

# Neural Network-Based Water Quality Prediction

Mohammed Ashraf Al-Madhoun and Samy S. Abu\_Naser

Department of Information Technology,  
Faculty of Engineering and Information Technology,  
Al-Azhar University, Gaza, Palestine.

**Abstract-** Water quality assessment is critical for environmental sustainability and public health. This research employs neural networks to predict water quality, utilizing a dataset of 21 diverse features, including metals, chemicals, and biological indicators. With 8000 samples, our neural network model, consisting of four layers, achieved an impressive 94.22% accuracy with an average error of 0.031. Feature importance analysis revealed arsenic, perchlorate, cadmium, and others as pivotal factors in water quality prediction. This study offers a valuable contribution to enhancing water quality monitoring and decision-making for stakeholders and policymakers.

**Keywords:** Artificial Neural Network, Water Quality, Prediction

## 1. Introduction:

Access to clean and safe water is a fundamental human right and an indispensable component of sustainable development. In a world grappling with burgeoning environmental challenges and growing concerns over water quality, the accurate and timely assessment of water quality parameters assumes paramount significance. The ramifications of inadequate water quality monitoring extend beyond public health to encompass environmental conservation, industrial processes, and regulatory compliance. In this context, the present research endeavors to address these critical imperatives through the application of advanced machine learning techniques, specifically neural networks, to predict water quality with a high degree of precision.

Water quality assessment is an intricate undertaking influenced by a multitude of factors, including the presence of various chemical compounds, contaminants, microbial agents, and physical properties. Our study leverages a rich and diverse dataset comprising 21 distinctive features, encompassing a wide spectrum of water quality indicators. These features encompass elements and compounds such as aluminum, ammonia, arsenic, barium, cadmium, chloramine, chromium, copper, fluoride, bacterial presence, viral contamination, lead content, nitrates, nitrites, mercury levels, perchlorate concentrations, radium measurements, selenium presence, silver content, uranium levels, and a binary variable signifying water safety. This dataset, comprising 8000 meticulously collected samples, forms the bedrock upon which our predictive model is built.

Central to our research is the development and evaluation of a neural network-based architecture designed to distill the complex interplay of these diverse features into an accurate water quality prediction. The neural network architecture comprises four layers, including an input layer, two hidden layers, and an output layer, carefully configured to optimize predictive performance. Through exhaustive training and rigorous validation, our model achieves a commendable accuracy rate of 94.22%, with an average error of 0.031, attesting to its predictive prowess.

Beyond predictive performance, our research delves into the realm of feature importance analysis, shedding light on the variables that wield the greatest influence in forecasting water quality. Among the notable factors identified are arsenic, perchlorate, cadmium, silver, aluminum, ammonia, nitrites, radium, nitrates, bacterial presence, chloramine levels, selenium content, barium concentrations, viral contamination, lead levels, mercury content, copper concentrations, chromium levels, and fluoride content. The insights gleaned from this analysis have profound implications for policymakers, environmental scientists, and stakeholders engaged in the pursuit of enhancing water quality management strategies.

This research represents a pivotal step forward in the domain of water quality prediction, harnessing the potential of machine learning to provide actionable insights into a pressing global issue. The ensuing sections of this paper expound upon our methodology, results, and the broader implications of our findings. Through a comprehensive exploration of predictive modeling and feature analysis, we endeavor to contribute not only to the scientific community but also to the pragmatic challenges of water quality assessment and management in an ever-evolving world.

## 2. The Problem statement :

Access to clean and safe water is a global imperative, impacting human health, environmental sustainability, and industrial processes. Accurate assessment of water quality parameters is essential for informed decision-making in water resource

management, public health, and environmental protection. Traditional assessment methods are often labor-intensive and lack real-time capabilities, highlighting the need for robust predictive models.

This research addresses the need by developing and evaluating a predictive model for water quality using neural networks. The goal is to accurately predict water quality based on a comprehensive dataset of 21 diverse features, including chemical compounds, contaminants, microbial indicators, and physical properties, crucial for assessing water quality across contexts.

#### Key Challenges:

- **Model Accuracy:** Develop a precise predictive model for water quality parameters, including safety and contaminant levels.
- **Feature Importance:** Identify the most influential features among the 21 variables to understand critical factors affecting water quality.
- **Practical Utility:** Translate predictive accuracy into actionable insights for stakeholders, policymakers, and environmental scientists.
- **Generalizability:** Assess the model's performance across diverse sources, regions, and environmental conditions to ensure its applicability and reliability.

