



# Forecasting water quality variable using deep learning and weighted averaging ensemble models

Mohammad G. Zamani<sup>1</sup> · Mohammad Reza Nikoo<sup>2</sup> · Sina Jahanshahi<sup>3</sup> · Rahim Barzegar<sup>4</sup> · Amirreza Meydani<sup>5</sup>

Received: 15 June 2023 / Accepted: 27 October 2023 / Published online: 24 November 2023  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

Water quality variables, including chlorophyll-a (Chl-a), play a pivotal role in comprehending and evaluating the condition of aquatic ecosystems. Chl-a, a pigment present in diverse aquatic organisms, notably algae and cyanobacteria, serves as a valuable indicator of water quality. Thus, the objectives of this study encompass: (1) the assessment of the predictive capabilities of four deep learning (DL) models — namely, recurrent neural network (RNN), long short-term memory (LSTM), gated recurrence unit (GRU), and temporal convolutional network (TCN) — in forecasting Chl-a concentrations; (2) the incorporation of these DL models into ensemble models (EMs) employing genetic algorithm (GA) and non-dominated sorting genetic algorithm (NSGA-II) to harness the strengths of each standalone model; and (3) the evaluation of the efficacy of the developed EMs. Utilizing data collected at 15-min intervals from Small Prespa Lake (SPL) in Greece, the models employed hourly Chl-a concentration lag times, extending up to 6 h, as models' inputs to forecast Chl-a ( $t+1$ ). The proposed models underwent training on 70% of the dataset and were subsequently validated on the remaining 30%. Among the standalone DL models, the GRU model exhibited superior performance in Chl-a forecasting, surpassing the RNN, LSTM, and TCN models by 8%, 2%, and 2%, respectively. Furthermore, the integration of DL models through single-objective GA and multi-objective NSGA-II optimization algorithms yielded hybrid models adept at effectively forecasting both low and high Chl-a concentrations. The ensemble model based on NSGA-II outperformed standalone DL models as well as the GA-based model across a range of evaluation indices. For instance, considering the *R*-squared metric, the study's findings demonstrated that the EM-NSGA-II stands out with exceptional effectiveness compared to DL and EM-GA models, showcasing improvements of 14% (RNN), 8% (LSTM), 6% (GRU), 8% (TCN), and 3% (EM-GA) during the testing phase.

**Keywords** Water quality forecasting · Ensemble model · Deep learning (DL) · Single- and multi-objective optimization algorithms · Non-dominated genetic algorithm (NSGA-II)

---

Responsible Editor: Marcus Schulz

✉ Mohammad Reza Nikoo  
m.reza@squ.edu.om

<sup>1</sup> Department of Water Resources Engineering, Faculty of Civil, Water and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran

<sup>2</sup> Department of Civil and Architectural Engineering, Sultan Qaboos University, Muscat, Oman

<sup>3</sup> Department of Water Resources Engineering, Faculty of Civil, Water and Environmental Engineering, University of Tehran, Tehran, Iran

<sup>4</sup> Groundwater Research Group (GRES), Research Institute on Mines and Environment (RIME), Université du Québec en Abitibi-Témiscamingue (UQAT), Amos, Québec, Canada

<sup>5</sup> Department of Geography and Spatial Sciences, University of Delaware, Newark, DE, USA

## Abbreviation

AI	Artificial intelligence
ANFIS	Adaptive neural-based fuzzy inference system
ANN	Artificial neural network
BPNN	Back propagation neural network
CD	Crowding distance
Chl-a	Chlorophyll-a
CNN	Convolutional neural networks
CEEMDAN	Complete ensemble empirical mode decomposition with adaptive noise
VMD	Variational mode decomposition
DL	Deep learning
DO	Dissolved oxygen
EC	Electrical conductivity
ELM	Extreme learning machine
EM	Ensemble model

EM-GA	Ensemble model based on genetic algorithm
EM-NSGA-II	Ensemble model based on non-dominated sorting genetic algorithm
FFNN	Feed-forward neural network
GA	Genetic algorithm
GRU	Gated recurrent unit
LSTM	Long short-term memory
LSSVM	Least-squares support vector machine
MAE	Mean absolute error
MARE	Mean absolute relative error
ML	Machine learning
MOO	Multi-objective
MP	Mutant population
MSE	Mean square error
ND	Non-dominated
NSE	Nash-Sutcliffe efficiency
NSGA-II	Non-dominated sorting genetic algorithm
OP	Offspring population
ORP	Oxidation-reduction potential
P	Population
pH	Power of hydrogen
PRMSE	Percentage root mean square error
RMSE	Root mean square error
RMAE	Root mean absolute error
RNN	Recurrent neural network
R-squared ( $R^2$ )	Coefficient of determination
Water Temp.	Water temperature
TCN	Temporal convolutional network
TDS	Total dissolved oxygen
SOO	Single objective
SPL	Small Prespa Lake
SVM	Support vector machine
SVR	Support vector regression
WQV	Water quality variable

## Introduction

Sustainable water management entails using water in a way that fulfills present human needs while also ensuring its availability for future generations (Azizi et al. 2022; Nova 2023). It requires careful consideration of three aspects, including economic efficiency, social quality, and environmental preservation (Jahanshahi and Kerachian 2019; Jahanshahi et al. 2023). The concept of sustainable development is intricately tied to water quality management. This integration is explicitly embraced within the sustainable development framework (Chapman and Sullivan 2022). In addition to playing an important role in environmental processes, water quality also has a profound effect on human and animal health. Water contamination can cause serious health problems, including gastrointestinal diseases, skin diseases,

and cancer, among others (Lin et al. 2022; Babuji et al. 2023). Therefore, monitoring the quality of water is a fundamental component of effectively managing water resources (Sinshaw et al. 2019; Uddin et al. 2021). Evaluating water quality involves the monitoring of diverse variables, with the resulting data finding application across multiple domains. For instance, the analysis of surface and groundwater quality through spatiotemporal data (Yan et al. 2022; Zamani et al. 2023b; Uddin et al. 2023a, b, c) aids in the identification of emerging contaminants, thereby facilitating the formulation of effective strategies for safeguarding and conserving environments (Nadiri et al. 2022). Additionally, WQVs can be integral in optimizing reservoir operations and water allocation (Nikoo et al. 2013), as well as in establishing dedicated networks for water quality monitoring (Jiang et al. 2020).

Globally, one of the noteworthy concerns considering water quality is the eutrophication of reservoirs and rivers, which poses significant economic, social, and environmental risks (Vinçon-Leite and Casenave 2019). This phenomenon entails water bodies becoming enriched with excessive plant nutrients, primarily phosphorous and nitrogen. The proliferation of these nutrients leads to an accelerated growth of algae, significantly deteriorating the quality of water across diverse sources (Boyd 2020; Uddin et al. 2022b). Chl-a is one of the commonly employed indicators for assessing eutrophication, offering the potential to supply an approximation of the algae concentration in a water body (Zhang et al. 2019). Thus, monitoring the levels of Chl-a using AI tools, as an innovative and efficient approach, can help manage water bodies more sustainably (Shin et al. 2020; Barzegar et al. 2020). Numerous research studies are available within the scope. For example, Li et al. (2017) utilized hybrid ANNs to forecast Chl-a levels, as a representative index of eutrophication, in several lakes in China. Guo et al. (2018) assessed the condition related to the nutrient status of a water supply reservoir by analyzing Chl-a data during different seasons. Papenfus et al. (2020) investigated the feasibility of utilizing spectral bands within remote sensing data to monitor Chl-a levels in lakes and reservoirs located in the USA by analyzing both in situ measurements and satellite imagery on a nationwide scale.

ML and its subset, DL, have transformed the field of AI with their applications. Such approaches have increasingly gained popularity in the field of water resources such as in hydrology (Khosravi et al. 2021; Meydani et al. 2022). Most traditional knowledge-driven methods are laborious and susceptible to prediction errors since they presume linear relationships among variables (Ma et al. 2014). Unlike traditional models, ML and DL models excel in managing complex and non-linear systems due to their ability to operate without the need for an initial hypothesis regarding the data structure or underlying physical processes (Shen et al. 2018; Barzegar et al. 2019; Sahraei et al. 2021; Uddin et al. 2022a). Ewuzie et al. (2022) highlighted the

increasing interest in the use of ML and DL, particularly DL, in the domain of modeling and predicting/forecasting water quality. Numerous studies have explored the capability of different AI models, encompassing ML and DL models, such as NB (Bhardwaj et al. 2022; Uddin et al. 2023a, b, c), RF (Sakaa et al. 2022; Virro et al. 2022; Zamani et al. 2023a), ANNs (Chen and Liu 2014; Dawood et al. 2021; Wang et al. 2022; Georgescu et al. 2023; Farshbaf Aghajani et al. 2023), RNN (Cho et al. 2014; Qi et al. 2020; Prasad et al. 2022; Ni et al. 2023), BPNN (Barzegar and Asghari Moghaddam 2016; Chen et al. 2023b), ANFIS (Gaya et al. 2020; Rizal et al. 2023), SVM (Haghiabi et al. 2018; Chou et al. 2018; Elkiran et al. 2019; El Bilali and Taleb 2020; Ortiz-Lopez et al. 2022; Uddin et al. 2023a, b, c), ELM (Zhu and Heddami 2020), MLR (Barzegar et al. 2017; Liu et al. 2019), CNN (Pyo et al. 2020; Chen et al. 2020b), LSTM (Li et al. 2022; Liang et al. 2020; Kouadri et al. 2021), GRU (Luo et al. 2019; Fu et al. 2021), and TCN (Chen et al. 2020a) for predicting/forecasting various WQVs, such as DO, salinity, fluoride, Chl-a (Choi et al. 2019), TDS, and EC (Li et al. 2019).

While individual ML/DL models can achieve satisfactory performance in predicting/forecasting WQVs, hybrid models combine diverse individual models to address limitations such as a restricted capacity to capture intricate patterns, limited generalization, challenges in handling noise and outliers, and constrained interpretability. These hybrid models harness the strengths of each singular model, including their simplicity, faster training, and robust interference capabilities (Barzegar et al. 2018; Chen and Dai, 2020; Bahrami et al. 2023). The main objective of an EM is to integrate standalone models in a way that enhances the precision of the outcomes of standalone models. The enhancement in the hybrid model is attained by maximizing the benefits of standalone models while minimizing their limitations (Rozinajová et al. 2018). In the past few years, hybrid procedures have found extensive application in the evaluation of water quality, modeling, and prediction/forecasting. Barzegar et al. (2018) investigated ELM and hybrid models designed to forecast EC values over multiple time steps and compared the results with the outcomes of an ANFIS and a wavelet-ANFIS model. Bui et al. (2020) explored four individual and twelve hybrid ML prediction models for WQVs in Iran. Barzegar et al. (2020) employed the CEEMDAN and VMD algorithms alongside ELM and LSSVM models to predict DO and Chl-a levels in the SPL in Greece. Cao et al. (2020) combined the GRU neural network and the K-means clustering method for the prediction of DO. Dehghani et al. (2021) developed four hybrid models by incorporating four metaheuristic optimization algorithms into SVR and observed improvement in their results. Alqahtani et al. (2022) presented individual ML algorithms alongside their amalgamation for the prediction of EC and TDS. This combination incorporated RF, GEP, and ANN. In a different

light, Sakaa et al. (2022) formulated a hybridized model utilizing sequential minimal optimization and SVM to predict WQVs. Wu and Wang (2022) developed an ensemble model by integrating ANN and LSTM and their combination with wavelet transform for WQV prediction in a river in China. Moreover, Zamani et al. (2023a) proposed an innovative fusion model leveraging the Bayesian maximum entropy fusion approach to precisely estimate DO and Chl-a concentrations across various locations and depths within a reservoir in Oman.

Considering the existing literature, it is evident that various data-driven algorithms have been employed to accurately model WQVs. However, as far as the authors are concerned, the current literature still lacks a comprehensive comparative analysis of widely established DL forecasting techniques and their ensembling utilizing single- and multi-objective optimization structures — specifically, GA and NSGA-II — for WQV forecasting. Moreover, the integration of DL models holds the potential to enhance the precision of forecasting WQVs, especially in situations characterized by incomplete or insufficient observed data, such as short-period time series. Given these considerations, this study seeks to contribute to the realm of WQV (e.g., Chl-a) forecasting. The foundational framework for this modeling endeavor revolves around the training of four distinct DL models — namely, RNN, LSTM, GRU, and TCN — using Chl-a concentration data. Subsequently, these trained DL models are integrated through the application of MOO algorithms, resulting in the creation of EMs that is poised to forecast Chl-a concentration. Notably, this study marks the first endeavor to utilize an NSGA-II-based ensemble DL model for Chl-a forecasting. This study primarily aims to (1) comprehensively examine and compare the capabilities of diverse individual DL algorithms (e.g., RNN, LSTM, GRU, and TCN) in effectively forecasting Chl-a concentration 1 h ahead in SPL, Greece; (2) combine these individual DL models into a harmonized ensemble model, leveraging the strengths of each constituent model through the application of meta-heuristic optimization algorithms, including GA and NSGA-II; (3) execute a comparative analysis, contrasting the results generated by the individual DL models with those yielded by the EMs; (4) compare the results obtained from the EMs, thereby highlighting and elucidating the proficiency of NSGA-II in seamlessly integrating the distinct individual DL models. Forecasting the concentration of Chl-a in the lake 1 h ahead holds practical significance in environmental monitoring, water management, and ecological research. This predictive capability aids in detecting harmful algal blooms, guiding eutrophication assessments, facilitating timely responses to water quality changes, ensuring safe recreational activities, and optimizing aquaculture and fisheries practices. By providing insights into Chl-a dynamics, this forecasting enhances decision-making for ecosystem health, water safety, and sustainable resource management.

## Methodology

The DL models are utilized to forecast the concentration of Chl-a in SPL in Greece. After assessing the effectiveness of these models using different statistical metrics, they are integrated into a hybrid model through the utilization of the GA and NSGA-II algorithms. The capability of these ensemble approaches to combine the results of diverse DL models is then compared to determine their effectiveness. The models are generated considering the standard practice for developing the DL forecasting model which is briefly reviewed in the following. This is a general framework suggested to ensure that the final hybrid model would work appropriately (Goodfellow et al. 2016; Chollet 2021; Géron 2022). Figure 1 depicts an illustrative depiction of the flowchart outlining the proposed methodology for the purpose of the Chl-a concentration. The framework comprises the subsequent stages.

Data collection and processing were initiated to assess water quality in this study. To obtain the required information, a multi-probe sensor CYCLOPS-7, developed by TURNER DESIGNS, was employed to measure various parameters in SPL. This step is shown in Fig. 1a.

