Deep Learning for Combined Water Quality Testing and Crop Recommendation

Tahani Alkhudaydi, Maram Qasem Albalawi, Jamelah Sanad Alanazi, Wejdan Al-Anazi, Rahaf Mansour Alfarshouti
Faculty of Computers and Information Technology
University of Tabuk, Tabuk, Saudi Arabia

***Abstract*—The field of agriculture and its specifics has been gaining more attention nowadays due to the limited present resources and the continuously increasing need for food. In fact, agriculture has benefited greatly from the advancements of artificial intelligence, namely, Machine Learning (ML). In order to make the most of a crop field, one must initially plan on what crop is best for planting in this particular field, and whether it will provide the necessary yield. Additionally, it's very important to constantly monitor the quality of soil and water for irrigation of the selected crop. In this paper, we are going to rely on Machine Learning and data analysis to decide the type of crop that we will use, and the quality of soil and water. To do so, certain parameters must be taken into consideration. For choosing the crop, parameters such as sun exposure, humidity, soil pH, and soil moisture will be taken into consideration. On the other hand, water pH, electric conductivity, content of minerals such as chloride, calcium, and magnesium are among the parameters taken into consideration for water quality classification. After acquiring datasets for crop and water potability, we implemented a deep learning model in order to predict these two features. Upon training, our neural network model achieved 97% accuracy for crop recommendation and 96% for water quality prediction. However, the SVM model achieves 96% for crop recommendation and 92% for water quality prediction.**

***Keywords*—Deep learning; irrigation; artificial intelligence; soil moisture**

I. INTRODUCTION

As humans became more advanced, they learned that plants do not only provide necessary food for both humans and animals, but it also plays a very important role in the fields of medicine, energy production, and the wellbeing of the entire planet. Agriculture is one of the most essential needs for humans to maintain a sustainable livelihood. Without agriculture, humans would not be getting the sufficient nutrients in their meals, and livestock will not have food to eat which means soon cows and sheep will no longer survive [1]. Thus, it is important to play close attention to crops and to constantly monitor their needs. One of the important needs is water, and not just any water. Irrigation water must have certain qualities in order to be used for planting the crops. So, a water quality assessment or classification system is necessary. In addition, it is essential to know which crops grow best in which environmental conditions, which means a crop prediction system becomes necessary as well.

In many fields such as aquaculture, livestock production, and food industry, water is a critical raw material. For this reason, not any kind of water can be used in any field. To illustrate, not all water is good enough for drinking, or watering plants, etc. [2]. Chatterjee described four different water qualities which are palatable water, infected water, potable water, and contaminated water. From these four types, only palatable and potable water are useable. The water can be classified based on several parameters that have different effects on the water quality.

Generally speaking, the water parameters can be divided into three categories: physical parameters, biological parameters, and chemical parameters as shown in Fig. 1. The physical parameters include total suspended solids (TSS), temperature, electric conductivity (EC), and turbidity. The biological parameters consider whether the water contains any microorganisms. On the other hand, the chemical parameters include Sulfate, pH, heavy metals, and total nitrogen [3].

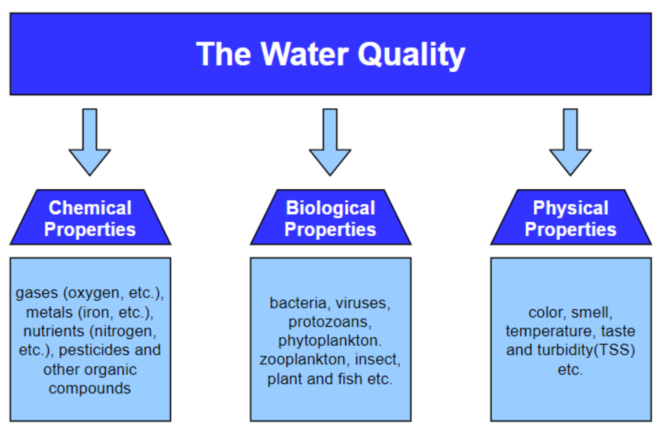

Fig. 1. Water quality parameters.

Predicting the crop is critical for sustainable intensification and efficient use of natural resources [4]. There are many factors that play a role and influence the yield production of crops including environmental conditions and management [5]. In Fig. 2, some of the factors are soil conditions such as pH and moisture play a role in determining the yield of a crop, in addition to weather conditions such as humidity, temperature, and rainfall. Furthermore, factors such as the genotype of the plant, the implemented water irrigation systems, and pesticide control also contribute greatly to how much yield a plant will produce [6].

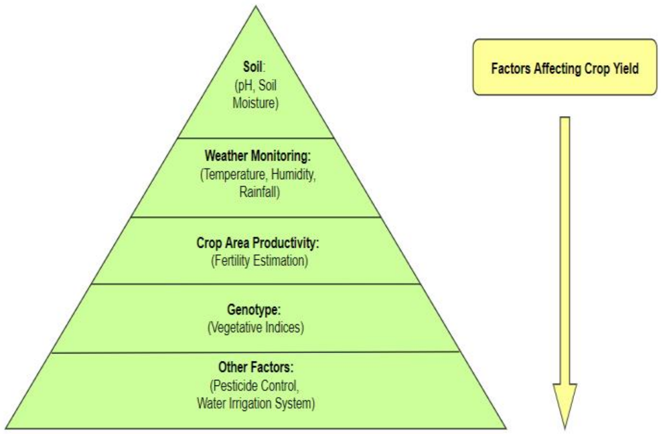

Fig. 2. Factors that influence the crop.

