International Journal of Physical Sciences Vol. 7(9), pp. 1516 - 1525, 23 February, 2012
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS11.1775
ISSN 1992 - 1950 ©2012 Academic Journals

Full Length Research Paper

Classification of cultivated soils in Al-Kharj, Saudi Arabia based on natural radionuclides using artificial neural networks

Saad A. Al-Hamed[1], Mohamed F. Wahby[1]*, Abdulwahed M. Aboukarima[2,3] and Ahmed A. Sayedahmed[1]

[1]Department of Agricultural Engineering, College of Food and Agriculture Sciences, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia.
[2]Agricultural Engineering Research Institute, Agricultural Research Center, Egypt.
[3]Community College, Huraimla, Shaqra University, P. O. Box 300, Huraimla 11962, Saudi Arabia.

Accepted 3 February, 2012

A great awareness of safe crop production is needed by selecting the suitable uncontaminated soil, and good seed types, etc. In this study, artificial neural network (ANN) was used to classify thirteen cultivated soils, in adjacent regions, in Al-Kharj Governorate, Saudi Arabia based on their natural radionuclide contents detected using gamma-ray spectrometer and geographical coordinates and altitude measured using GPS system. A total of 3497 patterns were collected. Specific activities (238-U, 40-K, 137-Cs and 232-Th in Bq/kg) were acted as inputs to an ANN classifier. In addition, geographical coordinates and altitude acted as inputs to the ANN classifier. The best predictive power for the classification of soils from the thirteen sites was achieved using 140 hidden neurons in the hidden layer of the ANN classifier. Most of soil data not included in the training data was correctly classified with an overall classification rate of 94.63%. The obtained results indicate that natural radionuclide contents detected by gamma-ray spectrometer and geographical coordinates and altitude measured by GPS in combination with ANN classifier is a viable tool for soil classification. The outcome of this research may have many benefits to environmental, soil and crop specialists in Saudi Arabia.

Key words: Soil, natural radionuclides, Saudi Arabia, artificial neural network, classification.

INTRODUCTION

A great awareness of safe crop production is needed by selecting the suitable production items like uncontaminated soil, good seed types, etc. Recently, high resolution sensor gamma-ray for detecting soil natural radionuclide contents is helping to make safe crop production a reality. These sensors are generating enormous amounts of data that can be processed, stored, and made available to the user community.

*Corresponding author. E-mail: wahby@ksu.edu.sa. Tel: 0096614678470.

Natural radioactive materials under certain conditions can reach hazardous radiological levels. So, it becomes necessary to study the natural radioactivity levels in the soil to assess the dose for the population in order to know the health risks and to have a baseline for future changes in the environmental radioactivity due to human activities (Singh et al., 2009).

Soils are formed by numerous processes of physical and chemical weathering. The classification of the soil is considered to be an important process due to the soil types must be the same in all locations during running agricultural activities for the results to be accurate. Soil classification deals with the systematic categorization of

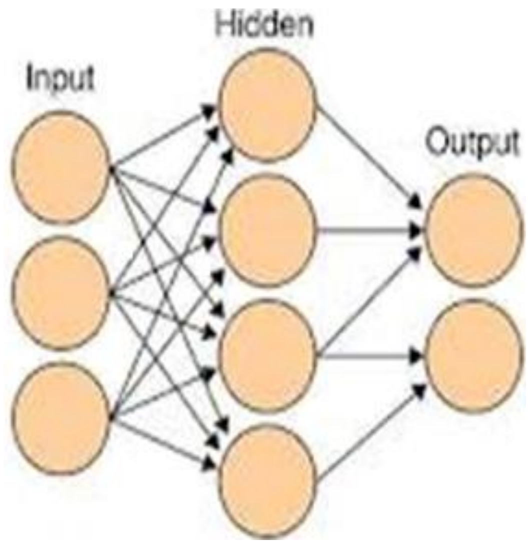

Figure 1. ANN structure.

soils based on distinguishing characteristics as well as criteria that state choices in use (Ramesh and Ramar, 2011).

A number of studies have been carried out on the application of data mining techniques for agricultural data sets. In the field of classification, artificial neural network (ANN), multivariate logistic regression, and multivariate discriminant analysis modeling could be building to perform the classification of soils from different criteria. However, soil classification was made more accurate, cheaper and easier with the implementation of ANN trained model (Omidiora et al., 2008).

ANN is a system based on the operation of biological neural networks that has different advantages (http://www.learnartificialneuralnetworks.com). When an element of the neural network fails, it can continue without any problem by their parallel nature.

A neural network learns and does not need to be reprogrammed. It can be implemented in any application without any problem. On the other hand, the neural network needs training to operate. The structure of a neural network has no specific rules for building. ANN is a best way to implement a solution due to their simplicity, design and universality. The real power of neural networks is evident when the trained network is able to produce good results for data which the network has never "seen" before (Ward System Group Inc., 2007).

