INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

*Corresponding author.

aaeldamarany@ju.edu.sa

An Intelligent Groundwater Management Recommender System

A A Abd El-Aziz[1,2]*, Khalaf Okab Alsalem[1], Mahmood A Mahmood[1,2]

1 Department of Information Systems, College of Computer and Information Sciences, Jouf University, KSA
2 Department of Information Systems & Technology, FGSSR, Cairo University, Egypt

Abstract

Objectives: To explore the area of groundwater that can assist to improve the accessibility of freshwater. **Methods :** We propose a machine-deep learning model based on a recommender system to manage and classify groundwater. **Finding**: The main goal of our proposed approach is to classify groundwater into multi-labels, which are drinking water (Excellent or Good) or Irrigation water (Poor or Very Poor) with guarantee a higher accuracy score. The recommender system is applied on the testing dataset and the accuracy of the deep learning technique was 91% and the accuracy of machine leaning technique was 84%.

Keywords: Groundwater Management; Intelligent System; Recommender Systems; Datamining; Machine Learning; Deep Learning

1 Introduction

Water is a common asset to serve people, creatures, plants, grass, and everyone, significant supporters of the government assistance and wellbeing of individuals. Lungs provide oxygen to the water which offers out to every piece of the body. Water secures a high percentage of the human body and standardized internal heat level. Water in seas and 1.7% of groundwater encase more than 70% of the earth. Groundwater addresses the biggest load of open freshwater and records for around 33% of freshwater withdrawals internationally.

In supplying fresh water for drinking, irrigation, and industrial operations, water supplies, including groundwater, play a critical role. Freshwater, found in the pore space of soil and stones, is sub-surface water or groundwater. There are approximately 790 billion cubic meters of water that seeps into the soil, approximately 430 billion cubic meters are expected to remain in the top layers of the soil and produce soil moisture that is necessary for vegetation growth. In the porous strata, the remaining 360 billion cubic meters percolate and reflect the real enrichment of underground water. The water which can be economically extracted from this is only around 255 cubic billion meters.

Sustainable groundwater management thus plays a major role in a country's overall growth. In recent decades, it has become clear in many countries around the world that groundwater is one of the most valuable natural resources. Groundwater has a range of important advantages as a source of water supply relative to surface water: as a rule, it is of higher quality, better protected from potential contaminants, including infection,

less susceptible to seasonal and perennial variability, and much more evenly dispersed than surface water across large regions. Very often, really often, in areas where there is no surface water, groundwater is available.

Recommender systems (RSs) are intelligent systems that use machine learning techniques and statistical methodologies to identify a set of client preferences. In the field of water, RSs will assist customers in selecting and implementing appropriate water management operations.

Sahraei Amir et al. [1] showed an examination that explored the relevance of Artificial Neural Network and Support Vector Machine to anticipate most extreme occasion water portions in streamflow utilizing precipitation, soil dampness, and air temperature as a bunch of informative information includes that are more direct and more affordable to gauge contrasted with stable isotopes of water, in the Schwingbach Environmental Observatory (SEO), Germany.

Janaki B. Mohapatra et al. [2] utilized Adaptive Neuro-Fuzzy Inference System, Deep Neural Network and Support Vector Machine for predicting occasional groundwater levels at the nation scale utilizing in situ groundwater-level and relevant meteorological data of 1996–2016.

Sudhakar Singha et al. [3] proposed a deep learning model for predicting groundwater quality. The proposed model was compared with three machine learning techniques. 226 groundwater samples are accumulated from horticulturally serious region; Arang of Raipur district, Chhattisgarh, India.

D. Hou et al. [4] established the Drinking Water Quality Early Warning and Control framework (DEWS), to respond to China's serious necessity for securing its urban drinking water. Water quality monitoring, early warning and dynamic water quality mishap capabilities are all available through the DEWS' web-based administration system. The usefulness of the DEWS was guided by control hypothesis and hazard assessment criteria as applied to the criticism control of urban water gracefully setups. The DEWS had been implemented in a few large Chinese cities and was shown to be effective in terms of water quality early warning and crisis dynamics. The suggested framework integrates water well-being technology and theories. Our model is different in that it determines whether the desalinated water is potable or arable. N.

