

# Comparative Analysis of Machine Learning and Deep Learning Models for Groundwater Potability Classification <sup>†</sup>

Ahmad Abubakar Suleiman <sup>1,2,\*</sup>, Arsalaan Khan Yousafzai <sup>3,4</sup> and Muhammad Zubair <sup>5</sup>

<sup>1</sup> Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia

<sup>2</sup> Department of Statistics, Aliko Dangote University of Science and Technology, Wudil 713281, Nigeria

<sup>3</sup> Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia

<sup>4</sup> Department of Civil Engineering, University of Engineering & Technology Peshawar 25000, Pakistan

<sup>5</sup> Department of Computer and information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia

\* Correspondence: ahmad\_22000579@utp.edu.my (A.A.S.)

† Presented at the title, place, and date.

**Abstract:** Ensuring access to safe drinking water is a critical concern, particularly in regions with limited resources. This study evaluates groundwater potability using a range of machine learning models, including logistic regression, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Random Forest, as well as deep learning models such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Feedforward Neural Networks (FNN), and Long Short-Term Memory (LSTM). We collected thirty groundwater samples from residential and industrial locations in Jaen, Kano State, Nigeria, focusing on nine crucial physicochemical parameters: electric conductivity, pH, total dissolved solids, calcium, magnesium, chloride, zinc, manganese, and copper. Machine learning models, such as logistic regression and random forest, achieved accuracy scores of 0.833. They were closely followed by deep learning models, such as ANN with an accuracy score of 0.833, and LSTM, which scored 0.666. KNN and SVC provided moderately accurate predictions, scoring 0.667, while CNN and FNN achieved lower scores of 0.333 and 0.5, respectively. This study represents a significant step toward ensuring safe drinking water for communities and preserving the sustainability of natural resources.

**Keywords:** groundwater; artificial intelligence; machine learning; deep learning; classification, logistic regression, random forest, artificial neural network, convolutional neural network

## 1. Introduction

Groundwater serves as a vital source of drinking water worldwide. To ensure the safety and purity of groundwater for domestic use, it is imperative to regularly assess its quality, a crucial step in enhancing the well-being of the growing global population [1]. Safe drinking water is a critical component of public health and environmental sustainability. Access to clean and potable groundwater is critical, especially in limited resources areas. In many of these places, numerous physicochemical parameters might impair groundwater quality, providing serious health hazards to residents. Traditional methods of determining groundwater potability require time-consuming and expensive laboratory studies. Moreover, the ground water quality can also be affected depending upon the type of pumping technique employed [2]. Several researchers have employed various techniques to assess the quality of groundwater for drinking purposes. For instance, multivariate statistics [3], automatic exponential smoothing model [4], explanatory analysis [5], theoretical probability models [6, 7], correlation and regression analyses [8], and other methods. However, groundwater quality is influenced by a complex interplay of physicochemical parameters, including electric conductivity (EC), pH, total dissolved solids (TDS), and concentrations of various ions and minerals. Ensuring the safety of this vital

**Citation:** To be added by editorial staff during production.

Academic Editor: Firstname Last-name

Published: date



**Copyright:** © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

resource necessitates the development of reliable predictive models that can rapidly and accurately classify groundwater samples as potable or non-potable.

The introduction of machine learning and deep learning techniques has opened new paths for efficient, accurate, and cost-effective potability prediction. These techniques have emerged as powerful tools, offering the potential to enhance and complement traditional methods for more accurate and efficient evaluation. While deep learning techniques have been widely used in computer vision applications, such as Optical Character Recognition (OCR) [9], face and skin detection/recognition [10, 11], visual pattern classification [12], and image classification [13], limited work has been done in the domain of water quality evaluation [14, 15]. This study embarks on a comprehensive evaluation of machine learning and deep learning models for groundwater potability classification. We explore the efficacy of machine learning models, such as logistic regression, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Random Forest, alongside deep learning models, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Feedforward Neural Networks (FNN), and Long Short-Term Memory (LSTM). These models are applied to groundwater samples collected from both industrial and residential locations, with a specific focus on nine important physicochemical parameters.