The development of DL models involves several steps that ensure a robust and effective solution to a given problem. These steps are depicted in Fig. 1b. First, it is crucial to define and understand the problem, as it guides the selection of an appropriate DL model. In this study, the objective is to forecast short-term Chl-a concentrations using historical data (i.e., lag times of Chl-a). The evaluation metrics, such as RMSE and  $R^2$ , are explained as they are selected to evaluate the models during the training and testing phases. Relevant data were collected and preprocessed, involving cleaning, handling missing data, data transformation, and feature extraction (i.e., input variable selection). Splitting the data is also of utmost importance to assess the performance of the model on data. It was done by separating the dataset into two primary sets: the training and testing datasets. Four models were selected for their ability to work with time-series data. Subsequently, the model underwent training using the preprocessed data through backpropagation to adjust the weights. Fine-tuning may be necessary to improve the model's performance by adjusting hyperparameters, modifying the architecture, or adding regularization techniques. Finally, the models were tested on the validation dataset to guarantee its capacity to apply to novel data, ensuring generalizability. Achieving acceptable effectiveness on the testing dataset is a critical indication of the model's capability to apply knowledge to unfamiliar data and its readiness for deployment in solving the given problem. To improve the results obtained from individual DL models, model ensembling strategies based on GA and NSGA-II

were developed, as illustrated in Fig. 1c. The comparison was based on several evaluation indices and aimed to demonstrate the efficacy of the model ensembling approach.

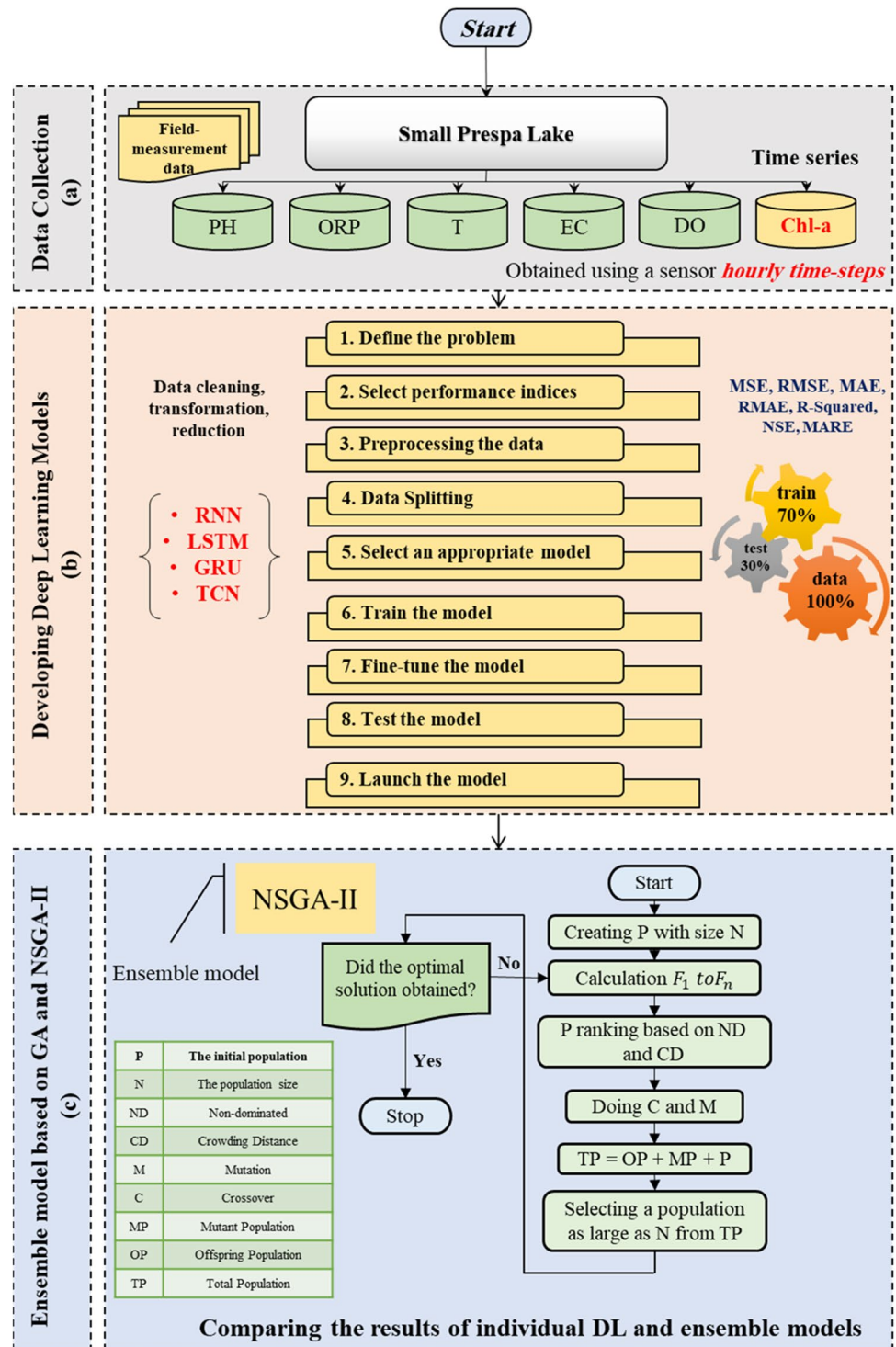
## RNN

An ANN known as a RNN falls within the realm of neural network models that can establish loops by connecting nodes, providing the ability for the output of one node to influence the output of the next node (Medsker and Jain 2001). By utilizing their internal state or memory, RNNs exhibit the capability to handle input signals of diverse lengths due to their temporal dynamic behavior, in contrast to feedforward neural networks (Carcano et al. 2008). RNNs are a type of network characterized by their infinite impulse response, while CNNs refer to networks with a finite impulse response. Both classes of networks have a temporal dynamic component. Finite impulse recurrent networks are structured as directed acyclic graphs, allowing them to be unfolded and replaced with exclusively FFNN. Conversely, directed cyclic graphs are manifested by infinite impulse recurrent networks, making them non-unrollable. Figure S1 (a & b) illustrates the architecture of the developed RNN in the current study, along with a detailed representation of the RNN structure.

## LSTM

The LSTM consists of a sequence of LSTM cells, with each of these cells having gates that control the flow of information within them (Fig. S2 (a & b)). The input gate selectively determines which data from the current input should be preserved and propagated through the cell state. Similarly, the forget gate decides which data should be discarded. Lastly, the output gate determines the specific data from the cell state that should be passed to the output. Alongside these gates, the LSTM cell includes a memory cell that accumulates information over time (Schmidhuber and Hochreiter 1997). This capacity allows the network to selectively retain or forget information as necessary, making it well-suited for processing lengthy sequences of data. During its forward pass, the LSTM network processes input sequentially, handling one element at a time. As each new input element is processed, the network updates its hidden state and memory cell. The final output is often generated by a fully connected layer, taking the final hidden state as input. The gates use sigmoid activation functions to regulate data transmission and a hyperbolic tangent function to process data entering the cell state (Greff et al. 2016). For a detailed mathematical implementation of LSTM, refer to Graves et al. (2012).

**Fig. 1** The visual representation of suggested forecasting framework



**GRU**

The GRU is a recurrent neural network architecture initially introduced by Chung and colleagues (Chung et al. 2014) (Fig. S3). Similar to an LSTM unit, a GRU unit comprises gating mechanisms that regulate the flow of information through these gates. However, unlike LSTM, the GRU unit

does not have a separate cell for storing information. The GRU structure consists of three gates — the candidate activation gate, the update gate, and the reset gate. The reset gate controls the extent to which the previous hidden state is overlooked, while the update gate manages the incorporation of new input data into the current state. By considering the previous hidden state and the current input, the candidate

activation function calculates the hidden state of the new candidate. For a comprehensive mathematical explanation of GRUs, refer to Haverkos et al. (2016).

## CNN

CNNs are a type of deep learning algorithm that has had a significant impact on computer vision, finding widespread use in such tasks as object detection, semantic segmentation, and image recognition (Sultana et al. 2020). The application of such sophisticated algorithms holds promise for enhancing water quality monitoring. These networks consist of multiple layers and encompass several crucial components, including fully connected layers, convolutional layers, and pooling layers. During the forward pass, the provided information undergoes a sequence of operations. It starts by being fed into the initial layer, which contains a convolutional layer responsible for extracting features from the input data. This extraction is achieved through the application of learned filters acquired during the training process. Following the convolutional layer, the output goes through a pooling layer that reduces spatial dimensionality, thereby helping to prevent overfitting. Finally, the output is passed through one or more fully connected layers that conduct classification or regression operations based on the extracted features. For further information on this methodology, refer to Chua and Roska (1993)

## TCN

TCNs were introduced by Bai et al. (2018) as a specialized neural network architecture designed for sequence modeling and prediction/forecasting tasks. TCNs employ 1D convolutional layers, which are adept at capturing temporal dependencies in data while using fewer parameters compared to traditional RNNs. This characteristic makes TCNs more efficient and easier to train. The core structure of a TCN consists of a series of 1D convolutional layers, optionally followed by a fully connected layer and an output layer. Each convolutional component employs fixed kernel sizes and strides, with the number of filters typically increasing as the sequence length decreases. The results from each convolutional component are then passed through a ReLU activation function and merged with the input to the subsequent layer using a skip connection. The skip connection facilitates the flow of information through the network without undergoing filtration by the convolutional layers, thereby addressing the vanishing gradient problem. For a detailed understanding of the implementation of a TCN model, refer to Fu et al. (2021). Figure S4 illustrates the structure of the TCN model used in this study.

## Ensemble models

EM integrates the obtained outcomes of different models. Such procedure combines the outputs from different algorithms or models to achieve more accurate results (Hall and Llinas 1997). Regarding the vector of weights  $w = [w_1, w_2, \dots, w_n]^T$  for the results generated by standalone models, the result of the EM ( $p_j$ ) is performed in the following manner:

$$Y_j = \sum_{i=1}^n w_i y_{ij}^* \quad (1)$$

Subject to:

$$\sum_{i=1}^n w_i = 1, \quad w_i \in [0, 1] \quad (2)$$

wherein  $n$  and  $y_{ij}^*$  are numbers and forecasted values by  $i$ th singular models, respectively (i.e., RNN, LSTM, GRU, and TCN). It is pertinent to highlight that the rationale behind opting for a variant of stacking, specifically weighted averaging, instead of pursuing boosting or bagging techniques, is attributed to factors encompassing model diversity (Zhou et al. 2023), complexity and interpretability (Ribeiro and dos Santos Coelho 2020), adaptability (Tang et al. 2022), ensemble size (Zounemat-Kermani et al. 2021), and the careful navigation of trade-offs (Card et al. 2019; Tan et al. 2019). The decision to utilize GA and NSGA-II as optimization methodologies for weighted averaging is underpinned by a multifaceted set of advantages. Firstly, GA and NSGA-II are renowned for their inherent flexibility, making them highly adaptable to a wide array of complex optimization challenges (Goldberg 1989; Katoch et al. 2021). Secondly, these algorithms are adept at generating a rich diversity of solutions, an attribute that is indispensable for thoroughly investigating various combinations of base learners (Deb et al. 2002). Thirdly, their robust design enables them to effectively manage and filter out noisy or irrelevant base learners through integrated feature selection and robustness mechanisms (Sivanandam et al. 2008). Lastly, both GA and NSGA-II demonstrate exceptional proficiency in fine-tuning the weights assigned to each base learner, thereby enhancing the overall efficacy of the ensemble model (Zhang and Li 2007; Xu et al. 2022; Dai et al. 2022).

## GA

A GA can function as an EM, which is a machine learning technique that combines multiple models using a GA optimization procedure to enhance the precision and effectiveness of an overall predictive water quality model (Barzegar et al.

2016; Chen et al. 2023a). In this approach, GA is employed to identify the most suitable subset of models from a pool of potential models, which are generally diverse and complementary in their forecasting capabilities (Babatunde et al. 2014). GA optimizes the model ensembling process by iteratively selecting the best combination of models and refining the amalgamation through crossover and mutation procedures. The GA process commences with a population of potential models, subject to fitness evaluation based on their accuracy and other performance metrics. Models with higher fitness are selected and bred to generate offspring, inheriting genetic traits from their parents. The offspring then undergo mutation and crossover operations to introduce diversity and explore new solutions (Mirjalili and Mirjalili 2019). Figure S5 illustrates the GA process used to determine optimal solutions for the Chl-a forecasting model.

### EM based on NSGA-II

NSGA-II, introduced by Deb et al. (2000), is a MOO algorithm that combines the principles of GAs with the concepts of dominance and CD. This optimization technique generates population members using crossover and mutation within the GA framework. These members are then categorized into different fronts based on their non-dominance. Subsequently, population members within each front are ranked using the CD operator. In this study, the decision variables within NSGA-II pertain to the weights assigned to different DL models. The optimization process revolves around two objective functions, as outlined below:

$$Z_1 = \frac{\sum_{j=1}^k \left| \frac{\sum_{i=1}^n w_i y_{ij}^* - y_j}{y_j} \right|}{k} \times 100 \quad (3)$$

$$Z_2 = \frac{\sum_{j=1}^m \left| \frac{\sum_{i=1}^n w_i x_i - x_j}{x_j} \right|}{m} \times 100 \quad (4)$$

where  $Z_1$  and  $Z_2$  represent the MARE between the EM output and the actual value assigned to the data used for both the training and testing phases. In accordance with the description of these two objective functions, NSGA-II is utilized to find weights that minimize these two objective functions. For clarity, the other variables in Eqs. 1–4 are elaborated on hereunder:

- $y_{ij}^*$  The  $i^{\text{th}}$  output value of each model on the TD
- $y_j$  The observed output value of the TD
- $x_{ij}^*$  The  $i^{\text{th}}$  forecasted value for each instance in the TD
- $x_j$  The recorded value of TD

$k$  The quantity of TD

$m$  The number of TD

### The structure of NSGA-II

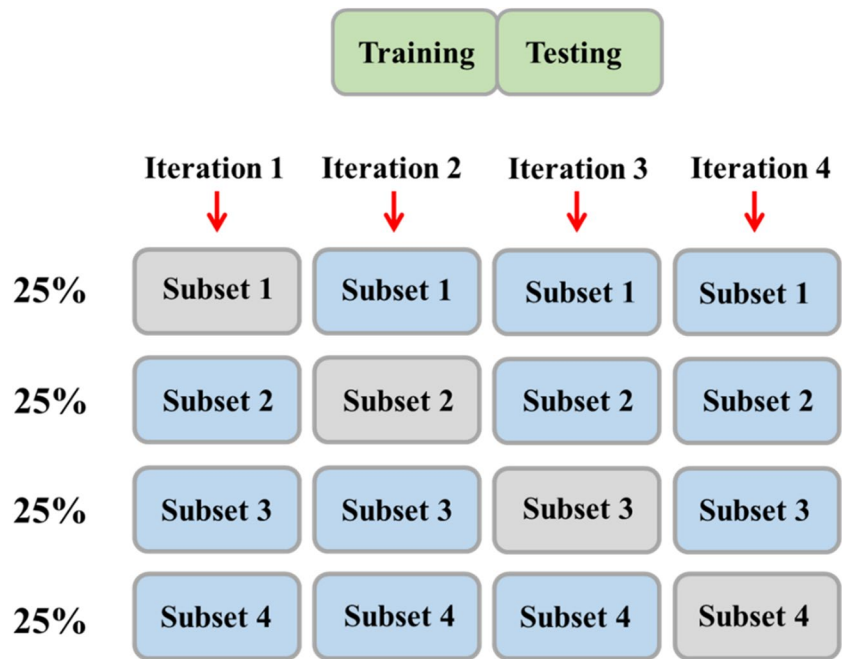
MOO techniques, derived from competing objectives, result in multiple optimal solutions in contrast to SOO methods. One particularly effective technique is NSGA-II, a robust method for optimizing multiple objectives (Deb et al. 2002; Zamani et al. 2022, 2023c). The process begins by creating an initial population of  $N$  individuals, followed by calculating the objective function values ( $Z_1$  to  $Z_n$ ). Subsequently, the initial population is evaluated and ranked based on the criteria of ND and CD. To enhance the overall response, the process employs two key procedures: mutation and crossover. These operators are used to generate an MP and an OP, respectively. The total population consists of three subpopulations: the original P, the MP, and the OP (Ehsani et al. 2022). This process is iteratively repeated to determine the optimal solution. The flowchart in Fig. S6 illustrates this iterative process.