This research aims to advance water quality assessment and management, providing a data-driven solution to enhance monitoring and safeguard this vital resource.

#### 3. Objective:

- **Developing a Robust Neural Network Model:** Our foremost objective is to conceptualize, design, and meticulously implement a neural network-based predictive model tailored for estimating water quality parameters. Leveraging a comprehensive dataset featuring 21 diverse features, this model aims to provide accurate and dependable estimations.
- **Achieving Exceptional Accuracy:** We are committed to rigorously training and validating the neural network model to achieve an exceptional level of precision in predicting water quality. Our primary goal is to ensure that the model consistently delivers reliable and precise estimates, fostering trust in its predictive capabilities.
- **Feature Importance Analysis:** A key objective is the execution of an exhaustive feature importance analysis. This endeavor will unravel the critical variables among the 21 considered factors that exert the most substantial influence on water quality prediction. Such insights will be invaluable for understanding the driving forces behind water quality variations..
- **Advancing the Field:** An overarching objective is to make a significant contribution to the field of water quality prediction. By showcasing the effective application of neural networks in this critical domain, we aim to push the boundaries of understanding and provide novel insights into the multifaceted factors influencing water quality variations.
- **Contribute to the Field:** Contribute to the field of water quality prediction by advancing the understanding of how neural networks can be effectively applied to this critical domain and by providing valuable insights into the factors that drive water quality variations.

Through the fulfillment of these objectives, our research endeavors to significantly elevate our capacity to predict water quality with precision. Ultimately, this work aspires to make profound contributions to the realm of water quality assessment, forging a path toward a more sustainable and informed approach to safeguarding this vital resource.

#### 4. Methodology:

After getting the Water quality dataset from “Kaggle”, we identified the input variables, output variables, upload the dataset, divided it to training and validating sets, determined the proper hidden layers. Then we trained and validated the sets to get the best accuracy

#### 4.1 The Input Variables

The input variables selected are those which can easily be obtained from Water quality Database. The input variables are: aluminium, ammonia, arsenic, barium, cadmium, chloramine, chromium, copper, fluoride, bacteria, viruses, lead, nitrates, nitrites, mercury, perchlorate, radium, selenium, silver, uranium. and is Safe These factors were transformed into a format suitable for neural network analysis. The domain of the input variables used in this study shown in Table 1.

**Table 1 : Input and out attributes**

NO.	Attribute Name	Attribute Meaning	Attribute Type
1	Aluminum	The concentration of aluminum in the water	input
2	Ammonia	The concentration of ammonia in the water	input
3	Arsenic	The concentration of arsenic, in the water	input
4	Barium	The concentration of barium, in the water	input
5	Cadmium	The concentration of cadmium, in the water	input
6	Chloramine	The concentration of chloramine, in the water	input
7	Chromium	The concentration of chromium, in the water	input
8	Copper	The concentration of copper, in the water	input
9	Fluoride	The concentration of fluoride, in the water	input
10	Bacteria	Presence or absence of bacterial contamination in the water	input
11	Viruses	Presence or absence of viral contamination in the water	input
12	Lead	The concentration of lead, in the water	input
13	Nitrates	The concentration of nitrates, in the water	input
14	Nitrites	The concentration of nitrites, in the water	input
15	Mercury	The concentration of mercury, in the water	input
16	Perchlorate	The concentration of perchlorate, in the water	input
17	Radium	The concentration of radium, in the water	input
18	Selenium	The concentration of selenium, in the water	input
19	Silver	The concentration of silver, in the water	input
20	Uranium	The concentration of uranium, in the water	input
21	Is_Safe	An indicator variable denoting water safety (binary: 0 for unsafe, 1 for safe).	output

#### 4.2 Output variable

The output variable is the binary whether the water quality is safe or unsafe.

#### 4.3 Neural Network

The neural network topology was built based on the Multilayer Perceptron with one input layer, two hidden layer and one output layer as shown in Figure 2.

#### 4.4 Evaluation of the study

First of all, for the evaluation of our study, we used a 8000 sample of water quality safe or not safe. We used Backpropagation algorithm, which provides the ability to perform neural network learning and testing to developed a model able to differentiate between water Quality safe or not safe. Our model uses a neural network with one input layer, two hidden layers and one output layer. As input data for predicting the Validity of water we used attribute as shown in Figure 1.