Bassiliades et al. [5] proposed a sophisticated framework for observing and forecasting water quality based on two distinct systems: Andromeda (ocean waters) and Interrisk (inland waters) (freshwaters and surface air). When a specific ecological boundary crosses specified contamination limitations, a fuzzy expert system was used to provide early warnings. Furthermore, machine learning and adaptive filtering approaches were used to predict specific water quality boundaries one day ahead of time in order to avoid unfavorable natural events. In our model, we utilized a recommender system, machine, and deep learning to analyze the groundwater.

Gino Sophia et al. [6] proposed a genetic task-based intelligent system that uses fitness values and a neural network to prepare the model. The fitness approach was used to develop new intelligent persons for water collection and dispersion from the existing population of water assets. The framework is significant for expectations about water usage, appropriation, and expansion of advancement execution by determining target capacity of various population types using decision-making procedures. However, the proposed model used the feature selection technique to select the attributes that effect on the decision label. Four criteria were employed by

F. S. Alahmadi [7] (pH, TDS, latitude, and longitude). Two methodologies for multivariate group evaluation use latitude and longitude to spatially partition the groundwater quality in Madinah, in western Saudi Arabia. The results showed that the two unsupervised machine learning strategies (k mean and k medoids) produced generally similar results, with the examination region partitioned into three principal locales: southwest, southeast, and north, with different designs. The study found that multivariate bunch analysis, as a reliable fact-gathering tool for water quality boundaries, can provide useful information in water quality management. However, in our model, we used the feature selection technique to select the attributes that effect on the decision label, and we used machine learning and deep learning techniques.

J. Inoue et al. [8] investigated the use of unsupervised machine learning in Cyber-Physical Systems strangeness detection (CPSs). They created a Deep Neural Networks (DNN) that responded to time-series data and evaluated the presentation of a probabilistic anomaly indicator against a one-class SVM. The information from the Secure Water Treatment (SWaT), a reduced but fully working crude water sanitization plant, is compared to these methodologies. This is different than our model that used recommender systems.

Through using machine learning methods to distinguish water-influenced habitations, N. Yuvaraj et al. [9] created a healthcare recommendation system. To define the water-influenced residences and provide critical recommendations, the authors used recommender systems. While in our model, we used a recommender system to analyze the groundwater and used the feature selection technique to select the attributes that effect on the validity of the groundwater for drinking or for irrigation.

S. Adnan et al. [10] provided a simple technique for classifying groundwater quality based on a variety of attractive elements. To construct the technology and create groundwater quality classifications, researchers used hierarchical clustering analysis

(HCA) and classification and regression tree (CART). The tool was given ten factors to work with as inputs. This tool could be used to determine groundwater quality prior to the construction of new water supplies; however, adding more groundwater quality factors will improve the results. The tool can also be used to prioritize groundwater quality areas, and its validity can be validated by conducting a full spatial-temporal analysis of groundwater quality in the same or a different location. The outputs of the tool are the groundwater quality classes, but our output is whether the groundwater is potable or arable.

A. S. Salman et al. [11] assessing the groundwater quality in the Northern Western zone of KSA utilizing the four traditional water facies diagram. The authors utilized the Principal Component Analysis (PCA) for deciding the elements controlling the groundwater chemistry and stochastic geostatistics to comprehend the spatial distribution of various significant components in the groundwater of the Tabuk-Madina zone were utilized.

R. Kamakshaiah and K. Kamakshaiah Seshadri [12] presented a study on groundwater quality, sources of groundwater pollution, groundwater quality diversity, and regional distribution of groundwater quality. Groundwater bodies and agent observing networks are the reason for groundwater quality assessment, which allows for the determination of the chemical status of the groundwater body. The water samples were tested for four Physico-chemical factors utilizing standard methods in the research facility and contrasted and the principles.

To assess drinking water quality, A. Al-Omran et al. [13] divided Riyadh governorate, Saudi Arabia, into five areas: Riyadh primary zone, Ulia, Nassim, Shifa, and Badiah areas. The basic water network, as well as the subterranean and upper residential tanks, was all used to collect water in each neighborhood. A mathematical method called Water Quality Index (WQI) is utilized to ease water quality explanation. The WQI was computed by four physico-chemical and microbial factors. This study reasoned that the density of weighty components of all examined water samples was discovered to be inside safe cutoff points. The outcomes likewise showed that the Riyadh main area had the noteworthy absolute number of microbes followed by Ulia, Albadyah, Shifa, and then Alnassim zone.