In order to achieve the objectives of the study, we first have to mention the problems that we are trying to address. As mentioned previously, two of the major issues or challenges in the proper maintenance of an agricultural field are the quality of irrigation water that is reaching the plants and that the plants need for proper growth, as well as the appropriate crop type since crops don't react and grow the same as each other in different climates and conditions. Thus, we propose a model that can attempt to solve the water quality analysis and crop recommendation problems at the same time. The contributions that we offer in our study can be summed as follows:

- We developed a system that can predict whether the available water is suitable for irrigation or not, and at the same time can perform crop recommendation based on crop prediction.
- Our model can provide reliable and accurate recommendations for farmers.
- The model that we propose is an inexpensive solution to solve the common problems that farmers face such as low crop.

In summary, this paper aims to address the challenges faced by farmers in selecting the appropriate crop type and evaluating the quality of irrigation water. We propose a model that utilizes deep learning and machine learning algorithms to predict the suitability of available water for irrigation and recommend the best crop type based on crop prediction. Our model offers a cost-effective solution to the common problems faced by farmers, providing reliable and accurate recommendations. The paper also includes a literature review on water quality studies and crop recommendation datasets, as well as a detailed description of the dataset used in our model. Additionally, we discuss the factors that influence crop growth and the contributions of our study.

II. Literature Review

Due to the importance of correctly selecting a crop type, and to continuously evaluate the quality of irrigation water, there's a huge pool of studies that revolve around these two topics.

A. Water Quality Studies

El Bilali et al. [7] designed and implemented a total of 8 Machine Learning models to statistically predict the irrigation water quality (IWQ) parameters that make water suitable for irrigation. Ten irrigation water quality parameters, namely Kelly ration (KR), sodium absorption ratio (SAR), adjusted SARa, Sodium percentage (Na^{+}%), exchangeable sodium percentage (ESP), residual sodium carbonate (RSC), total dissolved solids (TDS), chloride (Cl^{-}), permeability index (PI), and magnesium absorption ratio (MAR) were measured by analysis of 264 samples that were gathered from the Bouregreg watershed, as well as 29 samples from Cherrate and 35 samples from Nfifikh watersheds in Morocco. As for the Machine Learning algorithms, the chosen ones were: multiple linear regression, artificial neural network (ANN), decision tree (DT), random forest (RF), K-nearest neighbor (KNN), support vector machine/regression (SVM/R), stochastic gradient descent (SGD), and adaptive boosting. Upon testing, the results show that all of the 8 models except for SVR and KNN, are capable of predicting only 8 IWQ parameters through the use of electrical conductivity and pH as input variables in Bouregreg watershed surface. To further confirm the results, the six validated ML models were generalized to the Cherrate and Nfifikh watersheds. The results of this generalization attempt revealed that the previous models can be generalized for three parameters in Cherrate watershed and 4 in Nfifikh watershed. Some of these models were not able to statistically predict the MAR and the PI, possibly because of the poor relationship between the EC and pH input variables and these two parameters. Additionally, it was revealed that the adaptive boosting model achieves better performance in comparison with other models in Bouregreg watershed. Thus, it is possible to confirm that ML models can help farmers to better manage the irrigation water quality through extensive analysis.

Ali Mokhtar et al. [8] purposed to study the quality index of the irrigation water of Bahr El-Baqr drain. The authors gathered data from the analysis of 105 water samples of 1L each at a depth of 1m collected from Bahr El-Baqr during the month of July in 2020, ten features were taken into consideration, including pH level, electrical conductivity, sodium concentration, potassium concentration, Ca^{2+}, Mg^{2+}, chloride, carbonate, bicarbonate and sulfate composition. These data were cleaned and modified in order to reach the best score and the most accurate prediction possible. Multiple regressions including principal component regression, stepwise regression, partial least squares regression and ordinary least squares regression, in addition to Machine Learning methods including random forest, extreme gradient boosting, and support vector machine were applied in order to go through the features and determine the features most responsible of identifying the quality index. After applying the root mean square error, the best performance was for the stepwise regression with the values of 0.21% and 0.03%. While after applying the scatter index, all models gave values less than 0.1% except RSC.

B. Crop Recommendation Studies

Bandara et al. [9] proposed a system that can be used for predicting what type of crop should be planted in a certain area within Sri Lanka with the help of artificial intelligence. The system in fact is a recommendation system based on the collection of multiple environmental factors that directly impact the growth and yield of a certain crop. These factors are collected via sensors, namely temperature and humidity sensor, soil moisture sensor, pH sensor, and sunlight sensor, and are then communicated through Arduino microcontroller to the database for storage and analysis. The study relied on datasets collected by the authorities, as well agricultural books and websites. The collected data is preprocessed before being used by two algorithms which are the support vector machine (SVM) and naïve Bayes (NB) algorithms to perform a prediction based on the input data. The parameters are used to generate a PLU code related to each crop as an output. The added value of this study is that it offers an option for feedback by the users, from which the system will self-train accordingly as a response. Initially the model generated a 92% accuracy which can be enhanced to 95% upon constant use and feedback from farmers.

Priyadharshini et al. [10] purposed to help farmers finding the best crop to be grown in their lands by creating a recommendation system. The authors gathered data from government website and from Kaggle including the yield dataset which includes 16 major crops, the cost of cultivation dataset indicating the cost of each crop, the modal price of crops dataset which gives the market prices of crops among two months, the soil nutrient content dataset which includes five features which are order state, nitrogen content, phosphorous content, potassium content, and average pH. And the rainfall temperature dataset which includes crops, min and max rainfall, min and max temperature, and pH values. These data were cleaned and modified so that the null numbers are replaced by -1 in order not to affect the prediction process. Seven different methods were applied in order to go through the features in order to achieve accuracy and precision, these methods are the linear regression, the neural network, decision tree, K nearest neighbor, K nearest neighbor with cross validation, naïve Bayes and support vector machine. The best performance was for the neural network, with an accuracy of 89.88%.