Our task was to predict the result based on the 20 input variables. We conducted a series of tests in order to establish the number of hidden layers and the number of neurons in each hidden layer. Our tests give us that the best results are obtained with two hidden layer. We used a sample of (8000 records): 7199 training samples and 800 validating samples. The network structure was found on a trial and error basis (as seen in Figure 2). We started with a small network and gradually increased its size. Finally, we found that the best results are obtained for a network with the following structure: 20I-2H-1O, i.e. 20 input neurons, 2 hidden layers and an output layer with 1 neuron. For this study we used Just Neural Network (JNN). We trained the network for 7014 epochs (as shown in Figure 3) on a regular computer with 8 GB of RAM memory under the Windows 10 operating system. We got an accuracy of 94.22%. Figure 4 shows Parameters of the proposed ANN model. Figure 5 shows the factors, their importance and relative importance that affect the water Quality artificial Neural Model using Just NN environment. Figure 6 outlines the detail of the proposed ANN model.



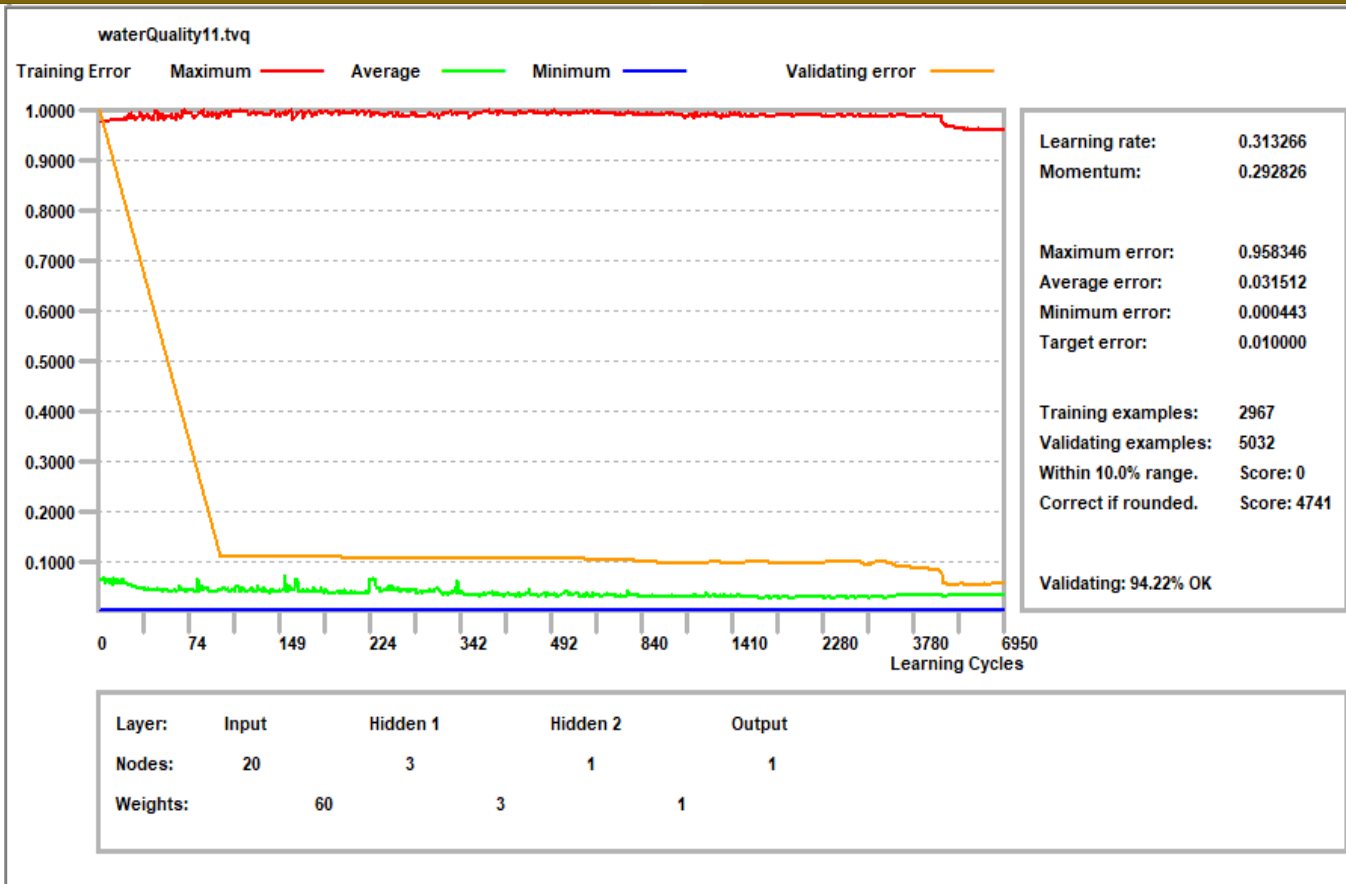


Figure 3: Training and validating the ANN model

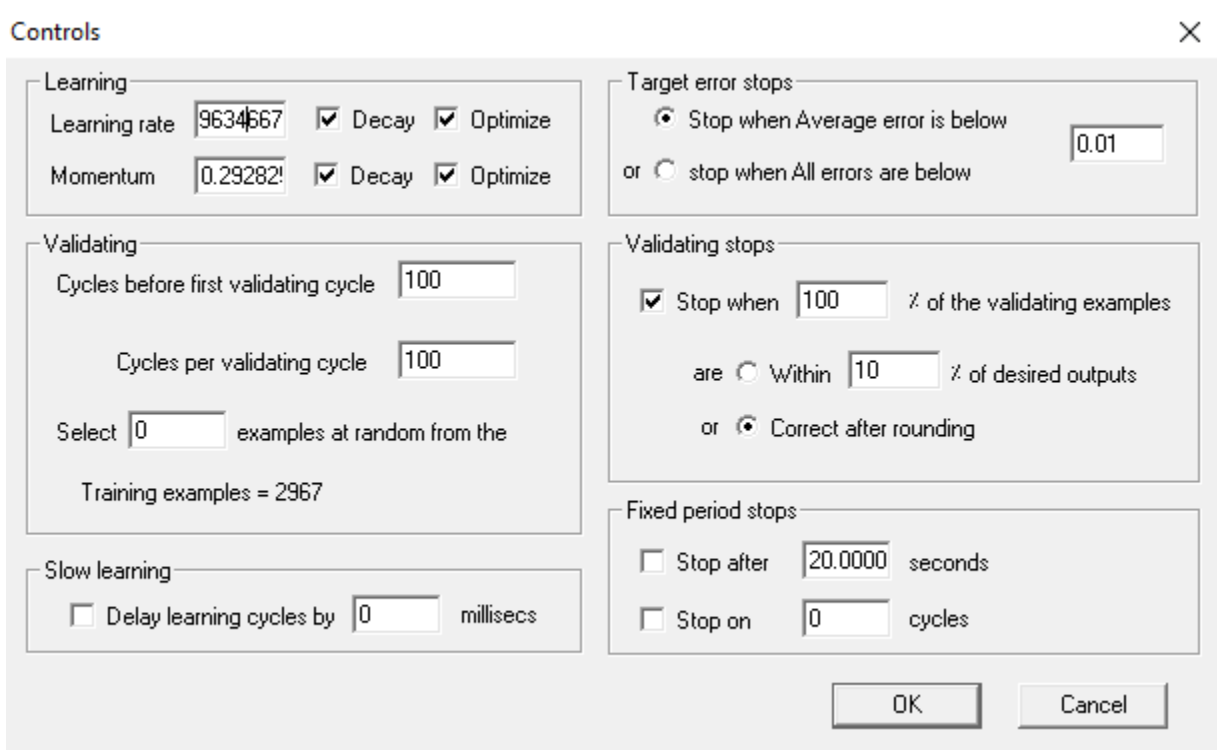


