Machine Learning Approach for an Automatic Irrigation System in Southern Jordan Valley

Anas H. Blasi
Computer Information Systems Department
Mutah University
Al Karak, Jordan

Mohammad A. Abbadi
Computer Science Department
Mutah University
Al Karak, Jordan

Rufaydah Al-Huweimel
Computer Science Department
Mutah University
Al Karak, Jordan

Abstract—**The agriculture sector is the most water-consuming sector. Due to the critical situation of available water resources in Jordan, attention should be paid to the issues of water demand and appropriate irrigation in order to spread the right management ways of modern irrigation to the farmers. The objectives of this paper are to improve the irrigation process and provide irrigation water to the highest possible extent through the use of artificial intelligence to construct a smart irrigation system that controls the irrigation mechanism using the necessary tools for sensing soil moisture and temperature, giving alerts of any change in the parameters entered as the baseline values for comparison, and installing system sensors buried at a depth of 3-5 inches below the roots to measure the moisture content in the soil. The sensors measure the humidity and temperature in the soil every ten minutes. They prevent the automatic irrigation process if the humidity is high, and permit it if the humidity is low. The smart automatic irrigation system model was built using the Decision Tree (DT) algorithm, which is a machine learning algorithm that trains the system on a part of the collected data to build the model that will be used to examine and predict the remaining data. The system had a prediction accuracy of 97.86%, which means that it may be successfully used in providing irrigation water for the agricultural sector.**

Keywords-irrigation; agriculture; artificial intelligence; sensor; machine learning; decision tree algorithm

I. INTRODUCTION

Jordan has a Mediterranean climate. Rainfall varies from 50mm in the desert to about 580mm (per year) in the northern highlands. Snow falls at short intervals on most mountain highlands in the north, central, and south of the kingdom and is sometimes accumulated [1]. Agricultural lands are widely spread, mainly in the Jordan Valley (northern, central, and southern). The total agricultural area in Jordan in 2017 was amounted to 1894324 dunums (1 dunum=1000m^2), whereas 864549.3 dunums are irrigated against 102,977,7.7 which depend on rainwater [2]. These lands contain field crops, while irrigated agriculture lands cover all the needs of productive trees, vegetables, and field crops also. Irrigated agriculture relies on water sources (from dams, groundwater, etc.). The available quantities of water in Jordan are somewhat low [3], especially as agricultural land is scattered in places with higher temperatures, which increases the demand. There is a need to look for an irrigation method that suits the water supply and the plants' needs. Traditionally, the first thing that comes to mind is

drip irrigation. Drip irrigation is a water-saving system that provides water by slowly spraying water into the soil or plant root which contributes in reducing the spread of herbs, plant diseases, and is suitable for all soil types, but it relies on pipes which need regular maintenance. Another element that plays a major role is the soil. The soil is a surface layer that covers the surface of the ground and consists of fragmented rock materials that have been changed because of the exposure to environmental, biological, and chemical factors [4]. The soil has several physical properties which mainly affect the growth of plants, the most important of these properties are [4]:

- Soil strength: represents the aggregation or arrangement of granules in the soil. The soil strength is classified to four types [4].

- Organic matter: contributes to the binding of granules in the soil texture and increases the ability to retain water. Good soil contains 1-6% organic matter.

- Soil texture: is the percentage of mineral grains in the soil from sand, silt, and clay and is considered as the most important property of the soil.

These characteristics are closely linked in the agricultural process because of their direct effect on the plants. The soil properties in Gour Alsafi and Al Karak are described in [3]. We can see in Table I that the soil type in Gour Alsafi is loamy sand soil, which means that the soft, coarse, and medium grain proportions are almost equal. A result of these homogeneous proportions is that the soil has medium water permeability, characterized by moderate ventilation and good fertility rate and is easier to deal with in terms of ease of cultivation and productivity. The weather conditions in Jordan range from extreme heat to extreme cold and from extreme dryness to excessive rainfalls, in addition to the irregularity in the rainfall distribution throughout the year. The idea behind this work is the application of machine learning in irrigation with less cost and effort, by designing an irrigation system model based on weather and soil type. The aim of this paper is to create an intelligent irrigation system that combines the advantages of traditional drip irrigation with a pipe control mechanism, which contributes to the reduction of diseases and the use of chemical pesticides. An automatic irrigation system is responsible to water the plants efficiently while providing a system that can reduce the needed manpower while saving water.