The primary aim of this study is to identify the most effective machine learning models for this specific task. The motivation behind this research lies in the fundamental importance of safe and potable groundwater for public health and sustainable water resource management, particularly where water scarcity and contamination are pressing concerns. Therefore, by evaluating and comparing machine learning models, this research offers a novel and practical approach to groundwater potability prediction, bridging the gap between data science and real-world challenges in water resource management and public health.

This paper is outlined as follows: Section 2 contains the materials and methods. Section 3 provides results and discussion. Section 4 gives the concluding remarks.

## 2. Materials and Methods

In this section, we investigate classification algorithms based on machine learning and deep learning and evaluate their effectiveness in classifying groundwater potability. Among the machine learning algorithms are logistic regression, KNN, SVC, and Random Forest. Deep learning models, on the other hand, include the ANN, CNN, FNN, and LSTM. The performance of each model was evaluated based on key metrics, including accuracy, precision, recall, and F1-Score. These metrics provide insights into the models' ability to classify groundwater samples correctly and are indicative of their overall performance. Python is used for all computations, training, and testing. The general flowchart of the investigation is depicted in Figure 1.

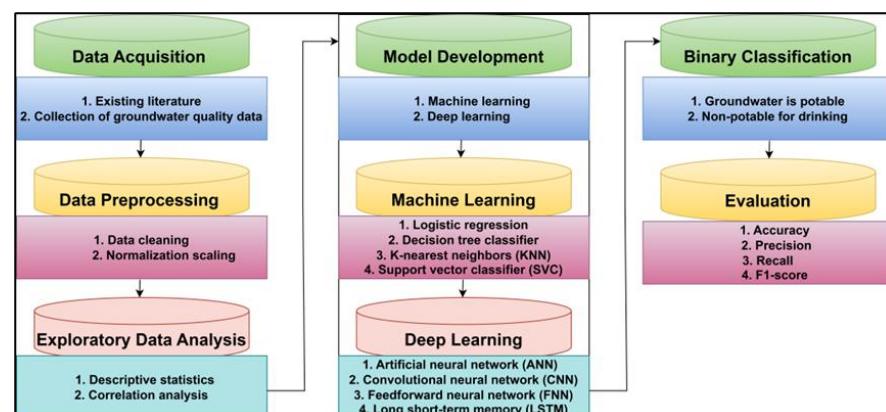
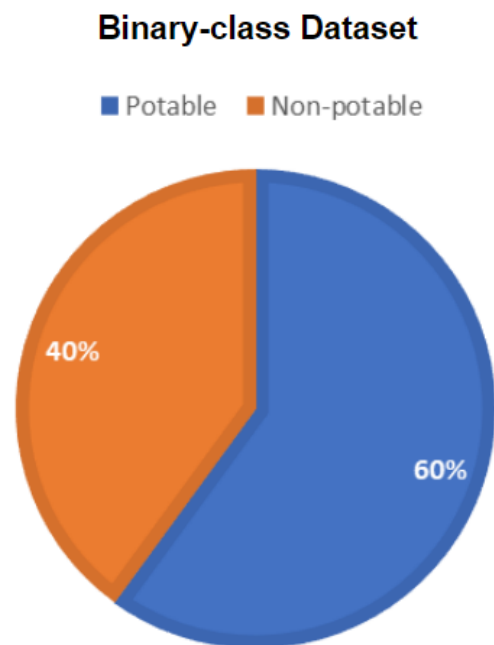


Figure 1. Study flowchart.

### 2.1. Water Sampling and Data Preprocessing

Thirty groundwater samples were randomly taken from open wells and boreholes in both industrial and residential areas during the study that was carried out in August 2020 in Jaen, Kano State, Nigeria. The Geographical Positioning System (GPS) was used to locate the sampling stations by determining their latitude and longitude coordinates. Following standard procedures, these samples were then meticulously maintained in sterile plastic bottles and kept in an icebox. For each groundwater sample, fifteen physicochemical parameters were measured. In this study, we are specifically focusing on nine of these parameters, such as electric conductivity (EC), pH, total dissolved solids (TDS), calcium, magnesium, chloride, zinc, manganese, and copper. Most of the parameters were expressed in milligrams per liter (mg/L), except for EC ( $\mu\text{S}/\text{cm}$ ), pH, and TDS (NTU).