The basic building block of ANN technology is the simulated neuron (depicted in Figure 1 as a circle). The network processes a number of inputs to produce an output (that is, the network's classifications). The neurons are connected by weights, (depicted as arrows in Figure 1).

Neurons are grouped into layers by their connection. In the input layer, neurons receive data. In output layer, a neuron contains the network's classifications. Neurons in between the input and output layers are in the hidden layer, which serves as a feature detector (Ward System Group Inc., 2007). The network begins by finding linear relationships between the inputs and the output. Weight values are assigned to the links between the input and output neurons. After those relationships are found, neurons are added to the hidden layer so that nonlinear relationships can be found. Input values in the first layer are multiplied by the weights and passed to the second (hidden) layer. Neurons in the hidden layer "fire" or produce outputs that are based upon the sum of weighted values passed to them. The hidden layer passes values to the output layer in the same fashion, and the output layer produces the desired results (predictions). The network "learns" by adjusting the interconnection weights between layers. The producing answers of the network are repeatedly compared with the correct answers, and each time the connecting weights are adjusted slightly in the direction of the correct answers. Additional hidden neurons are added as necessary to capture features in the data set (Ward System Group Inc., 2007).

Dragovic and Onjia (2006, 2007a, b) used different data analysis methods to recognize and classify soils of unknown geographic origin. A total of 103 soil samples collected from different regions in Serbia and Montenegro were differentiated into classes. Their radionuclide (226-Ra, 238-U, 235-U, 40-K, 134Cs, 137-Cs, 232-Th and 7-Be) were used as the inputs in different pattern recognition methods. For the classification of soil samples, the prediction ability of linear discriminant analysis, *k*-nearest neighbors, soft independent modeling of class analogy and ANN were 82.8, 88.6, 60.0 and 92.1%, respectively. When they used principal component analysis for classification, classification rate of 86.0% correctly classified samples was achieved.

This research has utilized existing data about soil natural radionuclide contents collected from scanned 13 soils in order to classify them. The soil properties which have been collected through project funded by the National Plan for Science, Technology and Innovation Program, King Saud University, Saudi Arabia provide a vast amount of information on the classification of soil based on natural radionuclide contents and geographical coordinates and altitude. The analysis of these soil data sets with ANN technique may yield outcomes useful to researchers in the agricultural, soil management and environment. The aim of this research is to use an ANN classifier to classify some cultivated soils in Al-Kharj Governorate, Saudi Arabia based on their natural radionuclide contents and geographical coordinates and altitude. The outcome of this research may have many benefits to environmental, soil and crop specialists in Saudi Arabia.

Table 1. Soil particle distribution, soil moisture content and soil bulk density of the scanned soils.

Soil code	Soil moisture content (%, db*)	Soil bulk density (g/cm^3)	Soil particle distribution			Soil type
			Sand (%)	Silt (%)	Clay (%)	
F1	1.82	1.53	82.2	9.9	7.9	Loamy sand
F2	4.58	1.72	83.2	9.8	7.0	Loamy sand
F3	10.41	1.55	86.4	8.8	4.8	Loamy sand
F4	10.90	1.57	86.2	8.6	5.2	Loamy sand
F5	5.18	1.62	75.3	16.7	8.0	Sandy loam
F6	1.54	1.61	78.0	16.6	5.4	Loamy sand
F7	10.28	1.67	75.3	15.2	9.5	Sandy loam
F8	5.45	1.52	75.3	16.7	8.0	Sandy loam
F9	4.33	1.78	71.8	17.2	11.0	Loamy sand
F10	3.87	1.66	72.3	17.6	10.1	Sandy loam
F11	5.36	1.72	85.7	7.3	7.0	Loamy sand
F12	8.14	1.56	77.3	16.7	6.0	Loamy sand
F13	5.96	1.73	76.7	16.1	7.2	Sandy loam

*db, Dry base.

MATERIALS AND METHODS

Soil sites

The soil sites are located at Al-Kharj, Saudi Arabia. Al-Kharj is a city and governorate in central Saudi Arabia. The city is located at around latitude of 24.148°N and longitude of 47.305°E. Some vegetables like carrots, cucumbers, tomatoes and lettuce are examples of produced vegetables. Date palm trees and oranges, melons and grapes are famous produced fruits in this area (http://en.wikipedia.org/wiki/Al-Kharj). Soil particle distribution, soil moisture content and soil bulk density of the scanned soils are shown in Table 1. Mean and standard deviation of latitude, longitude and altitude of the scanned soils are shown in Table 2.