A. Khater Asma et al. [14] examined the water quality of various brands of Bottled Drinking Water (BDW) used in Saudi Arabia and compared the results to BDW standards. The authors analyzed the level of BDW quality using seven parameters.

Mallick, J et al. [15] concentrate on various water-quality factors, for groundwater quality assessment and spring weakness evaluation due to poisons/foreign substances present in groundwater. The current review gives an encompassing comprehension of various groundwater quality issues and subsequently distinguishes the holes of the past examinations and recognizes the perspectives of things to come research measurements. They portray the distinctive groundwater quality issues identified with poison levels of the fluoride, nitrate, and weighty metals and radionuclides in Saudi Arabia. A greater part of the groundwater poisons are of normal beginning, yet there is huge wastewater profluent release in the locale that is additionally liable for pollution of springs with substantial metals.

Pawlicka, A. et al. [16] have introduced the consequences of a wide, deliberate investigation of the potential utilization of recommender systems in cybersecurity. A few hundred recent works were set apart as conceivably important and afterward painstakingly broke down. A few papers introducing the executions of recommender systems in cybersecurity were found and portrayed. All things considered, the review showed that recommender systems could without a doubt be applied to help the human digital safeguard in their choices, and add to a safer, more secure cyberspace.

Fayyaz, Z et al. [17] introduced a point-by-point study of recommender systems that present various kind of recommender systems as shared sifting, content-based, segment based, utility-based, information based, and mixture based. Distinctive blend procedures of mixture based frameworks are additionally introduced and ordered into weighted, blended, exchanging, include mix, highlight expansion, course, also, meta-level. They introduced four primary difficulties that influence the exhibition of a suggestion framework, including cold-start, information sparsity, adaptability and variety, and measurements used to assess its execution.

However, our proposed model is different from than above techniques, where we used a feature selection technique on the training and testing dataset and compare between the results of a model built by a recommender system and machine learning and the results of a model built by using a recommender system and deep learning.

Table 1 summarizes the differences between our contribution and the recent related work:

Table 1. A Comparison between our Proposed Approach and the recent related work

Reference	Methodology
[9]	Proposed framework coordinates between technologies and theories for water wellbeing.
[10]	A fuzzy expert framework was utilized to give early cautions when certain ecological boundary surpasses certain contamination limits. Moreover, machine learning and adaptive filtering methods were utilized for one-day ahead expectations of specific water quality boundaries to forestall unfortunate natural circumstances.
[11]	Used fitness value and neural network to propose an intelligent system made out of genetic tasks. However, the proposed model used the feature selection technique to select the attributes that effect on the decision label.
[12]	Used four factors (pH, TDS, latitude, and longitude) and k mean and k medoids.
[13]	Used a Deep Neural Networks (DNN), which adapted to time-series data that actualized a probabilistic anomaly indicator and looked at its presentation against a one-class SVM.
[14]	Used recommender system and machine learning to propose a healthcare recommendation system by performing machine learning algorithms to recognize the water-influenced habitations.
[15]	Used Hierarchal clustering analysis (HCA) and classification and regression tree (CART) to implement a simple tool for groundwater quality classification based on many stylish constituents and create groundwater quality classes.
[16]	Used four traditional water facies diagram and utilized the Principal Component Analysis (PCA) for deciding the elements controlling the groundwater chemistry and stochastic geostatistics to comprehend the spatial distribution of various significant components in the groundwater of the Tabuk-Madina zone in KSA.
[17]	utilized standard methods in the research facility and contrasted and the principles
[18]	A mathematical method called Water Quality Index (WQI) is utilized to ease water quality explanation.
[19]	Used seven factors to analyze the quality level of water of different brands of Bottled Drinking Water (BDW) utilized in KSA and matched the quality levels to the BDW standards.
Proposed Technique	Used a machine learning model, deep learning model and a recommender system to management and classify groundwater. We used one deep learning model called Multi-Class Cross-Entropy Loss. The proposed approach classified groundwater into multi-labels, which are drinking water (Excellent or Good) or Irrigation water (Poor or Very Poor) with guarantee a higher accuracy score.

