



Quantitative improvement of streamflow forecasting accuracy in the Atlantic zones of Canada based on hydro-meteorological signals: A multi-level advanced intelligent expert framework

Mozhdeh Jamei^a, Mehdi Jamei^{b,c,d,*}, Mumtaz Ali^{e,f}, Masoud Karbasi^{b,g},
Aitazaz A. Farouqe^{b,f,*}, Anurag Malik^h, Saad Javed Cheema^b, Travis J. Esauⁱ, Zaher
Mundher Yaseen^{j,k}

^a Khuzestan Water and Power Authority, Ahvaz, Iran

^b Canadian Centre for Climate Change and Adaptation, University of Prince Edward Island, St Peters Bay, PE, Canada

^c Faculty of Civil Engineering and Architecture, Shahid Chamran University of Ahvaz, Ahvaz, Iran

^d New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Nasiriyah 64001, Iraq

^e UniSQ College, University of Southern Queensland, QLD 4350, Australia

^f Faculty of Sustainable Design Engineering, University of Prince Edward Island, Charlottetown, PE C1A4P3, Canada

^g Water Engineering Department, Faculty of Agriculture, University of Zanjan, Zanjan, Iran

^h Punjab Agricultural University, Regional Research Station, Bathinda, Punjab, India

ⁱ Department of Engineering, Faculty of Agriculture, Dalhousie University, Truro, NS B2N 5E3, Canada

^j Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

^k Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Keywords:

Streamflow forecasting
Hydro-meteorological drivers
Multivariate variational mode decomposition
CNN-BiGRU
Boruta-CART
Multi-temporal

ABSTRACT

Developing reliable streamflow forecasting models is critical for hydrological tasks such as improving water resource management, analyzing river patterns, and flood forecasting. In this research, for the first time, an emerging multi-level TOPSIS (technique for order preference by similarity to the ideal solution)-based hybridization comprised of the Boruta classification and regression tree (Boruta-CART) feature selection, multivariate variational mode decomposition (MVMD), and a hybrid Convolutional Neural Network (CNN) Bidirectional Gated Recurrent Unit (CNN-BiGRU) deep learning was adopted to multi-temporal (one and three days ahead) forecast the daily streamflow in the Rivers of Prince Edward Island, Canada. For this aim, in the first step, the Boruta-CART feature selection technique determines the most effective lagged components among all the antecedent two-day information (i.e., t-1 and t-2) of hydro-meteorological features (from 2015 to 2020), including the water level, mean air temperature, heat degree days, total precipitation, dew point temperature, and relative humidity in the Bear and Winter Rivers of Prince Edward Island, Canada. Afterwards, a multivariate variational mode decomposition (MVMD) decomposes the input time series to decrease the complexity and non-linearity of the non-stationary ones before feeding the deep learning (DL) models. Here, the CNN-GRU was employed as the primary DL model, along with the kernel extreme machine method (KELM), random variational function link (RVFL), and hybrid CNN bidirectional recurrent neural network (CNN-BiRNN) as the comparative models. A TOPSIS scheme applying several performance measures like the correlation coefficient (R), root mean square error (RMSE), and reliability was designed for the robustness assessment of the hybrid (MVM-CNN-BiGRU, MVM-CNN-BiRNN, MVM-RVFL, and MVM-KELM) and standalone models. The computational outcomes revealed that in the Bear River, the MVM-CNN-BiGRU, owing to its best forecasting performance (one day ahead: TOPSIS score 1, $R = 0.960$, $RMSE = 0.098$, and reliability = 65.082; three days ahead: TOPSIS score = 0.999, $R = 0.924$, and $RMSE = 0.33$) outperformed the other hybrid models, followed by the MVM-CNN-BiRNN, MVM-RVFL, and MVM-KELM, respectively. Moreover, in the Winter River, the MVM-CNN-BiGRU in terms of (one-day ahead: TOPSIS score = 0.890, $R = 0.955$, $RMSE = 0.274$, and reliability = 34.004; three-days ahead: TOPSIS score = 0.686, $R = 0.924$, and $RMSE = 0.330$) was superior to the other models. The provided expert system

* Corresponding authors at: Canadian Centre for Climate Change and Adaptation, University of Prince Edward Island, St Peters bay, PE, Canada.