Data collection and procedure

Gamma-ray spectrometer, the Mole (Egmond van et al., 2008) that is used commercially for high resolution mapping of physical and chemical soil properties is used to scan thirteen cultivated soils to detect natural radionuclides concentration in the surface layer. Latitude, longitude and altitude of all the scanned soils were determined using a global positioning system (Garmin GPS 60) which is a satellite based positioning and navigation system that provides position with accuracy less than 15 m. The Mole detector, GPS and laptop are placed on a fabricated iron carriage as illustrated in Figure 2. The carriage is pulled manually or by a vehicle over the field. The average scan speed for all soil sites is 1.43 m/s. The Mole system stores the data to enable post processing of them at a later stage using Gamman software. A calibration file is provided with the Mole system to complete data processing gathered from the scanned soils.

The Mole can scan soil surface layer in few seconds to get different radionuclides properties like 238-U, 40-K, 137-Cs and 232-Th in Bq/kg. The thirteen soils have been scanned during October 2011. Basic statistics of detected activity concentration of 40-K, 238-U, 232-Th and137-CS in the scanned thirteen soils are shown in Table 3.

Development of ANN classifier

The detected natural radionuclide contents in the scanned soils and geographical coordinates and altitude were processed with Gamman software and the data are converted to Excel spreadsheet. Then, the data were converted into comma delimited (CSV) format file for the Excel spread sheet and this file was used by NeuroShell Classifier for classification purposes. The software NeuroShell Classifier (release 3), developed by Ward System Group Inc. (2007) was used in this study. It is a neural network using a proprietary algorithm. Details of the structure of it were unavailable. The best hidden neurons were selected based on the highest overall accuracy for the training data set. Two random data sets were created and the dataset was grouped as 2790 and 707 patterns for training and testing purposes, respectively. Seven features (238-U, 40-K, 137-Cs, 232-Th, latitude, longitude and altitude were used as inputs to the developed ANN classifier.

Evaluation of ANN classifier performance

Classification is learning a function that maps (classifies) a data item into one of several predefined classes. The aim of the classification task is to discover some kind of relationship between the input attributes and the output class, so that the discovered knowledge can be used to predict the class of a new unknown object (Growe, 1999). NeuroShell Classifier has some performance criteria to evaluate the classifier. They are true-positive ratio, false-positive ratio, true-negative ratio, and false-negative ratio. The classifier outcome was compared with the known visual class, and performance of the classifier was judged based on accuracy of prediction. According to Shahine et al. (2002), classification accuracy can be calculated as the following:

$$Classification\ Accuracy\ (\%) = 100 \times \frac{number\ of\ correct\ predictions}{total\ number\ of\ actual} \quad (1)$$

Table 2. Mean and standard deviation of latitude, longitude and altitude of the scanned soils.

Soil code	Mean			Standard deviation		
	Latitude (°N)	Longitude (°E)	Altitude (m)	Latitude (°N)	Longitude (°E)	Altitude (m)
F1	24.324	47.131	465.459	0.00044	0.00029	3.698
F2	24.329	47.131	464.904	0.00034	0.00031	1.428
F3	24.182	47.220	447.016	0.00023	0.00029	1.865
F4	24.186	47.218	455.092	0.00036	0.00020	0.697
F5	24.255	47.257	444.791	0.00012	0.00061	0.949
F6	24.255	47.254	440.903	0.00007	0.00029	0.689
F7	24.253	47.271	444.975	0.00043	0.00011	1.193
F8	24.271	47.236	450.983	0.00004	0.00040	1.731
F9	24.208	47.570	399.956	0.00043	0.00007	1.900
F10	24.210	47.573	395.311	0.00029	0.00010	2.183
F11	24.204	47.562	401.684	0.00004	0.00036	0.778
F12	24.200	47.239	442.643	0.00010	0.00008	1.079
F13	24.196	47.235	446.192	0.00013	0.00017	0.819

Figure 2. The mole detector, GPS and laptop are placed on a fabricated carriage and pulled by a vehicle over the field.

RESULTS AND DISCUSSION

From Table 3, the maximum specific activity of 40-K, 137-CS, 232-Th and 238-U are 340.61, 17.12, 38.67 and 35.05 Bq/kg, respectively. Simple correlations as shown in Table 4 were analyzed to determine the relationships among soil characteristics and mean specific activities detected in the scanned soils. The results showed weak negative associations between all pairs of 40-K, 238-U, 137-Cs and clay content as shown by their negative correlations (r = -0.274, -0.095 and -0.175), respectively. The 232-Th concentration, detected in the investigated soils, showed a moderate negative correlation (r = - 0.667) with clay content. Meanwhile, it is positively proportional

Table 3. Basic statistics of specific activity concentration of 40-K, 238-U, 232-Th and137-CS in scanned thirteen soils.

