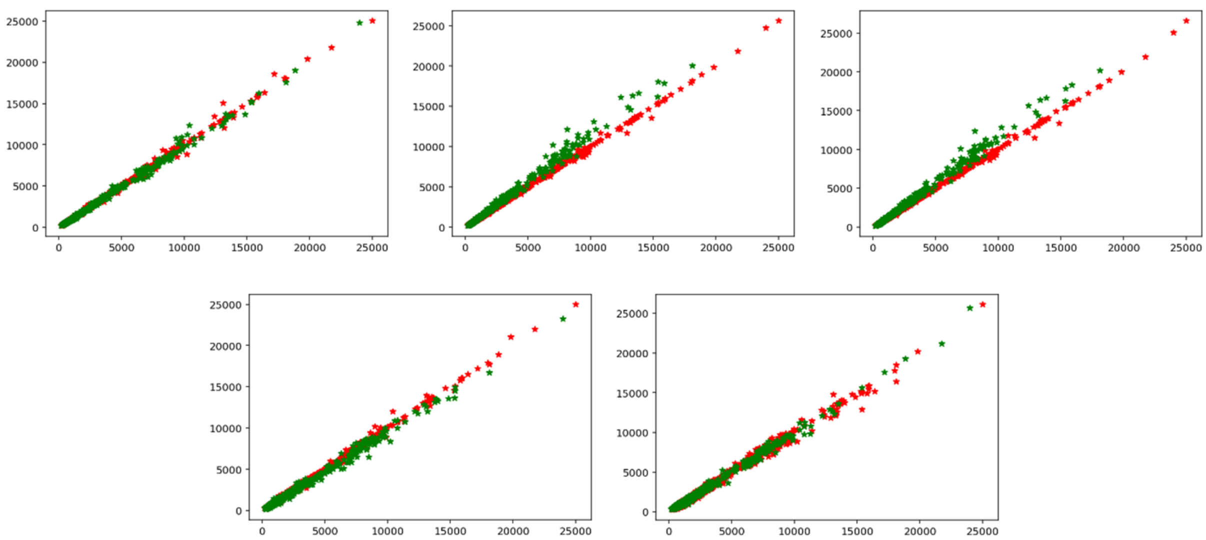

Fig. 7. Result of training and test to predict EC parameter (mS/cm) for each proposed deep neural network; from top left to right (Net 1 to Net 5). The red and green dots are training and test results respectively. The horizontal axes are the actual amount and horizontal axes are the predicted amount (mS/cm).