E-mail addresses: mozhdehjamei@gmail.com (M. Jamei), Mehdi.jamei59@gmail.com (M. Jamei), m.karbasi@znu.ac.ir (M. Karbasi), tesau@dal.ca (T.J. Esau).

<https://doi.org/10.1016/j.ecoinf.2023.102455>

Received 16 November 2023; Received in revised form 24 December 2023; Accepted 30 December 2023

Available online 4 January 2024

1574-9541/© 2023 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

could be vital in the local flood decision-making process, in the absence of streamflow information as input modeling, during the flood seasons to reduce flood damage in residential areas.

Nomenclature			
ANN	Artificial neural network	NIS	Worst Performance
BiGRU	Bidirectional Gated Recurrent Unit	NSE	Nash–Sutcliffe model efficiency
BiRNN	Bidirectional Recurrent Neural Network	PEI	Prince Edward Island
Boruta-CART	Boruta-Classification and Regression Trees	PIS	Best Performance
CNN	Convolutional Neural Network	Q_{flow}	Streamflow
COV	Coefficient of Variation	R	Correlation Coefficient
ELM	Extreme Learning Machine	RMSE	Root Mean Squared Error
IMFs	Intrinsic Mode Functions	RNN	Recurrent Neural Network
KELM	Kernel Extreme Learning Machine	RVFL	Random Vector Functional Link
KGE	Kling-Gupta efficiency	Std. D	Standard Deviation
LSTM	Long Short-Term Memory	SquD	Squared Chi-square Distance
MAPE	Mean Absolute Percentage Error	TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
ML	Machine learning	$U_{95\%}$	Uncertainty coefficient
MVMD	Multivariate Variational Mode Decomposition	UniRNN	Unidirectional Recurrent Neural Network
		VMD	Variational Mode Decomposition

1. Introduction

An accurate streamflow (Q_{flow}) forecasting is one of the essential hydrological modeling procedures to provide a sustainable design method for water bodies, implement control measures for flood disasters, and analyze river patterns for operational reasons (Bakhshi Ostadkalayeh et al., 2023; Moazam et al., 2023). Q_{flow} has been influenced by many such as seasonal patterns, variations in temperature and precipitation at the local and regional levels, and the spatial and temporal variability of the watersheds (Beyaztas et al., 2021). In order to achieve high accuracy for the Q_{flow} data, predictive models are developed based on pertinent hydrological input variables. The literature has suggested that the regression-based techniques cannot effectively model

the Q_{flow} data due to the inherent non-linear correlations between the associated variables (input and output parameters) during the simulation. Therefore, conceptual machine learning (ML) and data pre-processing methods are highly recommended and considered reliable tools to effectively analyze and extract the non-linear behaviours between the predictor-predictand relationships (Liu et al., 2022; Yaseen et al., 2018). For instance, data pre-processing methods can be combined with ML models at a prior stage of the learning process, where more comprehensive information can be supplied for the prediction/forecasting matrix (Nourani et al., 2014). This integration increases the modeling capability of a standalone prediction model used for Q_{flow} modeling.

Q_{flow} data-based modeling is a strategy that brings an essential era to the hydrologist and decision-makers (Papacharalampous and Tyrallis,

