Water Practice & Technology

IWA PUBLISHING

Water Practice & Technology Vol 19 No 6, 2364 doi: 10.2166/wpt.2024.139

Machine learning-based ensemble model for groundwater quality prediction: A case study

Annie Jose * and Srinivas Yasala

Centre for Geotechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India
*Corresponding author. E-mail: anniejose1996@gmail.com

AJ, 0000-0002-0653-317X

ABSTRACT

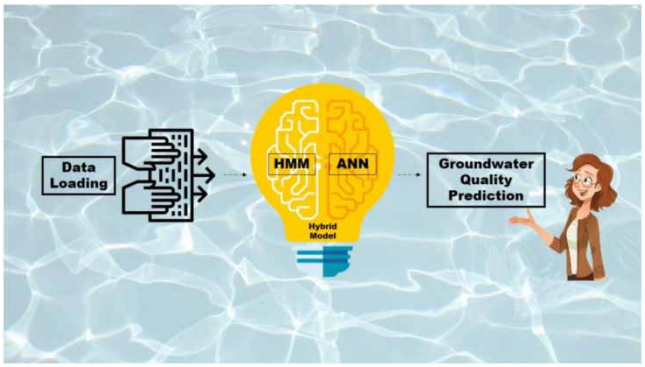
Groundwater quality is vital for public health and environmental sustainability. As managing large datasets is challenging for traditional methods, this study combines the hidden Markov model (HMM) and the artificial neural network (ANN), a machine learning-based ensemble model to predict groundwater quality in Kanyakumari District, Tamil Nadu, India. In order to train the model, the acquired data is cleaned and normalized. HMM is used to find hidden patterns while the ANN architecture is used to forecast groundwater quality categories. Accuracy, precision, sensitivity, and F1-scores calculation are necessary to evaluate the model's performance. The effectiveness of the approach can be analyzed by *k*-fold cross-validation scores. The study demonstrates the effectiveness of the HMM–ANN approach in groundwater quality prediction with an accuracy of 97.41%. Thus, the research contributes to groundwater quality assessment by offering a unique methodology that facilitates informed decision-making for water resource management and environmental conservation.

Key words: artificial neural network (ANN), groundwater quality, hidden Markov model (HMM), Kanyakumari district, prediction, public health

HIGHLIGHTS

- Introduces a novel approach for groundwater quality prediction.
- Demonstrates improved accuracy and reliability due to the integration of two machine learning models.
- Focuses on a specific study area to address a region-specific environmental concern.
- Combines different fields such as hydrogeology, machine learning, and environmental science for a better understanding of groundwater dynamics.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Water is an essential natural resource for human life and growth that cannot be substituted by any other substance. Water must thus be preserved in adequate amounts and of high quality. Water is a resource that can

be found in a region either as surface water or as groundwater (Listyani & Putranto 2022). Groundwater is a vital component of the natural water resources system, and human exploitation of this subsurface water dates back thousands of years. Advances in drilling and pumping technology have made it easier to use groundwater and have made it possible to access a vast reservoir of incredibly deep water. It also contains health-promoting minerals in addition to these benefits. In areas where surface water availability is limited, using groundwater could be highly beneficial in addressing a water shortage (Hanoon *et al.* 2021).

Recently, groundwater research in different nations has mainly focused on the quality of groundwater as it has a direct effect on the health of the citizens. Our reckless lifestyle causes an increasing number of soluble chemicals to be released into the groundwater ecosystems creating unsteadiness in the groundwater quality (Freeze & Cherry 1979). The quality of groundwater is influenced by multiple factors, increasing its complexity. In order to reduce the toxicity of newly formed compounds through their interactions, continuous research should be undertaken. Human activities are increasing the impacts on groundwater quality and complicating environmental conditions (Li *et al.* 2022). Therefore, groundwater quality research requires more monitoring data.

An understanding of the groundwater system should serve as the foundation for designing a baseline monitoring network. After it is established, it should also offer the necessary information to update and validate the conceptual model of the groundwater body under study. Using this framework, we can also interpret the primary temporal and spatial patterns in water quality (Preziosi *et al.* 2021). In the realm of groundwater quality research, the quest for innovation and cost-effective methods is paramount. Traditional approaches have certainly played a crucial role in understanding groundwater quality; but in a rapidly changing world, there is a growing need for more adaptive and sophisticated techniques to address emerging challenges. Therefore, a genuine attempt has been made in this work to analyze and predict the groundwater quality in the study area using machine learning-based methods.

In this study, a hybrid model integrating the hidden Markov model (HMM) and the artificial neural network (ANN) is constructed to predict the groundwater quality in the study area enhancing its capacity to adjust to the changes in the environment. Conventional modeling methods rely on statistical or time series analytic techniques, assuming a relationship between variables. These models often yield subpar prediction outcomes. Artificial intelligence (AI)-based predictive models, however, do not require data or correlation, outperforming statistical techniques (Raheja *et al.* 2021). In this study, HMMs, introduced in the 1970s, are a dynamic Bayesian network based on Bayes theory.

The HMM can be modeled as a stochastic model of discrete events and as a variant of the Markov chain, which is a chain of connected states or events where each subsequent state is solely dependent on the system's current state. An HMM's states are hidden and they are only deducible from the symbols (observations) that are provided. They have been used in biological sequence analysis and voice recognition since the late 1980s (Awad & Khanna 2015; Franzese & Iuliano 2019). These days, they are employed in many other contexts such as Nguyen (2017) has also studied the implementation of HMM for stock price prediction; hence HMM, their stochastic modeling shows how they can be used to solve real-world problems, especially for this study, a hand in the hybrid model involving groundwater quality studies.