Basic statistics	Unit	40-K	238-U	232-Th	137-CS
Mean	Bq/kg	157.20	17.50	14.68	4.45
Minimum	Bq/kg	18.02	2.90	2.41	0.005
Maximum	Bq/kg	340.61	35.05	38.67	17.12
Standard deviation	Bq/kg	52.68	4.60	5.25	3.02
Coefficient of variation	%	33.51	26.27	35.79	67.88
No. of data	-	3497	3497	3497	3497

Table 4. Correlation coefficients among soil characteristics and mean specific activities detected in the scanned soils.

Parameter	Soil moisture content	Soil bulk density	Clay	Silt	Sand	40-K	238-U	232-Th	137-Cs
Soil moisture content	1								
Soil bulk density	-0.174	1							
Clay	-0.300	0.492	1						
Silt	-0.267	0.114	0.500	1					
Sand	0.314	-0.267	-0.748	-0.949	1				
40-K	0.433	-0.348	-0.274	-0.028	0.121	1			
238-U	0.150	-0.191	-0.095	-0.011	0.043	0.559	1		
232-Th	0.632	-0.464	-0.667	-0.535	0.654	0.518	0.365	1	
137-Cs	0.464	-0.488	-0.175	-0.001	0.065	0.609	0.131	0.315	1

Table 5. Correlation coefficients among inputs to ANN classifier.

Parameter	Latitude	Longitude	Altitude
Latitude	1		
Longitude	-0.476	1	
Altitude	0.508	-0.992	1
40-K	-0.370	-0.346	0.339
238-U	-0.525	0.189	-0.178
232-Th	-0.529	-0.231	0.231
137-Cs	0.045	-0.463	0.496

to the sand content. Also, weak positive associations between all pairs of 40-K, 238-U, 137-Cs and sand content as shown by their positive correlations (r = 0.121, 0.043 and 0.065), respectively. Positive associations between all pairs of 40-K, 238-U, 137-Cs and 232-Th in soil moisture content are shown by their positive correlations with r = 0.433, 0.150, 0.632 and 0.464, respectively. Detected activity of natural 40-K is proportional to the activity of 238-U, 232-Th and 137-Cs with moderate positive association (r = 0.559, 0.518, 0.609), respectively. Weak positive association between association between 238-U and 137-Cs (r = 0.131) was seen. Weak positive association between 232-Th and 137-Cs (r = 0.315) was noticed.

Table 5 illustrates correlation coefficients between seven inputs to ANN classifier. The results showed moderate negative associations between all pairs of 40-K, 238-U, 232-Th and latitude as shown by their negative correlations (r = -0.370, -0.525 and -0.529), respectively. The 137-Cs concentration, detected in the investigated soils, showed an increasing tendency with altitude. This result is in agreement with findings by Kubica et al. (2007). Demonstrated increase in 137-Cs activity in correlation to altitude might be elucidated due to the humic development level (Kubica et al., 2007). Furthermore, for 40-K, the changes in its specific activity detected in the scanned sites in respect to the longitude were noticed with weak negative correlation (r = -0.346). For 238-U, the changes in its specific activity detected in the scanned sites in respect to the longitude were noticed with weak positive correlation (r = 0.189). For 232-Th, the changes in its specific activity detected in the scanned