### Cross-validation approach for DL models

Cross-validation serves as a crucial approach for assessing the model's predictive capacity on new and unseen data, offering essential insights into its ability to generalize beyond the training set. This methodology provides an estimation of the model's potential performance when applied to unfamiliar data points (Zhou 2012). To mitigate concerns, a strategy involves withholding a portion of the dataset during the training phase. Prior to commencing the training process, this subset of data is segregated and kept separate. Following the completion of the training phase, the remaining data is then utilized to assess the algorithms' proficiency. This foundational principle underlies a broad spectrum of model evaluation techniques collectively referred to as cross-validation.

The process of partitioning the dataset holds the power to influence the outcomes of machine learning algorithms. Numerous methods have been proposed in previous research endeavors to implement the concept of cross-validation. Nevertheless, these methodologies all share fundamental elements that encapsulate a common essence (Wang et al. 2023). Among the various available techniques, this study opted for the hold-out procedure due to its straightforward and uncomplicated nature. Figure 2 visually presents the employed cross-validation approach used in this study.

**Fig. 2** The utilization of CV methodology



**Model performance evaluation**

This study evaluated the forecasting accuracy using various statistical metrics, including MSE, RMSE, MAE, RMAE,  $R^2$ , NSE, and MARE %. The MSE and its variant, RMSE, delve into the squared differences between forecasted and actual values, effectively capturing both the magnitude and direction of errors. They range from 0 to  $\infty$ , with lower values indicating closer alignment between forecasting and observed values (Uddin et al. 2023a, b, c). These metrics emphasize the significance of larger errors due to the squaring process, making them particularly sensitive to outliers. While providing a comprehensive view of overall forecasting accuracy, the squared nature of these metrics might obscure the understanding of the directional bias of errors. Consequently, they might not readily distinguish between underestimation and overestimation tendencies. The RMAE introduces a dimension of normalization, allowing for comparisons relative to the scale of the observed data, thereby aiding in discerning proportionate inaccuracies (Hajikarimi et al. 2022). The  $R^2$  metric operates on a scale from 0 to 1. It measures the extent to which the model’s forecasting values can account for the variability in the recorded data. A value nearing 1 indicates a stronger alignment of the model with the data, highlighting a greater level of concurrence between forecasted and recorder values. However,  $R^2$  does not explicitly indicate the direction of errors, focusing more on the degree of fit. The NSE measures how well the model’s forecasting outperforms a reference (usually the mean of observed values). Spanning from negative infinity to 1, it shows values approaching 1 to signify enhanced performance. When negative values appear, they indicate that the man of

recorder values serves as a more effective predictor than the model, potentially implying overestimation tendencies (Dargi et al. 2023; Ghadermazi et al. 2022). The MARE, expressed as a percentage, provides a relative insight into the accuracy of forecasting in relation to the actual values. The following statistical indices are elaborated upon below.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Chl_F^i - Chl_O^i)^2 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Chl_F^i - Chl_O^i)^2} \tag{6}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Chl_F^i - Chl_O^i| \tag{7}$$

$$RMAE = \frac{1}{n} \sum_{i=1}^n |Chl_F^i - Chl_O^i|^2 \tag{8}$$

$$R^2 = \frac{[\sum_{i=1}^n (Chl_O^i - \overline{Chl_O}) Chl_F^i - \overline{Chl_F}]^2}{\sum_{i=1}^n (Chl_O^i - \overline{Chl_O})^2 \sum_{i=1}^n (Chl_F^i - \overline{Chl_F})^2} \tag{9}$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Chl_O^i - Chl_O^i)^2}{\sum_{i=1}^n (Chl_O^i - \overline{Chl_O})^2} \tag{10}$$



$$MARE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Chl_O^i - Chl_F^i}{Chl_O^i} \right| \times 100 \quad (11)$$

In Eqs. 5–11,  $Chl_F$  and  $\overline{Chl}_F$  are forecasted Chl-a and forecasted Chl-a average, respectively.  $Chl_O$  and  $\overline{Chl}_O$  are observed Chl-a values and their corresponding average, while  $n$  demonstrates the total number of samples.

The study employed the Taylor diagram to evaluate model accuracy. This visual tool provides a concise representation of model alignment with observed data across multiple metrics. It captures accuracy, precision, and bias dimensions, aiding informed decisions. Components like correlation, standard deviation ratio, and centered root-mean-square difference showcase model behavior. Reference data is at the center, model forecasting along radial lines. Distance indicates correlation, angle reflects root-mean-square difference, and standard deviation ratio quantifies variability. Insights identify models excelling in specific dimensions, aiding trade-off decisions. The Taylor diagram complements quantitative metrics, holistically assessing performance and aiding model selection. It enhances our understanding of model behavior relative to recorder data.

## Case study

### Study area

Positioned in northeastern Greece near the borders of Albania and North Macedonia, Small Prespa (Mikri Prespa) is one of two interconnected lakes covering approximately 47 Km<sup>2</sup>, situated at an elevation of 853 m above sea level (Tziritis 2014). The lake primarily replenishes itself through surface runoff originating from various water bodies, as well as lateral sub-surface flow resulting from interconnected aquifer systems within the region. The climate in the area is continental, characterized by hot summers and chilly winters. The landscape transforms into lush greenery during the spring months of March to May, with average temperatures ranging from 10 to 20 °C and increased rainfall. During the winter months, Small Prespa receives a substantial amount of precipitation throughout the year, with the highest rainfall occurring from December to February. Average monthly precipitation during this period ranges from 120 to 140 mm. In addition, snowfall is common during this time of the year (Fijani et al. 2019).

The lake's water quality is compromised by pollution and human activities. The concentration of Chl-a, the primary pigment for photosynthetic activity in most algae and cyanobacteria, serves as a key indicator of phytoplankton abundance in aquatic environments. Monitoring Chl-a concentrations in Small Prespa provide valuable insights into

the lake's productivity, overall health, and the potential risk of harmful algal blooms. This information is essential for assessing the lake's ecological well-being and taking necessary measures to preserve water quality and protect the surrounding ecosystem's health (Van der Schriek, 2020). Figure 3 illustrates the location of Small Prespa, located in Greece.

### Data source

During the period from June 1, 2012, to May 31, 2013, WQVs were meticulously obtained at 15-min gaps. Such data encompassed an array of parameters intricately tied to the chemical and physical attributes of the water, including EC, ORP, pH, water temp., DO, and Chl-a concentration. These measurements were meticulously captured through the utilization of a sensor that boasted multiple probes, each contributing to the comprehensive understanding of water quality dynamics. Strategically located at a depth of 1.5 m below the water's surface along the northern shoreline of the lake, the sensor's consistent placement at this depth, approximately 3 m below the water's surface, was consistently achieved throughout the entire year. By carefully selecting the location for the installation, the acquired data represented the holistic aspects of the lake's water quality. Moreover, ensuring the equipment's safety was of paramount importance, safeguarding against any potential external influences that could disrupt or damage the equipment. The accuracy of the sensor measurements was ensured by rigorous calibration procedures prior to deployment. In this process, sensor readings were meticulously compared with standards or reference measurements, enabling potential systematic errors or biases to be identified and corrected. It was necessary to regularly calibrate the sensor during the study period in order to ensure its accuracy and reliability. Verifying the accuracy of the sensor measurements required the comparison of sensor data with measurements derived from well-established laboratory methods. A benchmark was established by collecting samples of water and analyzing them thoroughly using trusted techniques in order to measure the precision of the sensor's measurements. Using these stringent criteria and adhering to a rigorous method, the data collection process was designed to mitigate the impact of potential errors and biases. Through this comprehensive approach, the data were strengthened in terms of reliability and accuracy, providing a substantial foundation upon which to conduct subsequent analyses and modeling efforts. Table 1 provides a statistical overview of the measured WQVs. Specifically, the Chl-a concentration spans a range from 0.82 to 16.97 µg/L, with a corresponding mean value of 2.66 µg/L. To forecast Chl-a concentration on an hourly basis, the dataset was transformed into hourly time

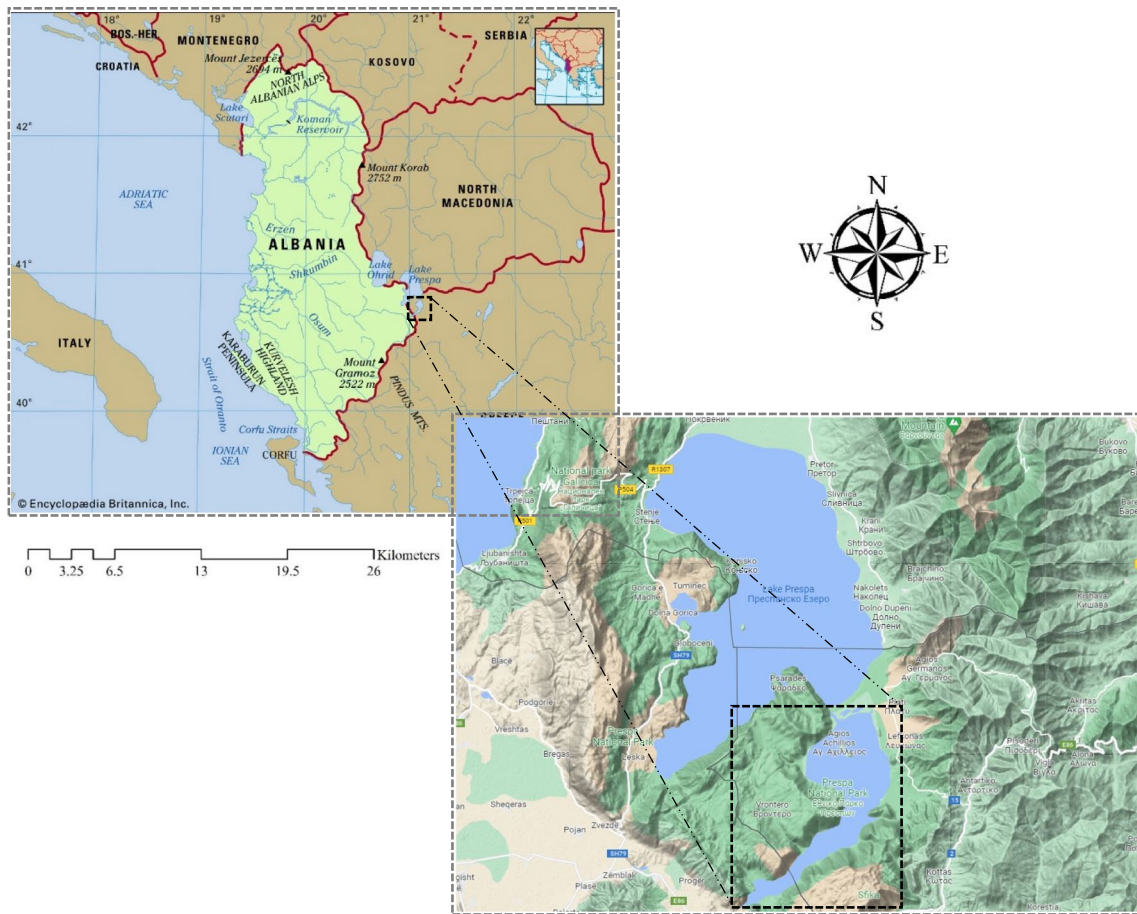


Fig. 3 The study area, situated in Greece (adapted from google.com/maps)

Table 1 A summary of statistical measures for the different QQVs in SPL

Water quality variable	Mean	Variance	Standard deviation	Minimum	Maximum	Skewness	Kurtosis
Chl-a (µg/L)	2.66	2.06	1.43	0.82	16.97	2.29	10.61
EC (µS/cm)	325.55	3277	57.24	219	484	0.86	-0.23
ORP (mV)	392.26	2766	52.59	52.59	473	-1.64	2.97
pH	8.04	0.04	0.18	0.18	8.6	-0.21	0.58
Temp. (°C)	15.48	66.31	8.14	3.02	28.95	-0.03	-1.49
DO (mg/L)	8.83	4.92	2.21	0.03	13.37	0.11	-1.21

steps, resulting in a reduction of data points from 34,825 to 8706. The transformation was performed using Google Colab’s resampling tool.

Figure 4(a–b) illustrates the temporal variations of QQVs collected in the SPL. As shown in Fig. 4(a–b), it highlights fluctuations in Chl-a, DO, pH, water temperature, EC, and ORP. Notably, water temperature, EC, and ORP in the study area exhibited significant fluctuations over the study period. In contrast, the pH value remained stable throughout this time. Additionally, the concentration of Chl-a peaked on October 7, 2012, reaching its maximum value, while it

decreased to nearly  $1 \mu\text{g/l}$  by the end of the water quality sampling period.