Additionally, the data set featured binary labels denoting the potability of each sample, with '1' signifying potable water and '0' indicating non-potable water. The data set also included binary labels indicating the potability of each sample, with '1' representing potable water and '0' representing non-potable water. The potability and non-potability standards for drinking water quality in Nigeria can be found in [16]. The class distribution of the groundwater data used for binary classification is shown in Figure 2.



**Figure 2.** Distribution of dataset for binary classification.

- To prepare the data for analysis, the following preprocessing steps were performed:
- i. Data cleaning: any missing or erroneous data points were identified and either corrected or removed from the dataset.
  - ii. Feature scaling: continuous variables were scaled to have a mean of 0 and a standard deviation of 1 to ensure that all features contributed equally to model training.
  - iii. Data split: the data set was divided into training and testing sets using an 80-20 split, ensuring that the same split was applied consistently across all models.

### 2.3. Model Selection and Training

For the comparative analysis, four machine learning models and four deep learning models were chosen.

### 2.3.1. Machine Learning Models

#### 1. Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms and comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. This concept of predictive modeling falls under the classification algorithm. It is used in this study to predict a potability class of groundwater from the set of predators. The logistic curve relates the independent variable,  $X$ , to the mean of the dependent variable,  $Y$ . This relationship can be expressed as:

$$P = \frac{\exp(a + bX)}{1 + \exp(a + bX)} = \frac{1}{1 + \exp-(a + bX)},$$

where  $P$  is the probability of a 1 (the proportion of 1s, the mean of  $Y$ ),  $\exp$  is the base of the natural logarithm,  $a$  and  $b$  are the parameters of the model.

#### 2. Random Forest Classifier

Random Forest is a popular ensemble learning technique used for both classification and regression tasks. It is an ensemble of decision trees, where each tree independently makes predictions, and the final prediction is determined by a majority vote (for classification) or averaging (for regression) of the individual tree predictions. The "random" aspect in Random Forest refers to two main sources of randomness:

- i. Bootstrap Aggregation (Bagging): Multiple subsets of the training data are created through bootstrapping, a resampling technique. Each decision tree in the forest is trained on a different subset, introducing diversity among the trees.
- ii. Feature Randomness: Random subsets of features are considered when splitting nodes in each tree. This ensures that not all features are used for every split, reducing the risk of overfitting and improving generalization.

#### 3. K-Nearest Neighbors

The K-Nearest Neighbor (KNN) algorithm is a popular machine learning technique used for classification and regression tasks. It relies on the idea that similar data points tend to have similar labels or values. During the training phase, the KNN algorithm stores the entire training dataset as a reference. When making predictions, it calculates the distance between the input data point and all the training examples using a chosen distance metric such as Euclidean distance. Next, the algorithm identifies the K nearest neighbors to the input data point based on their distances. In the case of classification, the algorithm assigns the most common class label among the K neighbors as the predicted label for the input data point. For regression, it calculates the average or weighted average of the target values of the K neighbors to predict the value for the input data point. With  $N$  sample size and  $p_i$  probability of every  $i$  sample, the nearest neighbor is expressed as:

$$p_i = \sum_{j \in K_i}^N p_{ij},$$

where  $K_i$  donates the set of points that fall within the same class as sample  $i$ , and  $p_{ij}$  donates the softmax over Euclidean distances within the embedded space as given by:

$$p_{ij} = \frac{\exp\left(-|Dx_i - Dx_j|^2\right)}{\sum_{m \neq i}^N \exp\left(-|Dx_i - Dx_m|^2\right)}, \quad p_{ii} = 0.$$

#### 4. Support Vector Classifier

Support vector classifier (SVC) is one of the most commonly used supervised algorithms that works for both regression and classification tasks, but generally, it works best for classification problems. SVC algorithms are used to find a hyperplane that best separates the two classes in N-dimensional spaces. A hyperplane is simply a line if there are only two input features, and a two-dimensional plane if there are only two input features. The maximum distance will exist between hyper-planes and the nearest elements of the classes in the best-case scenario when the data are perfectly separable. SVC often aims to approximate this scenario as accurately as possible. Although various classes of manifolds are used instead of hyperplanes in nonlinear SVC, the basic principle is the same. In nonlinear SVC, hyperplanes are replaced with different classes of manifolds, but the principle remains the same.