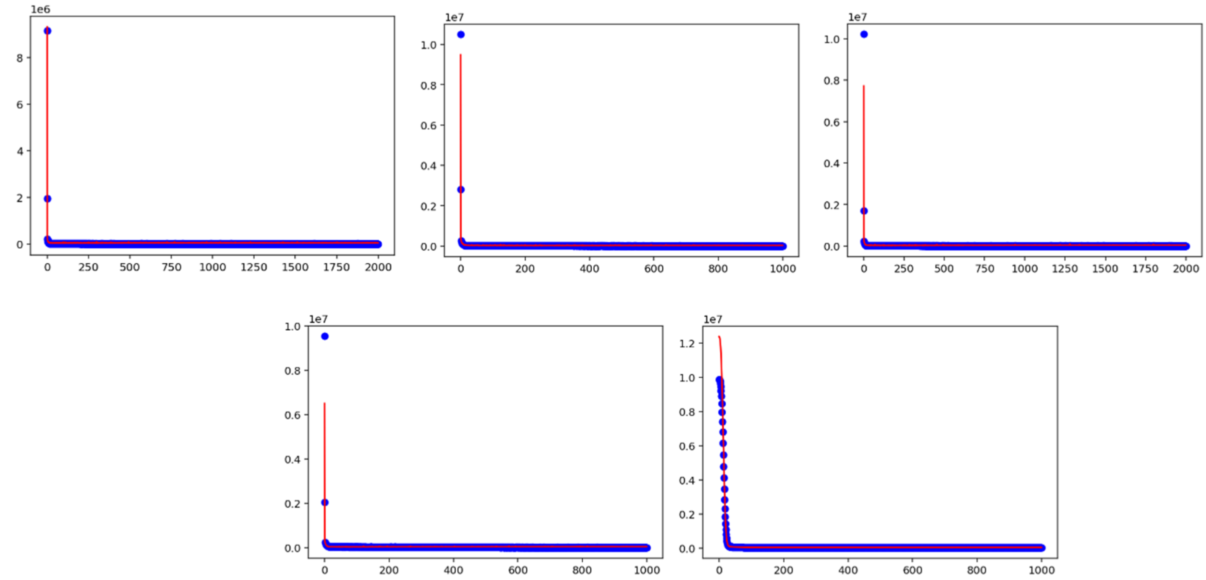

Fig. 8. Result of training and test loss for EC for each of the developed deep neural networks; from top left to right (Net 1 to Net 5). The red and blue dots are training and test results respectively. The horizontal axes are loss amount, and the horizontal axes are epoch numbers.

deep learning models to find the best model we have provided the result of the accuracy for each of the proposed deep neural networks with an emphasis on loss function, depicted in Figure 8. The result illustrated the minimum loss that occurred after approximately 200 epochs for each network. Hence, it is acceptable and doable to reduce the number of epochs up to 200 for each network and decrease the analysis time of proposed deep networks. In addition, it is depicted that a higher amount of iteration to achieve the minimum loss function has happened in Net 5.

Our deep neural networks were developed based on the available database of hydrochemical parameters of Tehran Province groundwater. The database is an updated version of the database used in our previous research (Nasrabadi and Abbasimaedeh 2014a and 2014 b). The achieved results showed the minimum error in our data prediction which is verifiable with the mentioned research when using NET 4. Based on the total results, we can conclude that for predicting EC, Net 5 performed a better estimation than the rest models (with high accuracy, less response time, and fewer required parameters as input to train the neural network), while other proposed deep learning models had almost acceptable accuracy.

CONCLUSIONS

To predict the Electrical Conductivity of groundwater in urban areas the dataset of groundwater parameters for Tehran city, the capital of Iran, was collected and considered as the input for deep neural networks to predict the EC in groundwater of urban areas. Previous research indicated that the most distinguishing factor of drinking water quality in urban areas was Electrical Conductivity in the urban area. Single-layer neural networks have been already used to estimate different quality parameters of drinking water in previous studies, while several deep neural networks were used to predict the Electrical Conductivity of Tehran urban areas in this study. Input parameters for proposed networks were selected based on the result of Pearson correlation, accessibility, cost-benefit, and time and location effects. Dense layers of deep neural networks were developed based on Keras and Tensor Flow libraries for artificial network configuration. In addition, the Relu activation function and Adam optimizer were used for the current proposed deep predictor networks. Furthermore, the MSE function was defined as the loss function. The result of the current study with an emphasis on the outcomes of the five proposed deep networks are summarized as follows:

- The developed deep neural networks based on Keras and Tensor flow could significantly and accurately predict the EC parameters of urban areas.
- Proposed deep networks in this study can predict EC with an emphasis on both locational and time, and quality parameters in urban areas. The achieved result showed it is possible to use both Net 4 and Net 5 to predict an accurate amount of EC parameters in urban areas
- The result of the proposed deep neural networks showed some of the developed networks could not predict with a high amount of accuracy in the high range of EC. The lower range of R-squared was detected in both Net 2 and Net 3 for both the training and testing stages.
- The amount of loss function outcome in proposed deep networks showed a minimum of about 200 epochs is enough to estimate the acceptable accuracy of the EC parameter in groundwater to achieve R-squared 0.99.
- The correlation between EC and other quality factors showed although there is a meaningful and positive relationship between Na, K, Mg, SO_4, Cl, and EC regarding easy to achieve and cheapest price of test SO_4, Cl, and pH was selected as the inputs of the developed and proposed networks 4 and 5. The maximum R-squared for proposed networks were 0.99 and they are enough strong to predict the high and low range of EC with an emphasis on the input ranges.

GRANT SUPPORT DETAILS

The present research did not receive any financial support.

CONFLICT OF INTEREST

The authors declare that there is not any conflict of interest regarding the publication of this manuscript.