The other hand of this hybrid model, ANN is a technology with a flexible mathematical framework that may identify complicated non-linear correlations between input and output data (Ubah *et al.* 2021). An output data could be predicted from input data using a network of artificial neurons, much as how neurons in the human brain process input signals and generate output signals. Additionally, by training and testing the model using previous data, the mathematical link between the independent variables (chemical parameters of water quality) and the dependent variable (WQI) was investigated for the purpose of modeling water quality. Additionally, the input data was retained in the network in order to use the knowledge that was gained from it for future WQI predictions (Khudair *et al.* 2018).

By combining HMM with ANN, this study takes advantage of the best features of both models. HMM is used to capture the sequences and state transitions, whereas ANN is capable of capturing intricate relationships within individual features. This combination enables a more comprehensive understanding of the dynamics of groundwater quality and also be able to improve the accuracy of groundwater quality models. In conclusion, the incorporation of HMM and ANN models in groundwater quality prediction provides a powerful tool for addressing environmental issues and helps improve society by improving predictive accuracy and assisting decision-makers in protecting water resources and public health.

1.1. Statement of the problem

The main focus of the study is to improve the quality of groundwater in Kanyakumari District, Tamil Nadu, India. To make this happen, several challenges should be tackled and that includes boosting the accuracy of the method, dealing with the intricacies of the data, ensuring adaptability, accounting for the regional differences, and crafting customized solutions tailored to this specific study area.

1.2. Research gap

In Kanyakumari District, Tamil Nadu, India, there has not been much research exploring a fresh and innovative approach by using data-driven techniques to predict groundwater quality. The primary goal of the study is to enhance the accuracy and reliability of prediction by introducing a unique hybrid model that combines HMM and ANN. This approach is quite novel in the world of hydrogeology and water quality prediction. By harnessing machine learning methods, this study aims to bridge the existing gap and bring a brand-new perspective to predicting groundwater quality in this particular study area.

1.3. Research objectives

The core objective of this research is to create a cutting-edge predictive model for evaluating groundwater quality in Kanyakumari District, Tamil Nadu, India. This study aims to blend two distinct machine learning techniques, HMM and ANN, in order to significantly improve the accuracy of the groundwater quality predictions.

1.4. Related works

The combination of HMM and ANN has found applications across diverse fields. Some of the related works are as follows:

(i) Trentin & Gori (2003) introduced a hybrid system that integrates HMM and ANN to enhance the speech recognition. In their work, they put forward algorithms and training techniques essential for the effective development of this hybrid system to examine and address the limitations inherent in previous approaches to ensure a more comprehensive understanding of the proposed model. By using the strengths of both models, they created a more effective and robust voice recognition system. Experimental results are presented to validate the efficacy of the hybrid system, and implementation issues related to its practical application are thoroughly discussed in their work. Ong & Ahmad (2011) also used the combination of HMM and ANN for the creation of a Malay language speech recognition system using standard Malay phonetics and phonology.

(ii) Kishna & Francis (2017) proposed a novel approach for recognizing cursive handwritten Malayalam characters using the HMM–ANN hybrid model. They address the intricacies associated with recognizing Malayalam characters, which may have diverse fonts and styles. They convert the handwritten text into editable text through the application of optical character recognition (OCR) technology and by involving ANN and HMM, they aim to overcome the challenges and contribute to the advancement of accurate and efficient OCR systems tailored for handwritten Malayalam script. The study not only outlines the methodology but also underscores the broader implications and practical applications of their approach in the realm of handwritten text recognition.

(iii) HMM and ANN integration was also used in the field of medicine. Estebanez *et al.* (2012) introduced a movement recognition system marking the initial steps toward the automation of a two-arm surgical robotic system in the laparoscopic surgical domain using the ANN–HMM hybrid model. They used ANN to encode the movements through Fourier spectra and used HMM to capture the intricate interactions between the surgical tools, and thus, encompassing tasks such as tissue cutting, suturing, and transporting. Sanghavi & Bojewar (2014) also proposed a work for improving patient treatment for cardiovascular disease by predicting its future progression, allowing for timely adjustments to medication using ANN as a classifier and incorporates wavelet transform for feature extraction, reducing the ECG dataset and usage of HMM as a predictor in conjunction with the ANN. The results highlight that the joint application of ANN and HMM significantly enhances efficiency compared to using ANN alone.

(iv) Nowadays, the fusion of HMM and ANN also been used in the weather prediction analysis (Acharya & Sahani 2022).

Despite the broad applications of this hybrid model, there has yet to be a research study in groundwater quality that integrates HMM and ANN. Thus, this research paper takes on the task of addressing and filling this specific gap in the current body of literature.