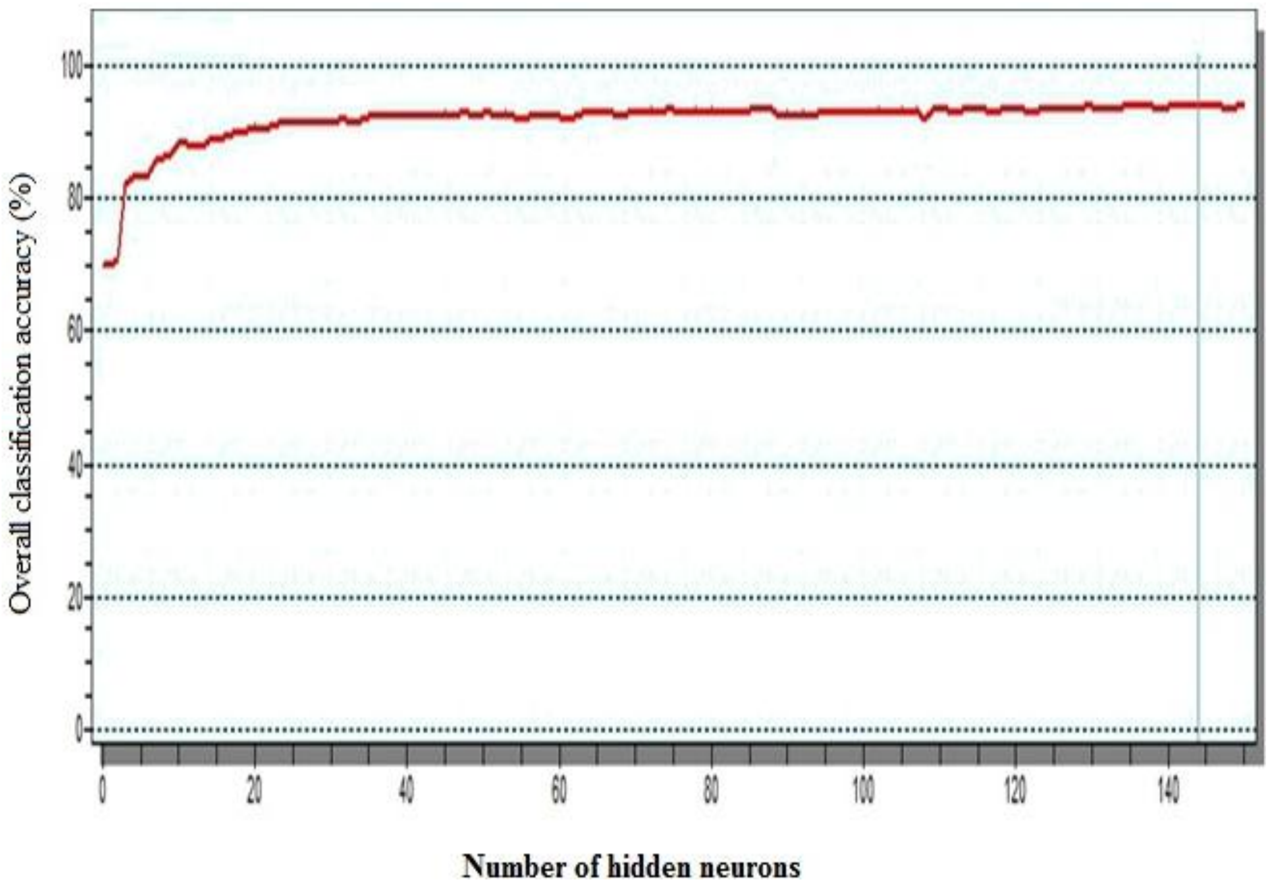

Figure 3. The obtained training graph from neuroshell classifier using training data set.

sites in respect to the longitude were noticed with weak negative correlation (r = -0.231). For 137-Cs, the changes in its specific activity detected in the scanned sites in respect to the longitude were noticed with weak negative correlation (r = -0.463). Specific activities of 40-K, 232-Th and 137-Cs are correlated with altitude with weak positive correlation (r = 0.339, 0.231, 0.496), respectively. Also, 238-U is correlated with altitude with weak negative correlation (r = -0.178).

The NeuroShell Classifier allows building of a powerful classification model quickly. In this research, data are randomized and 80% of the data (2790 rows) were randomly selected for training the classifier and the rest 20% (707 rows) for testing. Number of hidden neurons trained was 150 and the optimal number of the hidden neurons was 144. Figure 3 depicts the obtained training graph from NeuroShell Classifier using training data set. The best overall classification accuracy was 94.63% (2631/2790*100) for training data set. Agreement matrix (sometimes also called confusion matrix or contingency table), true-positive, false positive, true-negative and false-negative ratios, sensitivity and specificity values are displayed in Table 6 for training data.

In agreement matrix, numbers appearing on the diagonals of the matrix corresponded to correctly classified cases, while off-diagonal entries represent misclassification. A matrix table makes it easy to identify where the misclassifications are occurring. The training sample classification rate indicates the model performance when the model is build with seen data values. This gives an idea if a given data set can be classified with ANN based approach. As indicated in Table 6, the summary result of ANN classifier using all seven indicators together showed that from the total train examples of 323 instances, 323 were correctly classified for F1 soil code. For F10 soil code, examples of 246 instances, 224 were correctly classified and 22 instances was misclassified and classified as F9 soil code. The percentage of correctly classified instances for each soil code in training data set was shown in Figure 4.

It was clear from Figure 4 that, the percentage of correctly classified instances for each soil in training data ranged between 80.48 and 100%. According to the results of network training, the network has successfully captured the relationship between the input parameters and output. The network with the specified structure gives

Table 6. Agreement matrix summary of randomly of training data.