REFERENCES

Abbasimaedeh, P. and Mehrdadi, N. (2012). Assessment of groundwater basins in Tehran quality with the World Health Organization Index. International Journal of Water Engineer, 64, 10 – 19.

Aggarwal, CC. (2015). Outlier analysis, Data mining. Springer, 237–263.

Asadpour, G.A. and Nasrabadi, T. (2011). Municipal and medical solid waste management in different districts of Tehran, Iran. Fresenius Environmental Bulletin, 20(12), 3241 – 3245.

Asghari Moghaddam, A., Nadiri, A. and Fijani, E. (2006). Ability to study different Models of Artificial Neural Networks to Evaluate Groundwater Water levels in the Hard Formation, Tenth. Conference of Geological Society, Tehran.

Chan, H.J. (2000). Effect of land use and urbanization on hydrochemistry and contamination of groundwater from Taejon area, Korea. J Hydrol, 253, 194–210.

Chapagain, S.K., Pandey, V.P., Shrestha, S., Nakamura, T. and Kazama, F. (2010). Assessment of deep groundwater quality in Kathmandu Valley using multivariate statistical techniques. Water Air and Soil Pollution, 210(1-4), 277 – 288.

Coppola, E., Szidarovszky, F., Poulton, M. and Charles, E. (2003). Artificial Neural Network Approach for Predicting Transient Water Levels in a Multi-Layered Groundwater system under variable state. Pumping and climate conditions, Hydrologic engineering, 8(6), 348 – 360.

Coulibaly, P., Anctil, F., Aravena, R. and Bobée, B. (2001). Artificial neural network modelling of water table depth fluctuations. Water resources research, 37(4), 885 – 896.

Dehghani, A.A., Asgari, M. and Mosaedi, A. (2009). Comparison of Geostatistics, Artificial Neural Networks and Adaptive Neuro-Fuzzy Inference System Approaches in groundwater level interpolation (case study: Ghazvin aquifer), Journal of Agriculture science natural resource, 16(1), 517 – 528.

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B. and Wentz, E. (2006). Principles of Kriging. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers. ((**2nd Quarter 2016, first digital).**

Esmaeili Varaki, M., Khayat Khalaghi, M. and Shafiei, M. (2004). Provide a model for intelligent water level fluctuations estimated alluvial groundwater aquifer using an artificial neural network. Articles first annual conference of Iran water resources management, 1 – 11.

Hosaini, M.T., Siosemarde, A., Fathi, P. and Siosemarde, M. (2007). Application of artificial neural networks (ANN) and multiple regressions for estimating and assessing the performance of dry farming wheat yield in Ghorveh Region, Kurdistan province. Agricultural research: Water and soil and plant, 7(1), 41 – 54.

Khalili, S.R., Davari, k. and Mousavi Baygi, M. (2008). Monthly Precipitation Forecasting Using Artificial Neural Networks: A Case Study for Synoptic Station of Mashad. Journal Water and Soil, Agricultural Science & Technology Ferdowsi University of Mashhad, 22(1), 39 – 99.

Kumar, M, Raghuwanshi, N., Singh, R., Wallender, W., and Pruitt, W. (2002). Estimating evaporate transpiration using artificial neural networks. Journal of Irrigation and drainage engineering ASCE 128, 4, 224 – 233.

Maedeh, PA, Mehrdadi, N., Bidhendi GRN. and Abyaneh HZ. (2013).Application of artificial neural network to predict Total dissolved solids variations in groundwater of Tehran Plain, Iran. Int J Environ Sustain, 2(1), 10-20.

Maghrebi, M., Noori, R., Partani, S., Araghi, A., Barati, R., Farnoush, H. and Haghighi, A. T. (2021). Iran's groundwater hydrochemistry. Earth and Space Science, 8, DOI: 10.1029/2021EA001793.

Mehrdadi, N., Hasanlou, H., Jafarzadeh, M.T., Hasanlou, H. and Abodolabadi, H. (2012). Simulation of low TDS and biological units of Fajr industrial wastewater Treatment plant using artificial neural network and principal component analysis hybrid method. Journal of water resource and protection, 4, 370 – 376.