Therefore, in this study, we propose a machine-deep learning model based on a recommender system to management and classify groundwater. The main goal of our proposed approach is to classify groundwater into multi-labels, which are drinking water (Excellent or Good) or irrigation water (Poor or Very Poor) with guarantee a higher accuracy score. The recommender system is applied on the testing dataset.

2 Methodology

DataSet Description

2.1 NWIS (National Water Information System)

This water information is gathered at over 1.5 million destinations and regional locales. This appropriated organization of PCs is known as the National Water Information System (NWIS). Many sorts of information are put away in NWIS, including complete data for site qualities, well-development subtleties, time-series information for gage tallness, streamflow, groundwater level, precipitation, physical and substance properties of water and water use information. Also, top streams, substance examinations for discrete examples of water, residue, and natural media are open inside NWIS[18].

Test data from the former GEOTHERM dataset are incorporated. Since NWIS data is kept on a State-by-State premise, test and site data close to borders were deficient and various inquiries were important to get all the data. NWIS site and test data are effectively in independent tables in a one-to-many relationship, so allocating sites to test data was excessive. NWIS names normally comprise the latitude/longitude coordinates of the site along with a number or name; we have adjusted the NWIS name in the 'Name' fields to make them more coherent. Some NWIS data fields were joined were functional (for instance, fields P00056 (flow rate, gallons per day), P00058 (flow rate, gallons per minute), and P00059 (flow rate, instantaneous gallons per minute) were consolidated and changed over to a single field 'Flow rate, liters/minute'. As referenced above, data with alphanumeric fields (" <"," >", etc.) were eliminated and set in isolated tables. A few strengths follow or isotope fields with very small data in them (by and large if fewer than 10 of 42,000 test data) were not held in the dataset.

Table 2. The Statistics of the Dataset.

Type	Class	No. of Objects
Drinking	Excellent	34516
	Good	9915
Irrigation	Poor	1892
	Very Poor	77
Total		46400

3 Machine and Deep Learning Techniques

3.1 Kernelized Support Vector Machine

kernelized support vector machine (SVM) provides more models that can go beyond linear decision boundaries. As with other supervised machine learning methods, SVM can be used for both classification and regression. Kernelized SVM takes the original input data space and transforms it to a new higher dimensional feature space, where it becomes much easier to classify the transform to data using a linear classifier. This idea of transforming the input data points to a new feature space where a linear classifier can be easily applied is a very general and powerful one. There are lots of different possible transformations that could be applied to data and the different kernels available for the kernelized SVM correspond to different transformations (i.e., the kernel is used for data transformation), such as radial basis function kernel (RBF) and polynomial kernel. Table 3 presents the pros and cons of the SVMs.

Table 3. Kernelized Support Vector Machine: Pros and Cons.

Pros	Cons
Perform well on dataset	Its speed decrease as the dataset size increases.
versatile: different Kernel functions can be utilized.	Needs data normalization and parameters tuning.
used for low and high dimensional data.	Difficult to interpret its predictions.

On the positive side, support vector machines perform well on a range of datasets and have been successfully applied on data that range from text to images and many more types. The support vector machine's also potentially very versatile, due to its ability to specify different kernel functions, including possible custom kernel functions depending on the data. Support vector machines also typically work well for both low and high-dimensional data. Including data with hundreds, thousands, or even millions of sparse dimensions. This makes it well suited to test classification.

On the negative side, as the training set size increases, the run time, speed and memory usage in the SVM training phase also increase. Hence, for a large dataset with hundreds of thousands, or millions of instances, an SVM may become less practical.

When applying a support vector machine to a real-world dataset, an SVM requires careful normalization of the input data as well as parameter tuning. The input should be normalized that all features have comparable units and round similar scales if they aren't already. Finally, it could be difficult to interpret the internal model parameters of a support vector machine, which means the applicability of support vector machines in scenarios where interpretation is important for people may be limited when they want to understand why a particular prediction was made [19].