Shilpa Mangesh Pande [11] and his colleagues in this paper proposed a prediction system in order to help farmers choose the most profitable crop with maximum yield and the best time for using fertilizers in order to reach the best results possible. The authors collected historical data for Maharashtra and Karnataka from different sources among which are indianwaterportal.com, data.gov.in and kaggle.com, while taking six features into consideration which are region, soil type, crop type, area, season, and year. These data were cleaned and modified so that the unavailable values are substituted with the mean values. Five different methods were applied in order to go through the features which are the support vector machine, artificial neural network, K nearest neighbor algorithms, random forest, and multivariate linear regression algorithms. The best performance was for the random forest algorithm with an accuracy of 95%.

Jadhav et al. [12] purposed to deal with the difficulties faced by farmers and find the best solutions and crops for farmers to grow in order to reach the best results possible. The authors gathered data from Kaggle while taking seven features into consideration; the features selected are the ratio of nitrogen content in soil, the phosphorous ratio, the potassium ratio, the temperature, humidity, rainfall and pH value. While these data were cleaned and modified so that only important features are selected using bar charts, scatter plots, box plots etc., also a UI was built so that the farmer can enter his data in order to get immediate results and recommended crops to grow. Four different methods were applied in order to go through the features leading to the ones responsible of finding the most accurate and the best result of which crop to grow, these methods are random forest, decision tree, logistic regression and XGBoost. The best performance was for the random forest algorithm, reaching an accuracy of 98.9%.

In conclusion, the literature review has highlighted the importance of water quality assessment and crop prediction in agriculture. However, the studies have gaps such as the lack of a comprehensive model for both water quality assessment and crop recommendation, and underutilization of deep learning and machine learning algorithms. Our proposed model addresses these gaps by using these algorithms to predict water quality and recommend crop types. Additionally, our model provides a cost-effective solution to common farmer problems with reliable and accurate recommendations. The limitations across the four references on crop recommendation systems include the lack of information on datasets used, need for further validation in real-world scenarios, generalizability of the systems, and lack of comparison with existing systems. Our study contributes to the literature by providing an efficient model for water quality assessment and crop recommendation.

III. METHODOLOGY

A. *Deep Learning Model*

There are in the least millions of neurons in a human brain, that interaction among each other and communicate information. The neuronal interactions usually occur via electrochemical signals. The parts of the neurons that are responsible for connecting them to others are known as a Synapse, and they are the location of the passage of electrochemical signals. Neural networks resemble the functioning of the human central nervous system [13], since deep neural networks comprise a huge number of processing units that are connected to each other [14]. Deep neural networks are an important concept of Machine Learning as they can process large amounts of data, and then use them to come up with a pattern from which it can learn [15]. In the majority of cases, neural networks are used for classification tasks because they have a robust and efficient capability of processing the datasets [16]. There are three basic layers in the artificial neural network, which are the input layer, hidden layers, and output layer (Fig. 3).

1) Input layer: In the input layer, each input is assigned a vector where the attributes are represented. However, an output must also be given in order to be able to evaluate the model based on its accuracy.

2) Hidden layer: The hidden layers consist of weights and thresholds that improve the attributes. Two main processes take place in the hidden layers which are the multiplication of weights and attributes, followed by sigmoid function that is used for output generation.

3) Output layer: The output layer is the level on which comparisons between the resultant output and the actual output take place. Depending on the similarities or differences, the feedback is given to the hidden layers. Additionally, permutation and combination of the weights and attributes occur in the output layer, to ensure better accuracy.

B. Forward propagation

In the case of forward propagation, the data flow is unidirectional in the sense of the output direction.in addition, no feedback is available which make the accuracy measurements difficult as a way of evaluating the model [17].

C. Back propagation

Back propagation is different from forward propagation since it is constantly training itself by looping the comparisons between the actual and the desired outputs. The comparison is then propagated to the error function which alters the weights of the hidden layers in order to significantly reduce the differences between the actual and the desired outputs.

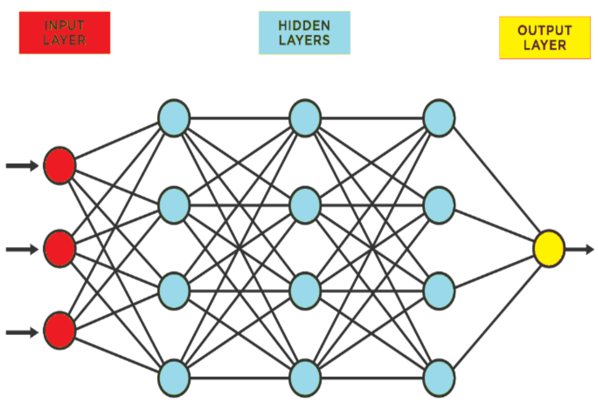

Fig. 3. Artificial neural network.