#### 2.3.2. Deep Learning Models

##### 1. Artificial Neural Networks (ANN)

ANN is a foundational deep learning model that consists of interconnected layers of artificial neurons, organized into input, hidden, and output layers. This model is highly versatile and excels at capturing complex patterns and relationships in data.

##### 2. Convolutional Neural Networks (CNN)

CNN is specifically designed for processing grid-like data, such as images. It comprises convolutional layers to automatically detect features and patterns in data. While originally developed for image analysis, CNNs can be adapted to various domains.

##### 3. Feedforward Neural Networks (FNN)

FNN, like ANN, is a type of feedforward neural network. It's characterized by layers of interconnected neurons, each connected to the next layer without feedback loops. FNNs are commonly used for regression and classification tasks.

##### 4. Long Short-Term Memory (LSTM)

LSTM is designed to handle sequences and time-series data. It excels at capturing long-range dependencies in data, making it a suitable choice for sequential data such as time-series measurements.

For each model, the following steps were executed:

1. Model Training: The model was trained on the training dataset using default hyperparameters.
2. Model Evaluation: The model's performance was evaluated on the testing dataset using accuracy score.

### 3. Results and Discussion

#### 3.1. Descriptive Statistics

Table 1 provides a statistical overview of the groundwater parameters that were measured. Table 1 shows that there are significant variances within these parameters, with especially high standard deviations suggesting the greatest variability. Conductivity ranges from a low of 209 to a high of 1490, with a mean of 672.73, indicating significant

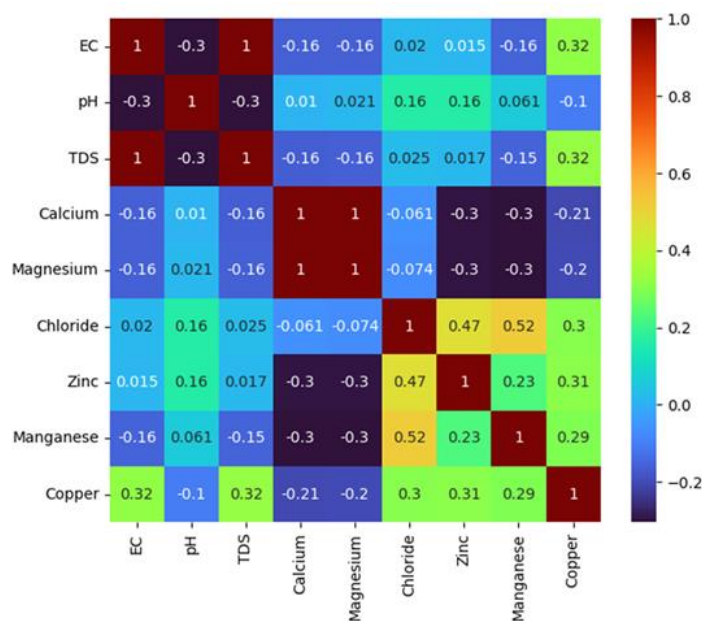
diversity in groundwater conductivity. The pH of the water ranges from 5.59 to 7.22, with a mean of 6.67, indicating the presence of acidity. Positive skewness implies that the data has a right tail, whereas negative skewness indicates that the data has a left tail. Kurtosis readings indicate that the groundwater has platykurtic distributions. These investigations not only help us understand groundwater composition, but they also have practical consequences for water quality monitoring and environmental management.

**Table 1:** Statistical summary of the groundwater parameters.

Parameters	Mean	median	Min	Max	Std. deviation	Skewness	Kurtosis
EC	672.73	640	209	1490	302.61	0.63	0.32
pH	6.67	6.73	5.59	7.22	0.41	-0.92	0.55
TDS	339.2	322.5	104	731	151.26	0.54	0.04
Calcium	0.59	0.32	0.07	1.75	0.5	1.21	0.14
Magnesium	0.25	0.15	0.03	0.74	0.21	1.19	0.14
Chloride	1.26	0.8	0	4.6	1.18	1.25	1.09
Zinc	0.12	0.11	0.01	0.7	0.13	3.35	14.58
Manganese	0.09	0.1	0	0.3	0.09	0.77	-0.17
Copper	0.11	0.1	0.04	0.3	0.06	1.88	3.64