3.2 K-Nearest Neighbors Classification

The K-Nearest Neighbors algorithm (KNN) can be used for classification and regression. KNN classifiers are an example of instance-based or memory-based supervised learning, which means that instance-based learning methods memorize the labeled examples seen in the training set and then utilize those learned examples to categorize new objects. The K in KNN stands for the number of closest neighbors that the classifier will obtain and utilize to create a prediction. When K is small, like one, the classifier does a decent job of learning the classes for individual points in the training set, but with a decision boundary, the training set becomes fragmented and variable. This is because when K = 1, individual data points are more susceptible to noise, outliers, mislabeled data, and other sources of volatility. The areas assigned to distinct classes get smoother and less fragmented as K increases, and they become more resilient to noise in individual points. As a result, the value of K has an impact on the classifier's accuracy. The KNN method, in particular, comprises three phases that can be specified:

- When given a new previously unseen instance of something to classify, a KNN classifier will investigate its set of memorized training examples to find the K examples that have the closest features.
- The classifier will look up the class labels for those k-Nearest Neighbor examples.
- The classifier will combine the labels of those examples to make a prediction for the label of the new object.

The mathematical model of KNN is the function h: X!Y. Hence, having an unknown observation x, h(x) can positively predict the identical output y[20].

3.3 Bagging

Bootstrap aggregating is called bagging. Bagging is an ensemble machine learning technique. It is used to enhance the precision and the stability of machine learning techniques utilized in regression and classification. Bagging diminishes difference and assists with staying away from overfitting. Although it is generally applied to decision tree strategies, it tends to be utilized with any sort of strategy. Bagging is a particular instance of the model averaging method[21].

3.4 Decision Stump

A decision stump is a Decision Tree that utilizes just a single feature for splitting. For discrete features, this normally implies that the tree comprises just of a single inside hub (i.e., the root has just left as replacement hubs). If the feature is numerical, the tree might be more mind-boggling. Decision stumps perform surprisingly well on some ordinarily utilized benchmark datasets from the UCI repository, which shows that learners with a high Bias and low Variance may perform well since they are less inclined to Overfitting. Decision stumps are additionally regularly utilized as frail learners in Ensemble Methods, such as boosting[22].

3.5 Multi-Class Cross-Entropy Loss function

cross-entropy is a deep learning method that regularly utilized loss functions for classification tasks. Cross-entropy loss estimates the performance of a classification model whose output is a likelihood value between 0 and 1. Cross-entropy loss increments as the anticipated likelihood diverge from the actual label. So, anticipating a likelihood of .012 when the actual observation label is 1 would be terrible and result in a high loss value. An ideal model would have a log loss of 0.

4 Recommender System

Nowadays, with the high accessibility of data, the wide utilization of social networks, and the fast development of the web caused a gigantic amount of data, it requires an intricate interaction to separate valuable information that can be introduced to the client to assist him with dealing with the information appropriately prompting settling on right choices. Research has been done that helps in overseeing data to produce this valuable information[23]. Numerous clients are keen on frameworks that recommend few items to them dependent on specific elements, and thus a framework is required to help clients in choosing an item or anything mulling over that the client may have less information about the space, we call this framework a recommender system (RS)[24] defines RS as an intelligent system that gives counsel to the client about a particular item supporting him in the decision-making cycle.

In RS, data mining is utilized to explain the aggregations of analysis methods that are utilized to conclude recommendation rules or construct recommendation models from huge data sets. RS combines techniques that form their recommendation utilizing knowledge from client features and activities. It is frequently dependent on the improvement of the client profile. These techniques include association rules, clustering and classification, and the creation of resemblance charts through various methods[25].

5 Proposed methodology

To know the validity of groundwater, whether, for drinking or irrigation, we propose a model that coordinates between a recommendation system and machine and deep learning techniques to analyze groundwater to find its factors and determine whether it is potable or arable. The proposed model consists of the following phases as shown in Figure 1 :

1. Collecting the Ground Water dataset.
2. Data splitting: The dataset is divided into a training dataset and a testing dataset.

3. Dimensionality decreasing: In this phase, we conquer the issue of high-dimensional datasets by selecting the important features of the training and testing datasets. The selection is done by using the Pearson Correlation Coefficient technique.

4. Building a Recommender System Phase: A recommender system is built to solve the sparsity problem of the testing dataset.

5. Applying machine learning techniques: For improving the accuracy, we use the following machine learning techniques: SVM, KNN, Bagging, and Decision Stump, and select the technique that gives the high accuracy.