2. METHODOLOGY

2.1. Study area

The study area is located in the southernmost part of India. Kanyakumari district covers 1,671.84 sq. km. The study area falls between 8°03′–8°35′ North latitude and 77°15′–77°36′ East longitude (District Statistical Handbook 2021-2022). Kanyakumari district experiences an average annual rainfall of 1,448.6 mm, with the majority of precipitation taking place during the northeast (NE) and southwest (SW) monsoon seasons. The yearly average minimum and maximum temperatures stand at 23.78 and 33.95 °C, respectively. Kanyakumari district is characterized by the presence of both porous and fissured geological formations. The significant aquifer systems in the region include unconsolidated and semi-consolidated formations and fractured crystalline rocks. Groundwater is present in nearly all geological formations within the district, and the recharge of groundwater is influenced by the degree of weathering (Rajammal *et al.* 2021). The construction of the study area map (Figure 1) was accomplished using QGIS software version 3.32, commonly referred to as QGIS Lima.

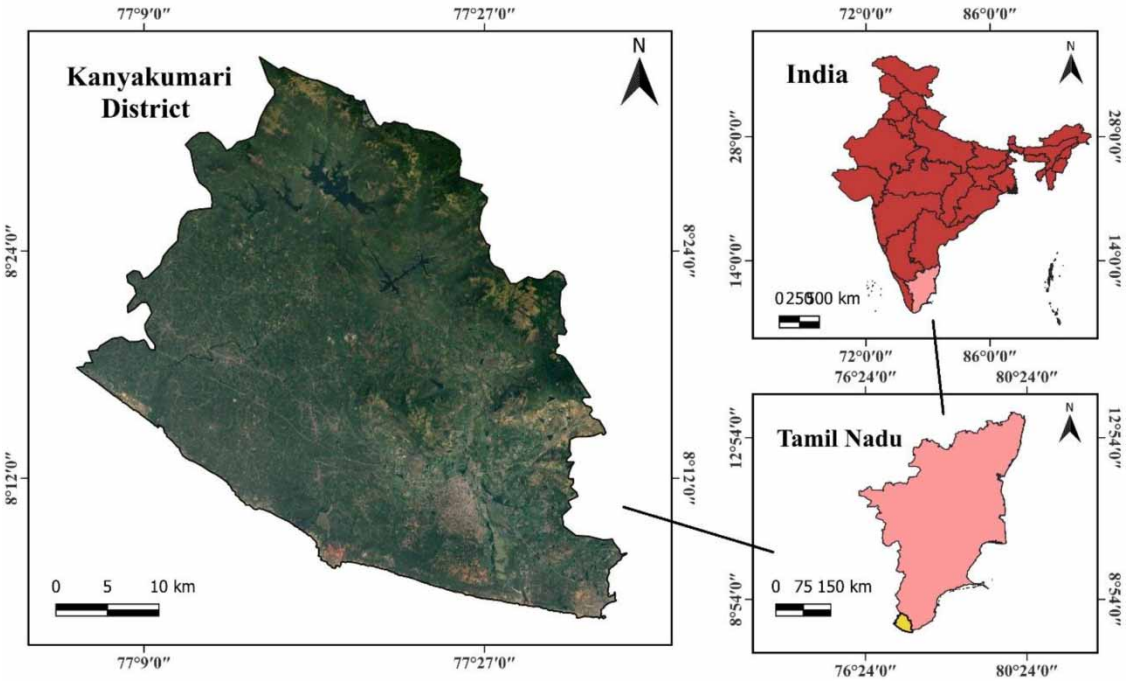


Figure 1 | Study area map – Kanyakumari District.

2.2. Calculation of ion balance error and drinking water quality index

The data is collected and the ion balance error (IBE) is calculated to assess the reliability of the measured ion concentrations. To ensure analytical precision in assessing the concentrations of total cations and total anions expressed in milliequivalents per liter (meq/L) for each sample, the IBE is calculated using the equation provided (Adimalla & Qian 2019). Then the drinking water quality index (DWQI) is calculated using the formula:

$$\text{IBE} = \frac{\sum \text{cation} - \sum \text{anion}}{\sum \text{cation} + \sum \text{anion}} \times 100 \quad (1)$$

The determined IBE was to be within the acceptable threshold of $\pm 10\%$ (Domenico & Schwartz 1990). If the value exceeds, the outliers are eliminated.

The DWQI serves as a means to assess the suitability of water for consumption (Jose & Srinivas 2023). In order to evaluate the fluctuations in the quality of drinking water in the study area, an equation introduced by

Vasanthavigar et al. in 2010 was employed.

$$\text{DWQI} = \sum_{i=1}^{n} \text{SI}i \tag{2}$$

$$\text{Sub-index, SI}i = Wi \times Qi \tag{3}$$

$$\text{Relative weight, } Wi = \frac{wi}{\sum_{i=1}^{n} wi} \tag{4}$$

where Wi is the weight of each parameter and 'n' is the number of parameters.

$$\text{Quality rating scale, } Qi = \frac{Ci}{Si} \times 100 \tag{5}$$

where Ci is the concentration of specific chemical parameter in each water sample (mg/L) and Si is the maximum recommended permissible limit (WHO 2011; BIS 2012).

2.3. HMM–ANN hybrid

Jupyter Notebook 6.5.4, an open-source web application is used for the coding process and data analysis. Python script is used to combine the HMM and ANN models that can be used for the groundwater quality prediction task. Figure 2 provides an overview of the process. Necessary libraries are imported and the dataset is loaded. The dataset is cleaned and pre-processed to handle the missing values. The features such as 'pH', 'EC', 'TDS', 'Na', 'K', 'Ca', 'Mg', 'TH = Ca + Mg', 'SO_4', 'Cl', 'HCO_3', 'CO_3', 'F', '$NO_2 + NO_3$', and 'DWQI' are selected and are standardized using standard scalar to ensure that all the variables are in a similar scale.