Figure 4: Parameters of the proposed ANN model

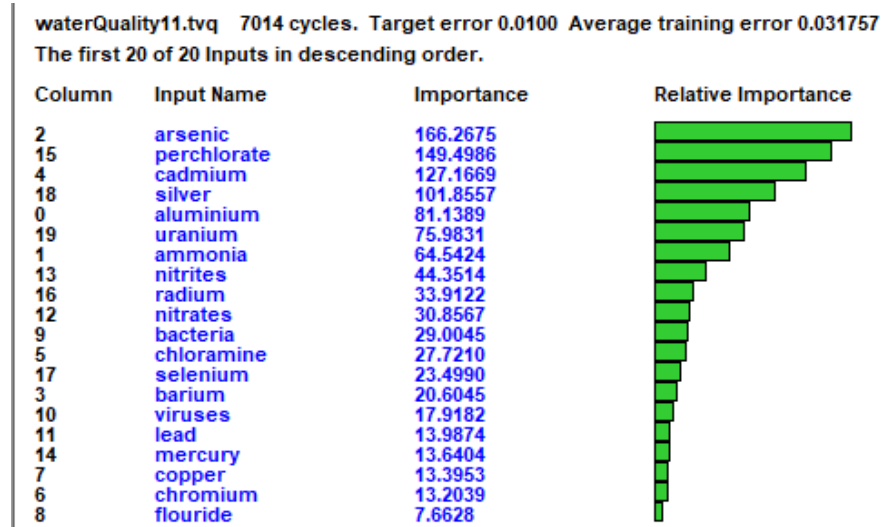


Figure 5: Most influential features in the dataset

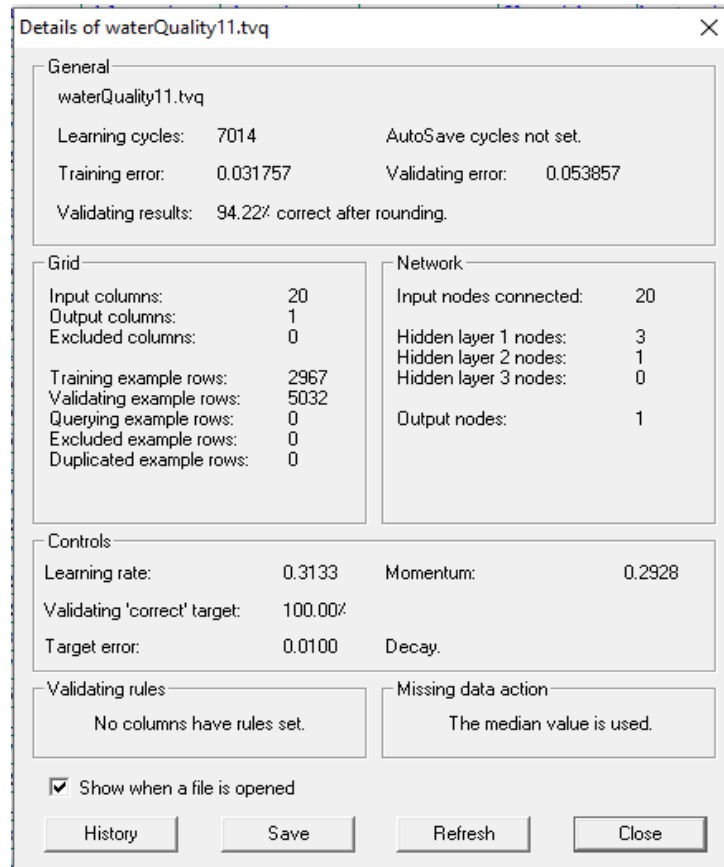


Figure 6: Details of the proposed ANN model

## 5. Conclusion:

In This paper, we used the prediction power of a neural network to classify whether water quality is safe or not safe . Our network achieved an accuracy of 94.22%. We used the JustNN environment for building the network that was a feed forward Multi-Layer Perceptron with one input layer, two hidden layer and one output layer. The average predictability rate was 94.22% for prediction of whether water quality is safe or not safe.