### 3.2. Correlation Analysis

Correlation analysis is a useful statistical tool for determining associations between various water quality parameters. We can get useful insights into the complex mechanisms that govern groundwater quality by identifying and these relationships [17, 18]. The correlation analysis results demonstrate substantial positive and negative relationships among the evaluated groundwater parameters, indicating similar patterns in their variations that are statistically significant at the 1% level (Figure 3). Notably, the extraordinarily strong correlation of 0.99 between Calcium and Magnesium implies a nearly linear relationship, implying that changes in Calcium and/or Magnesium concentration have a strong relationship within groundwater samples. Furthermore, conductivity and TDS have a strong positive association of 0.99. Hence, the groundwater quality of the study area is influenced by the complex interplay of these parameters.



**Figure 3.** Correlation matrix for the groundwater parameters.

*3.1. Model Performance Evaluation*

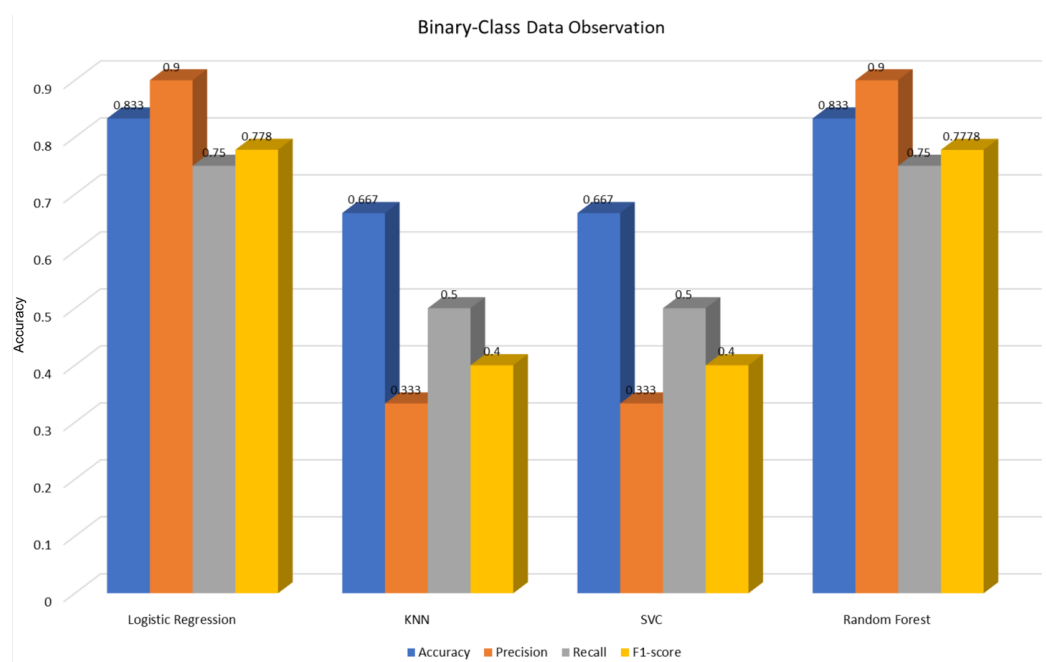
The comparative analysis of machine learning and deep learning models for groundwater potability prediction yielded insightful results. The primary evaluation metrics, including accuracy, precision, recall, and F1-Score, were employed to assess the performance of each model, providing a comprehensive view of model effectiveness.

Table 2 shows that four distinct machine learning models were employed to assess groundwater potability. Logistic Regression and Random Forest yielded the highest accuracy of 0.833, accompanied by precision scores of 0.900. Logistic Regression maintained a recall of 0.750, demonstrating reliable predictive accuracy, while Random Forest achieved a recall of 0.750. KNN and SVC delivered moderate accuracy with scores of 0.667. Among these models, Random Forest and Logistic Regression maintained a balanced F1-Score of 0.778. These findings underscore the potential of machine learning models for precise groundwater classification, highlighting their relevance in water quality assessment.