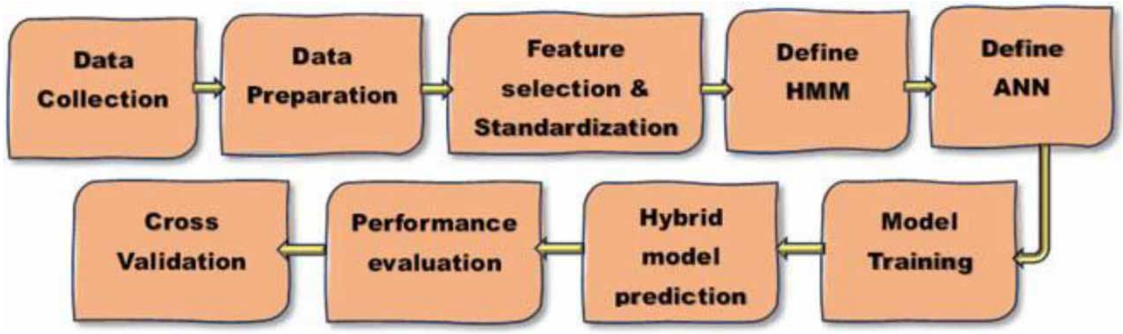


Figure 2 | Overall framework for this study.

First, the HMM model is employed to capture temporal dependencies in the groundwater quality data. The HMM's criteria for training the data are specified. The HMM produces sequences by the coexistence of two stochastic processes: the transition between states and the emission of an output sequence that possesses both output independence and Markov properties. The term 'hidden Markov model' refers to the fact that the sequence of state transitions is a hidden process, meaning that the variable states are seen through a series of emitted symbols rather than being directly observed. States, state probabilities, transition probabilities, emission probabilities, and starting probabilities are then used to build a HMM. They constitute an HMM's architecture (Franzese & Iuliano 2019). According to Jurafsky & Martin (2009), an HMM is specified by the following components:

States (S): The model's unobservable, hidden variables are denoted by *S*.

Observations (O): These are the variables that are visible and may be directly observed or measured at each time step. Observations reveal details about the underlying hidden state.

Transition Probabilities (A): These are the chances to transition from one hidden state to another. These are sometimes expressed as a matrix, with each member (*a_ij*) representing the likelihood of shifting from state *i* to state *j*.

Emission Probabilities (B): These are the likelihood of emitting an observation from a hidden condition. This can be expressed as a matrix, similar to transition probabilities, where each member (b_ik) indicates the likelihood of emitting observation k from state i.

Initial State Probabilities (π): These are the likelihood of starting the sequence in a specific hidden state. It is frequently expressed as a vector, with each element (π_i) representing the likelihood of starting in state i.

These elements use a set of equations to describe the dynamics of the system (Rabiner & Juang 1986) as follows:

State Transition Probability: This equation represents the probability of transitioning from state i to state j at time t.

$$P(q\_t = S\_j | q\_(t-1) = S\_i) = a\_ij \tag{6}$$

Observation Probability: This equation represents the probability of emitting observation k at time t, given that the system is in state i.

$$P(o\_t = O\_k | q\_t = S\_i) = b\_ik \tag{7}$$

Initial State Probability: This equation represents the probability of starting in state i at time 1.

$$P(q\_1 = S\_i) = \pi\_i \tag{8}$$

In this study, the Gaussian emission probabilities are used to estimate the hidden states based on the training data. Then, the conditions for the ANN model (Figure 3) are defined to predict groundwater quality categories. ANNs are groups of neurons that collaborate to carry out a certain job; they are inspired by biological neural networks. After receiving input and combining it with factors like weights and bias, a neuron produces its output by feeding the outcome via a non-linear activation function. Since neurons are arranged in layers, information passes across one or more hidden layers of neurons to go from the input layer to the output layer. The difference between the expected and projected output for various input data points is used to calculate the network's performance (Bedi *et al.* 2020).

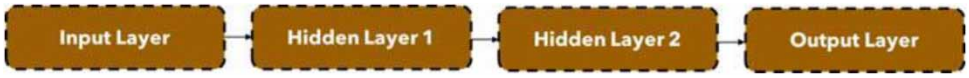


Figure 3 | Outline of the ANN architecture.

For this study, ANN architecture consists of an input layer, two hidden layers (128 and 64 neurons), and an output layer (Figure 3) with three neurons corresponding to categories 0, 1, and 2 which indicates excellent water, good water, and poor water. The dataset is trained for HMM and ANN separately. After that, the hybrid model – HMM and ANN models – works in parallel. The HMM estimates hidden states, while the ANN predicts groundwater quality categories. The hybrid model uses a majority voting rule to enhance the overall prediction accuracy. A confusion matrix based on Figure 4 is generated to visualize the model's performance in terms of True positives (TP), True negatives (TN), False positives (FP), and False negatives (FN).