6. Applying deep learning methods: We apply the CNN with the Multi-Class Cross-Entropy Loss function.

7. Evaluating the Model: The performance of the proposed model is evaluated using the accuracy, mean absolute error, and root mean squared error.
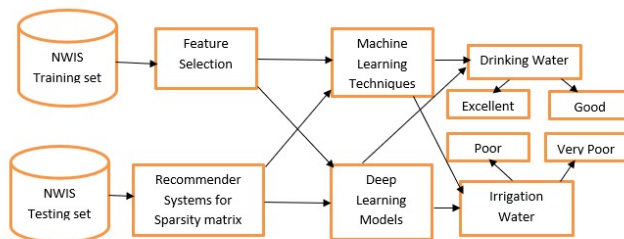

Fig 1. The Proposed Model

5.1 Data Splitting

The dataset is divided into 75% training dataset and 25% testing dataset. The training dataset is complete, but the testing dataset has a sparsity problem.

5.2 Dimensionality Decreasing

The Pearson Correlation Coefficient technique was applied to reduce the dimensional of the dataset. The Pearson correlation Coefficient technique selects the features that affect the decision label. The dataset has eleven features, which are Station ID, STATION TYPE, SAMPLE DATE, Dissolved Calcium, Dissolved Chloride, Dissolved Magnesium, Dissolved Potassium, Dissolved Sodium, pH, Total Dissolved Solids, and Class. After applying the feature selection techniques, the features are reduced to six features, which are SAMPLE DATE, Dissolved Calcium, Dissolved Magnesium, Dissolved Potassium, pH, and Total Dissolved Solids. The six features effect on the decision label.

5.3 Building a Recommender System

The proposed recommender system algorithm is used on the testing dataset to solve the sparsity problem of the dataset. The proposed recommender system is presented in the following algorithm:

Algorithm 1 The Steps of the Proposed Model

Input: Dataset (DT) with sparsity data of Testing Set and (DR)Dataset of Training set without sparsity data.

Output: Dataset (DU) without sparsity data.

Begin:

1. Create a new table (DU).
2. for Ai j 2 DT do:
3. if Ai j is null then:
4. compute the similarity score for each element in (DR) by using cosine similarity:
   $$\text{sim}(u, V) = \frac{u \cdot v}{\|u\|\|v\|}$$
5. if sim(u,v) $\geq$ 0.85 then :
6. Insert sim(u,v) into (simData).

7. End If.
8. End If.
9. compute the average of (simData) and insert into corresponding DU.
10. End For.

End

5.4 Classification by Machine Learning Techniques

The machine learning techniques, SVM, KNN, Bagging, and Decision Stump are used to classify the dataset. Moreover, we evaluate the performance of each machine learning technique by calculating the accuracy, mean absolute error and root mean squared error. Finally, we select the technique that gives the high accuracy and minimum mean absolute error and root mean squared error.

5.5 Classification by Deep Learning Techniques

A deep learning method called CNN with the Multi-Class Cross-Entropy Loss function is used to classify the dataset. In addition, we evaluate its performance by calculating the accuracy, mean absolute error and root mean squared error.

6 Results and Discussion

This paper proposes an intelligent recommender model based on machine and deep leaning techniques. The recommender system is applied on the testing dataset to solve the sparsity problem of the testing dataset. The proposed model analyzes the groundwater to find its factors and determines whether it is potable or arable. If the groundwater is suitable for drinking, it will be classified as excellent or good. If the groundwater is suitable for irrigation, it will be classified as poor or very poor. The experiments are performed by Weka 3.8.338 for machine learning techniques and python for the deep learning method on a PC with a 1.8 GHz Intel Core i7 processor and 16GB RAM.

According to the methodology, after splitting the data, decreasing the dimensionality by Pearson correlation coefficient method, and applying the recommender system on the testing dataset to solve the sparsity problem. We apply the machine learning techniques (SVM, KNN, bagging, and decision stump) and deep learning by using the multi-class cross-Entropy loss function. Table 3 shows a comparison between the machine learning techniques. As shown in Table 4 the accuracy of the decision stump is 91%and the accuracy of bagging is 84%. Since the mean absolute errors of decision stump and bagging are 0.0785 and 0.026 respectively, and the root means squared errors f decision stump and bagging are 0.1173 and 0.0976, respectively. The multi-class cross-entropy loss has an accuracy of 85% and the mean absolute error and the root mean squared error are 0.0295 and 0.1034, respectively. Hence, the decision stump is the recommended machine learning classification technique.