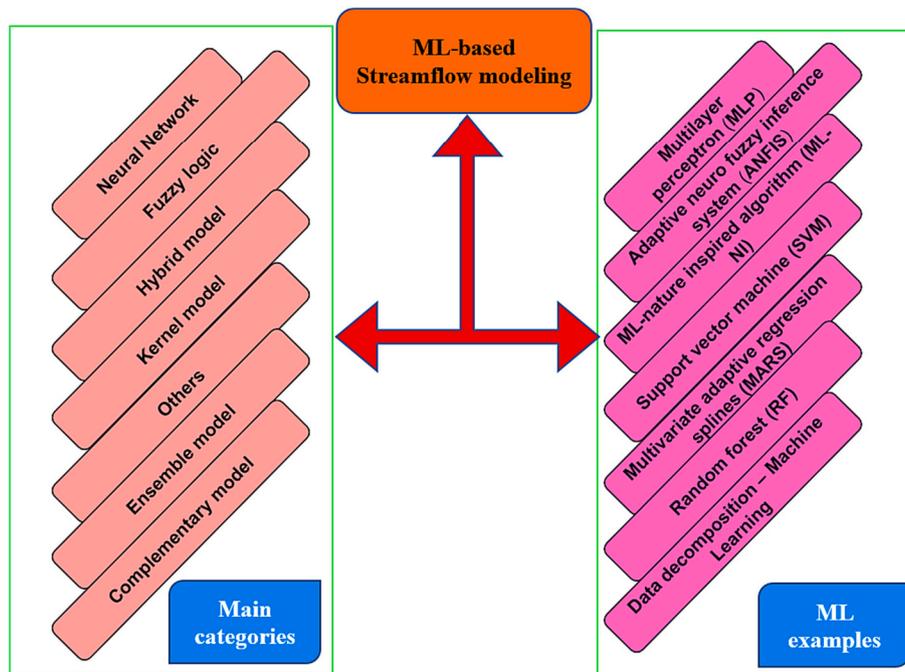


Fig. 1. The types of machine learning models applied for streamflow simulation over the literature.