Subsequently, the following performance metrics (accuracy, precision, recall, and F1-score) are computed to evaluate the model:

(i) Accuracy evaluates the overall model accuracy.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{9}$$

(ii) Precision, recall, and F1-score are the measures to evaluate the performance for each groundwater quality category (Class 0, Class 1, and Class 2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

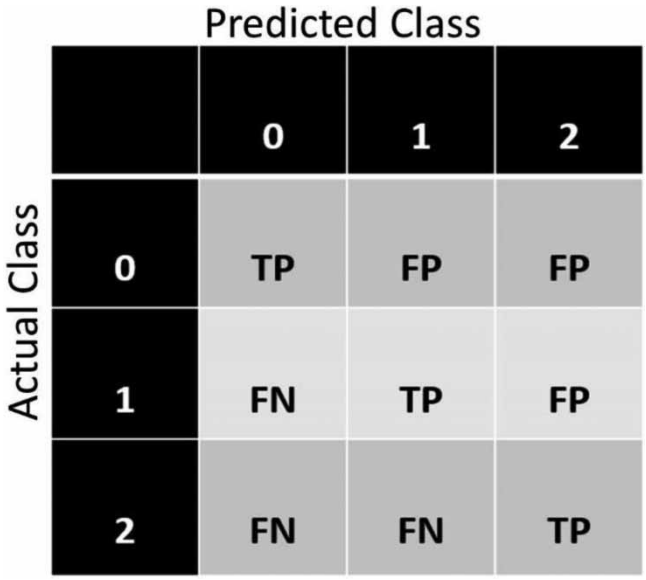


Figure 4 | Architecture of the confusion matrix for the multi-class classification.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

where TP indicates True Positive, FP indicates False Positive, TN indicates True Negative, and FN indicates False Negative.

(iii) Macro-average metrics are to compute macro-average precision, recall, and F1-score to assess the overall performance across classes and the weighted-average metrics are to calculate the weighted-average precision, recall, and F1-score to account for class imbalances.

$$\text{macro\_average} = \text{sum}(\text{metric\_per\_class})/\text{n\_classes} \tag{13}$$

$$\text{weighted\_average} = \text{sum}(\text{metric\_per\_class} * \text{class\_support})/\text{sum}(\text{class\_support}) \tag{14}$$

where metric_per_class is the metric (e.g., precision, recall, and F1-score) calculated for each class, n_classes are the number of classes, and class_support is the number of samples in each class.

In addition to this, cross-validation is one of the most important aspects of the study. Cross-validation is a model validation approach that is used to verify the efficacy of a machine learning model. It is used to limit issues such as overfitting and underfitting and to gain an understanding of how the model will generalize to an independent dataset. This is accomplished by separating the entire dataset into two sets: training and testing (Darapureddy *et al.* 2019). In this study, the k-fold cross-validation approach with $k = 5$ is applied. As a result, the entire dataset is partitioned into five folds and iterated five times. Cross-validation helps in evaluating how well the models predict groundwater quality on unseen data. After that, various visualizations are created to interpret and present the results effectively.

3. RESULTS AND DISCUSSION

3.1. Drinking water quality index

According to Verma *et al.* (2019), the calculated water quality index for drinking is categorized as Excellent (<50), Good (50–100), Poor (100–200), Extremely poor (200–300), and Unfit for consumption (>300). Of the examined samples in this study area, only Excellent (77%), Good (18%), and Poor water (5%) are observed. These categories are labeled as Class 0, Class 1, and Class 2 for data training and testing purposes.

3.2. Accuracy, precision, recall, and F1-score

The hybrid model HMM–ANN achieved an overall accuracy of almost 97.41%, demonstrating its ability to predict groundwater quality. Figure 5 shows the performance scores obtained during the research in detail. Precision scores indicate the reliability of the model's positive predictions for each class. Class 1 demonstrates perfect precision, while the lower precision in Class 2 suggests that when the model predicts this class, it may have a higher likelihood of false positives. Recall scores indicate how well the model captures all actual positive instances. Class 0 and Class 2 have high recall, indicating that the model effectively identifies all true positive instances in these classes. F1-scores take into account both precision and recall. Class 0 and Class 1 have a high F1-score, indicating a strong balance between precision and recall but Class 2's F1-score is relatively lower, highlighting potential room for improvement in precision–recall trade-offs for that class. Macro-average metrics compute the average precision, recall, and F1-score across all classes. The values provide a good overall performance assessment. Weighted-average metrics take into account class imbalances by giving more weight to classes with larger sample sizes. The high weighted precision, recall, and F1-score indicate that the model performs well on a weighted average across all classes. The model demonstrates strong classification performance with high accuracy, indicating its ability to make correct predictions.

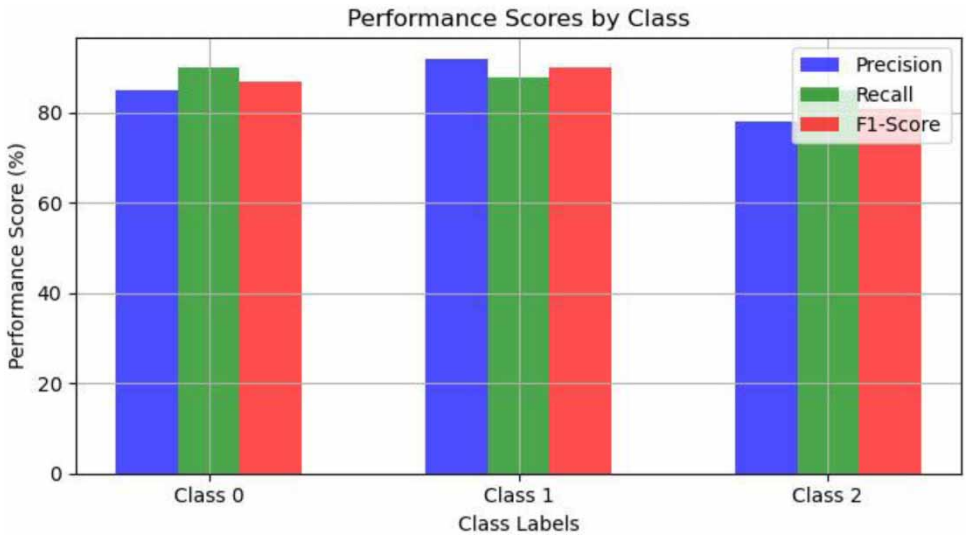


Figure 5 | Results of evaluation metrics.