Figure 4 displays a graphical representation of the comparative performance metrics across various machine learning model algorithms. The visual presentation aids in readily identifying the highest accuracy scores achieved by different model algorithms.

**Table 2.** Machine learning model performance for groundwater potability classification.

Machine learning algorithm	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.833	0.900	0.750	0.778
KNN	0.667	0.333	0.500	0.400
SVC	0.667	0.333	0.500	0.400
Random Forest	0.833	0.900	0.750	0.778



**Figure 4.** Comparative performance metrics of machine learning algorithms.

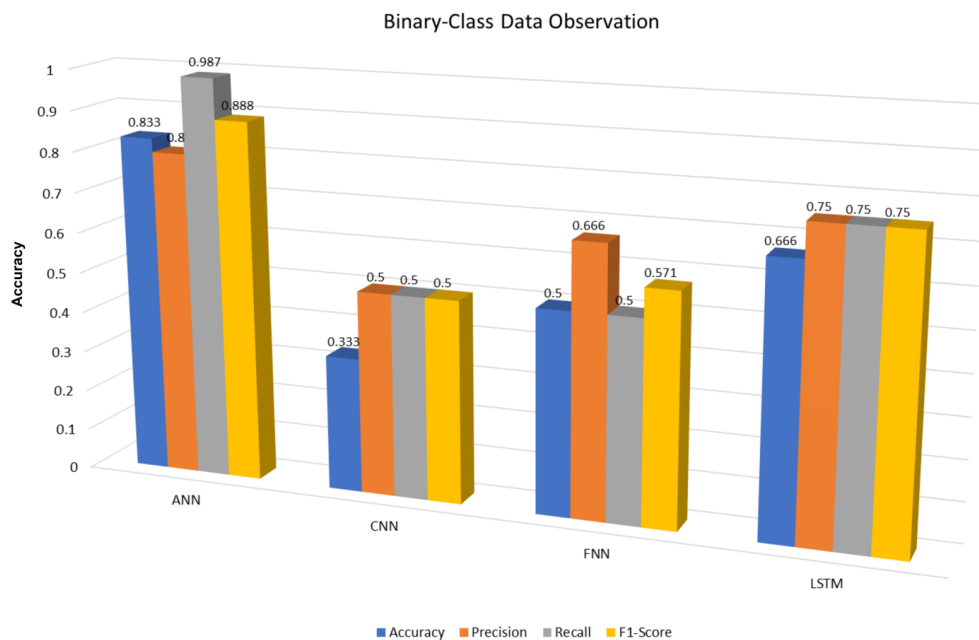


Table 3 presents the outcomes of the deep learning models applied for groundwater potability classification assessment. Among these models, ANN achieved the highest accuracy of 0.833, along with a remarkable precision score of 0.800. ANN exhibited an impressive recall of 0.987, signifying its exceptional ability to correctly identify potable groundwater. LSTM also delivered reliable results with an accuracy score of 0.666 and a balanced precision of 0.750. Conversely, CNN and FNN displayed lower accuracy scores of 0.333 and 0.500, respectively. These results highlight the diverse performance of deep learning models and their potential for accurate groundwater classification, especially notable in the case of ANN and LSTM, underlining their significance in groundwater quality classification assessment.

Figure 5 provides a graphical illustration of the comparative performance metrics across diverse deep learning model algorithms. This visual representation facilitates the quick identification of the highest accuracy scores attained by different model algorithms.

**Table 3:** Deep learning model performance for groundwater potability classification.

Deep learning algorithm	Accuracy	Precision	Recall	F1-Score
ANN	0.833	0.800	0.987	0.888
CNN	0.333	0.5000	0.500	0.500
FNN	0.5000	0.666	0.500	0.571
LSTM	0.666	0.750	0.750	0.750



**Figure 5.** Comparative performance metrics of deep learning algorithms.

### 3.2. Discussion

The results from our comprehensive evaluation of machine learning and deep learning models have provided valuable insights into their effectiveness in predicting groundwater potability. In the realm of machine learning, Logistic Regression and Random Forest have demonstrated their capability by achieving the highest accuracy scores of 0.833. Logistic Regression, in particular, maintained a commendable recall of 0.750, emphasizing



its consistent predictive accuracy. Similarly, Random Forest showcased its potential with a recall of 0.750. KNN and SVC exhibited moderate accuracy levels, both scoring 0.667. Notably, Random Forest and Logistic Regression achieved balanced F1-Scores of 0.778, highlighting their reliability.