The correlation plot of all measured QQVs is presented in Fig. 5. There is no strong correlation between Chl-a concentrations and other measured QQVs. This correlation analysis contributes to a deeper understanding of the complex interactions governing aquatic environments. It reveals how different parameters are interconnected and offers insights into potential cause-and-effect relationships (Zhou et al. 2016; Ly et al. 2018; Bui et al. 2020). Regarding the age of the data, no

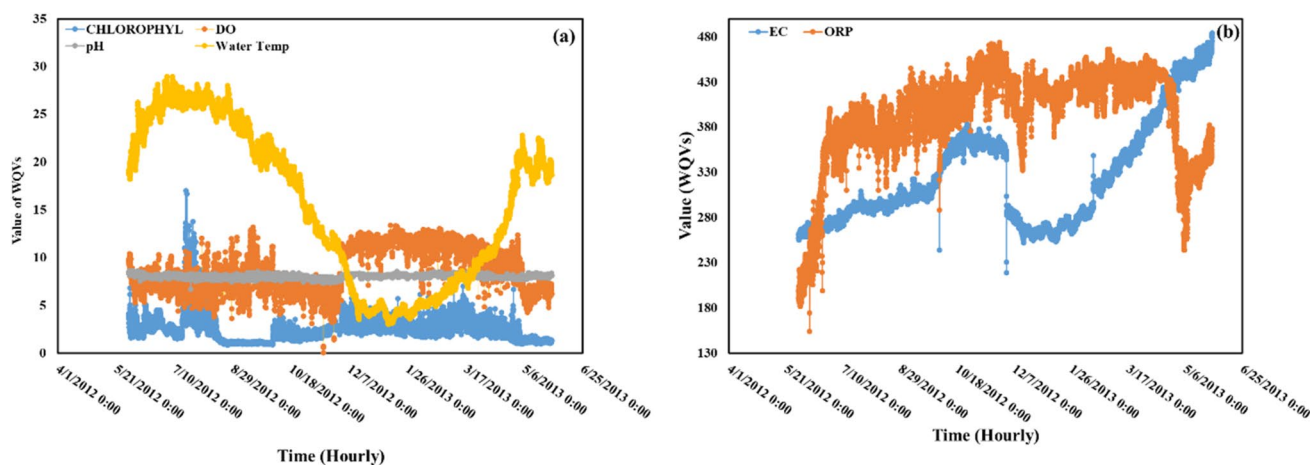


Fig. 4 Temporal variations of WQVs (a) Chl-a, DO, pH, and water temperature and (b) EC and ORP

Fig. 5 Correlation plot among WQVs

	Chlorophyll-a	EC	DO	ORP	pH	Water Temp.
Chlorophyll-a	1.00	-0.32	0.20	0.08	0.11	-0.13
EC	-0.32	1.00	-0.29	0.06	-0.18	0.04
DO	0.20	-0.29	1.00	0.30	0.49	-0.72
ORP	0.07	0.06	0.30	1.00	-0.33	-0.56
pH	0.10	-0.18	0.49	-0.33	1.00	-0.21
Water Temp.	-0.13	0.04	-0.72	-0.56	-0.21	1.00

more recent water quality data for Small Prespa Lake has been made available. Although it is true that using up-to-date data can be more useful than that of in the past in the study area, the main purpose of our research is presenting a framework that can capable of forecasting water quality variable, namely, Chl-a using weighted averaging approaches and time-series water quality dataset. It is worth mentioning that various recent studies used such water quality data (Tziritis (2014); Fijani et al. (2019); and Barzegar et al. (2020)), and one of the main goals of the present study is the comparison of these studies with the present study.

### Model development

We established four DL models (RNN, LSTM, GRU, and TCN) individually, along with their combination through the GA and NSGA-II ensemble methods. These ensembling techniques were applied to forecast Chl-a concentrations and enhance the predictive outcomes of the individual DL models. Accordingly, the hourly Chl-a dataset was separated into two portions: 75% of the data was allocated for the training phase, while the remaining 25% was

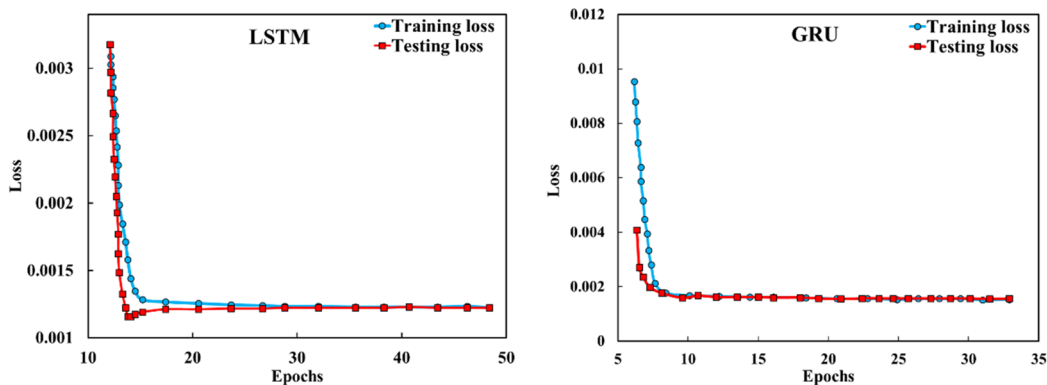
designated for the testing phase. Furthermore, the models were implemented on a personal computer operating with the Windows 10 operating system. The computer was equipped with an Intel(R) Core(TM) i7-10750H processor running at 2.60 GHz and 16 GB of RAM. For development, Python 3.9.7 was used, and the DL models were built using the Google Colab IDE and the Keras development framework.

## DL models

DL models were developed for forecasting Chl-a in the next time step (i.e., Chl-a (t+1)) in Small Prespa Lake using Chl-a at different lag times as inputs. Within the scope of this study, univariate forecasting was employed as the chosen methodology. Univariate forecasting using lag times offers certain advantages over multivariate forecasting in specific contexts. Here are some key advantages of univariate forecasting using lag times: (1) simplicity and ease of implementation: univariate forecasting focuses on forecasting a single variable's future values based solely on its past values. This approach is often simpler to implement compared to multivariate methods that require handling multiple variables and their potential interactions. (2) Reduced complexity: univariate models are less complex than multivariate models since they involve fewer variables. This simplicity can make univariate models easier to interpret and require less computational resources. (3) Less data preprocessing: univariate models require only the historical data of the variable being forecasted, reducing the need for extensive data preprocessing and alignment that's often necessary in multivariate methods. (4) Data availability: univariate forecasting can be advantageous when historical data for other correlated variables might be limited or unavailable. In the context of our study, we observe an intriguing facet: the absence of a substantial correlation between Chl-a and other measured WQVs. This observation underscores the complexity of the relationships governing these variables, implying that univariate forecasting might be a valuable avenue to explore for forecasting Chl-a in this specific scenario. To develop the models, at first, the entire Chl-a dataset was divided into distinct training and testing periods. The training data served a dual purpose: validating and comparing the effectiveness of individual models developed over the training period. In Python, the scikit-learn library was utilized for normalization and scaling, which ranged from 0 to 1. This was achieved through the application of normalization and minimum-maximum scaling techniques. This process helps mitigate abrupt changes in gradients, leading to smoother convergence during training of DL models. During the training of DL models, a function was iteratively refined through trial and error,

ultimately selecting the model with the lowest RMSE. Here, lag times ranging from 1 to 6 h were considered: input variables included Chl-a (t), Chl-a (t-1), ..., Chl-a (t-6), and the forecasting focused on one step ahead Chl-a concentration Chl-a (t+1).

To find the best-performing DL model, various model structures were trained. Models with a one-layer input were selected for modeling due to demonstrating the lowest RMSE during the training period, outperforming more complex structures. An essential goal of this research is to comprehensively compare DL models alongside other approaches. For the LSTM, GRU, and RNN models, the parameters of each layer were approximately consistent, comprising 50 neurons with the ReLU as the activation function. The return sequences parameter, which pertains to the return of the hidden state  $a^{<t>}$ , was set to true due to the time-dependent nature of the data. Different activation functions were tested for all models, including eLU, Tanh, Softmax, SeLU, Softplus, and ReLU. Among these, the ReLU function yielded the lowest RMSE, demonstrating its suitability for analyzing time series data. At each training step, the dropout layer introduces randomness by setting input units to zero at a specified rate, aiming in mitigating the issue of overfitting. The remaining inputs, those not set to zero, are adjusted by scaling factor of  $\frac{1}{(1-rate)}$  to maintain the overall sum of all inputs, where the rate is the proportion of inputs that are set to 0. For example, if 20% of the inputs are set to 0, then the rate would be 0.2, and the scaling factor would be 1.25. The objective of this study is to evaluate and compare the outcomes obtained from individual DL models with those of the EMs. To ensure an equitable comparison, a uniform dropout layer value of 0.001 was applied to all DL models. This value was chosen due to its superiority over other values, as it effectively mitigated overfitting, enhanced generalization, and expedited the convergence of the DL model (Barzegar et al. 2021). Furthermore, the optimization of other parameters for the DL models was performed using a trial-and-error approach. For instance, in the LSTM model, multiple layers were utilized, with the input layer incorporating lag times of Chl-a as inputs to each subsequent stage. To construct this model, an LSTM-based hidden layer was employed, utilizing a ReLU activation function. The hidden layer comprised 64 and 32 units dedicated specifically to Chl-a. After the LSTM layer, a dropout layer with a rate of 0.001 was added to help reduce the degree of overfitting. Subsequently, a fully connected layer known as "Dense" was introduced. This model was developed using the "Adam" optimizer and the "MSE" loss function. The "Adam" optimizer is a gradient-based algorithm that adjusts dynamically. Additionally, a learning rate of 0.01 was employed in the optimization process.



**Fig. 6** Loss function of the DL models for Chl-a forecasting

Notably, the number of epochs was set to 100, and a validation split of 0.15 was utilized. Figure 6 illustrates plotted graphs depicting the loss function of the developed DL algorithms, namely LSTM and GRU. In this representation, the *x*-axis corresponds to the training iterations or epochs, while the *y*-axis denotes the loss function’s values. Notably, the convergence observed in these graphs indicates the models’ approach towards optimal performance. Both LSTM and GRU models exhibited commendable predictive performance in forecasting Chl-a concentration. Furthermore, the consistent decrease in loss values across both the training and testing sets attests to the models’ suitability, indicating a well-fitted and well-generalized behavior for both algorithms.

One of the principal aims of this study is to establish a Chl-a forecasting model using TCN in order to thoroughly investigate the capabilities of this model. Given the inherent complexity of DL models, the task of identifying the optimal TCN network structure and hyperparameters holds significant importance. The selected values for the residual block, kernel size, and number of filters were {1, 2, and 3}, {4, 8, and 8}, and {32, 64, 128}, respectively.

Furthermore, the input data size was set to six, indicating that a combined set of six-time intervals, comprising both present and past data, was utilized for forecasting future values. Subsequently, all structures were meticulously explored, considering various permutations of the following parameters: batch size options {32, 64, and 128}, and epoch values of {20, 50, and 100}. In total, 243 experiments were conducted on the dataset as part of this research endeavor. The objective was to pinpoint the optimal network structure and hyperparameter values across diverse scenarios. Table 2 provides an in-depth summary of the optimal parameters for the trained DL algorithms. The hyperparameter configuration by the trial and error procedure for the developed models brings forth a multitude of distinct advantages. These encompass the fine-tuning of models to align with the domain-specific characteristics of the dataset, the maintenance of controlled model complexity to mitigate overfitting risks, the efficient utilization of computational resources compared to the automated techniques, the harnessing of domain expertise for informed parameter choices, and adeptness at adapting to scenarios with limited data to thwart overfitting.

**Table 2** The summary of DL algorithms in the current study

Model	Target variable	Structure
RNN, LSTM, GRU	Chl-a (t+1)	1 RNN, LSTM, GRU layer (50 neurons, ReLU activation, return sequences = True) + 1 Dropout layer (0.001) + 1 Dense layer (2 neurons, linear activation) + Compile (MSE loss, Adam, Learning rate = ‘0.001’, metrics = ‘MSE’, ‘RMSE’, ‘MAE’, ‘RMAE(%)’, ‘RMSE(%)’, ‘R-Squared’, ‘NSE’) + Fit (epochs = 100, validation_split = 0.15)
TCN	Chl-a (t+1)	Convolutional layer (64 filters + 2 filter sizes + ReLU activation + padding ‘same’ + 1 stride) + Maxpooling (1 pooling size) + Flatten + Dense layer (24 neurons, linear activation) + 1 Dropout layer (0.01) + 1 Dense layer (0.001) + 1 Dense layer (1 neuron) + compile (MSE loss, Adam, Learning rate = ‘0.001’)

**Table 3** The optimal parameter values for the GA and NSGA-II algorithms

GA		NSGA-II	
Parameter	Value	Parameter	Value
Stall generations	40	Initial population size	100
Population size	40	Maximum population size	200
Termination tolerance on the constraint violation	0.1	Probability of crossover	0.9
Generations	80	Probability of mutation	0.2

**Ensemble-based singular algorithms**

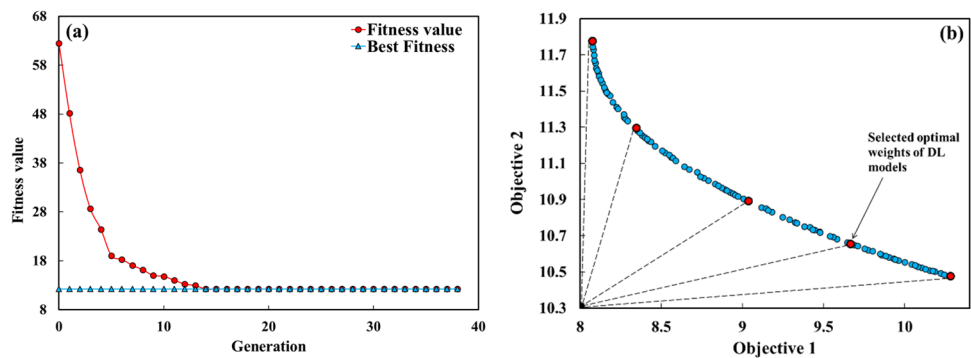
The ensembling of DL models using GA and NSGA-II was utilized to enhance the optimal values of EMs parameters. In this approach, both GA and NSGA-II algorithms were utilized to discover the optimal combination of weights or coefficients for individual DL models. For GA, a MATLAB-based optimization method inspired by natural selection was developed. It involves generating a population of candidate solutions and then iteratively evolving them using principles from genetics and evolution. In terms of the objective function for GA, the evaluation metric, MARE, was minimized to achieve its minimum value. Table 3 presents the parameters of the GA algorithm along with its optimal value. Additionally, Fig. 7(a) illustrates the optimal solutions achieved for forecasting Chl-a using GA. This result emerged when

the fitness value and best fitness exhibited equal values under specific circumstances.