D. Activation Function

1) Sigmoid function: It can be used whenever the values are between 0 and 1. Where 0 represents the minimal probability of an event and 1 represents the maximum probability. Thus, for event predictions, the sigmoid function is very suitable. In addition, it is possible to find the derivative of the sigmoid function, being a curve between two points.

$$f(x) = \frac{1}{1+e^{-(x)}} \tag{1}$$

2) SoftMax function: It is similar to the sigmoid function, yet the output values in the SoftMax function are divided and can be summed up to a total of 1 [18]. Thus, it is like a probability distribution of the output values.

$$softmax(z_i) = \frac{exp(z_i)}{\sum_j exp(z_j)} \tag{2}$$

3) Rectified linear unit (ReLU) function: It is beginning to replace the sigmoid function. In this case, whenever the output value is below zero, it will be rounded up so that the output is zero. The output and input values are considered equal when the input value is greater than zero.

$$\mathrm{RELU}(x) = \begin{cases} 0 \text{ if } x < 0 \\ x \text{ if } x >= 0 \end{cases} \tag{3}$$

E. Support Vector Machines

Support vector machines (SVM) is a supervised machine learning method that is commonly used for solving classification and regression problems. It was initially developed by Vapnik and Chervonenkis, and has its roots in Statistical Learning Theory. SVM is designed to learn structure from given data and can handle both continuous and categorical variables. The model represents different classes in a hyperplane within a multi-dimensional space, and its objective is to categorize a dataset into different classes by identifying the maximum marginal hyperplane (MMH) [19].

SVM employs kernel functions to transform input data into a desired form. For non-linear problems, the kernel trick technique is utilized in SVM with the aid of slack variables and additional dimensions, which transforms the data into a higher dimensional space. SVM utilizes several types of kernels, which are listed in Table I.

TABLE I. DESCRIPTION OF WATER QUALITY PARAMETERS

Kernel Type	***Equation***
Polynomial	$k(x,y) = (ax^T y + c)^2$
Linear	$k(x,y) = (ax^T y + c)$
Sigmoid	$k(x,y) = tanh(ax^T y + c)$
Laplacian kernel	$k(x,y) = \exp\left(-\frac{\parallel x-y \parallel}{2\sigma}\right)$
Radial Basis Function (RBF)	$k(x,y) = \exp\left(-\frac{\parallel x-y \parallel^2}{2\sigma^2}\right)$

F. Proposed System Workflow

The following figure describes the workflow of the overall system including the steps required in both of its subsystems.

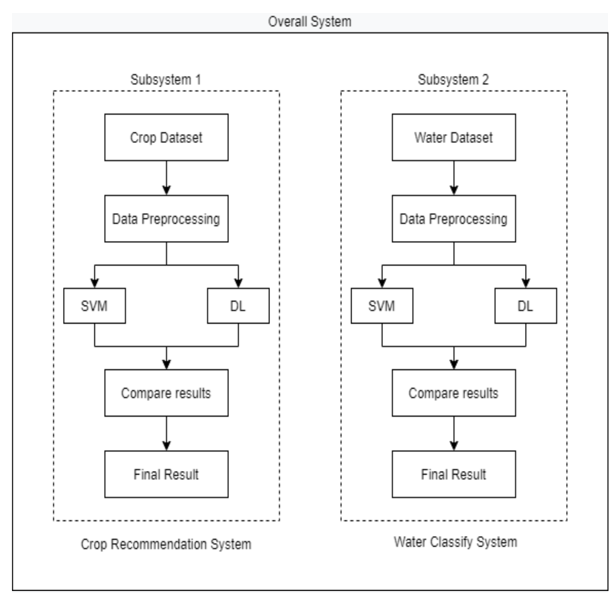

Fig. 4. The proposed workflow of our proposed prediction system.

Fig. 4 describes the general process performed in order to create our proposed prediction model. Our overall model is concerned with two separate tasks which are crop recommendation, and water quality classification. Thus, we decided to divide the system into two subsystems: subsystem one for crop recommendation, and subsystem two for water quality.

Mainly, similar processes take place in both of the subsystems, which slight differences. Initially, two different datasets are used for the different subsystems. In subsystem one, a crop dataset is acquired from Kaggle, where data selection and preprocessing take place. After that, a deep learning algorithm, namely neural network is implemented to see the results it can achieve, while SVM machine learning algorithm is also used for the same purpose. The objective is to determine for the case of crop recommendation, which perform better: neural network or SVM? The answer to this question can be obtained by comparing their results and thus determining whose result will be taken into consideration when recommending crops.

In subsystem two, a water quality assessment dataset is used and subjected to preprocessing, where it is later fed to a neural network and an SVM algorithm, in order to compare which one of them performs better in terms of classifying the quality of water which will be used for irrigating the crops.

G. Dataset Description and Preprocessing

1) Dataset Description

a) Crop recommendation dataset: For subsystem one, we use a dataset comprising the soil-specific attributes which are collected from online sources [20]. The crops considered in our model include 'rice', 'maize', 'chickpea', 'kidney beans', 'pigeon peas', 'moth beans', 'moonbeam', 'blackgram', 'lentil', 'pomegranate', 'banana', 'mango', 'grapes', 'watermelon', 'muskmelon', 'apple', 'orange', 'papaya', 'coconut', 'cotton', 'jute', 'coffee'.

Fig. 5 illustrates the number of instances or repetitions of each crop available in the training dataset. These crops differ in the climate that they need to grow and have good yields. For instance, the rice crop requires a lot of water, thus it is suitable for planting in areas with a lot of rainfall. The coffee crop is suitable to be planted in tropical regions. Jute for example is a crop that requires rainfall but also special soil conditions, thus it only grows in specific regions around India. Black gram crop conversely requires hot and humid climate to grow well and provide the best yield. Therefore, the different crop types require different climate conditions for their prosperity.

The five different attributes or parameters that were considered in the crop yield prediction dataset can be visualized in Fig. 6. For instance, the majority of data about the Phosphorous content fall between 45 and 60 units, whereas the Potassium values are always below 50 units. The pH value falls almost always between 6 and 7, humidity is above 80% most of the times, and the rainfall is between 40 to 120 units in most cases.