3.3. Confusion matrix

The confusion matrix (Figure 6) shows the model's performance for each class in excellent detail. True Positive rates in Class 0 and Class 2 are high, showing that the model accurately detects the majority of cases in these classes. Class 1 performs reasonably well; however, there are some False Negatives and one False Positive.

3.4. ROC curves

From Figure 7, it is observed that all three classes (Class 0, Class 1, and Class 2) display substantial ROC curve regions, which suggests good prediction accuracy. Particularly for Class 2, the model obtains a perfect area of 1, indicating that it is particularly good at identifying and forecasting this class of groundwater quality. These high ROC curve regions show how well the HMM–ANN model works for categorizing and forecasting various groundwater quality categories. The overall performance of the suggested technique is demonstrated by the micro-average ROC curve area of 0.98. This micro-average takes into consideration the accumulation of forecasts across all classes, highlighting the model's capacity to produce precise and reliable predictions independent of the specific water quality category.

3.5. Prediction graph

A visual aid for comparing the model's predictions to the actual values in machine learning and data analysis is the prediction graph. This graph can direct future studies and model refinements to ensure reliable and consistent

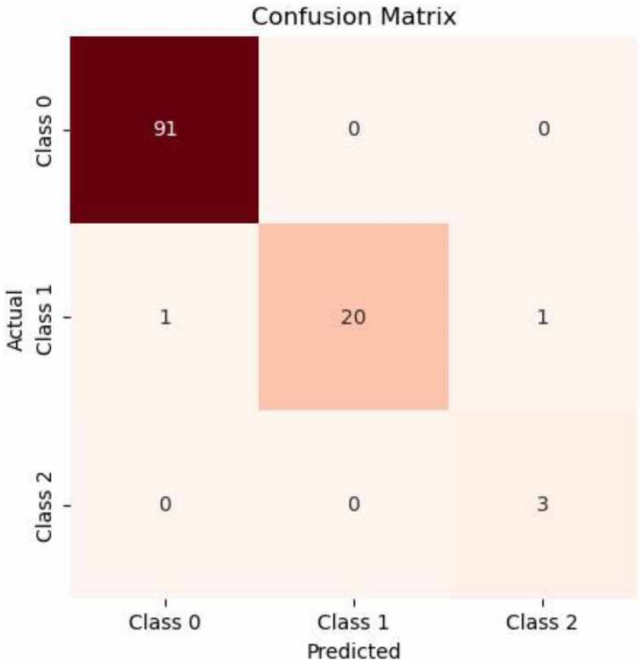


Figure 6 | Results of confusion matrix.

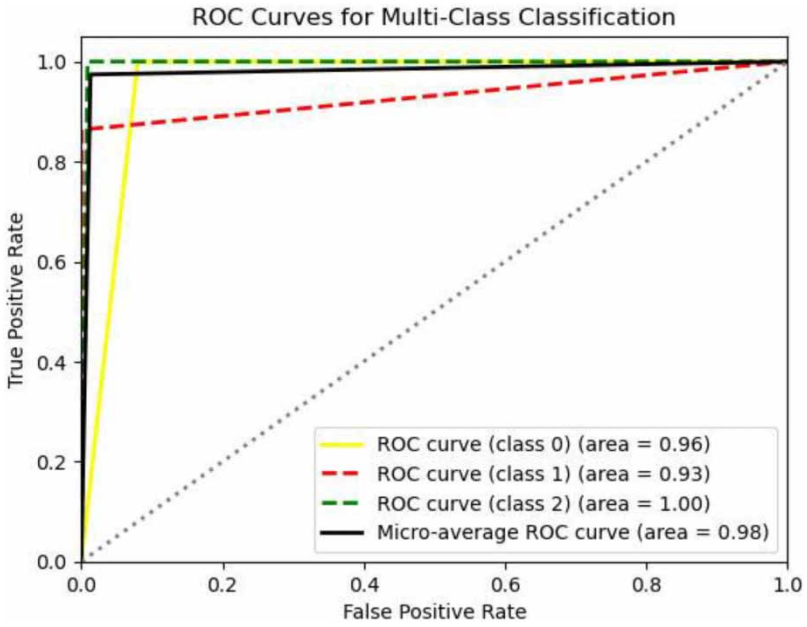


Figure 7 | ROC curve generated for the classes.

classification results. When looking at the graph (Figure 8), one can see that the model predictions are accurate when the 'Predicted' line (blue line) closely matches the 'Actual' line (orange line). With 97.41% accuracy, the prediction graph shows a reliable result as well.