Corresponding author: Anas H. Blasi

<div align="center">TABLE I.   GOUR ALSAFI SOIL PROPERTIES</div>

Soil layers	Layer depth (cm)	Soil texture	Sand %	Silt %	Clay %	SAT %	WP %	Field capacity %	KSAT (mm/hr)	Salinity (MS/cm)
Layer 1	0-20	Loamy sand	76.2	20.2	3.6	27.8	2.7	6.5	27.6	9.3
Layer 2	20-46	Loamy sand	84.4	12.9	2.7	29.0	2.1	4.3	42.7	2.7
Layer 3	46-87	Loamy sand	88.9	8.9	2.1	30.0	1.9	2.8	65.6	1.8

II.   RELATED WORK

Many studies have been published in relation with the use of artificial intelligence applications in agriculture and irrigation. Through the review of these papers, the absence of Arab countries on the research in this field has been noticed, since there is only one Arab country that has recently started working on this, which is the United Arab Emirates, so the current paper may be considered as one of the first that look at the field of artificial intelligence in agriculture in Arab countries. Authors in [5] presented a fully automated drip irrigation system in India. This system takes an image of the plant to check its height. If the plant is short, it spreads a fertilizer on it, if there is no need of fertilization, it checks the temperature and soil moisture, if they are less than the set point, the system opens the water motor. All irrigation orders are controlled by the farmer since the sensor reads the moisture and temperature in the soil, assesses the plants need for water, and then sends a short message to the farmer. The system uses a fuzzy logic algorithm to water the plants based on a set of rules. The idea in [6] was to design an automated irrigation system based on scheduling irrigation and fertilization. The irrigation depends on the soil moisture using special sensors. The system will start automatically when the level of moisture is low and it will automatically stop when it reaches a suitable level. Additionally, the system handled the fertigation process by injecting fertilizers through the pipes of the system thrice a week by using a water pump which supplies fertilizer mixed with water to the plants. The authors applied the Bernoulli's theorem in order to find the pressure of the water supply, the flow rate of water supply, and the water losses. The system utilizes a microcontroller to convert and manipulate the data obtained from the soil moisture sensors to get the actual humidity values. The control unit will instruct the pump to start working based on the obtained data and the pump will stop working also based on the obtained data.

Authors in [7], considered the role of artificial intelligence in improving the agriculture sector and dealing with the huge amount of data obtained daily (soil reports, plant needs for fertilizer), and the use of robots in improving crop harvesting. The authors focused on the use of image-based insight generation to conduct a complete field analysis of agricultural land by combining computer vision technology, Internet of Things (IoT), and drone data to monitor crops and fields, and used all these in combinations in disease detection, field management, crop readiness identification, identification of the optimal mix for agronomic products, crop health monitoring, and automation techniques in irrigation. India was taken as an example of the AI agricultural state, because at present in India, in the state of Andhra Pradesh, Microsoft Corporation is working with farmers rendering farm advisory services using the Cortana Intelligence Suite which includes Machine Learning and Power BI, which transform data into intelligent actions. The soil's water needs vary from one area to another, so the irrigation decisions must be based on local needs and rules. Authors in [8] tackled this problem by categorizing all what we consider as ambiguous in the agriculture sector in general, and the rate of irrigation in particular, considering the type of soil, the type of crops, place and time, irrigation method, and fuzzy logic. All these factors were integrated with GIS for decision-making in the irrigation process thus obtaining an intelligent irrigation system called the fuzzy inference system, an intelligent deduction system based on accurate irrigation knowledge, capable of creating guidance maps to control the speed of rotation of the irrigation central axis whose inputs are processed images based on soil characteristics and precise quantities of irrigation using mathematical equations to process the ambiguous data.

Authors in [9] discussed the groundwater, its characteristics, and its suitability for irrigating crops through an experiment in the Nanded, Maharashtra, India, indicating the area characteristics in terms of climate, rainfall, and soil quality. An artificial neural network/back propagation algorithm was used to study the soil concentration in salts and minerals, the physical and chemical analysis, and the design of the optimal model to predict the validity of groundwater for irrigation. Fifty representative groundwater samples were collected from different locations in rural and agricultural areas. The model consisted of 13 input neurons (salts and minerals), 7 hidden neurons, and 5 output variables to calculate the suitability of groundwater for irrigation in the study area. The result is a high accurate spatial distribution map of the measured and the predicted values. The proposed ANN model can be applied in measuring the ground water suitability for irrigation. This model can be considered by local development authorities for better management of groundwater resources. Authors in [10] presented a proposal for a photoelectric irrigation system based on a sensor buried in the root zone that transmits soil readings to an Arduino board which controls the water pump. The irrigation is done automatically according to the level of soil moisture, and saves electrical energy as the pumps are powered by solar energy through the use of solar cells and thus is compatible with remote areas where there is no electricity grid. Authors in [11] aimed to build an automatic irrigation system in Indonesia that depends on a programmed chip that controls the launch of water sprinklers to irrigate crops based on the soil moisture. The system works using many sensors and boards and is an extraction from previous research studies in the same field. Authors in [12] also used photovoltaic panels to produce the needed electrical energy. In their system, a humidity sensor detects the moisture level in the soil and sends the information to an Arduino controller which in turn detects the water level in the tank with an ultrasonic sensor. Depending on the water level, the Arduino controls the water flow through valves, an LCD monitor displays the information

about soil moisture content and water level, and that information can be conveyed to the end user in the form of short messages through a GSS module.