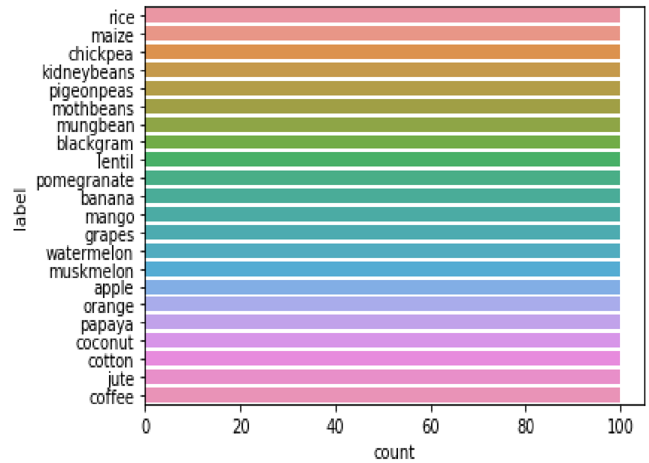

Fig. 5. Number of how many times each crop is present in the training dataset.

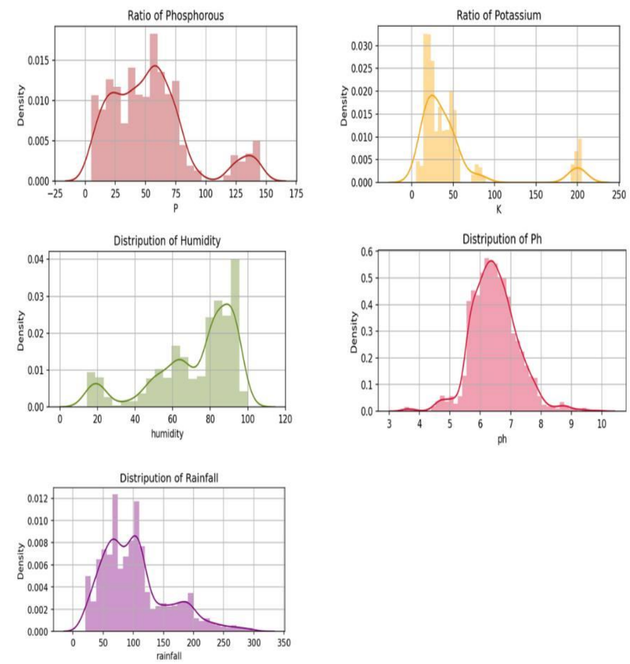

Fig. 6. Distribution of the crop prediction parameters.

b) Water potability dataset: The water quality dataset is acquired from Kaggle [21]. This dataset includes metrics for 3276 water bodies that differ in their water quality. This dataset is often used to determine the potability of the tested water. The parameters according to which the water is assessed are shown in Table II:

The nine different parameters that were considered in the water potability dataset are shown in Fig. 7. The ratio between the parameters being suitable or not is different and this difference in each of the features is used to predict the water quality. For instance, the difference between the values of turbidity being suitable for irrigation or insuitable for irrigation determines if this parameter in particular classifies the tested water as suitable for irrigation or not.

TABLE II. DESCRIPTION OF WATER QUALITY PARAMETERS

Parameter	Description
pH value	representing the acid-base balance in the water, since this factor can be quite harmful to human or natural life if it were significantly off.
Hardness	representing the amount of calcium and magnesium salts present in water.
Total dissolved solids	representing the collection of organic and inorganic minerals that can be dissolved in water.
Chloramines	representing the measure of disinfectants that remain in the water after its treatment.
Sulfate	it is a mineral that can be found groundwater and its concentration varies depending on location.
Conductivity	representing the degree to which the water is capable of conducting electricity based on its minerals content.
Organic carbon	representing the organic matter that can be found in water as a result of decaying natural or synthetic matter.
Trihalomethanes	representing the chemicals that can be detected in water after chlorine treatment.
Turbidity	describing the light emitting properties of water depending on the amount of solid matter in its suspension.
Potability	describing the safety of water for human drinking.

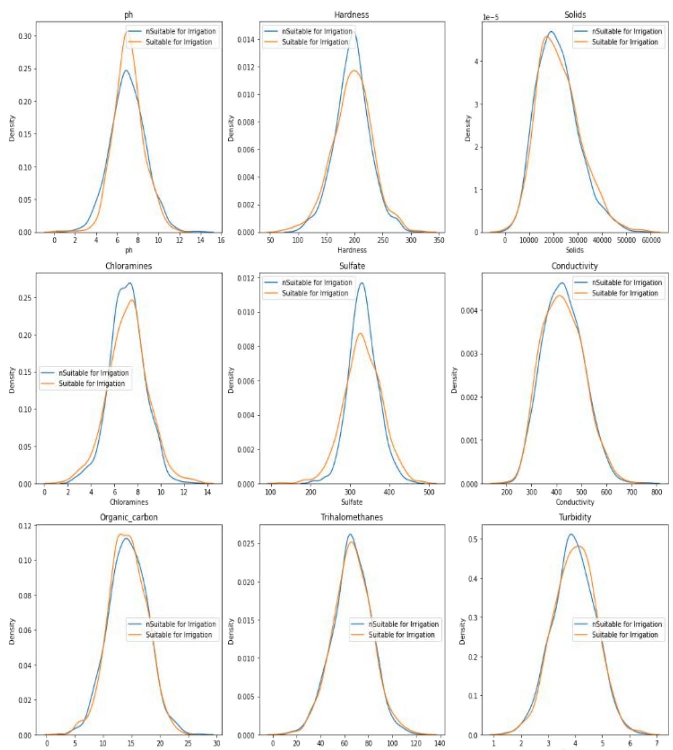

Fig. 7. Water quality parameter distribution.

