IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 December 2023, accepted 3 January 2024, date of publication 9 January 2024, date of current version 24 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3351754

RESEARCH ARTICLE

Comparative Analysis of Artificial Intelligence Methods for Streamflow Forecasting

YAXING WEI[1], HUZAIFA BIN HASHIM[1], SAI HIN LAI[1,2], KAI LUN CHONG[3], YUK FENG HUANG[4], ALI NAJAH AHMED[5], MOHSEN SHERIF[6,7], AND AHMED EL-SHAFIE[1]

[1]Department of Civil Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia
[2]Department of Civil Engineering, Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak 94300, Malaysia
[3]Faculty of Engineering and Quantity Surveying, INTI International University (INTI-IU), Nilai, Negeri Sembilan 71800, Malaysia
[4]Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kajang, Selangor 43000, Malaysia
[5]School of Engineering and Technology, Sunway University, Petaling Jaya 47500, Malaysia
[6]Civil and Environmental Engineering Department, College of Engineering, United Arab Emirates University, Al Ain, United Arab Emirates
[7]National Water and Energy Center, United Arab Emirates University, Al Ain, United Arab Emirates

Corresponding author: Huzaifa Bin Hashim (huzaifahashim@um.edu.my)

This work was supported by the National Water and Energy Center, United Arab Emirates University, United Arab Emirates, for APC.

ABSTRACT Deep learning excels at managing spatial and temporal time series with variable patterns for streamflow forecasting, but traditional machine learning algorithms may struggle with complicated data, including non-linear and multidimensional complexity. Empirical heterogeneity within watersheds and limitations inherent to each estimation methodology pose challenges in effectively measuring and appraising hydrological statistical frameworks of spatial and temporal variables. This study emphasizes streamflow forecasting in the region of Johor, a coastal state in Peninsular Malaysia, utilizing a 28-year streamflow-pattern dataset from Malaysia's Department of Irrigation and Drainage for the Johor River and its tropical rainforest environment. For this dataset, wavelet transformation significantly improves the resolution of lag noise when historical streamflow data are used as lagged input variables, producing a 6% reduction in the root-mean-square error. A comparative analysis of convolutional neural networks and artificial neural networks reveals these models' distinct behavioral patterns. Convolutional neural networks exhibit lower stochasticity than artificial neural networks when dealing with complex time series data and with data transformed into a format suitable for modeling. However, convolutional neural networks may suffer from overfitting, particularly in cases in which the structure of the time series is overly simplified. Using Bayesian neural networks, we modeled network weights and biases as probability distributions to assess aleatoric and epistemic variability, employing Markov chain Monte Carlo and bootstrap resampling techniques. This modeling allowed us to quantify uncertainty, providing confidence intervals and metrics for a robust quantitative assessment of model prediction variability.

INDEX TERMS Artificial neural network, deep learning convolutional neural network, Bayesian statistic, streamflow, time series, uncertainty analysis.

ABBREVIATIONS AND ACRONYMS

A	Approximation Component.
AdaBoost	Adaptive Boosting.
AI	Artificial Intelligence.
ANFIS	Adaptive Neuro-Fuzzy Inference Systems.
ANN	Artificial Neural Network.
ARIMA	Autoregressive Integrated Moving Average.
CNN	Convolutional Neural Network.
COD	Chemical Oxygen Demand.
DW	Detail Component.
ELM	Elman Neural Network.
GP	Gaussian Processes.
GRU	Gated Recurrent Units.
LSTM	Long Short Term Memory.

The associate editor coordinating the review of this manuscript and approving it for publication was Massimo Cafaro.

MLP	Multi-Layer Perceptron.
MSE	Mean Square Error.
NSE	Nash-Sutcliffe Efficiency.
NH_4^+	Ammonium Concentration.
PCA	Principal Component Analysis.
RF	Random Forest.
RFBN	Radial Basis Function Neural Networks.
RMSE	Root Mean Square Error.
RNN	Recurrent Neural Network.
SVM	Support Vector Machine.
SVR	Support Vector Regression.
SWAT	Soil & Water Assessment Tool.
SWE	Snow Water Equivalent.
TSL	Total Sediment Load.
TSS	Total Suspended Solids.
WT	Wavelet Transform.
WWTP	Wastewater Treatment Plant.
XGB	eXtreme Gradient Boosting.