3.6. Cross-validation scores

Cross-validation results suggest that the model has a mean accuracy of 0.95, demonstrating excellent prediction potential. The standard deviation of 0.02 suggests that there is great consistency across different subsets, showing stability and are not influenced by the random fluctuations in the dataset. As a result, when tested on different

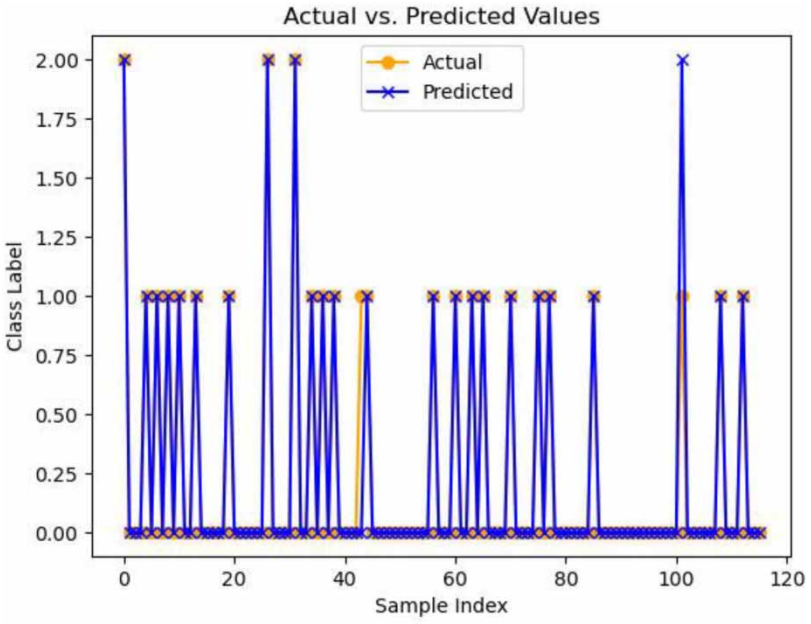


Figure 8 | Prediction graph generated by the HMM–ANN model.

partitions of data, the model's accuracy is not significantly affected. These findings show that the model is reliable in forecasting groundwater quality in Kanyakumari District, producing consistent and explicit results.

4. CONCLUSIONS

The study concludes with some exceptionally positive findings. The created model has been shown to be incredibly effective in estimating and evaluating groundwater quality. The model demonstrates its capacity to generate precise forecasts for various classes of groundwater quality, with an overall accuracy of 97%. Furthermore, the model's effectiveness in differentiating between different water quality categories is demonstrated by the evaluation metrics – accuracy, precision, recall, and F1-score metrics for each class. The fact that Class 0 and Class 1 both attained pristine F1-score is very distinctive. These findings demonstrate the model's capacity to reduce false positives and false negatives in addition to producing accurate predictions. The weighted-average and macro-average measures demonstrate that the model consistently performs well overall across all classes. The weighted F1-score of 0.97 and the macro F1-score of 0.92 demonstrate how well the model predicts the whole dataset while taking class imbalances into assessment. With remarkable areas under the curve for every class and a micro-average ROC curve area of 0.98, the ROC curve study further demonstrates the resilience of the model. The model consistently produces high-quality predictions, as evidenced by its 95% mean accuracy and 0.02 standard deviation on cross-validation. Since the model provides trustworthy predictions, the study holds great promise for the management of the water resources and the environment in the study area. This will help to ensure that groundwater supplies are preserved and used sustainably. Further research and implementation of the approach in the real world help to maximize its potential for addressing water quality issues in the study area.

FUNDING

No funding was received to assist with the preparation of this manuscript.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Acharya, U. & Sahani, G. 2022 Weather prediction analysis by using hybrid Markov model and artificial neural network. In: *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*. https://doi.org/10.1109/icais53314.2022.9743055.

Adimalla, N. & Qian, H. 2019 Groundwater quality evaluation using water quality index (WQI) for drinking purposes and human health risk (HHR) assessment in an agricultural region of Nanganur, south India. *Ecotoxicology and Environmental Safety* **176**, 153–161. https://doi.org/10.1016/j.ecoenv.2019.03.066.

Awad, M. & Khanna, R. 2015 Hidden Markov model. In: *Efficient Learning Machines*, pp. 81–104. https://doi.org/10.1007/978-1-4302-5990-9_5.

Bedi, S., Samal, A., Ray, C. & Snow, D. 2020 Comparative evaluation of machine learning models for groundwater quality assessment. *Environmental Monitoring and Assessment* **192**(12). https://doi.org/10.1007/s10661-020-08695-3.

Bureau of Indian Standards (BIS) 2012 *Drinking Water – Specification*.

Darapureddy, N., Karatapu, D. N. & Battula, D. T. K. 2019 Research of machine learning algorithms using k-fold cross validation. *International Journal of Engineering and Advanced Technology* **8**(6s), 215–218. https://doi.org/10.35940/ijeat.f1043.0886s19.

District Statistical Handbook (2021–2022) Department of Economics and Statistics, Kanyakumari District, Government of Tamil Nadu | Land of Cash Crops | India. n.d. Available from: https://kanniyakumari.nic.in.

Domenico, P. A. & Schwartz, F. W. 1990 *Physical and Chemical Hydrogeology*. Wiley, New York, pp. 824.

Estebanez, B., Del Saz-Orozco, P., Rivas, I., Bauzano, E., Muñoz, V. & García-Morales, I. 2012 Maneuvers recognition in laparoscopic surgery: Artificial neural network and hidden Markov model approaches. In: *The Fourth IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics Roma, Italy*. https://doi.org/10.1109/biorob.2012.6290734.