2) Data preprocessing: The acquired data are not always clean, in fact they often include missing values, null values, and noise which make the dataset unfit to be used in ML or DL algorithms. Thus, data preprocessing is performed to clean the data and prepare it to be a suitable input for the algorithms. Preprocessing includes many tasks such as the removal of outliers or flawed data, as well as replacing the missing values if present. There are two techniques that can be performed to resolve the missing data, among which is deletion. Deletion means the removal of the entire row where the missing data reside. In the case of deletion, this might lead to significant reduction in the size of the dataset if it was already small. The other technique includes filling the missing spaces with the average or mean values of the attributes.

a) Crop recommendation system data preprocessing: The crop recommendation dataset did not contain any empty values. The data was entered to Standard Scaler in order to standardize the deatures since the input from the dataset have very different characteristics range. Feature standardization takes place through subtracting the mean and then scaling to unit variance, where unit variance is achieved by dividing all the values by the standard deviation.

b) Water data preprocessing: The water quality classification dataset contained some null values, thus these null value rows were dropped from the data so that the performance of the algorithms is not negatively affected by them. Just like the crop recommendation dataset, the data was entered to StandardScaler.

H. Experimental Set Up

In order to achieve the best results, the ML and DL algorithms must be trained and tested under a variety of scenarios. In this study, we trained models on the data so that it will predict the crop that can be grown based on various given parameters such as the soil nutrients and environmental factors. We give different set of input parameters and based on them we train the data to predict the exact crop to be grown. We fit the data to the X, Y training values and make predictions on the X test data. We trained the model for 100 epochs. The model with the lowest loss is considered as the best model and that model is used for evaluation and testing.

The DL model's performance is evaluated in the python environment. TensorFlow is a free deep-learning framework tool it offered by google. It provides a library of various models for data preprocessing, classification, clustering, forecasting, visualization, etc. The collab in which our experiments were conducted contains many powerful features that help the developer and researchers in the development and research process.

IV. RESULTS

A. Correlation for Water

Seaborn heat map function (fig. 8) was used to determine the correlations between the different factors. Each two factors affect each other to a certain degree, which is referred to as correlation.

The correlation matrix illustrates that each feature is strongly correlated to itself only (+1 score), and not to other features. In fact, the water potability/quality factors don't show any string correlations between each other, except for the two parameters pH and hardness, where a weak correlation exists (0.08).

These correlation results imply that dimension reduction is not possible in the water quality data due to the lack of correlation between its parameters.

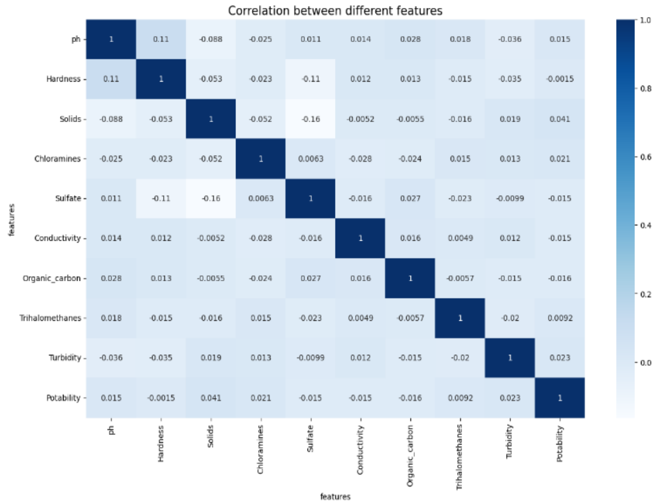

Fig. 8. Correlation between different features.

In the water quality classification dataset, all of the parameters are regarded as independent features, whereas the water potability parameter is the only dependent factor.

B. Evaluation Metrics

There are several metrics that can be used to evaluate the performance of the models such as precision, recall, f1 score, and accuracy [see (4), (5), (6) and (7)]. Recall is another term used for sensitivity, which resembles the true positive value, which is also the portion of the correctly classified inputs as positive among the entire inputs that should have been classified as positive. Precision is the portion of the true positive classifications over the entirety of the positive results. F-measure is the harmonic mean of the precision and recall and sums up the predictive performance of a model.

$$Recall = TP / (TP + FN) \quad (4)$$

$$Precision = TP / (TP + FP) \quad (5)$$

$$F\text{-}Measure = 2\ Precision \cdot Recall\ Precision + Recall \quad (6)$$

$$Accuracy = (TN + TP) / (TN + TP + FN + FP) \quad (7)$$

Where, True positive is designated by TP. True negative is designated by TN. False positive is designated by FP. False negative is designated by FN. Area under curve AUC is also a beneficial metric, where the values must be between 0 and 1, such that the higher the AUC value, the better the performance. If the model can discriminate between the instances of two classes perfectly, then AUC would be 1. Conversely, if the model fails to distinguish between any instances, the AUC would be 0.

C. System Results Evaluation

The evaluation metrics were obtained for the proposed water quality and crop recommendation for irrigation system during the training phase. The subsystem was evaluated based on accuracy, precision, f1 score, recall, and loss.

For the crop recommendation system, both the DL and SVM models achieved high levels of accuracy, precision, recall, and F1 score, with only slight variations between the two models. As shown in Fig. 9, the DL model achieved a slightly higher accuracy of 0.975, compared to the SVM model's accuracy of 0.968. Both models achieved high precision scores of 0.97 and high recall scores of 0.97 and 0.98, respectively. The F1 score was also high for both models, at 0.97. These results suggest that both the DL and SVM models were effective in predicting crop recommendations, with the DL model performing slightly better than the SVM model in terms of accuracy.