The optimization process of NSGA-II entails generating a set of potential solutions that are at a superior level, followed by evaluating the optimization criteria for each individual solution. From this set, the ND solutions are selected. The solutions chosen through this algorithm constitute the optimal Pareto front, offering a spectrum of trade-off solutions that balance multiple objectives.

Considering objective functions in NSGA-II, minimizing the MARE between EM output and actual value for training the model was considered. Figure 7(b) illustrates the compromises between two distinct objective functions pertaining to the EMs designed for Chl-a forecasting. In Fig. 7(b), the presentation showcases individuals within the population found on a particular non-dominated front, where none of the members exhibit dominance over others. Within the NSGA-II model, a count of 100 members and generations was adopted. Careful consideration was given to the number of members and generations within the NSGA-II model, resulting in a curated list of five members. This selection was meticulously ranked based on the CD criteria. The utilization of the CD criteria serves as a pivotal aspect of the model’s design, aimed at augmenting the diversity and representation of non-dominated solutions spanning the Pareto front. This strategic approach bolsters the robustness and efficacy of the NSGA-II algorithm, especially when tackling complex MOO problems. By prioritizing a well-distributed

**Fig. 7** The Ems’ results: **a** the MARE between the forecasted and observed Chl-a values during the training and testing periods and **b** the tradeoff between two considered objective functions



**Table 4** The initial quintet acquired through NSGA-II

Rank	Decision variables				Objective functions	
	$W_1$	$W_2$	$W_3$	$W_4$	$Z_1$	$Z_2$
1	0.21	0.22	0.30	0.27	8.08	11.78
2	0.00	0.02	0.51	0.47	10.28	10.48
3	0.08	0.09	0.42	0.40	9.04	10.89
4	0.04	0.05	0.47	0.44	9.67	10.66
5	0.15	0.15	0.36	0.34	8.35	11.30

$W_1$ , the weight of the RNN model;  $W_2$ , the weight of the LSTM model;  $W_3$ , the weight of the GRU model;  $W_4$ , the weight of the TCN model;  $Z_1$ , the first objective function;  $Z_2$ , the second objective function

set of solutions that showcase Pareto optimality, the NSGA-II algorithm achieves a refined equilibrium between competing objectives, resulting in optimal outcomes that resonate with the inherent intricacies of real-world scenarios. Table 3 presents the parameters of NSGA-II and their corresponding optimal values. Additionally, Table 4 displays the five optimal solutions generated by NSGA-II, indicating the optimal weights for each DL model and the optimal values of the objective functions.

### Results

DL models were developed based on hourly time steps of Chl-a data in SPL, located in Greece. The test data were employed to evaluate the performance of each individual DL model and EMs, respectively. Table 5 provides a

comparative analysis of the evaluation metrics for the developed DL models, as well as the model ensembling, for both the training and testing periods. Additionally, Figs. 6 and 7 provide graphical comparisons of some of these evaluation metrics.

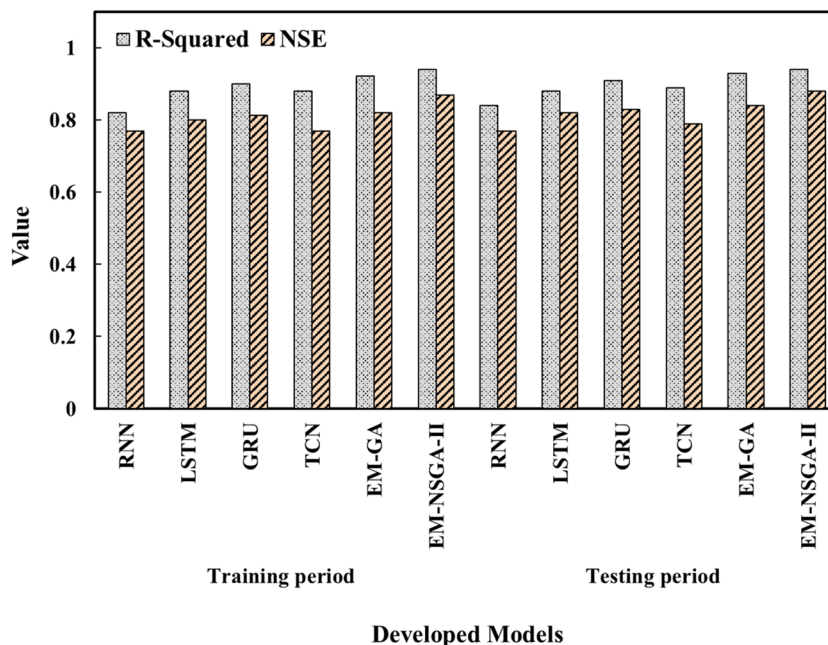
Figure 8 illustrates the comparison of evaluated indices for the models developed throughout the training and testing periods. Considering  $R^2$  metric, EM exhibited the highest accuracy in comparison to the other DL models. With the  $R^2$  metric ranging from 0 to 1, these findings indicate that the EM outperformed the other models in terms of accuracy. Specifically, these differences amounted to 14%, 6%, 4%, and 7% for RNN, LSTM, GRU, and TCN, respectively. During the testing period, the Nash-Sutcliffe efficiency (NSE) values ranged from 0.72 to 0.84, classifying the performance as good ( $0.65 \leq NSE \leq 0.75$ ) to very good ( $0.75 < NSE \leq 1.00$ ) according to the classification by Moriasi et al. (2007).

**Table 5** Performance of the developed DL (RNN, LSTM, GRU, and TCN) and ensemble-DL models for Chl-a forecasting

Variable	Metrics	Training					Testing						
		RNN	LSTM	GRU	TCN	EM-GA	EM-NSGA-II	RNN	LSTM	GRU	TCN	EM-GA	EM-NSGA-II
Chl-a (t+1)	I	0.83	0.79	0.75	0.801	0.69	0.61	0.87	0.83	0.79	0.83	0.74	0.658
	II	0.76	0.71	0.71	0.725	0.70	0.70	0.80	0.75	0.75	0.79	0.74	0.724
	III	0.34	0.33	0.32	0.34	0.32	0.30	0.46	0.45	0.43	0.45	0.43	0.432
	IV	12.81	11.80	10.96	11.75	10.88	10.10	11.02	11.00	8.14	10.24	10.91	9.51
	V	25.84	25.24	25.01	26.01	24.94	24.76	24.61	24.24	24.13	23.41	23.01	22.83
	VI	0.82	0.88	0.901	0.88	0.921	0.94	0.84	0.88	0.91	0.89	0.93	0.94
	VII	0.77	0.80	0.812	0.77	0.82	0.87	0.77	0.82	0.83	0.79	0.84	0.88
	VIII	12.83	12.52	11.24	12.71	12.24	10.28	12.21	11.14	10.81	11.21	11.83	9.16

I: MSE; II: RMSE; III: MAE; IV: RMAE (%); V: PRMSE (%); VI: R-squared; VII: NSE; VIII: MARE (%)

**Fig. 8** The assessment of the models' accuracy — R-squared as well as NSE — for DL and EMs



This assessment pertains to the forecasting of Chl-a using both individual models and EM. Better model performance is indicated when the RMSE is closer to 0. Across both the training and testing phases, EM consistently outperformed the other DL models. Further analysis of this index, considering the nature of individual DL models, revealed that RNN had the weakest performance due to challenges related to gradient vanishing and exploding. In contrast, the development of EM capitalized on the strengths of each individual model, resulting in the lowest RMSE.

All the models assessed in Table 5 exhibit remarkably effective performance, evident from their high  $R^2$  values, which are close to 1. Notably, EM-NSGA-II stands out with exceptional effectiveness compared to DL and EM-GA models, showcasing improvements of 14% (RNN), 8% (LSTM), 6% (GRU), 8% (TCN), and 3% (EM-GA) during the testing phase.

Figure 8 illustrates the comparison of efficiency among the developed models. From both Fig. 8 and Table 5, it is evident that the RMSE, MAE, and MSE values for the models are all below 0.3, signifying highly accurate models for forecasting Chl-a concentration. Regarding model performance assessed through errors, the results indicate that most of the DL and EM models demonstrated similar performance in forecasting Chl-a concentration during both the training and testing periods. However, the EM-NSGA-II model distinctly stood out as the top performer (Fig. 9).

In Fig. 10, the spider plot visually represents the performance metrics employed in the present study. These diagrams pertain to the evaluation of the developed models during the testing phase of DL and EMs development. Notably, based on this analysis, it is evident that the EM-NSGA-II algorithm consistently outperformed the other developed models, which include RNN, LSTM, GRU, TCN, and

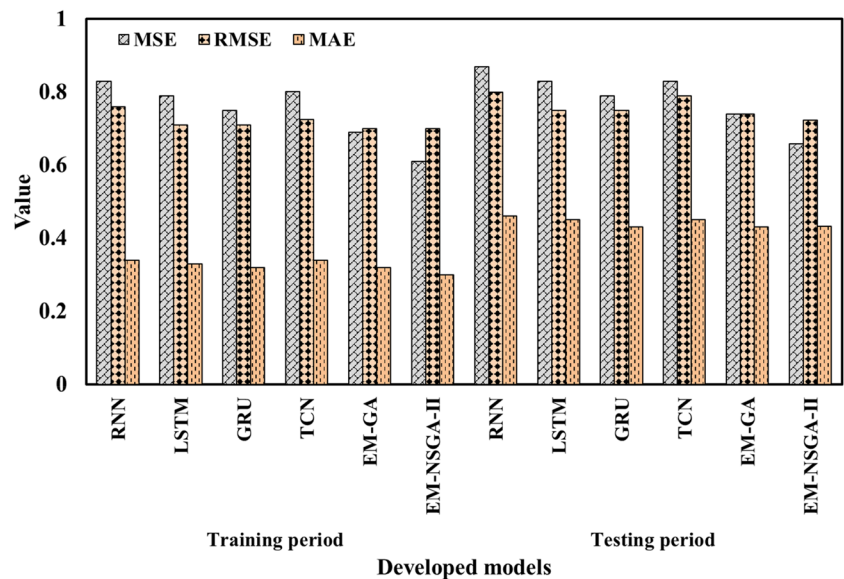
GA-based EMs, across a spectrum of evaluation criteria. This superior trend was pervasive across nearly all algorithms. Conversely, the RNN model exhibited the least accuracy in forecasting Chl-a concentration, as evidenced by its highest values for MSE, MAE, and RMSE, coupled with the lowest values for NSE and R-Squared. As for the other individual DL and EM-GA models, their efficacy for Chl-a forecasting is generally satisfactory, with the EM-GA model demonstrating a marginally superior performance compared to the other developed models.

The time series plots in Fig. 11(a) depict the forecasted versus observed Chl-a concentrations during TSP. The plots provide a visual comparison of the actual and estimated Chl-a values for four DL models (e.g., RNN, LSTM, GRU, and TCN) and two EM models (e.g., EM-GA and EM-NSGA-II). Among the individual DL models, GRU is better suited to capturing the low and high concentrations of the observed Chl-a compared to RNN and LSTM, likely due to its specialized gating mechanism.

According to Fig. 11(b), the  $R^2$  metric demonstrates the model's tendency to underestimate positive errors while overestimating negative errors. In the scatter plots shown in Fig. 11(b), points positioned above the line of equivalence indicate that the predicted Chl-a concentration is lower than the actual Chl-a concentration. The proximity of points to the line of equality directly corresponds to the quality of forecasting performance. The high similarity between actual and estimated Chl-a concentrations, coupled with the majority of points positioned close to the line of equality, signifies that the GRU model outperforms the other DL models in terms of the  $R^2$  metric.

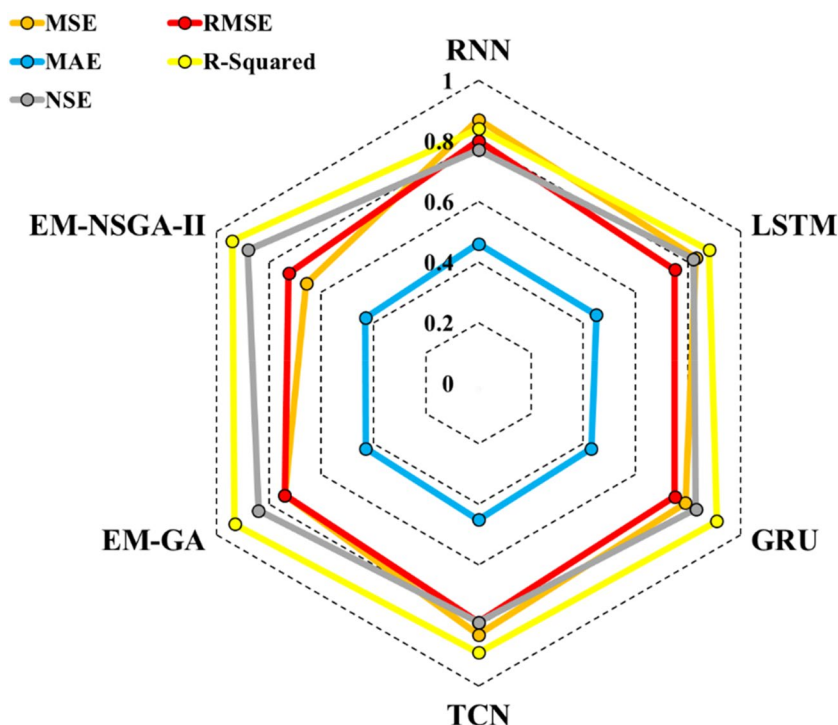
The study demonstrates that EMs perform better than individual DL models in forecasting Chl-a concentrations. The comparison of time series plots between observed and

**Fig. 9** The assessment of the models' accuracy — MSE, RMSE, and MAE — for DL algorithms and EMs





**Fig. 10** Spider diagram illustrating the model's performance in Chl-a forecasting



forecasted Chl-a levels shows that EMs, in general, are more effective in handling complex datasets and producing accurate forecasting. The EM-NSGA-II, which uses two objective functions, was found to be more accurate than the EM-GA in forecasting both low and high concentrations of Chl-a. However, the NSGA-II-based EM model required more training time than the GA-based EM model. Despite this, the increased accuracy of the NSGA-II-based EM highlights the importance of using multiple objective functions in EM models to enhance their forecasting efficiency. Overall, the study suggests that the combination of DL models through EMs provides a powerful approach for forecasting the Chl-a concentration. Furthermore, the use of optimization techniques such as NSGA-II and GA can further enhance the accuracy and robustness of EMs.