Nasrabadi, T., Nabi Bidhendi, GR., Karbassi, AR. and Mehrdadi, N. (2010). Evaluating the efficiency of sediment metal pollution indices in interpreting the pollution of Haraz River sediments, southern Caspian Sea basin. Environ Monit Assess, 171,395–410.

Nasrabadi, T. and Abbasi Maedeh, P. (2014a). Groundwater quality assessment in southern parts of Tehran plain. Iran. Environ Earth Sci, 71, 2077–2086.

Nasrabadi, T. and Abbasi Maedeh, P. (2014b). Groundwater quality degradation of urban areas (case study: Tehran city, Iran). Int. J. Environ. Sci. Technol, 11, 293–302.

Nolan, B.T. (2001). Relating nitrogen sources and aquifer susceptibility to nitrate in shallow groundwaters of the United States. Groundwater, 39, 290–299.

Noori, R., Farahani, F., Jun, C., Aradpour, S., Bateni, S.M., Ghazban, F., Hosseinzadeh, M., Maghrebi, M., Naseh, M.R.V. and Abolfathi, S. (2022). A non-threshold model to estimate carcinogenic risk of nitrate-nitrite in drinking water. J. Clean. Prod., 363., DOI 10.1016/j.jclepro.2022.132432.

Noori R, Maghrebi M, Mirchi A, Tang Q, Bhattarai R, Sadegh M, Noury M, Torabi Haghighi A, Kløve B, Madani K. Anthropogenic depletion of Iran's aquifers. Proc Natl Acad Sci U S A. 2021 Jun 22;118(25):e2024221118. DOI: 10.1073/pnas.2024221118.

Pandey, V.P. and Kazama, F. (2011). Hydrogeologic characteristics of groundwater aquifers in Kathmandu Valley, Nepal. Environmental Earth Sciences. 62(8), 1723 – 1732.

Pandey, V.P. and Kazama, F. (2012). Groundwater storage potential in the KathmanduValley's shallow and deep aquifers In Shrestha S., Pradhananga D., Pandey V.P. Kathmandu Valley Groundwater Outlook, AIT/SEN/ CREEW/ICRE-UY, 31 – 38.

Pandey, V.P., Shrestha, S., Chapagain, S.K. and Kazama, F. (2011). A framework for measuring groundwater sustainability. Environmental Science & Policy, 14(4), 396 – 407.

Pandey, V.P., Shrestha, S. and Kazama F. (2012). Groundwater in the Kathmandu Valley: development dynamics, consequences and prospects for sustainable management. European Water, 37, 3 – 14.

Pazand, K., Hezarkhani, A., Ghanbari, Y. and Aghavali, N. (2012). Geochemical and quality assessment of groundwater of Marand Basin, East Azarbaijan Province, northwestern Iran. Environ Earth Sci, 67(4), 1131–1143.

Sutton, R.S., and Barto, R.G. (2018). Reinforcement Learning An Introduction, Second, MIT Press Cambridge.

Stigter, T.Y., Van Ooijen, S.P.J., Post, V.E.A., Appelo, C.A.J. and Carvalho Dill, A.M.M. (1998). A hydrogeological and hydrochemical explanation of the groundwater composition under irrigated land in a Mediterranean environment, Algarve, Portugal. J Hydrol, 208, 262–279.

Sun, L., Yan, H., Xin, K. and Tao, T. (2019). Contamination source identification in water distribution networks using convolutional neural network. Environmental Science and Pollution Research, 26, 36786–36797.

T. Taulli. (2019). Artificial Intelligence Basics. Springer.

Taherion, M. (2006). Artificial neural network and its application in environmental engineering. First conference on environmental engineering., Tehran: Tehran Univ.

Vizintin, G., Souvent, P., Veselic, M. and Curk, B. (2009). Determination of urban groundwater pollution in alluvial aquifer using linked process models considering urban water cycle. J Hydrol, 377, 261–273.

Zhang, Z. and Sabuncu, M. (2018). Generalized cross-entropy loss for training deep neural networks with noisy labels. Advances in neural information processing systems, 31, 8778–8788.