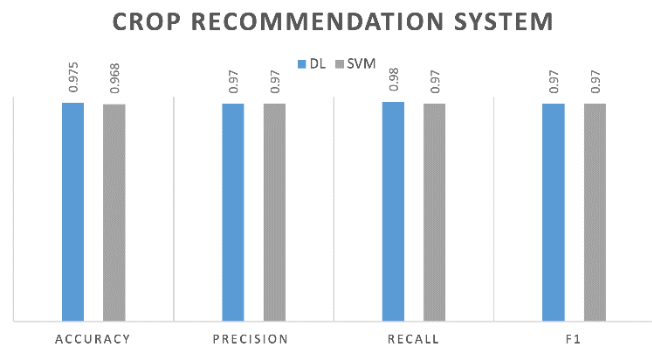

Fig. 9. Crop recommendation system.

As shown in Fig. 10, for the water classify system, the DL model achieved a higher level of accuracy (0.964) compared to the SVM model (0.927), as well as higher precision and recall scores (0.94 for both). The SVM model achieved a lower precision score of 0.914 and a lower recall score of 0.9. The F1 score for both models were relatively similar, with the DL model achieving 0.94 and the SVM model achieving 0.9. These results suggest that the DL model outperformed the SVM model in predicting water potability, with a higher accuracy, precision, and recall score. Overall, the DL model was more effective in classifying water samples as potable or non-potable compared to the SVM model.

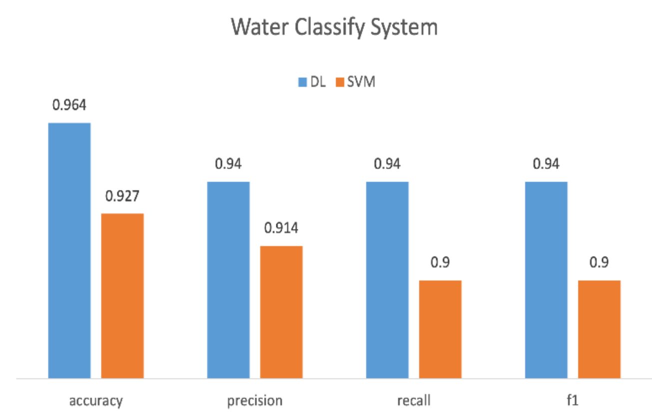

Fig. 10. Water classify system.

As a result, the SVM algorithm was able to achieve the accuracy less than neural network model by scoring 97% for crop recommendation for DL as shown in Fig. 11, and 96% for SVM. On the other hand, the water classify subsystem achieve 96% accuracy for DL as shown in Fig. 12, and 92% for SVM.

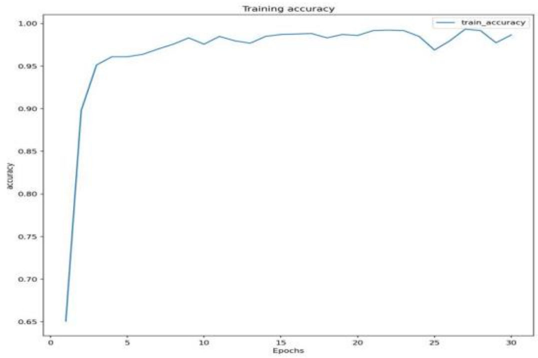

Fig. 11. Training accuracy for crop recommendation using DL.

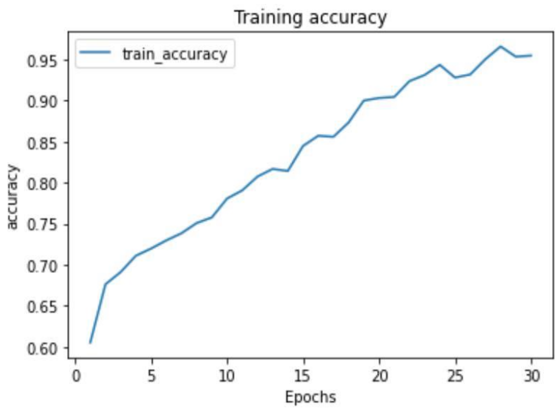

Fig. 12. Training accuracy for water classify using DL.

Compared to the related works mentioned in the literature review, our work stands out for providing a solution for both crop recommendation and water potability prediction, while some of the studies focused on only one of these tasks. While our work and the related studies all utilize datasets for crop recommendation systems, our approach stands out due to the unique nature of our dataset. Our dataset consists of soil-specific attributes that were collected from online sources, providing a comprehensive and informative dataset for crop recommendations. In contrast, the datasets used in the related studies are either unspecified or contain a limited number of crops. For example, [10] only contains 16 crops. Furthermore, our dataset includes 23 crops, which is a more extensive and diverse selection compared to some of the other studies that do not mention the number or types of crops considered. Therefore, our dataset is more comprehensive and suitable for accurate crop recommendations.

Additionally, our work used a variety of parameters for prediction, such as sun exposure, humidity, soil pH, and soil moisture, as well as water pH, electric conductivity, and content of minerals such as chloride, calcium, and magnesium. Some of the references also used similar parameters, but the methods varied, including decision trees, random forests, and Naive Bayes classifiers. Results from the four references show that accuracy ranged from 90% to 96.7% using different machine learning algorithms. However, the proposed system in our work achieved a higher accuracy for both crop recommendation and water quality prediction, demonstrating the effectiveness of the proposed model.

V. Conclusion

Two of the major issues or challenges in the proper maintenance of an agricultural field are the quality of irrigation water that is reaching the plants and that the plants need for proper growth, as well as the appropriate crop type since crops don't react and grow the same as each other in different climates and conditions.