Through the previous review of the published papers related to the use of artificial intelligence in the field of agriculture, it was noted that the potential of machine learning techniques in this field is big and its application can contribute in the agriculture development. This field is very active and promising new solutions are being developed continuously to meet the challenges. Nevertheless, the respective efforts in Arab countries are quite limited, and this work comes to cover that gap.

III. RESEARCH METHODOLOGY

There are many data mining methods that can be applied to projects related to modern science data. One of the most common of these is the Knowledge Discovery in Database (KDD) method, which has been adopted in this paper. KDD is the process of discovering useful knowledge from a set of data [13], through a set of steps described below.

A. Data Collection

The data used in this paper are real data obtained from the Directorate of Agriculture of the Southern Jordan Valley / Department of Agricultural Extension [3] and the National Center for Agricultural Research and Extension / Soil Survey Project [4]. In addition, weather data were acquired from Taqs Al-Arab website and the Department of Meteorology [1]. The data were compiled as a table in a CSV file containing 1498 records and 15 attributes such as: plant name, type of plant (fruits, leaves, roots, etc.), protected/exposed which is indicating the nature of the environment in which the plant is grown, land area (dunums), soil type which describes the nature of the soil with regard to its ability to retain water and moisture, month which is the planting season, temperature, morning weather humidity, midnight humidity, noon weather humidity, soil moisture, soil salinity which is the percentage of salts, evaporation transpiration (m^3/donum/month) which is the amount of water that one dunum of agricultural land loses in a month as a result of its evaporation, and average daily water consumption (mm) which is the amount of water a plant needs to reach the desired degree of hydration.

B. Data Selection and Preparation

From the collected data, a target data set was created, by selecting a subset of variables (soil humidity, soil type, soil salinity, and temperature) which directly affect the irrigation process. The selected data were processed, analyzed and correlated using Python programming language. The data preprocessing was done for the collected dataset from numerical to categorical data, since this model was built using the Decision Tree (DT) algorithm which is a very simple, understandable, easy to implement, and also gives high accurate results. It is widely known and used in many businesses to support decision making and risk analysis [14, 15]. This algorithm works only on categorical data types.

C. Decsion Tree Algorithm

Data Mining (DM) is the extraction of useful information from a large data set. In that sense, DM is also known as Knowledge Discovery (KD) or Knowledge Extraction (KE). One of the main purposes of DM is that the information gathered from it helps to predict hidden patterns, future trends and behaviors, and boosts decision making. It can be applied to any type of data, e.g. data warehouses, transactional databases, relational databases, multimedia databases, spatial databases, time-series databases, and the World Wide Web. In the process of DM, the large data sets are first sorted, then patterns are identified, and relationships are established to perform data analysis and solve problems. This process is called data classification which is the process of finding a model that describes and distinguishes data classes and concepts. Classification is the problem of identifying to which of a set of categories (subpopulations), a new observation belongs to, on the basis of a training set of data containing observations and whose categories membership is known. The data set was split into two groups. At first, the training dataset consisting of 67% of the data, was used to train the machine classifier and then create a general classification model using the DT algorithm with the available training set. The second is the testing dataset which consisted of the 33% of the data, which entered the model for testing based on the training result. The constructed model was used to predict class labels and hence estimate the accuracy of the classification rules. After dividing the data into the previous groups, the training process begun by measuring the temperature and degree of soil moisture through sensors buried at a depth of 3-5 inches below the roots. The sensors measure the humidity and temperature in the soil every 10 minutes, preventing the automatic irrigation process if the humidity is high, and permitting it if the humidity is low. During the testing phase, the testing data were the input for the designed model, and the comparison results between the prediction and the result were recorded. Python programming language is considered the most suitable and has been used in this study. However, the prediction model will be built using sklearn with some important libraries and methods such as accuracy_score, train_test_split, DecisionTreeClassifier and panads.