These findings underscore the potential of machine learning models in precisely classifying groundwater, emphasizing their significance in water quality assessment. Machine learning models are well-poised for applications in real-world scenarios where accurate predictions are crucial. The strength of machine learning models lies in their ability to handle data with low correlations among parameters.

On the other hand, deep learning models have also made significant contributions to the groundwater potability prediction task. ANN delivered an accuracy score of 0.833, rivaling the performance of Logistic Regression and Random Forest. LSTM demonstrated its potential with an accuracy score of 0.666. Although CNN and FNN achieved lower accuracy scores of 0.333 and 0.5, respectively, they remain promising in specific applications.

However, it is important to note that while accuracy is a valuable metric, it may not be the sole determinant of model suitability. Factors such as computational efficiency, interpretability, and the specific needs of the application must be considered when selecting a model for real-world deployment. Moreover, the potential for further model refinement through hyperparameter tuning and the exploration of ensemble techniques should not be overlooked. These avenues may enhance the predictive capabilities of the models and offer improved potability classifications for groundwater resources in the study area.

### 3. Concluding Remarks

This study has provided an extensive comparative assessment of machine learning and deep learning models for predicting groundwater potability in the Jaen district of Kano State, Nigeria. Our findings offer valuable insights into the selection and performance of these models, particularly in situations characterized by low correlations among groundwater parameters. Logistic Regression and the Decision Tree classifier have emerged as standout performers, each achieving an impressive accuracy score of 0.833. These models not only exhibit robust predictive capabilities but also offer interpretability, positioning them as promising candidates for practical applications in water resource management and public health initiatives. On the other hand, our exploration of deep learning models yielded a range of outcomes. ANN exhibited a remarkable accuracy score of 0.833, underlining its potential for accurate predictions. LSTM followed closely with a score of 0.666, demonstrating strong predictive abilities. CNN and FNN delivered slightly lower scores, emphasizing the need for further investigation and refinement. However, we emphasize that the selection of a suitable model should consider various factors, such as computational efficiency and the specific requirements of the application. Future research directions include the fine-tuning of models and the exploration of ensemble techniques to enhance predictive accuracy. These ongoing efforts hold the promise of advancing the precision of groundwater potability predictions, contributing to the overarching goal of ensuring clean and safe drinking water for communities and the sustainable management of vital natural resources.

**Author Contributions:** Conceptualization, A.A.S.; Methodology, A.A.S.; Software, A.A.S.; Validation, A.A.S., A.K.Y., and M.Z.; Writing—original draft, A.A.S.; A.K.Y., and M.Z.; Data curation, A.A.S.; Visualization, M.Z., A.K.Y., and A.A.S. All authors have read and agreed to the published version of the manuscript.