Franzese, M. & Iuliano, A. 2019 Hidden Markov models. In: *Encyclopedia of Bioinformatics and Computational Biology*, pp. 753–762. https://doi.org/10.1016/b978-0-12-809633-8.20488-3.

Freeze, R. A. & Cherry, J. A. 1979 *Groundwater*. Prentice Hall, p. 384.

Hanoon, M. S., Ahmed, A. N., Fai, C. M., Birima, A. H., Razzaq, A., Sherif, M., Sefelnasr, A. & El-Shafie, A. 2021 Application of artificial intelligence models for modeling water quality in groundwater: Comprehensive review, evaluation and future trends. *Water, Air, & Soil Pollution* **232**, 10. https://doi.org/10.1007/s11270-021-05311-z.

Jose, A. A. & Srinivas, Y. 2023 An assessment on the quality of groundwater in Chennai's urbanised areas. *International Journal of Global Warming* **29**(4), 406. https://doi.org/10.1504/ijgw.2023.130159.

Jurafsky, D. & Martin, J. H. 2009 *Speech and Language Processing*. Prentice Hall, pp. 168–172.

Khudair, B. H., Jasim, M. M. & Alsaqqar, A. S. 2018 Artificial neural network model for the prediction of groundwater quality. *Civil Engineering Journal* **4**(12), 2959. https://doi.org/10.28991/cej-03091212.

Kishna, N. P. T. & Francis, S. 2017 Intelligent tool for Malayalam cursive handwritten character recognition using artificial neural network and hidden Markov model. In: *2017 International Conference on Inventive Computing and Informatics (ICICI)*. https://doi.org/10.1109/icici.2017.8365201.

Li, P., Wu, J. & Shukla, S. 2022 Achieving the one health goal: Highlighting groundwater quality and public health. *Water* **14**(21), 3540. https://doi.org/10.3390/w14213540.

Listyani, R. A. T. & Putranto, T. T. 2022 Groundwater quality assessment for drinking and clean water in Bagelen and its surrounding area. *Sustinere: Journal of Environment and Sustainability* **6**(2), 121–131. https://doi.org/10.22515/sustinerejes.v6i2.188.

Nguyen, N. 2017 An analysis and implementation of the hidden Markov model to technology stock prediction. *Risks* **5**(4), 62. https://doi.org/10.3390/risks5040062.

Ong, H. F. & Ahmad, A. M. 2011 Malay language speech recogniser with hybrid hidden Markov model and artificial neural network (HMM/ANN). *International Journal of Information and Education Technology* 114–119. https://doi.org/10.7763/ijiet.2011.v1.19.

Preziosi, E., Rotiroti, M., Condesso de Melo, M. T. & Hinsby, K. 2021 Natural background levels in groundwater. *Water* **13**(19), 2770. https://doi.org/10.3390/w13192770.

Rabiner, L. & Juang, B. 1986 An introduction to hidden Markov models. *IEEE ASSP Magazine* **3**(1), 4–16. https://doi.org/10.1109/massp.1986.1165342.

Raheja, H., Goel, A. & Pal, M. 2021 Prediction of groundwater quality indices using machine learning algorithms. *Water Practice and Technology* **17**(1), 336–351. https://doi.org/10.2166/wpt.2021.120.

Rajammal, T. S. J., Balasubramaniam, P. & Kaledhonkar, M. J. 2021 Assessment of groundwater quality in Kanyakumari district, Tamil Nadu, using ionic chemistry. *Current Science* **121**(5), 676. https://doi.org/10.18520/cs/v121/i5/676-684.

Sanghavi, M. A. & Bojewar, S. M. 2014 Classification of cardiovascular disease from ECG using artificial neural network and hidden Markov model. *IOSR Journal of Computer Engineering* **16**(3), 92–98. https://doi.org/10.9790/0661-16349298.

Trentin, E. & Gori, M. 2003 Robust combination of neural networks and hidden Markov models for speech recognition. *IEEE Transactions on Neural Networks* **14**(6), 1519–1531. https://doi.org/10.1109/tnn.2003.820838.

Ubah, J. I., Orakwe, L. C., Ogbu, K. N., Awu, J. I., Ahaneku, I. E. & Chukwuma, E. C. 2021 Forecasting water quality parameters using artificial neural network for irrigation purposes. *Scientific Reports* **11**(1). https://doi.org/10.1038/s41598-021-04062-5.

Vasanthavigar, M., Srinivasamoorthy, K., Vijayaragavan, K., Rajiv Ganthi, R., Chidambaram, S., Anandhan, P., Manivannan, R. & Vasudevan, S. 2010 Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. *Environmental Monitoring and Assessment* **171**(1–4), 595–609. https://doi.org/10.1007/s10661-009-1302-1.

Verma, P., Singh, P. K., Sinha, R. R. & Tiwari, A. K. 2019 Assessment of groundwater quality status by using water quality index (WQI) and geographic information system (GIS) approaches: A case study of the Bokaro district, India. *Applied Water Science* **10**(1). https://doi.org/10.1007/s13201-019-1088-4.

World Health Organisation (WHO) 2011 *Guidelines for Drinking Water Quality*, 4th edn. ISBN: 978-92-4-154995-0.

First received 24 January 2024; accepted in revised form 16 May 2024. Available online 28 May 2024