IV. RESULTS AND EVALUATION

It is useful to control water consumption, especially in agriculture sector, since it is the most water consuming. Also this may affect the production quality, as controlling the amount of water reduces fungi and plant diseases and thus increases yield and profit. In this paper, an automatic irrigation system has been designed that measures soil moisture and temperature, in addition to some characteristics affecting the irrigation process such as soil quality and salinity. Using this system the farmer can monitor the crops without the need of extra manpower, saving the time and cost required for the irrigation process. Although somewhat inexpensive, the smart irrigation system is economically feasible and highly accurate. The confusion matrix is often used to describe the performance of a classification model on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand. Table I shows the system's confusion matrix which is an evaluation method for the DT model results. We note that the sum of True Positive (TP) and True Negative (TN) predictions reaches a percentage as high as the 98% of total predictions.

TABLE I.   CONFUSION MATRIX

N=1497	Predicted "NO"	Predicted "YES"
Actual "NO"	240	6
Actual "YES"	26	1225

TABLE II.   CROSS-VALIDATION RESULTS

	Accuracy	Precision	Recall	F1 score	MSE
0.0	0.98	0.9	0.98	0.94	0.0213
1.0	0.98	1.0	0.98	0.99	

The classification report of Table II displays the accuracy, precision, recall, F1 score, and Mean Square Error (MSE), of the model. In order to support easier interpretation and problem detection, the metrics are defined in terms of TP, False Positive (FP), TN, and False Negative (FN) predictions. Using this terminology, the metrics are defined as [16-18]:

- Precision is the ability of a classifier to not label an instance as positive when it is actually negative. For each class it is defined as the ratio of TP to the sum of TP and FP. The irrigation system model has a high precision equal to 0.90.

- Recall is the ability of a classifier to find all positive instances. It is defined as the ratio of TP to the sum of TP and FN. The proposed model has a recall value of 0.98.

- The F1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0. Generally speaking, F1 scores are lower than accuracy measures as they embed precision and recall into their computation. As a rule of thumb, the weighted average of F1 should be used to compare classifier models, not global accuracy.

- MSE measures the average squared difference between the estimated and true values. It is a risk function, corresponding to the expected value of the squared error loss, always nonnegative, and the values closer to zero are better. The MSE for the irrigation system model is equal to 0.0213.

- Accuracy is the most commonly used metric to judge a model which is the sum of TP with TN divided on the sum of all predictions (TP, TN, FP, FN). The proposed model has a high accuracy, equal to 0.98.

V.   CONCLUSION AND FUTURE WORK

The provision of water in the agriculture sector is an important research area because agriculture is the most water consuming sector and the available water resources are scarce. Attention should be paid to the water needs of plants and to the appropriate irrigation scheduling in order to effectively transfer to the farmers, modern and optimum ways of irrigation management. In this paper, the most prominent agricultural area of Jordan was covered, i.e. the southern Jordan Valley and the needed data to design a model for the irrigation process based on soil properties and temperatures were collected. This system helps in providing the quantities of water needed for irrigation while improving productivity. There are other agricultural regions in Jordan that have different characteristics from the southern Jordan Valley, especially in temperature, so we may not be able to generalize the results of this model to all

regions. This model can be improved in the future in order to be able to be applied to other regions in Jordan. This model can be further developed to include scheduling of the fertilization process, or to be able to heat the pipes to solve the problem of frost in the winter, and to use irrigation pipes to improve soil aeration and inject organic materials needed to improve agricultural production combining the automatic irrigation system with image processing through a set of cameras to monitor plant growth in terms of size and disease detection. In the future, different data sets will be applied from different regions of Jordan. In addition, other machine learning algorithms can be utilized, such as Artificial Neural Networks (ANNs) [19, 20], fuzzy logic [21], k-NN [22], and other data mining techniques [23, 24], and the results will be compared with the DT results.

REFERENCES

[1] "ArabiaWeather," *ArabiaWeather*. https://www.arabiaweather.com/ (accessed Dec. 16, 2020).

[2] "Annual Report 2015," Ministry of Water & Irrigation, Water Authority-Jordan Valley Authority, 2015.

[3] "Soil Survey Project -Gour Al-Safi/Al-Karak," The National Center for Agricultural Research and Extension.

[4] R. R. Weil and N. C. Brady, *The Nature and Properties of Soils*, 15th edition. Columbus, OH, USA: Pearson, 2016.

[5] S. Jadhav and S. Hambarde, "Android based Automated Irrigation System using Raspberry Pi," *International Journal of Science and Research*, vol. 5, no. 6, pp. 2345–2351, Jun. 2016, https://doi.org/ 10.21275/v5i6.NOV164836.