In conclusion, the study proposed a binary model based on deep learning to address the challenges of water quality analysis and crop recommendation. The model was divided into two subsystems that relied on data collected from separate sources for training and testing. The performance of the model was evaluated using various metrics, including a confusion matrix, accuracy, recall, and precision. The neural network achieved high accuracy rates of 97% and 96% for crop recommendation and water quality prediction, respectively, while SVM achieved 96% and 92% accuracy. The results suggested that the binary model had the potential to serve as an effective tool for addressing the complex issues of water quality analysis and crop recommendation simultaneously.

The crop recommendation dataset used in the study has some limitations, including missing or incomplete soil-specific attributes, reliance on online sources for data collection, and a limited number of crops. Meanwhile, the water potability dataset may be limited by a lack of representativeness in sampled water bodies and incomplete or missing data, which could affect the accuracy of water potability predictions.

In the future, we might add to our dataset a wider collection of crops that the system can choose from based on the selected parameters. Additionally, we can implement other algorithms in the future to check if the accuracy can be improved, or if tweaking the algorithms, a bit can add more efficiency. We can also integrate IoT systems in order to be able to collect more data from the field for both the water quality prediction and the crop recommendation features.

References

[1] R. Lal, "Soil science and the carbon civilization," Soil Science Society of America Journal, vol. 71, pp. 1425-1437, 2007.

[2] H. D. Nguyen, T. Q. D. Nguyen, H. N. Thi, B. Q. Lap, and T. T. H. Phan, "The use of machine learning algorithms for evaluating water quality index: A survey and perspective," in 2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), Phu Quoc, Vietnam, 2022, pp. 1-6.

[3] N. Koralay and Ö. Kara, "Forestry activities and surface water quality in a watershed," European Journal of Forest Engineering, vol. 4, pp. 70-82, 2018.

[4] B. Phalan, R. Green, and A. Balmford, "Closing yield gaps: Perils and possibilities for biodiversity conservation," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 369, p. 20120285, 2014.

[5] D. Paudel, H. Boogaard, A. de Wit, S. Janssen, S. Osinga, C. Pylianidis, *et al.*, "Machine learning for large-scale crop forecasting," Agricultural Systems, vol. 187, p. 103016, 2021.

[6] G. Pratyusha and A. Sinha, "Prediction of crop using deep learning techniques: A concise review," in Recent advances in computer based systems, processes and applications, A. Namburu and S. S. Barpanda, Eds. Boca Raton: CRC Press, 2020, pp. 145-159.

[7] A. T. Ali El Bilali. (2020). Prediction of irrigation water quality parameters using machine learning. University of Hassan II, Casablanca, Faculty of Sciences and Techniques of Mohammedia, Morocco.

[8] Mokhtar Ali, El-Ssawy Wessam, He Hongming, Al-Anasari Nadhir (2022). Using Machine Learning Models to Predict Hydroponically Grown Lettuce Yield. Frontiers in Plant Science.

[9] P. Bandara, T. Weerasooriya, R. T.H, W. J. M. Nanayakkara, D. M.A.C, and P. M.G.P, "Crop recommendation system," International Journal of Computer Applications, vol. 175, pp. 22-25, 2020.

[10] A, Priyadharshini & Chakraborty, Swapneel & Kumar, Aayush & Pooniwala, Omen. (2021). Intelligent Crop Recommendation System using Machine Learning. 843-848. 10.1109/ICCMC51019.2021.9418375.

[11] Shafiulla Shariff, Shwetha R B, Ramya O G, Pushpa H, Pooja K R, 2022, Crop Recommendation using Machine Learning Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICEI – 2022 (Volume 10 – Issue 11),

[12] A. Jadhav, N. Riswadkar, P. Jadhav, and Y. Gogawale, "Crop recommendation system using machine learning algorithms," International Journal of Engineering Research & Technology vol. 9, pp. 1227-1230, 2022.

[13] S. Mishra, D. Mishra, and G. H. Santra, "Applications of machine learning techniques in agricultural crop production: A review paper," Indian Journal of Science and Technology, vol. 9, pp. 1-14, 2016.

[14] V. V. Erokhin and V. E. Eliseeva, "Influence of neural network architecture on its performance," Soft Measurements and Computing, vol. 8, pp. 5-17, 2022.

[15] M. I. C. Rachmatullah, J. Santoso, and K. Surendro, "A novel approach in determining neural networks architecture to classify data with large number of attributes," IEEE Access, vol. 8, pp. 204728-204743, 2020.

[16] K. Sulaiman, L. Hakim Ismail, M. Adib Mohammad Razi, M. Shalahuddin Adnan, and R. Ghazali, "Water quality classification using an artificial neural network (ANN)," IOP Conference Series: Materials Science and Engineering, vol. 601, p. 012005, 2019.

[17] T. R. Zhangirov, A. B. Krevchik, D. V. Zuev, A. A. Liss, A. V. Ekalo, and N. Y. Grigoryeva, "Forward propagation neural network weighting analysis as a model estimation method," in 2021 II International Conference on Neural Networks and Neurotechnologies (NeuroNT), Saint Petersburg, Russia, 2021, pp. 10-12.

[18] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," arXiv preprint arXiv:1710.05941, 2017.

[19] K. Lata, "Analysis of SVM and RNN-LSTM on Crop Datasets," in 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020, pp. 1-5.

[20] A. Ingle, "Crop Recommendation Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset. [Accessed: Mar. 31, 2023].

[21] A. Kadiwal, "Water Potability," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/adityakadiwal/water-potability. [Accessed: Mar. 31, 2023].