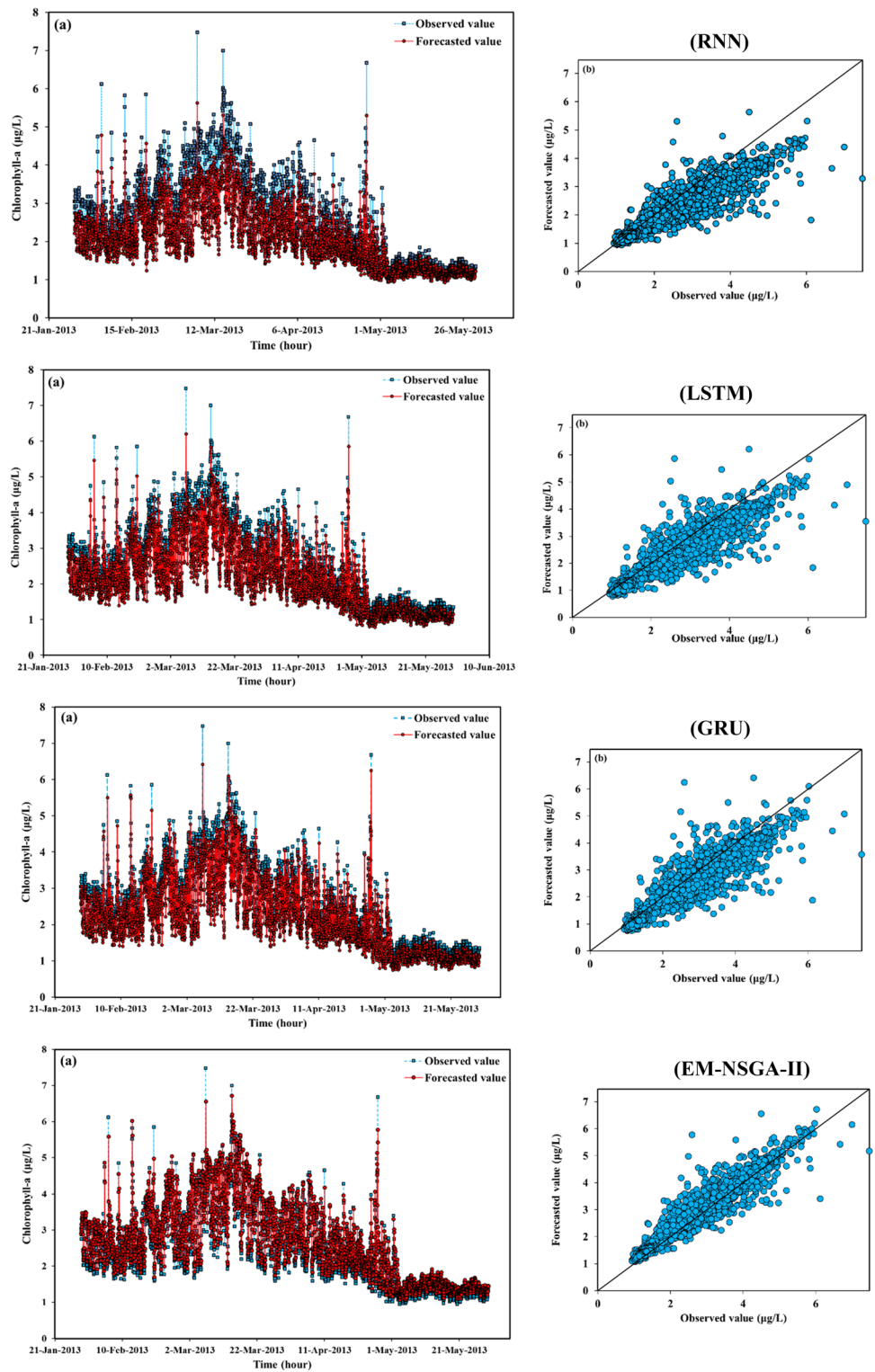
To evaluate the effectiveness of the models, a Taylor diagram (Fig. 12) was employed. This diagram offers a visual representation for comparing models and determining their accuracy. It illustrates how well a model's forecasts align with observed data and aids in identifying the most realistic model. As per the diagram, the EM-NSGA-II model exhibited the best results for Chl-a forecasting. Upon examination, the EM-NSGA-II model displayed a strong positive relationship, as evidenced by a correlation coefficient of 0.98, along with a standardized deviation of 0.93. These values indicate a robust correlation and a close match between actual and forecasted Chl-a levels. This substantiates the EM-NSGA-II model as the most accurate and realistic representation of Chl-a forecasting among the tested models. Conversely, the

RNN model yielded the least favorable results, with a correlation coefficient of 0.91, indicating a substantial positive association, and a normalized standard deviation of 1.06, reflecting moderate variability. These metrics suggest a lower correlation and greater variability between observed and forecasted Chl-a concentrations. The relatively poorer performance of the RNN model suggests that it might not be the most suitable choice for Chl-a concentration forecasting.

## Discussions

While the developed DL models revealed that the RNN exhibited relatively weaker performance compared to other individual DL models in Chl-a forecasting, it's important to note that the RNN model still demonstrated acceptable accuracy. This can be attributed to RNN's inherent capability to retain information from past inputs and incorporate them in processing new inputs, making it suitable for time series analysis. However, the RNN model encounters challenges when forecasting high values of Chl-a in an accurate manner. This limitation can be attributed to the gradient vanishing problem that arises during the backpropagation process. When gradients become exceedingly small, they tend to vanish, resulting in prolonged training periods and suboptimal performance. Conversely, gradient exploding occurs when gradients become overly large during backpropagation, causing weights to update too rapidly and introducing instability during training. Despite these challenges, the

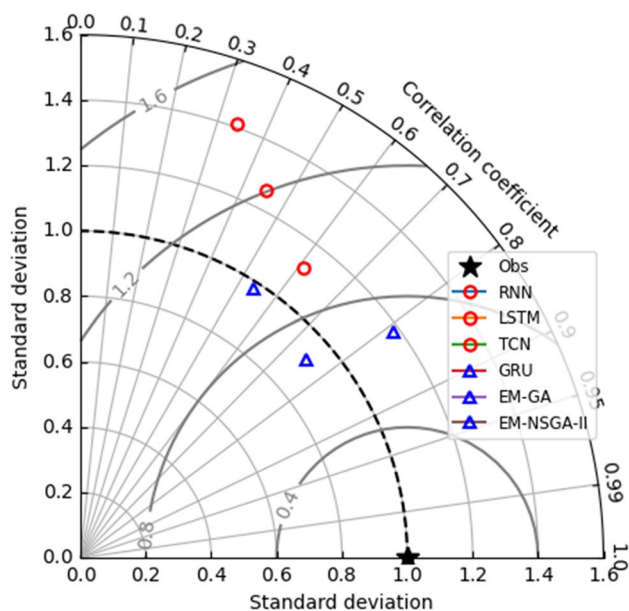
**Fig. 11** The hydrographs of actual and estimated values during the TSP for DL models, including RNN, LSTM, GRU, and EM-NSGA-II as EMs



RNN model’s capacity for temporal memory and sequence processing underscores its viability for certain aspects of Chl-a forecasting.

The LSTM model, a variant of RNN, stands out as a powerful tool for forecasting time series data. Given the

current parameters, it proves to be a valuable choice for forecasting upcoming Chl-a levels. In contrast to the basic RNN model, LSTM offers numerous advantages. These include its ability to handle dependencies over extended timeframes, its resilience against the vanishing gradient



**Fig. 12** Graph illustrating the performance of the models in estimating Chl-a using the Taylor diagram

issue, and its flexibility in accommodating input sequence lengths. Consequently, the LSTM model tends to deliver superior performance compared to RNN, particularly when it comes to forecasting high Chl-a values. Furthermore, the accuracy of the GRU model surpasses the results achieved by both RNN and LSTM. Notably, the primary distinction between LSTM and GRU lies in their architectural design. While LSTM employs three gates to regulate information flow and a cell state to store information, GRU uses reset and update gates along with a hidden state acting as a memory unit. The simplicity and reduced parameter count of GRU make it ideal for faster training and more suitable for smaller datasets. However, LSTM has demonstrated its prowess in tasks that require handling long-term dependencies. Consequently, the choice between the two models hinges on the specific task and dataset characteristics. Differentiating itself from the other DL models developed in this study, the TCN leverages temporal convolutions for sequential data processing, diverging from the recurrent connections employed by other models. TCNs employ convolutional layers to capture higher-level temporal dependencies by convolving the input sequence with a set of filters in a sliding window manner. EMs, combining the predictive outputs of RNN, LSTM, GRU, and TCN, stand as a robust approach for enhancing Chl-a concentration forecasts. This ensemble of models leads to more accurate Chl-a forecasting, leveraging the strengths of each constituent model. The use of EMs in the present study outperformed the individual DL models for several reasons:

1. Reducing model bias and variance: Individual models may suffer from bias or variance, which can limit their accuracy. By combining the forecasting values of multiple models, the bias and variance can be reduced, leading to more accurate forecasting. Therefore, in the present study, EM-GA and EM-NSGA-II had better performance in comparison with the individual models.
2. Capturing diverse perspectives: Individual models may have different strengths and weaknesses. By combining models with diverse perspectives, an EM can capture a wider range of information and make more accurate forecasting. For instance, while the RNN model is an acceptable model to predict time series analysis, it cannot forecast high value of Chl-a concentration due to the gradient vanishing and exploding. Traditional RNNs can suffer from gradient vanishing and exploding gradients, which can hinder their ability to learn and converge to a solution. LSTM and GRU models solve this problem through the use of gated cells that choose to keep or discard information selectively as time progresses. While LSTMs are recognized for their capacity to capture extended relationships over time, GRUs have a simpler architecture that makes them faster to train and less prone to overfitting. The EMs used in this study combine the strengths of RNNs, LSTMs, GRUs, and TCNs to produce more accurate forecasting of Chl-a concentrations.
3. Handling complex data: Some datasets may be too complex for a single model to capture all the nuances. An EM can leverage the strengths of multiple models to handle complex data and make more accurate forecasting.
4. Robustness: Ensemble models tend to be more robust to noise or outliers in the data, as the forecasting of individual models can cancel each other out. This can lead to more stable and reliable forecasting.

The findings of the current study, which highlight the superiority of EMs over individual DL models in forecasting Chl-a concentrations, align with several existing studies in literature. For instance, Barzegar et al. (2020) showcased the enhanced performance of a model combining LSTM and CNN when predicting Chl-a and DO, outperforming individual CNN and LSTM models. Interestingly, the current study's results demonstrated that the developed EMs exhibited even better Chl-a forecasting performance than the CNN-LSTM combination proposed by Barzegar et al. (2020). Similarly, the investigation conducted by Barzegar et al. (2018) indicated that ML models employing wavelet functions yielded improved forecasts for EC compared to individual models. Nonetheless, the current study developed EMs surpassed the accuracy achieved even by the combination of ML models with wavelet transforms. Another study

by Gao et al. (2020) concluded that their developed EMs outperformed a hybrid model, showcasing the advantages of ensemble strategies. Furthermore, Song et al. (2023) integrated metaheuristic optimization algorithms with LSTM to optimize DO forecasts, with their results demonstrating the superior performance of EMs based on GA and NSGA-II. In yet another context, Wu and Wang (2022) presented a fusion model incorporating ANN, discrete wavelet transform, and LSTM to forecast DO, yielding improved results compared to individual models. Remarkably, our study's EMs for Chl-a concentration forecasting delivered even higher accuracy in terms of evaluation indices, further emphasizing the strength of our proposed ensemble approach. Uddin et al. (2023a, 2023b, 2023c) introduced the Irish Water Quality Index (IEWQI) framework, which was developed to assess the quality of transitional and coastal waters. The primary goal was to enhance the methodology and establish a tool relevant to environmental regulators with the aim of tackling water pollution. The study's findings underscored the promising effectiveness and reliability of this index as a more accurate means of evaluating the quality of transitional and coastal waters. Lingxuan Chen et al. (2023b) developed a hybrid algorithm for WQVs forecasting in rivers, which outperformed individual models. It is worth mentioning that, considering the evaluation indices, the EMs developed in the current research lead to better results in comparison with their model.

### The application of developed models in water resource management

DL and EMs show great potential for forecasting WQVs, including Chl-a. Chl-a serves as a widely employed measure of water quality, as it is a proxy for the abundance of phytoplankton in the water, which can impact aquatic ecosystems and human health. Traditional approaches for Chl-a forecasting involve statistical models or physics-based models that require a significant quantity of input variables, including meteorological and hydrological data, and WQVs. However, DL and EMs can provide a more accurate and efficient alternative to traditional models by leveraging the power of neural networks to learn complex relationships between data.

In addition, EMs can be used to improve the accuracy of Chl-a forecasting by combining data from multiple sources, such as satellite images, in situ measurements, and environmental parameters. For example, an EM can combine satellite images of the water surface with in situ measurements of Chl-a concentrations to provide more accurate and comprehensive forecasts. DL and EMs have benefits in their ability to manage extensive and intricate datasets, making them especially valuable in modeling Chl-a levels in water bodies. Another advantage is that these models can learn from historical data to forecast future Chl-a concentrations,

which can be useful for monitoring water quality over time and detecting potential problems before they become more serious.

Integrating a hybrid DL model into water quality systems can significantly improve the management of water resources by providing accurate and timely data on Chl-a concentrations in water bodies. The occurrence of algal blooms resulting from high Chl-a levels can be mitigated by implementing an early warning system. The hybrid DL model can help water managers mitigate the risks of harmful algal blooms, which can result in fish kills, unsafe drinking water, and even human and animal illness or death. Additionally, the model can function as a tool for supporting decision-making in forecasting and managing water quality by providing information on when and where Chl-a levels will be high. The hybrid DL model helps water managers minimize harmful algal bloom risks, make informed decisions, and enhance monitoring efficiency. Traditional monitoring methods require frequent, costly sampling, while the hybrid DL model allows for continuous monitoring and real-time data analysis, reducing the need for manual sampling.

We acknowledge limitations in this study, most notably concerning the period of the dataset used. The data, which only extends up to 2013, imposes constraints on the applicability of our findings. Specifically, our model does not account for potential changes in the lake's ecosystem that may have occurred post-2013 — a significant concern given the dynamic nature of ecosystems and their susceptibility to both natural and anthropogenic influences over time. While this limitation is mitigated to some extent by the current lack of more recent, publicly accessible data, it remains a constraint. Our study, therefore, serves as the most current analysis possible within these data limitations. On the methodological side, our approach does not employ a multivariate forecasting strategy, which means it is not influenced by other potentially impactful variables, such as climate data or additional water quality parameters. While this simplifies our model, potentially increasing its robustness given the age of the data, it is a limitation that future research could address. Nevertheless, the methodology and analyses presented here could be applied directly to newer datasets as and when they become available, thereby validating and enhancing the utility and relevance of our current findings.