Table 4. A Comparison Analysis between Machine Learning Techniques and Deep Learning.

Model	Accuracy	MAE	MSE
SVM	0.4468	0.02287	0.2512
KNN	0.699	0.0321	0.1791
Bagging	0.8403	0.0264	0.0976
Decision Stump	0.9161	0.0785	0.1173
Multi-class Cross-Entropy Loss	0.8504	0.0295	0.1034

7 Conclusion

In this paper, we proposed a machine-deep learning model based on a recommender system to management and classify groundwater. To this end, we used one deep learning model called Multi-Class Cross-Entropy Loss. The main goal of our proposed approach was to classify groundwater into multi-labels, which are drinking water (Excellent or Good) or Irrigation water (Poor or Very Poor) with guarantee a higher accuracy score. Before evaluating the proposed method, we compared the performances of different machine learning models. The models were trained and tested using a database comprising 46,400 records. The results showed that the Decision Stump model achieved the best accuracy score compared to other machine

learning models. To evaluate the deep learning model, we applied multi-class classification. Our results showed that the Multi-Class Cross-Entropy Loss deep learning model improves the error rate compared to the Decision Stump machine learning model, but the accuracy score is less than the Decision stump accuracy score. On the other hand, our model-derived that the efficiency of the result of deep learning is not better than deep learning in numeric classification problems, so machine learning can be giving better results for numerical datasets than deep learning. The limitation of this approach is the dataset must be preprocessed before the approach execute. Also, the approach uses chemical parameters of water ground, so need sensors the water analysis into chemical parameters. Moreover, development out approach to solve this limitation in future work.

Acknowledgement

References

1) Sahraei A, Chamorro A, Kraft P, Breuer L. Application of Machine Learning Models to Predict Maximum Event Water Fractions in Streamflow. *Frontiers in Water*. 2021;3:52. Available from: https://dx.doi.org/10.3389/frwa.2021.652100.

2) Mohapatra JB, Jha P, Jha MK, Biswal S. Efficacy of machine learning techniques in predicting groundwater fluctuations in agro-ecological zones of India. *Science of The Total Environment*. 2021;785:147319. Available from: https://dx.doi.org/10.1016/j.scitotenv.2021.147319.

3) Singha S, Pasupuleti S, Singha SS, Singh R, Kumar S. Prediction of groundwater quality using efficient machine learning technique. *Chemosphere*. 2021;276:130265. Available from: https://dx.doi.org/10.1016/j.chemosphere.2021.130265.

4) Hou D, Song X, Zhang G, Zhang H, Loaiciga H. An early warning and control system for urban, drinking water quality protection: China's experience. *Environmental Science and Pollution Research*. 2013;20(7):4496–4508. Available from: https://dx.doi.org/10.1007/s11356-012-1406-y.

5) Bassiliades N, Antoniades I, Hatzikos E, Vlahavas I, Koutitas G. An Intelligent System for Monitoring and Predicting Water Quality. In: Proc. Of the European conference TOWARDS eENVIRONMENT. 2009;p. 534–542.

6) Sophia SGG, Sharmila VC, Suchitra S, Muthu TS, Pavithra B. Water management using genetic algorithm-based machine learning. *Soft Computing*. 2020;24(22):17153–17165. Available from: https://dx.doi.org/10.1007/s00500-020-05009-0.

7) Alahmadi FS. Groundwater Quality Categorization by Unsupervised Machine Learning in Madinah. *Proceedings of the International Geoinformatics Conference (IGC2019)*. 2019. Available from: www.researchgate.net/publication/332244617.

8) Inoue J, Yamagata Y, Chen Y, Poskitt CM, Sun J. Anomaly Detection for a Water Treatment System Using Unsupervised Machine Learning. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE. 2017. Available from: https://ieeexplore.ieee.org/document/8215783.