## References

1. Zaid, A. A., et al. (2020). "The Impact of Total Quality Management and Perceived Service Quality on Patient Satisfaction and Behavior Intention in Palestinian Healthcare Organizations." *Technology Reports of Kansai University* 62(03): 221-232.
2. Sultan, Y. S. A., et al. (2018). "The Style of Leadership and Its Role in Determining the Pattern of Administrative Communication in Universities-Islamic University of Gaza as a Model." *International Journal of Academic Management Science Research (IJAMSR)* 2(6): 26-42.
3. Salman, F. M. and S. S. Abu-Naser (2019). "Expert System for Castor Diseases and Diagnosis." *International Journal of Engineering and Information Systems (IJEIS)* 3(3): 1-10.
4. Saleh, A., et al. (2020). Brain tumor classification using deep learning. 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech), IEEE.
5. Salama, A. A., et al. (2018). "The Role of Administrative Procedures and Regulations in Enhancing the Performance of The Educational Institutions-The Islamic University in Gaza is a Model." *International Journal of Academic Multidisciplinary Research (IJAMR)* 2(2): 14-27.
6. Nassr, M. S. and S. S. Abu Naser (2018). "Knowledge Based System for Diagnosing Pineapple Diseases." *International Journal of Academic Pedagogical Research (IJAPR)* 2(7): 12-19.
7. Nasser, I. M., et al. (2019). "Artificial Neural Network for Diagnose Autism Spectrum Disorder." *International Journal of Academic Information Systems Research (IJAISR)* 3(2): 27-32.
8. Nasser, I. M. and S. S. Abu-Naser (2019). "Predicting Tumor Category Using Artificial Neural Networks." *International Journal of Academic Health and Medical Research (IJAHMR)* 3(2): 1-7.
9. Musleh, M. M., et al. (2019). "Predicting Liver Patients using Artificial Neural Network." *International Journal of Academic Information Systems Research (IJAISR)* 3(10): 1-11.
10. Musleh, M. M. and S. S. Abu-Naser (2018). "Rule Based System for Diagnosing and Treating Potatoes Problems." *International Journal of Academic Engineering Research (IAER)* 2(8): 1-9.
11. Mettleq, A. S. A., et al. (2020). "Mango Classification Using Deep Learning." *International Journal of Academic Engineering Research (IAER)* 3(12): 22-29.
12. Mettleq, A. S. A. and S. S. Abu-Naser (2019). "A Rule Based System for the Diagnosis of Coffee Diseases." *International Journal of Academic Information Systems Research (IJAISR)* 3(3): 1-8.
13. Masri, N., et al. (2019). "Survey of Rule-Based Systems." *International Journal of Academic Information Systems Research (IJAISR)* 3(7): 1-23.
14. Madi, S. A., et al. (2018). "The Organizational Structure and its Impact on the Pattern of Leadership in Palestinian Universities." *International Journal of Academic Management Science Research (IJAMSR)* 2(6): 1-26.
15. Madi, S. A., et al. (2018). "The dominant pattern of leadership and Its Relation to the Extent of Participation of Administrative Staff in Decision-Making in Palestinian Universities." *International Journal of Academic Management Science Research (IJAMSR)* 2(7): 20-43.
16. Kashkash, K., et al. (2005). "Expert system methodologies and applications-a decade review from 1995 to 2004." *Journal of Artificial Intelligence* 1(2): 9-26.
17. Hilles, M. M. and S. S. Abu Naser (2017). "Knowledge-based Intelligent Tutoring System for Teaching Mongo Database." *EUROPEAN ACADEMIC RESEARCH* 6(10): 8783-8794.
18. Elzamily, A., et al. (2015). "Classification of Software Risks with Discriminant Analysis Techniques in Software planning Development Process." *International Journal of Advanced Science and Technology* 81: 35-48.
19. Elsharif, A. A. and S. S. Abu-Naser (2019). "An Expert System for Diagnosing Sugarcane Diseases." *International Journal of Academic Engineering Research (IAER)* 3(3): 19-27.
20. Elqassas, R. and S. S. Abu-Naser (2018). "Expert System for the Diagnosis of Mango Diseases." *International Journal of Academic Engineering Research (IAER)* 2(8): 10-18.
21. El-Mashharawi, H. Q., et al. (2020). "Grape Type Classification Using Deep Learning." *International Journal of Academic Engineering Research (IAER)* 3(12): 41-45.
22. El Talla, S. A., et al. (2018). "The Nature of the Organizational Structure in the Palestinian Governmental Universities-Al-Aqsa University as a Model." *International Journal of Academic Multidisciplinary Research (IJAMR)* 2(5): 15-31.
23. El Talla, S. A., et al. (2018). "Organizational Structure and its Relation to the Prevailing Pattern of Communication in Palestinian Universities." *International Journal of Engineering and Information Systems (IJEIS)* 2(5): 22-43.
24. Dheir, I. and S. S. Abu-Naser (2019). "Knowledge Based System for Diagnosing Guava Problems." *International Journal of Academic Information Systems Research (IJAISR)* 3(3): 9-15.