### Conclusions

Monitoring and forecasting water quality in water bodies is crucial for managing water resources, as it has a substantial influence on environmental processes and the welfare of both humans and animals. This study aims to develop an EM using single- (GA) and multi-objective (NSGA-II) optimization algorithms that utilize DL models separately, with

the goal of forecasting the Chl-a concentration in the SPL located in Greece. Data on Chl-a levels and other WQVs were collected using a sensor from June 1, 2012, to May 31, 2013. To forecast the Chl-a concentration, four different DL models — RNN, LSTM, GRU, and TCN — were developed, and their results were assessed and compared with each other. Subsequently, EMs based on the GA and NSGA-II algorithms were developed to improve the obtained results of individual models. The study revealed that the GRU model exhibited superior performance in comparison with other individual DL models in Chl-a forecasting due to its simpler structure compared to LSTM and its ability to address the problems of RNNs. Additionally, EMs, which are based on SOO and MOO algorithms, demonstrated superior results when compared to individual DL models.

Based on the insights gained and the results achieved in this study, several promising avenues for future research in water management and water quality forecasting can be outlined:

1. Subsequent studies could delve into the application of decomposition tools before implementing DL models. Among these tools, continuous wavelet transform, fast Fourier transform, and VMD, among others, hold potential.
2. Future research could delve into exploring additional hybrid models such as RNN-LSTM, RNN-TCN, GRU-TCN, WT-GRU-TCN, LSTM-TCN-WT, and GRU-TCN-VMD, and comparisons drawn with the present study could be insightful.
3. Understanding and modeling uncertainty emerges as a pivotal endeavor within the realm of DL and EMs. Addressing this facet can significantly enhance the reliability of forecasting and ensuing decisions. Within this context, various types of uncertainty, including aleatoric, epistemic, and model uncertainty, warrant thorough consideration.

The primary contribution of this study lies in the formulation of an ensemble approach that harnesses the complementary strengths of distinct DL models, effectively addressing the complexities tied to forecast dynamic water quality variables. Our approach demonstrates heightened predictive performance compared to individual models, highlighting the potency of synergistic model amalgamations. Additionally, the availability of the water quality data used for the training and testing phase, model assumption, transferability, temporal and spatial variability, model interpretability, computational resources, and so on can be considered as limitations and implication of the research.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11356-023-30774-4>.

**Acknowledgements** Data collection was conducted as part of the “Vodafone World of Difference” program, a philanthropic volunteer initiative facilitated by Vodafone Greece. The Society for the Protection of Prespa (SPP), in collaboration with the Prespa Municipality, played a crucial role as the initiating host organization for the project. The authors of this study express their heartfelt gratitude to all individuals’ mentioned above for their invaluable contributions and steadfast support. Furthermore, special recognition is extended to Evangelos Tziritis for generously sharing the data that made this research possible.

**Author contribution** M.G.Z.: conceptualization; software; validation; investigation; resources; data curation; writing, original draft; visualization. M.R.N.: project administration, supervision, methodology, software, conceptualization, methodology, investigation, writing — review and editing. S.J.: conceptualization, writing — review and editing. R.B.: resources, writing — review and editing. A.R.M.: visualization, writing — review and editing

**Data availability** The data that support the findings of this study are available from the corresponding author upon a reasonable request.

## Declarations

**Ethics approval** Not applicable

**Consent to participate** Not applicable

**Consent for publication** Not applicable

**Conflict of interest** The authors declare no competing interests.

## References

- Azizi K, Diko SK, Saija L, Zamani MG, Meier CI (2022) Integrated community-based approaches to urban pluvial flooding research, trends and future directions: A review. *Urban Clim* 44:101237
- Babatunde OH, Armstrong L, Leng J, Diepeveen D (2014) A genetic algorithm-based feature selection. <http://ro.ecu.edu.au/theses/1733>
- Babuji P, Thirumalaisamy S, Duraisamy K, Periyasamy G (2023) Human health risks due to exposure to water pollution: a review. *Water* 15(14):2532. <https://doi.org/10.3390/w15142532>
- Bahrami M, Talebbeydokhti N, Rakhshandehroo G, Nikoo MR, Adamowski JF (2023) A fusion-based data assimilation framework for runoff prediction considering multiple sources of precipitation. *Hydrological Sciences Journal*, (just-accepted). <https://doi.org/10.1080/02626667.2023.2180375>
- Bai S, Kolter JZ, Koltun V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271. <https://doi.org/10.48550/arXiv.1803.01271>
- Barzegar R, Asghari Moghaddam A (2016) Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. *Model Earth Syst Environ* 2:1–13. <https://doi.org/10.1007/s40808-015-0072-8>
- Barzegar R, Aalami MT, Adamowski J (2020) Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch Env Res Risk A* 34(2):415–433. <https://doi.org/10.1007/s00477-020-01776-2>
- Barzegar R, Asghari Moghaddam A, Adamowski J, Fijani E (2017) Comparison of machine learning models for predicting fluoride

- contamination in groundwater. *Stoch Env Res Risk A* 31:2705–2718. <https://doi.org/10.1007/s00477-016-1338-z>
- Barzegar R, Asghari Moghaddam A, Adamowski J, Ozga-Zielinski B (2018) Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stoch Env Res Risk A* 32:799–813. <https://doi.org/10.1007/s00477-017-1394-z>
- Barzegar R, Ghasri M, Qi Z, Quilty J, Adamowski J (2019) Using bootstrap ELM and LSSVM models to estimate river ice thickness in the Mackenzie river basin in the Northwest Territories, Canada. *J Hydrol* 577:123903
- Barzegar R, Moghaddam AA, Baghban H (2016) A supervised committee machine artificial intelligent for improving DRAS-TIC method to assess groundwater contamination risk: a case study from Tabriz plain aquifer, Iran. *Stoch Env Res Risk A* 30:883–899
- Barzegar R, Aalami MT, Adamowski J (2021) Coupling a hybrid CNN-LSTM deep learning model with a boundary corrected maximal overlap discrete wavelet transform for multiscale lake water level forecasting. *J Hydrol* 598:126196
- Bhardwaj A, Dagar V, Khan MO, Aggarwal A, Alvarado R, Kumar M, Proshad R (2022) Smart IoT and machine learning-based framework for water quality assessment and device component monitoring. *Environ Sci Pollut Res* 29(30):46018–46036
- Boyd CE (2020) Eutrophication. Water quality: an introduction, 311–322. [https://doi.org/10.1007/978-3-030-23335-8\\_15](https://doi.org/10.1007/978-3-030-23335-8_15)
- Bui DT, Khosravi K, Tiefenbacher J, Nguyen H, Kazakis N (2020) Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Sci Total Environ* 721:137612. <https://doi.org/10.1016/j.scitotenv.2020.137612>
- Carcano EC, Bartolini P, Muselli M, Piroddi L (2008) Jordan recurrent neural network versus IHACRES in modelling daily streamflows. *J Hydrol* 362(3–4):291–307
- Card D, Zhang M, Smith NA (2019) Deep weighted averaging classifiers. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 369–378)
- Chapman DV, Sullivan T (2022) The role of water quality monitoring in the sustainable use of ambient waters. *One Earth* 5(2):132–137. <https://doi.org/10.1016/j.oneear.2022.01.008>
- Chen WB, Liu WC (2014) Artificial neural network modeling of dissolved oxygen in reservoir. *Environ Monit Assess* 186:1203–1217. <https://doi.org/10.1007/s10661-013-3450-6>
- Chen X, Dai Y (2020) Research on an improved ant colony algorithm fusion with genetic algorithm for route planning. In: 2020 IEEE 4th Information technology, networking, electronic and automation control conference (ITNEC) 1:1273–1278. IEEE. <https://doi.org/10.1109/ITNEC48623.2020.9084730>
- Chen H, Chen A, Xu L, Xie H, Qiao H, Lin Q, Cai K (2020a) A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources. *Agric Water Manag* 240:106303. <https://doi.org/10.1016/j.agwat.2020.106303>
- Chen K, Chen H, Zhou C, Huang Y, Qi X, Shen R, Liu F, Zuo M, Zou X, Wang J, Zhang Y (2020b) Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Res* 171:115454. <https://doi.org/10.1016/j.watres.2019.115454>
- Chen L, Wu T, Wang Z, Lin X, Cai Y (2023a) A novel hybrid BPNN model based on adaptive evolutionary artificial bee colony algorithm for water quality index prediction. *Ecol Indic* 146:109882. <https://doi.org/10.1016/j.ecolind.2023.109882>
- Chen L, Wu T, Wang Z, Lin X, Cai Y (2023b) A novel hybrid BPNN model based on adaptive evolutionary artificial bee colony algorithm for water quality index prediction. *Ecol Indic* 146:109882
- Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv: 1406.1078. <https://doi.org/10.48550/arXiv.1406.1078>
- Choi JH, Kim J, Won J, Min O (2019) Modelling chlorophyll-a concentration using deep neural networks considering extreme data imbalance and skewness. In 2019 21st International Conference on Advanced Communication Technology (ICACT) (pp. 631–634). IEEE. <https://doi.org/10.23919/ICACT.2019.8702027>
- Chollet F (2021) Deep learning with Python. Simon and Schuster
- Chou JS, Ho CC, Hoang HS (2018) Determining quality of water in reservoir using machine learning. *Ecological Inform* 44:57–75. <https://doi.org/10.1016/j.ecoinf.2018.01.005>
- Chua LO, Roska T (1993) The CNN paradigm. *IEEE Trans Circuits Syst I: Fundamental Theory Appl* 40(3):147–156
- Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555
- Dai M, Yang H, Yang F, Zhang Z, Yu Y, Liu G, Feng X (2022) Multi-strategy Ensemble Non-dominated sorting genetic Algorithm-II (MENSAGA-II) and application in energy-enviro-economic multi-objective optimization of separation for isopropyl alcohol/diisopropyl ether/water mixture. *Energy* 254:124376
- Dargi M, Khamsehchi E, Mahdavi Kalatehno J (2023) Optimizing acidizing design and effectiveness assessment with machine learning for predicting post-acidizing permeability. *Sci Rep* 13(1):11851
- Dawood T, Elwakil E, Novoa HM, Delgado JFG (2021) Toward urban sustainability and clean potable water: prediction of water quality via artificial neural networks. *J Clean Prod* 291:125266
- Deb K, Agrawal S, Pratap A, Meyarivan T (2000) A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In *Parallel Problem Solving from Nature PPSN VI: 6th International Conference Paris, France, September 18–20, 2000 Proceedings 6* (pp. 849–858). Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-45356-3\\_83](https://doi.org/10.1007/3-540-45356-3_83)
- Deb K, Pratap A, Agarwal S, Meyarivan TAMT (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6(2):182–197
- Dehghani R, Torabi Poudeh H, Izadi Z (2021) Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques. *Modeling Earth Systems and Environment*, pp.1–15. <https://doi.org/10.1007/s40808-021-01253-x>
- Ehsani M, Moghadas Nejad F, Hajikarimi P (2022) Developing an optimized faulting prediction model in jointed plain concrete pavement using artificial neural networks and random forest methods. *Intl J Pavement Eng*, 1–16. <https://doi.org/10.1080/10298436.2022.2057975>
- El Bilali A, Taleb A (2020) Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. *J Saudi Soc Agric Sci* 19(7):439–451. <https://doi.org/10.1016/j.jssas.2020.08.001>
- Elkiran G, Nourani V, Abba SI (2019) Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *J Hydrol* 577:123962. <https://doi.org/10.1016/j.jhydrol.2019.123962>
- Ewuzie U, Bolade OP, Egbedina AO (2022) Application of deep learning and machine learning methods in water quality modeling and prediction: a review. *Current Trends and Advances in Computer-Aided Intelligent Environmental Data Engineering*, pp.185–218. <https://doi.org/10.1016/B978-0-323-85597-6.00020-3>
- Farshbaf Aghajani H, Karimi S, Hatefi Diznab M (2023) An experimental and machine-learning investigation into compaction of the cemented sand-gravel mixtures and influencing factors. *Transp Infrastruct Geotechnol* 10(5):816–855