[6] A. Norazizan, M. Muzni, M. Ramsaid, M. Ismail, and M. Mazalan, "Design and Development of the Irrigation System for Planting (Part 1)," *Mechanical System Design*, May 2016, https://doi.org/ 10.13140/RG.2.1.4088.1529.

[7] V. Dharmaraj and C. Vijayanand, "Artificial Intelligence (AI) in Agriculture," *International Journal of Current Microbiology and Applied Sciences*, vol. 7, no. 12, pp. 2122–2128, 2018, https://doi.org/10.20546/ijcmas.2018.712.241.

[8] W. R. Mendes, F. M. U. Araújo, and S. Er-Raki, "Integrating Remote Sensing Data into Fuzzy Control System for Variable Rate Irrigation Estimates," *Irrigation - Water Productivity and Operation, Sustainability and Climate Change*, Jun. 2019, https://doi.org/ 10.5772/intechopen.87023.

[9] V. M. Wagh, D. B. Panaskar, A. A. Muley, S. V. Mukate, Y. P. Lolage, and M. L. Aamalawar, "Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India," *Modeling Earth Systems and Environment*, vol. 2, no. 4, pp. 1–10, Dec. 2016, https://doi.org/10.1007/s40808-016-0250-3.

[10] S. A. Hamoodi, A. N. Hamoodi, and G. M. Haydar, "Automated irrigation system based on soil moisture using arduino board," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 3, pp. 870–876, Jun. 2020, https://doi.org/10.11591/eei.v9i3.1736.

[11] I. Prasojo, P. T. Nguyen, O. Tanane, and N. Shahu, "Design of Ultrasonic Sensor and Ultraviolet Sensor Implemented on a Fire Fighter Robot Using AT89S52," *Journal of Robotics and Control (JRC)*, vol. 1, no. 2, pp. 55–58, Jan. 2020, https://doi.org/10.18196/jrc.1212.

[12] C. A. Bolu, J. Azeta, F. Alele, E. O. Daranijo, P. Onyeubani, and A. A. Abioye, "Solar Powered Microcontroller-based Automated Irrigation System with Moisture Sensors," *Journal of Physics: Conference Series*, vol. 1378, Dec. 2019, Art. no. 032003, https://doi.org/10.1088/1742-6596/1378/3/032003.

[13] "What is Knowledge Discovery in Databases (KDD)? - Definition from Techopedia," *Techopedia*. http://www.techopedia.com/definition/25827/ knowledge-discovery-in-databases-kdd (accessed Dec. 16, 2020).

[14] F. Li, Y. Y. Li, and C. Wand, "Uncertain data decision tree classification algorithm," *Journal of Computer Applications*, vol. 29, no. 11, pp. 3092–3095, Dec. 2009.

[15] https://informatic-ar.com/id3_algorithm/ (accessed Dec. 16, 2020)

[16] "scikit-learn: machine learning in Python — scikit-learn 0.23.2 documentation." https://scikit-learn.org/stable/ (accessed Dec. 16, 2020).

[17] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, "Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking," *IEEE Access*, vol. 8, pp. 90847–90861, 2020, https://doi.org/10.1109/ACCESS.2020.2994222.

[18] M. Hossin and M. N. Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 1–11, Nov. 2019, https://doi.org/10.5281/zenodo.3557376.

[19] A. Blasi, "Performance Increment of High School Students using ANN model and SA algorithm," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 11, pp. 2417–2425, Jan. 2017.

[20] R. A. Aroud, A. H. Blasi, and M. A. Alsuwaiket, "Intelligent Risk Alarm for Asthma Patients using Artificial Neural Networks," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, 58/30 2020, https://doi.org/10.14569/IJACSA.2020.0110612.

[21] A. Blasi, "Scheduling Food Industry System using Fuzzy Logic," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 19, pp. 6463–6473, Oct. 2018.

[22] M. A. A. Lababede, A. H. Blasi, and M. A. Alsuwaiket, "Mosques Smart Domes System using Machine Learning Algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 40/30 2020, https://doi.org/10.14569/IJACSA.2020.0110347.

[23] M. Alsuwaiket, A. H. Blasi, and K. Altarawneh, "Refining Student Marks based on Enrolled Modules' Assessment Methods using Data Mining Techniques," *Engineering, Technology & Applied Science Research*, vol. 10, no. 1, pp. 5205–5210, Feb. 2020, https://doi.org/10.48084/etasr.3284.

[24] M. Alsuwaiket, A. H. Blasi, and R. A. Al-Msie'deen, "Formulating Module Assessment for Improved Academic Performance Predictability in Higher Education," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4287–4291, Jun. 2019, https://doi.org/10.48084/etasr.2794.