Soil code		Actual													Total
		F1	**F10**	**F11**	**F12**	**F13**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	
Classified as	F1	323	0	0	0	0	0	0	0	0	0	0	0	0	323
	F10	0	224	0	0	0	0	0	0	0	0	0	0	22	246
	F11	0	0	132	0	0	0	0	0	0	0	0	0	23	155
	F12	0	0	0	173	2	0	0	0	0	1	0	0	0	176
	F13	0	0	0	11	177	0	2	0	0	0	0	0	0	190
	F2	0	0	0	0	0	256	0	0	0	0	0	0	0	256
	F3	0	0	0	1	2	0	222	0	0	0	0	0	0	225
	F4	0	0	0	0	0	0	0	200	0	0	0	0	0	200
	F5	0	0	0	0	0	0	0	0	216	19	13	0	0	248
	F6	0	0	0	0	0	0	0	0	0	77	0	0	0	77
	F7	0	0	0	0	0	0	0	0	12	0	185	0	0	197
	F8	0	0	0	0	0	0	0	0	2	0	0	244	0	246
	F9	0	3	46	0	0	0	0	0	0	0	0	0	202	251
Total		323	227	178	185	181	256	224	200	230	97	198	244	247	2790
True positive ratio		1	0.9868	0.7416	0.9351	0.9779	1	0.9911	1	0.9391	0.7938	0.9343	1	0.8178	
False positive ratio		0	0.0086	0.0088	0.0012	0.005	0	0.0012	0	0.0125	0	0.0046	0.0008	0.0193	
True negative ratio		1	0.9914	0.9912	0.9988	0.995	1	0.9988	1	0.9875	1	0.9954	0.9992	0.9807	
False negative ratio		0	0.0132	0.2584	0.0649	0.0221	0	0.0089	0	0.0609	0.2062	0.0657	0	0.1822	
Sensitivity (%)		100	98.68	74.16	93.51	97.79	100	99.11	100	93.91	79.38	93.43	100	81.78	
Specificity (%)		100	99.14	99.12	99.88	99.50	100	99.88	100	98.75	100	99.54	99.92	98.07	

reliable results during the training stage. Based on the obtained results, a soil can be classified based on its natural radionuclide contents detected using gamma-ray spectrometer and geographical coordinates and altitude measured using GPS system. The ANN classifier could be very useful helper for environmental and soil specialists. In this research, some soils were incorrectly assigned to one or two wrong classes. This misidentification of the soils indicates that other soil characteristics may influence the pattern of specific activities of the four detected natural radionuclides. The true-positive ratios (also called sensitivity) for all soils are illustrated in Table 6 for training data set. However, sensitivity is usually expressed as percentage, ranging from 0% (very bad classification) to 100% (perfect classification). The sensitivity values for soil codes F1, F10, F11, F12, F13, F2, F3, F4, F5, F6, F7, F8 and F9 were 100, 98.68, 74.16, 93.51, 97.79,100, 99.11, 100, 93.91, 79.38, 93.43, 100 and 81.78%, respectively in training data set. The true-negative ratios (also called specificity) for all soils are illustrated in Table 6 for training data set.

However, specificity is usually expressed as percentages, ranging from 0% (very bad classification) to 100% (perfect classification).

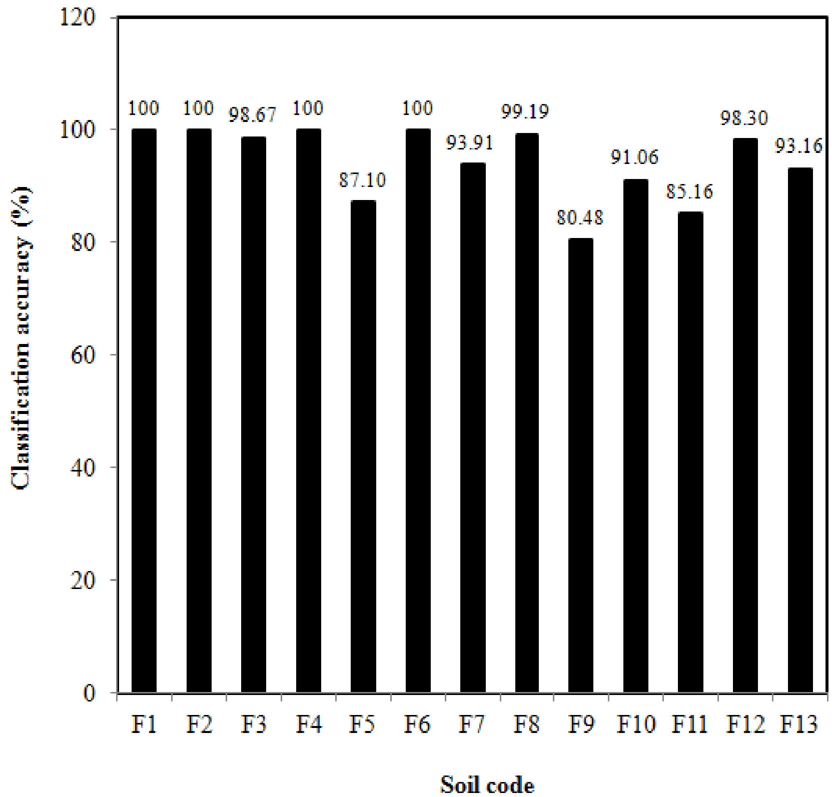

Figure 4. The percentage of correctly classified instances for each soil code in training data set.