- Fijani E, Barzegar R, Deo R, Tziritis E, Skordas K (2019) Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. *Sci Total Environ* 648:839–853. <https://doi.org/10.1016/j.scitotenv.2018.08.221>
- Fu Y, Hu Z, Zhao Y, Huang M (2021) A long-term water quality prediction method based on the temporal convolutional network in smart mariculture. *Water*, 13(20), p.2907. <https://doi.org/10.3390/w13202907>
- Gao X, Ren B, Zhang H, Sun B, Li J, Xu J, Li K (2020) An ensemble imbalanced classification method based on model dynamic selection driven by data partition hybrid sampling. *Expert Syst Appl* 160:113660
- Gaya MS, Abba SI, Abdu AM, Tukur AI, Saleh MA, Esmaili P, Wahab NA (2020) Estimation of water quality index using artificial intelligence approaches and multi-linear regression. *Int. J. Artif. Intell.* ISSN, 2252, 8938
- Ghadermazi P, Re A, Ricci L, Chan SHJ (2022) Metabolic engineering interventions for sustainable 2, 3-butanediol production in gas-fermenting *Clostridium autoethanogenum*. *mSystems* 7(2):e011111–e011121
- Georgescu PL, Moldovanu S, Iticescu C, Calmuc M, Calmuc V, Topa C, Moraru L (2023) Assessing and forecasting water quality in the Danube river by using neural network approaches. *Sci Total Environ* 879:162998
- Géron A (2022) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc.
- Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley
- Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press
- Graves A (2012) Sequence transduction with recurrent neural networks. arXiv preprint arXiv:1211.3711
- Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2016) LSTM: A search space odyssey. *IEEE Trans Neural Netw Learn Syst* 28(10):2222–2232
- Guo J, Zhang C, Zheng G, Xue J, Zhang L (2018) The establishment of season-specific eutrophication assessment standards for a water-supply reservoir located in Northeast China based on chlorophyll-a levels. *Ecol Indic* 85:11–20. <https://doi.org/10.1016/j.ecolind.2017.09.056>
- Hajikarimi P, Ehsani M, Haloui YE, Tehrani FF, Absi J, Nejad FM (2022) Fractional viscoelastic modeling of modified asphalt mastics using response surface method. *Constr Build Mater* 317:125958
- Hall DL, Llinas J (1997) An introduction to multisensor data fusion. *Proc IEEE* 85(1):6–23. <https://doi.org/10.1109/5.554205>
- Haverkos BM, Pan Z, Gru AA, Freud AG, Rabinovitch R, Xu-Welliver M, Porcu P (2016) Extranodal NK/T cell lymphoma, nasal type (ENKTL-NT): an update on epidemiology, clinical presentation, and natural history in North American and European cases. *Curr Hematol Malignancy Reports* 11:514–527
- Jahanshahi S, Kerachian R (2019) An evidential reasoning-based sustainability index for water resources management. *Hydrol Sci J* 64(10):1223–1239. <https://doi.org/10.1080/02626667.2019.1628347>
- Jahanshahi S, Kerachian R, Emamjomehzadeh O (2023) A leader-follower framework for sustainable water pricing and allocation. *Water Resour Manage* 1–18. <https://doi.org/10.1007/s11269-023-03428-w>
- Jiang J, Tang S, Han D, Fu G, Solomatine D, Zheng Y (2020) A comprehensive review on the design and optimization of surface water quality monitoring networks. *Environ Model Softw* 132:104792. <https://doi.org/10.1016/j.envsoft.2020.104792>
- Katoch S, Chauhan SS, Kumar V (2021) A review on genetic algorithm: past, present, and future. *Multimed Tools Appl* 80:8091–8126
- Khosravi K, Golkarian A, Booij MJ, Barzegar R, Sun W, Yaseen ZM, Mosavi A (2021) Improving daily stochastic streamflow prediction: comparison of novel hybrid data-mining algorithms. *Hydrol Sci J* 66(9):1457–1474
- Kouadri S, Pande CB, Panneerselvam B, Moharir KN, Elbeltagi A (2021) Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. *Environ Sci Pollut Res*, 1–25. <https://doi.org/10.1007/s11356-021-17084-3>
- Li L, Jiang P, Xu H, Lin G, Guo D, Wu H (2019) Water quality prediction based on recurrent neural network and improved evidence theory: a case study of Qiantang river, China. *Environ Sci Pollut Res* 26:19879–19896. <https://doi.org/10.1007/s11356-019-05116-y>
- Li W, Wei Y, An D, Jiao Y, Wei Q (2022) LSTM-TCN: dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network. *Environ Sci Pollut Res* 29(26):39545–39556. <https://doi.org/10.1007/s11356-022-18914-8>
- Li X, Sha J, Wang ZL (2017) Chlorophyll-a prediction of lakes with different water quality patterns in China based on hybrid neural networks. *Water* 9(7):524. <https://doi.org/10.3390/w9070524>
- Liang Z, Zou R, Chen X, Ren T, Su H, Liu Y (2020) Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. *J Hydrol* 581:124432. <https://doi.org/10.1016/j.jhydrol.2019.124432>
- Lin L, Yang H, Xu X (2022) Effects of water pollution on human health and disease heterogeneity: a review. *Front Environ Sci*, 975
- Liu P, Wang J, Sangaiah AK, Xie Y, Yin X (2019) Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Sustainability* 11(7):2058. <https://doi.org/10.3390/su11072058>
- Luo W, Zhu S, Wu S, Dai J (2019) Comparing artificial intelligence techniques for chlorophyll-a prediction in US lakes. *Environ Sci Pollut Res* 26:30524–30532. <https://doi.org/10.1007/s11356-019-06360-y>
- Ly A, Marsman M, Wagenmakers EJ (2018) Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica* 72(1):4–13
- Ma Z, Song X, Wan R, Gao L, Jiang D (2014) Artificial neural network modeling of the water quality in intensive *Litopenaeus vannamei* shrimp tanks. *Aquaculture* 433:307–312. <https://doi.org/10.1016/j.aquaculture.2014.06.029>
- Medsker LR, Jain LC (2001) Recurrent neural networks. *Design Appl* 5:64–67
- Meydani A, Dehghanipour A, Schoups G, Tajrishy M (2022) Daily reservoir inflow forecasting using weather forecast downscaling and rainfall-runoff modeling: application to Urmia Lake basin, Iran. *J Hydrol Region Stud* 44:101228. <https://doi.org/10.1016/j.ejrh.2022.101228>
- Mirjalili S, Mirjalili S (2019) Genetic algorithm. *Evolutionary algorithms and neural networks: theory and applications*, 43–55. [https://doi.org/10.1007/978-3-319-93025-1\\_4](https://doi.org/10.1007/978-3-319-93025-1_4)
- Nadiri AA, Sedghi Z, Barzegar R, Nikoo MR (2022) Establishing a data fusion water resources risk map based on aggregating drinking water quality and human health risk indices. *Water* 14(21):3390. <https://doi.org/10.3390/w14213390>
- Ni Q, Cao X, Tan C, Peng W, Kang X (2023) An improved graph convolutional network with feature and temporal attention for multivariate water quality prediction. *Environ Sci Pollut Res* 30(5):11516–11529
- Nikoo MR, Karimi A, Kerachian R, Poorsepahy-Samian H, Daneshmand F (2013) Rules for optimal operation of reservoir-river-groundwater systems considering water quality targets:

- application of M5P model. *Water Resour Manag* 27:2771–2784. <https://doi.org/10.1007/s11269-013-0314-3>
- Nova K (2023) AI-enabled water management systems: an analysis of system components and interdependencies for water conservation. *Eigenpub Rev Sci Technol* 7(1):105–124 <https://studies.eigenpub.com/index.php/erst/article/view/12>
- Ortiz-Lopez C, Bouchard C, Rodriguez M (2022) Machine learning models with potential application to predict source water quality for treatment purposes: a critical review. *Environ Technol Rev* 11(1):118–147. <https://doi.org/10.1080/21622515.2022.2118084>
- Papenfus M, Schaeffer B, Pollard AI, Loftin K (2020) Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs. *Environ Monit Assess* 192(12):808. <https://doi.org/10.1007/s10661-020-08631-5>
- Prasad DVV, Venkataramana LY, Kumar PS, Prasannamedha G, Harshana S, Srividya SJ, Indraganti S (2022) Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Sci Total Environ* 821:153311
- Pyo J, Park LJ, Pachepsky Y, Baek SS, Kim K, Cho KH (2020) Using convolutional neural network for predicting cyanobacteria concentrations in river water. *Water Res* 186:116349. <https://doi.org/10.1016/j.watres.2020.116349>
- Qi C, Huang S, Wang X (2020) Monitoring water quality parameters of Taihu lake based on remote sensing images and LSTM-RNN. *IEEE Access* 8:188068–188081
- Ribeiro MHD, dos Santos Coelho L (2020) Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl Soft Comput* 86:105837
- Rizal NNM, Hayder G, Yusoff S (2023) River water quality prediction and analysis—deep learning predictive models approach. In: *Sustainability challenges and delivering practical engineering solutions: resources, materials, energy, and buildings*. Springer International Publishing, Cham, pp 25–29. [https://doi.org/10.1007/978-3-031-26580-8\\_5](https://doi.org/10.1007/978-3-031-26580-8_5)
- Rozinajová V, Ezzeddine AB, Lóderer M, Loeb J, Magyar R, Vrablcová P (2018) Computational intelligence in smart grid environment. In *Computational intelligence for multimedia big data on the cloud with engineering Applications* (pp. 23–59). Academic Press. <https://doi.org/10.1016/B978-0-12-813314-9.00002-5>
- Sahraei A, Chamorro A, Kraft P, Breuer L (2021) Application of machine learning models to predict maximum event water fractions in streamflow. *Front Water* 3:652100
- Sakaa B, Elbeltagi A, Boudibi S, Chaffai H, Islam ARMT, Kulimushi LC, Choudhari P, Hani A, Brouziyne Y, Wong YJ (2022) Water quality index modeling using random forest and improved SMO algorithm for support vector machine in Saf-Saf river basin. *Environ Sci Pollut Res* 29(32):48491–48508. <https://doi.org/10.1007/s11356-022-18644-x>
- Schmidhuber J, Hochreiter S (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Shen C (2018) A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour Res* 54(11):8558–8593. <https://doi.org/10.1029/2018WR022643>
- Shin Y, Kim T, Hong S, Lee S, Lee E, Hong S, Lee C, Kim T, Park MS, Park J, Heo TY (2020) Prediction of chlorophyll-a concentrations in the Nakdong river using machine learning methods. *Water* 12(6):1822. <https://doi.org/10.3390/w12061822>
- Sinshaw TA, Surbeck CQ, Yasarer H, Najjar Y (2019) Artificial neural network for prediction of total nitrogen and phosphorus in US lakes. *J Environ Eng* 145(6):04019032. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001528](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001528)
- Sivanandam SN, Deepa SN, Sivanandam SN, Deepa SN (2008) Genetic algorithms (pp. 15–37). Springer Berlin Heidelberg
- Song Y, Shen C, Wang Y (2023) Multi-objective optimal reservoir operation considering algal bloom control in reservoirs. *J Environ Manage* 344:118436
- Sultana F, Sufian A, Dutta P (2020) Evolution of image segmentation using deep convolutional neural network: a survey. *Knowl-Based Syst* 201:106062. <https://doi.org/10.1016/j.knsys.2020.106062>
- Tan TY, Zhang L, Lim CP, Fielding B, Yu Y, Anderson E (2019) Evolving ensemble models for image segmentation using enhanced particle swarm optimization. *IEEE access* 7:34004–34019
- Tang A, Wang C, Zhang D, Zhang K, Zhou Y, Zhang Z (2022) A multi-model real covariance-based battery state-of-charge fusion estimation method for electric vehicles using ordered weighted averaging operator. *Int J Energy Res* 46(12):17273–17284
- Tziritis EP (2014) Environmental monitoring of Micro Prespa Lake basin (Western Macedonia, Greece): hydrogeochemical characteristics of water resources and quality trends. *Environ Monit Assess* 186(7):4553–4568
- Uddin MG, Nash S, Olbert AI (2021) A review of water quality index models and their use for assessing surface water quality. *Ecol Indic* 122:107218
- Uddin MG, Nash S, Diganta MTM, Rahman A, Olbert AI (2022a) Robust machine learning algorithms for predicting coastal water quality index. *J Environ Manage* 321:115923
- Uddin MG, Nash S, Rahman A, Olbert AI (2022b) A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment. *Water Res* 219:118532
- Uddin MG, Nash S, Rahman A, Olbert AI (2023a) A sophisticated model for rating water quality. *Sci Total Environ* 868:161614
- Uddin MG, Nash S, Rahman A, Olbert AI (2023b) Assessing optimization techniques for improving water quality model. *J Clean Prod* 385:135671
- Uddin MG, Rahman A, Nash S, Diganta MTM, Sajib AM, Moniruz-zaman M, Olbert AI (2023c) Marine waters assessment using improved water quality model incorporating machine learning approaches. *J Environ Manage* 344:118368
- van der Schriek T, Giannakopoulos C, Varotsos KV (2020) The impact of future climate change on bean cultivation in the Prespa Lake catchment, northern Greece. *Euro-Mediterr J Environ Integr* 5:1–10
- Vinçon-Leite B, Casenave C (2019) Modelling eutrophication in lake ecosystems: a review. *Sci Total Environ* 651:2985–3001
- Virro H, Kmoach A, Vainu M, Uuemaa E (2022) Random forest-based modeling of stream nutrients at national level in a data-scarce region. *Sci Total Environ* 840:156613
- Wang G, Jia QS, Zhou M, Bi J, Qiao J, Abusorrah A (2022) Artificial neural networks for water quality soft-sensing in wastewater treatment: a review. *Artif Intell Rev* 55(1):565–587
- Wang Y, Khodadadzadeh M, Zurita-Milla R (2023) Spatial+: a new cross-validation method to evaluate geospatial machine learning models. *Int J Appl Earth Obs Geoinf* 121:103364. <https://doi.org/10.1016/j.jag.2023.103364>
- Wu J, Wang Z (2022) A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory. *Water* 14(4):610. <https://doi.org/10.3390/w14040610>
- Cao X, Yao J, Xu Z, Meng D (2020) Hyperspectral image classification with convolutional neural network and active learning. *IEEE Trans Geosci Remote Sens* 58(7):4604–4616. <https://doi.org/10.1109/TGRS.2020.2964627>
- Xu J, Ancil F, Boucher MA (2022) Exploring hydrologic post-processing of ensemble streamflow forecasts based on affine kernel dressing and non-dominated sorting genetic algorithm II. *Hydrol Earth Syst Sci* 26(4):1001–1017



- Yan T, Shen SL, Zhou A (2022) Indices and models of surface water quality assessment: review and perspectives. *Environ Pollut*, 119611. <https://doi.org/10.1016/j.envpol.2022.119611>
- Zamani MG, Moridi A, Yazdi J (2022) Groundwater management in arid and semi-arid regions. *Arab J Geosci* 15(4):362. <https://doi.org/10.1007/s12517-022-09546-w>
- Zamani MG, Nikoo MR, Niknazar F, Al-Rawas G, Al-Wardy M, Gandomi AH (2023a) A multi-model data fusion methodology for reservoir water quality based on machine learning algorithms and bayesian maximum entropy. *J Clean Prod* 416:137885
- Zamani MG, Nikoo MR, Rastad D, Nematollahi B (2023b) A comparative study of data-driven models for runoff, sediment, and nitrate forecasting. *J Environ Manag* 341:118006. <https://doi.org/10.1016/j.jenvman.2023.118006>
- Zamani MG, Saniei K, Nematollahi B, Zahmatkesh Z, Poor MM, Nikoo MR (2023c) Developing sustainable strategies by LID optimization in response to annual climate change impacts. *J Clean Prod* 416:137931
- Zhang Q, Li H (2007) MOEA/D: a multiobjective evolutionary algorithm based on decomposition. *IEEE Trans Evol Comput* 11(6):712–731
- Zhang F, Li J, Shen Q, Zhang B, Tian L, Ye H, Wang S, Lu Z (2019) A soft-classification-based chlorophyll-a estimation method using MERIS data in the highly turbid and eutrophic Taihu Lake. *Int J Appl Earth Obs Geoinf* 74:138–149. <https://doi.org/10.1016/j.jag.2018.07.018>
- Zhou ZH (2012) Ensemble methods: foundations and algorithms. CRC press
- Zhou H, Deng Z, Xia Y, Fu M (2016) A new sampling method in particle filter based on Pearson correlation coefficient. *Neurocomputing* 216:208–215
- Zhou H, Yan P, Huang Q, Wu D, Pei J, Zhang L (2023) Weighted average selective ensemble strategy of deep convolutional models based on grey wolf optimizer and its application in rotating machinery fault diagnosis. *Expert Syst Appl* 234:121076
- Zounemat-Kermani M, Batelaan O, Fadaee M, Hinkelmann R (2021) Ensemble machine learning paradigms in hydrology: a review. *J Hydrol* 598:126266
- Zhu S, Heddam S (2020) Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). *Water Qual Res J* 55(1):106–118

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.