**Data Availability Statement:** All data in this paper is available from the corresponding author upon request.

**Acknowledgements:** The authors extend their appreciation to Universiti Teknologi PETRONAS for providing research facilities. Special gratitude is owed to the Tertiary Education Trust Fund (TETFund) for funding this research through the institutional-based grant allocated to Aliko Dangote University of Science and Technology, Wudil, Nigeria for the collection of experimental data. Lastly, the authors would like to convey special thanks to the organizers and sponsors of this conference.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- [1] A. A. Suleiman, U. A. Abdullahi, A. Suleiman, R. B. Yunus, and S. A. Suleiman, "Assessment of Groundwater Quality Using Multivariate Statistical Techniques," in *Intelligent Systems Modeling and Simulation II: Machine Learning, Neural Networks, Efficient Numerical Algorithm and Statistical Methods*: Springer, 2022, pp. 567-579.
- [2] M. Adil, M. Arshad, and M. Aslam, "Low Cost Water Pumping for Sustainable Irrigation Using Renewable Energy Based Ram Pump," in *5th International Mechanical Engineering Congress, Karachi, Pakistan, May, 2015*, pp. 9-10.
- [3] E. O. Thomas, "Evaluation of groundwater quality using multivariate, parametric and non-parametric statistics, and GWQI in Ibadan, Nigeria," *Water Science*, vol. 37, no. 1, pp. 117-130, 2023.
- [4] A. Nsabimana, J. Wu, J. Wu, and F. Xu, "Forecasting groundwater quality using automatic exponential smoothing model (AESM) in Xianyang City, China," *Human and Ecological Risk Assessment: An International Journal*, vol. 29, no. 2, pp. 347-368, 2023.
- [5] A. A. Suleiman, A. Ibrahim, and U. A. Abdullahi, "Statistical explanatory assessment of groundwater quality in Gwale LGA, Kano state, northwest Nigeria," *Hydrospatial Analysis*, vol. 4, no. 1, pp. 1-13, 2020.
- [6] V. V. Singh, A. A. Suleiman, A. Ibrahim, U. A. Abdullahi, and S. A. Suleiman, "Assessment of probability distributions of groundwater quality data in Gwale area, north-western Nigeria," *Annals of Optimization Theory and Practice*, vol. 3, no. 1, pp. 37-46, 2020.
- [7] A. Ibrahim, A. A. Suleiman, U. A. Abdullahi, and S. A. Suleiman, "Monitoring Groundwater Quality using Probability Distribution in Gwale, Kano state, Nigeria," *Journal of Statistical Modeling & Analytics (JOSMA)*, vol. 3, no. 2, 2021.
- [8] A. A. Suleiman, U. A. Abdullahi, A. Suleiman, S. A. Suleiman, and H. U. Abubakar, "Correlation and regression model for physicochemical quality of groundwater in the Jaen District of Kano State, Nigeria," *Journal of Statistical Modeling & Analytics (JOSMA)*, vol. 4, no. 1, 2022.
- [9] S. U. Rehman, S. Tu, Y. Huang, and O. U. Rehman, "A benchmark dataset and learning high-level semantic embeddings of multimedia for cross-media retrieval," *IEEE Access*, vol. 6, pp. 67176-67188, 2018.
- [10] S. U. Rehman, S. Tu, Y. Huang, and Z. Yang, "Face recognition: A novel un-supervised convolutional neural network method," in *2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS)*, 2016: IEEE, pp. 139-144.
- [11] S. ur Rehman *et al.*, "Unsupervised pre-trained filter learning approach for efficient convolution neural network," *Neurocomputing*, vol. 365, pp. 171-190, 2019.
- [12] S. Tu, Y. Huang, and G. Liu, "CSFL: A novel unsupervised convolution neural network approach for visual pattern classification," *Ai Communications*, vol. 30, no. 5, pp. 311-324, 2017.
- [13] S. U. Rehman, S. Tu, O. U. Rehman, Y. Huang, C. M. S. Magurawalage, and C.-C. Chang, "Optimization of CNN through novel training strategy for visual classification problems," *Entropy*, vol. 20, no. 4, p. 290, 2018.
- [14] D. V. V. Prasad *et al.*, "Analysis and prediction of water quality using deep learning and auto deep learning techniques," *Science of The Total Environment*, vol. 821, p. 153311, 2022/05/15/ 2022, doi: <https://doi.org/10.1016/j.scitotenv.2022.153311>.
- [15] Y. Im, G. Song, J. Lee, and M. Cho, "Deep Learning Methods for Predicting Tap-Water Quality Time Series in South Korea," *Water*, vol. 14, no. 22, p. 3766, 2022. [Online]. Available: <https://www.mdpi.com/2073-4441/14/22/3766>.

- 
- [16] U. NSDWQ, "Nigerian standard for drinking water quality," *Nigerian Industrial Standard, NIS*, vol. 554, pp. 13-14, 2007.
- [17] A. A. *et al.*, 5, no. 02, pp. 288-297, 2021.
- [18] S. F. Salleh, A. A. Suleiman, H. Daud, M. Othman, R. Sökkalingam, and K. Wagner, "Tropically Adapted Passive Building: A Descriptive-Analytical Approach Using Multiple Linear Regression and Probability Models to Predict Indoor Temperature," *Sustainability*, vol. 15, no. 18, p. 13647, 2023.