25. Dahouk, A. W. and S. S. Abu-Naser (2018). "A Proposed Knowledge Based System for Desktop PC Troubleshooting." *International Journal of Academic Pedagogical Research (IJAPR)* 2(6): 1-8.
26. Barhoom, A. M. and S. S. Abu-Naser (2018). "Black Pepper Expert System." *International Journal of Academic Information Systems Research (IJAISR)* 2(8): 9-16.
27. Ashqar, B. A. M. and S. S. Abu-Naser (2019). "Identifying Images of Invasive Hydrangea Using Pre-Trained Deep Convolutional Neural Networks." *International Journal of Academic Engineering Research (IAER)* 3(3): 28-36.
28. Anderson, J., et al. (2005). "Adaptation of Problem Presentation and Feedback in an Intelligent Mathematics Tutor." *Information Technology Journal* 5(5): 167-207.
29. AlZamily, J. Y. and S. S. Abu-Naser (2018). "A Cognitive System for Diagnosing Musa Acuminata Disorders." *International Journal of Academic Information Systems Research (IJAISR)* 2(8): 1-8.
30. Al-Shawwa, M. and S. S. Abu-Naser (2019). "Knowledge Based System for Apple Problems Using CLIPS." *International Journal of Academic Engineering Research (IAER)* 3(3): 1-11.
31. Alshawwa, I. A., et al. (2020). "Analyzing Types of Cherry Using Deep Learning." *International Journal of Academic Engineering Research (IAER)* 4(1): 1-5.
32. Al-Nakhla, M. A. and S. S. Abu Naser (2017). "Adaptive Intelligent Tutoring System for learning Computer Theory." *EUROPEAN ACADEMIC RESEARCH* 6(10): 8177-8782.
33. Almrshidi, S. H. and S. S. Abu Naser (2017). "Design and Development of Diabetes Intelligent Tutoring System." *EUROPEAN ACADEMIC RESEARCH* 6(9): 8770-8778.
34. Almasri, A., et al. (2019). "Intelligent Tutoring Systems Survey for the Period 2000-2018." *International Journal of Academic Engineering Research (IAER)* 3(5): 21-37.
35. Almasri, A., et al. (2018). "The Organizational Structure and its Role in Applying the Information Technology Used In the Palestinian Universities-Comparative Study between Al-Azhar and the Islamic Universities." *International Journal of Academic and Applied Research (IJAAAR)* 2(6): 1-22.
36. Al-Habil, W. I., et al. (2017). "The Impact of the Quality of Banking Services on Improving the Marketing Performance of Banks in Gaza Governorates from the Point of View of Their Employees." *International Journal of Engineering and Information Systems (IJEIS)* 1(7): 197-217.
37. Alhabbasha, M. I., et al. (2016). "An Intelligent Tutoring System for Teaching Grammar English Tenses." *EUROPEAN ACADEMIC RESEARCH* 6(9): 7743-7757.
38. AlFerjany, A. A. M., et al. (2018). "The Relationship between Correcting Deviations in Measuring Performance and Achieving the Objectives of Control-The Islamic University as a Model." *International Journal of Engineering and Information Systems (IJEIS)* 2(1): 74-89.
39. Al-Bastami, B. G. and S. S. Abu Naser (2017). "Design and Development of an Intelligent Tutoring System for C# Language." *EUROPEAN ACADEMIC RESEARCH* 6(10): 8795.
40. Alajrami, M. A. and S. S. Abu-Naser (2018). "Onion Rule Based System for Disorders Diagnosis and Treatment." *International Journal of Academic Pedagogical Research (IJAPR)* 2(8): 1-9.
41. Al Shobaki, M., et al. (2018). "Performance Reality of Administrative Staff in Palestinian Universities." *International Journal of Academic Information Systems Research (IJAISR)* 2(4): 1-17.
42. Al Shobaki, M. J., et al. (2018). "The Level of Organizational Climate Prevailing In Palestinian Universities from the Perspective of Administrative Staff." *International Journal of Academic Management Science Research (IJAMSR)* 2(5): 33-58.
43. Al Shobaki, M. J., et al. (2017). "Learning Organizations and Their Role in Achieving Organizational Excellence in the Palestinian Universities." *International Journal of Digital Publication Technology* 1(2): 40-85.
44. Al Shobaki, M. J., et al. (2017). "Impact of Electronic Human Resources Management on the Development of Electronic Educational Services in the Universities." *International Journal of Engineering and Information Systems* 1(1): 1-19.
45. Al Shobaki, M. J., et al. (2016). "The impact of top management support for strategic planning on crisis management: Case study on UNRWA-Gaza Strip." *International Journal of Academic Research and Development* 1(10): 20-25.
46. Al Shobaki, M. J. and S. S. Abu Naser (2016). "The reality of modern methods applied in process of performance assessments of employees in the municipalities in Gaza Strip." *International Journal of Advanced Scientific Research* 1(7): 14-23.