9) Yuvaraj N, Anusha K, Meagavarsha R. Healthcare Recommendation System for Water AffectedHabitations using Machine Learning Algorithms. *International Journal of Pure and Applied Mathematics*. 2018;118(20):3797–3809. Available from: https://acadpubl.eu/hub/2018-118-21/articles/21e/10.pdf.

10) Adnan S, Iqbal J, Maltamo M, Bacha MS, Shahab A, Valbuena R. A Simple Approach of Groundwater Quality Analysis, Classification, and Mapping in Peshawar, Pakistan. *Environments*. 2019;6(12):123. Available from: https://dx.doi.org/10.3390/environments6120123.

11) Salman AS, Zaidi FK, Hussein MT. Evaluation of groundwater quality in northern Saudi Arabia using multivariate analysis and stochastic statistics. *Environmental Earth Sciences*. 2015;74(12):7769–7782. Available from: https://dx.doi.org/10.1007/s12665-014-3803-7.

12) Kolli K, Seshadri R. Ground Water Quality Assessment using Data Mining Techniques. *International Journal of Computer Applications*. 2013;76(15):39–45. Available from: https://dx.doi.org/10.5120/13324-0885.

13) Al-Omran A, Al-Barakah F, Altuquq A, Aly A, Nadeem M. Drinking water quality assessment and water quality index of Riyadh, Saudi Arabia. *Water Quality Research Journal*. 2015;50(3):287–296. Available from: https://dx.doi.org/10.2166/wqrjc.2015.039.

14) Khater AEM, Al-Jaloud A, El-Taher A. Quality Level of Bottled Drinking Water Consumed in Saudi Arabia. *Journal of Environmental Science and Technology*. 2014;7(2):90–106. Available from: https://dx.doi.org/10.3923/jest.2014.90.106.

15) Mallick J, Singh CK, Almesfer MK, Singh VP, Alsubih M. Groundwater Quality Studies in the Kingdom of Saudi Arabia: Prevalent Research and Management Dimensions. *Water*. 1266;13(9):1266–1266. Available from: https://doi.org/10.3390/w13091266.

16) Pawlicka A, Pawlicki M, Kozik R, Choraś RS. A Systematic Review of Recommender Systems and Their Applications in Cybersecurity. *Sensors*. 2021;21(15):5248. Available from: https://dx.doi.org/10.3390/s21155248.

17) Fayyaz Z, Ebrahimian M, Nawara D, Ibrahim A, Kashef R. Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Applied Sciences*. 2020;10(21):7748. Available from: https://dx.doi.org/10.3390/app10217748.

18) . . Available from: https://help.waterdata.usgs.gov/faq/additional-background.

19) Sharma RK, Kalita HK, Borah P. Analysis of Machine Learning Techniques Based Intrusion Detection Systems. *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*. 2016;p. 485–493.

20) Keller JM, Gray MR, Givens JA. A fuzzy K-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*. 1985;SMC-15(4):580–585. Available from: https://dx.doi.org/10.1109/tsmc.1985.6313426.

21) uhlmann PB. Bagging, Boosting and Ensemble Methods. In: Handbooks of Computational Statistics book series. 2011;p. 985–1022. Available from: https://www.springer.com/series/7286.

22) Sammut C, Webb GI, editors. Decision Stump. Springer US. 2010.

23) Khaled S, Mahmood MA, Azab AE, Hefny HA. A Survey of Recommender Systems and Geographical Recommendation Techniques. In: GISApplications in the Tourism and Hospitality Industry. 2018;p. 249–274. Available from: https://www.igi-global.com/chapter/a-survey-of-recommender-systems-and-geographical-recommendation-techniques/202691.

24) Mahmood MA, Al-Shammari ET, El-Bendary N, Hassanien AE, Hefny HA. Recommender system for ground-level Ozone predictions in Kuwait. In: 2013 Federated Conference on Computer Science and Information Systems. 2013;p. 107–110. Available from: https://ieeexplore.ieee.org/document/6643983.

25) Mahmood MA, El-Bendary N, Platoš J, Hassanien AE, Hefny HA. An Intelligent Multi-agent Recommender System. *Innovations in Bio-inspired Computing and Applications*. 2014;p. 201–213. Available from: https://link.springer.com/chapter/10.1007/978-3-319-01781-5_19.