I. INTRODUCTION

Streamflow monitoring is critical for estimating the availability and distribution of water resources for human water demands, which is essential for agricultural irrigation, industrial operations, and municipal water supply planning. Water resource pressures rise and environmental concerns grow, prompting an urgent need to address a critical question: Why is streamflow important, and how can a nuanced understanding of its patterns and dynamics be leveraged for sustainable water management, ecological health, and infrastructure planning in the face of evolving environmental and societal demands?

More broadly, the significance of streamflow extends beyond the context of human water demands. Terrestrial aquatic ecosystems are intricately linked to streamflow dynamics, and their health depends on consistent adequate flow [1]. Streamflow can impact the prevalence and transmission of illnesses in aquatic ecosystems. While streamflow does not directly cause diseases, it does play a crucial role in generating habitat conditions that can impact pathogen dynamics and their consequences for aquatic animals. Streamflow facilitates the transport of pathogens, including bacteria, viruses, and parasites, through aquatic environments [2]. Increased flow can disperse pathogens over greater distances, potentially affecting a broader range of species and ecosystems. Streamflow influences water quality parameters such as temperature, oxygen level, and nutrient concentrations [3]. Changes in these factors due to variations in streamflow can stress aquatic organisms, making them more susceptible to diseases.

Understanding streamflow patterns not only is essential for ecological health but also forms the basis for sedimentary budget forecasting in fluvial flows [4]. Streamflow modeling becomes a key tool in this regard, as it simulates hydrodynamic processes within river systems, including flow velocity, discharge, and channel morphology. These models make it possible to identify areas prone to erosion or sediment deposition [5]. Inadequate streamflow management can lead to increased erosion along riverbanks and within the river channel. Without sufficient flow to transport sediment downstream, sediments may accumulate due to elevated net sedimentation rates. This can adversely affect water quality, aquatic habitats, and infrastructure [6].

Streamflow modeling is a vital technique for managing water resources, especially in the early detection of flood dangers [7], [8], [9]. Several types of advanced models can operate across the range of climate zones. Especially during flood events, fast efficient replication of streamflow is crucial to the forecasting process and is accomplished by hydrodynamic models [10], [11]. For instance, Mahdian et al. [12] employed the Soil and Water Assessment Tool model to simulate and analyze streamflow dynamics under various climate and land use scenarios. Their findings help clarify how changes in climate and human activities, such as deforestation and urbanization, can impact streamflow and sediment inputs to ecosystems, in their case the Anzali wetland ecosystem. However, these complex models require precise river geometry data, which is not always accessible. In contrast, artificial intelligence (AI) tools, such as artificial neural networks and deep learning, have abolished the necessity for detailed knowledge of river geometry [13]. Furthermore, their modeling can be nonlinear, piecewise, or discontinuous, among other types of relationships [14].

II. LITERATURE REVIEW

Existing literature reveals successful demonstrations of AI-based models in hydrological forecasting, leading to improved accuracy and predictive capabilities in hydrological and water resource management. Models range from widely used artificial neural networks (ANNs) to state-of-the-art algorithms, such as deep learning, which include convolutional neural networks (CNNs), long short-term memory (LSTM), and generative adversarial networks.

Recurrent neural networks (RNNs), which include LSTM, gated recurrent units (GRU), and standard RNN, are widely employed for time series forecasting due to their proficiency in handling sequential data. Sahoo et al. [15] illustrated their effectiveness, finding that RNN outperforms radial basis function neural networks (RFBN) for streamflow forecasting. However, standard RNNs can encounter issues related to vanishing or exploding gradients, potentially affecting their performance. Samantaray et al. [16] demonstrated this challenge by showing that support vector machines (SVMs) and adaptive neuro-fuzzy inference systems (ANFISs) outperformed RNN in rainfall modeling.

In response to the vanishing/exploding gradient problem, LSTM has gained prominence. LSTM models exhibit a lower susceptibility to these pitfalls and offer enhanced performance. For instance, Bala et al. [17] achieved success using LSTM for rainfall prediction, with the LSTM approach surpassing the Elman neural network (ELM) and autore-