The specificity values for soil codes F1, F10, F11, F12, F13, F2, F3, F4, F5, F6, F7, F8 and F9 were 100, 99.14, 99.12, 99.88, 99.50, 100, 99.88, 100, 98.75, 100, 99.54, 99.92 and 98.07%, respectively in training data set.

The developed ANN classifier model still needs to be tested with the unseen data. Hence, the testing sample classification rate represents the model performance for unseen data. The best overall classification accuracy was 94.63% (669/707*100) for testing data set. Agreement matrix, true-positive, false positive, true-negative and false-negative ratios, sensitivity and specificity values are displayed in Table 7 for testing data. As indicated in Table 7, the summary result of neural classifier using all seven indicators together showed that from the total test examples of 63 instances, 53 were correctly classified for F5 soil code, 2 instances was misclassified and classified as F6 soil code and 8 instances was misclassified and classified as F7 soil code with classification accuracy of 84.13 (53/63*100). For F8 soil code, examples of 55 instances, 55 were correctly classified with classification accuracy of 100%. The percentage of correctly classified instances for each soil in testing data set was shown in Figure 5. It was clear that, the percentage of correctly classified instances for each soil code in testing data ranged between 84.13 and 100%. The higher the percentage of correctly classified the better the model's predictive power.

The sensitivity values for soil codes F1, F10, F11, F12, F13, F2, F3, F4, F5, F6, F7, F8 and F9 were 100, 98.21, 79.49, 93.33, 98.11, 100, 94.55, 100, 92.98, 90, 82.61, 100 and 86.21%, respectively in testing data set. The specificity values for soil codes F1, F10, F11, F12, F13, F2, F3, F4, F5, F6, F7, F8 and F9 were 100, 99.39, 99.40, 99.85, 99.24, 100, 99.85, 100, 98.46, 100, 99.39, 100 and 98.61%, respectively in testing data set.

Conclusion

The overall correct classification performance was over 92% in this research for different training/testing data sets indicating that the detected specific activities of natural radionuclide contents of 238-U, 40-K, 137-Cs and 232-Th in Bq/kg and geographical coordinates and altitude with ANN is a viable tool for soil classification in adjacent regions in Al-Kharj Governorate, Saudi Arabia. Some soils were incorrectly assigned to one or two wrong classes. This misidentification of the soils indicates that

Table 7. Agreement matrix summary of randomly tested data.

Soil code		Actual													Total
		F1	F10	F11	F12	F13	F2	F3	F4	F5	F6	F7	F8	F9	
Classified as	F1	86	0	0	0	0	0	0	0	0	0	0	0	0	86
	F10	0	55	0	0	0	0	0	0	0	0	0	0	4	59
	F11	0	0	31	0	0	0	0	0	0	0	0	0	4	35
	F12	0	0	0	42	1	0	0	0	0	0	0	0	0	43
	F13	0	0	0	2	52	0	3	0	0	0	0	0	0	57
	F2	0	0	0	0	0	62	0	0	0	0	0	0	0	62
	F3	0	0	0	1	0	0	52	0	0	0	0	0	0	53
	F4	0	0	0	0	0	0	0	75	0	0	0	0	0	75
	F5	0	0	0	0	0	0	0	0	53	2	8	0	0	63
	F6	0	0	0	0	0	0	0	0	0	18	0	0	0	18
	F7	0	0	0	0	0	0	0	0	4	0	38	0	0	42
	F8	0	0	0	0	0	0	0	0	0	0	0	55	0	55
	F9	0	1	8	0	0	0	0	0	0	0	0	0	50	59
Total		86	56	39	45	53	62	55	75	57	20	46	55	58	707
True positive ratio		1	0.9821	0.7949	0.9333	0.9811	1	0.9455	1	0.9298	0.9	0.8261	1	0.8621	
False positive ratio		0	0.0061	0.006	0.0015	0.0076	0	0.0015	0	0.0154	0	0.0061	0	0.0139	
True negative ratio		1	0.9939	0.994	0.9985	0.9924	1	0.9985	1	0.9846	1	0.9939	1	0.9861	
False negative ratio		0	0.0179	0.2051	0.0667	0.0189	0	0.0545	0	0.0702	0.1	0.1739	0	0.1379	
Sensitivity (%)		100	98.21	79.49	93.33	98.11	100	94.55	100	92.98	90	82.61	100	86.21	
Specificity (%)		100	99.39	99.40	99.85	99.24	100	99.85	100	98.46	100	99.39	100	98.61	