47. Al Shobaki, M. J. and S. S. Abu Naser (2016). "Performance development and its relationship to demographic variables among users of computerized management information systems in Gaza electricity Distribution Company." *International Journal of Humanities and Social Science Research* 2(10): 21-30.
48. Al Shobaki, M. J. and S. S. Abu Naser (2016). "Decision support systems and its role in developing the universities strategic management: Islamic university in Gaza as a case study." *International Journal of Advanced Research and Development* 1(10): 33-47.
49. Ahmed, A. A., et al. (2018). "The Impact of Information Technology Used on the Nature of Administrators Work at Al-Azhar University in Gaza." *International Journal of Academic Information Systems Research (IJAISR)* 2(6): 1-20.
50. Abu-Saqer, M. M., et al. (2020). "Type of Grapefruit Classification Using Deep Learning." *International Journal of Academic Information Systems Research (IJAISR)* 4(1): 1-5.
51. Abu-Saqer, M. M. and S. S. Abu-Naser (2019). "Developing an Expert System for Papaya Plant Disease Diagnosis." *International Journal of Academic Engineering Research (IAER)* 3(4): 14-21.
52. Abu-Nasser, B. S. and S. S. Abu Naser (2018). "Rule-Based System for Watermelon Diseases and Treatment." *International Journal of Academic Information Systems Research (IJAISR)* 2(7): 1-7.
53. Abu-Naser, S. S., et al. (2011). "An intelligent tutoring system for learning java objects." *International Journal of Artificial Intelligence & Applications (IJAA)* 2(2): 86-77.
54. Abu-Naser, S. S. and M. J. Al Shobaki (2016). "Computerized Management Information Systems Resources and their Relationship to the Development of Performance in the Electricity Distribution Company in Gaza." *EUROPEAN ACADEMIC RESEARCH* 6(8): 6969-7002.
55. Abu-Naser, S. S. and M. A. Al-Nakhla (2016). "A Ruled Based System for Ear Problem Diagnosis and Treatment." *World Wide Journal of Multidisciplinary Research and Development* 2(4): 25-31.
56. Abu-Naser, S. S. (2016). "ITSB: An Intelligent Tutoring System Authoring Tool." *Journal of Scientific and Engineering Research* 3(5): 63-71.
57. Abu-Naser, S. S. (2009). "Evaluating the effectiveness of the CPP-Tutor, an Intelligent Tutoring System for students learning to program in C++." *Journal of Applied Sciences Research* 5(1): 109-114.
58. Abu-Naser, S. S. (2008). "JEE-Tutor: An Intelligent Tutoring System for Java Expression Evaluation." *Information Technology Journal* 7(3): 528-532.
59. AbuEloun, N. N. and S. S. Abu Naser (2017). "Mathematics intelligent tutoring system." *International Journal of Advanced Scientific Research* 2(1): 11-16.
60. Abu Naser, S. S., et al. (2017). "Trends of Palestinian Higher Educational Institutions in Gaza Strip as Learning Organizations." *International Journal of Digital Publication Technology* 1(1): 1-42.
61. Abu Naser, S. S., et al. (2016). "Measuring knowledge management maturity at HEI to enhance performance-an empirical study at Al-Azhar University in Palestine." *International Journal of Commerce and Management Research* 2(5): 55-62.
62. Abu Naser, S. S. and M. J. Al Shobaki (2016). The Impact of Management Requirements and Operations of Computerized Management Information Systems to Improve Performance (Practical Study on the employees of the company of Gaza Electricity Distribution). First Scientific Conference for Community Development.
63. Abu Naser, S. S. (2008). "Developing an intelligent tutoring system for students learning to program in C++." *Information Technology Journal* 7(7): 1055-1060.
64. Abu Naser, S. S. (2006). "Intelligent tutoring system for teaching database to sophomore students in Gaza and its effect on their performance." *Information Technology Journal* 5(5): 916-922.
65. Abu Naser, S. S. (1999). "Big O Notation for Measuring Expert Systems complexity." *Islamic University Journal Gaza* 7(1): 57-70.
66. Abu Naser, S. S. (1993). A methodology for expert systems testing and debugging, North Dakota State University, USA.
67. Abu Nada, A. M., et al. (2020). "Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach." *International Journal of Academic Information Systems Research (IJAISR)* 4(8): 6-9.
68. Abu Nada, A. M., et al. (2020). "Age and Gender Prediction and Validation Through Single User Images Using CNN." *International Journal of Academic Engineering Research (IAER)* 4(8): 21-24.
69. Abu Amuna, Y. M., et al. (2017). "Understanding Critical Variables for Customer Relationship Management in Higher Education Institution from Employees Perspective." *International Journal of Information Technology and Electrical Engineering* 6(1): 10-16.
70. Abu Amuna, Y. M., et al. (2017). "Strategic Environmental Scanning: an Approach for Crises Management." *International Journal of Information Technology and Electrical Engineering* 6(3): 28-34