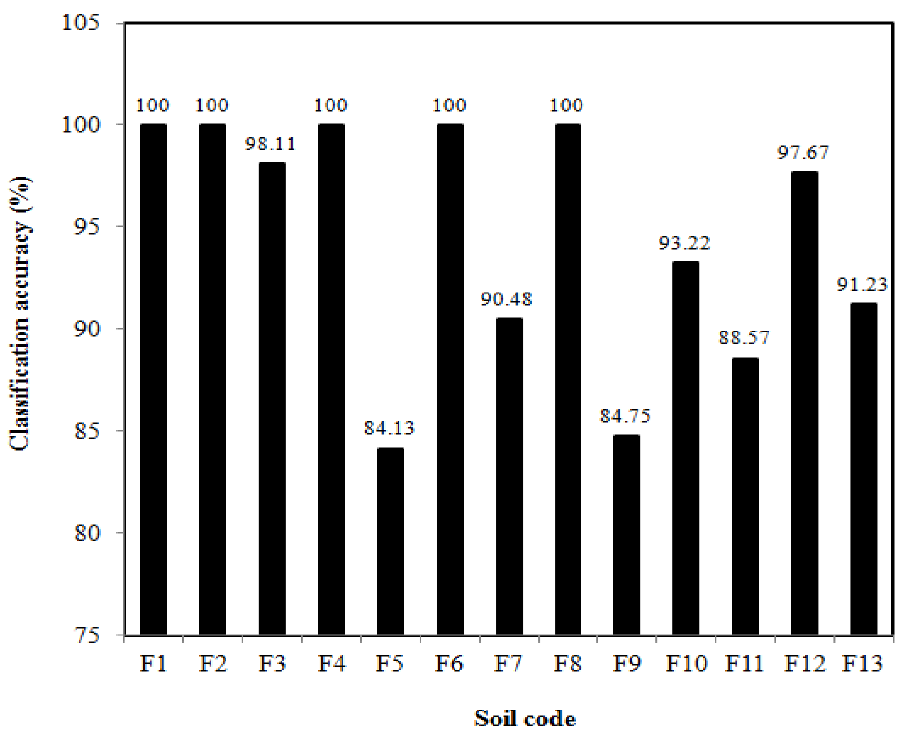

Figure 5. The percentage of correctly classified instances for each soil code in testing data set.

other soil properties may affect the pattern of specific activities of the four detected natural radionuclides.

ACKNOWLEDGMENTS

Authors express their gratitude to the National Plan for Science, Technology and Innovation Program, King Saud University, Saudi Arabia for financially supporting this research effort as a part of the project entitled "Modeling of energy consumption during seed bed preparation operation based on soil mechanical properties", No. 09-SPA876-02.

REFERENCES

Dragovic S, Onjia A (2006). Classification of soil samples according to their geographic origin using gamma-ray spectrometry and principal component analysis. J. Environ. Radioactivity, 89: 150-158.

Dragovic S, Onjia A (2007a). Classification of soil samples according to geographic origin using gamma-ray spectrometry and pattern recognition methods. Appl. Rad. Isotopes, 65: 218-224.

Dragovic S, Onjia A (2007b). Artificial neural network data analysis for classification of soils based on their radionuclide content. Russian J. Phys. Chem. A, 81: 1477-1481.

Egmond van FM, Loonstra EH, Limburg J (2008). Gamma-ray sensor for topsoil mapping: The Mole. 1st Global workshop on High Resolution Digital Soil Sensing and Mapping, Sydney Australia, 10 p.

Growe GA (1999). Comparing algorithms and clustering data: components of the data mining process. MSc. Thesis, Department of Computer Science and Information Systems, Grand Valley State University, Mackinac Hall, Allendale, Michigan, USA.

Kubica B, Kwiatek WM, Stobiński M, Skiba S, Skiba M, Gołaś J, Kubica M, Tuteja-Krysa, M, Wrona A, Misiak R, Krzan Z (2007). Concentrations of 137-Cs, 40-K radionuclides and some heavy metals in soil samples of Chochołowska Valley from Tatra National Park. Polish J. Environ. Stud., 16: 723-729.

Omidiora EO, Oyediran GO, Olabiyisi SO, Arulogun OT (2008). Classification of soils of Central Western Nigeria using neural network rule extraction and decision table. Agric. J., 3: 305-312.

Ramesh V, Ramar K (2011). Classification of agricultural land soils a data mining approach. Int. J. Comput. Sci. Eng. (IJCSE), 3: 379-384.

Shahine MA, Tollner EW, Gitaitis RD, Sumner DR, Maw BW (2002). Classification of sweet onions based on internal defects using image processing and neural networks techniques. Trans. ASAE, 45: 1613-1618.

Singh J, Singh H, Singh S, Bajwa BS, Sonkawade RG (2009). Comparative study of natural radioactivity levels in soil samples from the Upper Siwaliks and Punjab, India using gamma-ray spectrometry. J. Environ. Radioactivity, 100: 94-98.

Ward System Group Inc. (2007). NeuroShell classifier, Release 3.0. Ward System Group Inc., Exective Park, West, 5 Hillcreat Drive, Frederic, MD 21703.