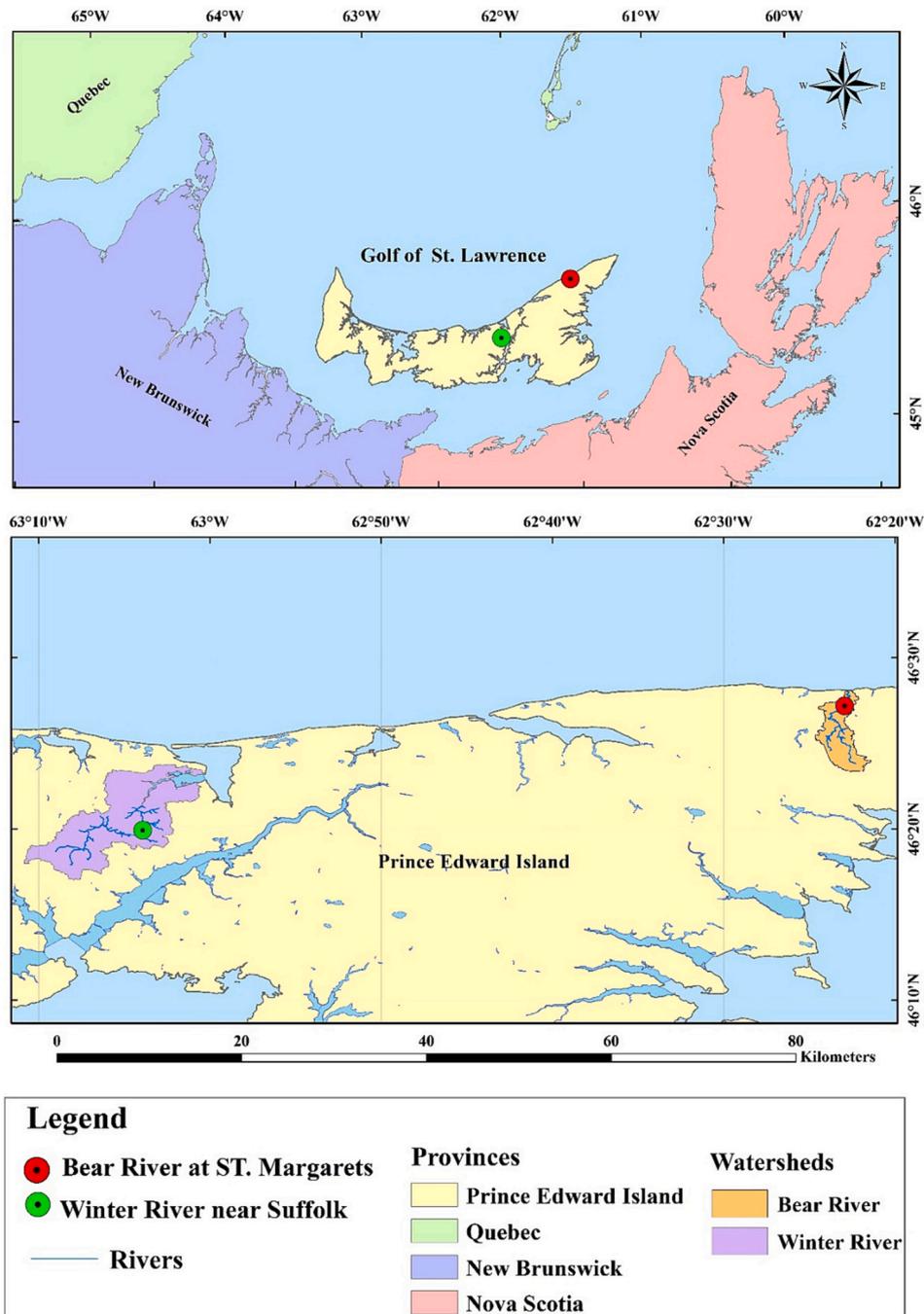


Fig. 2. Location of the Bear River and Winter River Stations in Prince Edward Island in Canada's Atlantic regions.

2022; Rozos et al., 2022). ML-based models could provide a robust technology for the extraction of data patterns from previous variables to estimate future events; they have been established as successful and popular tools for predictions and forecasting tasks (Allawi et al., 2022; Kilinc et al., 2023). Fig. 1 exhibits the well-established ML models over the literature for Q_{flow} modeling. However, most recently, significant advancement has been conducted in the domain of ML models for hydrological processes (Mosaffa et al., 2022).

ML-based models demonstrated exceptional results with excellent agreement with the actual/observed Q_{flow} data, specifically for the examined hydrological process (such as Q_{flow}) (Fan et al., 2023; Zhang et al., 2023). The modeling of Q_{flow} was adopted not only based on univariate modeling strategy; however, many modeling studies on Q_{flow} -related multivariate modeling strategy where several variables (i.e.,

evaporation, temperature, wind, humidity, etc.) have been carried out using various AI-based models for several reasons (Al-Areeq et al., 2022; Hamzah et al., 2020; Ibrahim et al., 2022), the ease of designing the prediction model, the related application, the cost-efficiency of the data needed for the model, the less complexity of the models in comparison to the physically-based hydrological models, the reproducibility and local-scale applications of the models (such as farms or irrigations). Being that the non-stationarities observed in the input-target data (such as those driven by seasonal changes, jumps, and temporal patterns) are discreetly considered in the design of the internal structure of ML models, the simulation of Q_{flow} has been made possible using several ML-approaches. Recent ML-based algorithms, e.g., deep learning or decision tree models for Q_{flow} prediction/forecasting, often gave better results than conventional models such as autoregressive and regression-based methods

Table 1Statistical descriptive of all the time series related to developing the forecasting model of streamflow (Q_{flow}) in two case studies of PEI province of Canada.

Station	Statistical indices	Minimum	Q25%	Q75%	Maximum	Mean	Std. D	C.O-V	Skewness	Kurtosis
Bear River	Level(m)	0.487	0.5668	0.698	1.475	0.6485	0.1163	17/94%	2.184	8.344
	Mean Temp (°C)	-19.7	-1.8	14.8	25.2	6.024	10.18	169/1%	-0.1191	-0.939
	Heat Deg Days (°C)	0	3.2	19.8	37.7	12.39	9.6	77/51%	0.2852	-0.9898
	Total Precip (mm)	0	0	2.8	74	3.266	7.227	221/3%	3.781	19.01
	Dew Point Temp (°C)	-24.88	-5.406	11.09	22.18	2.214	10.27	463/9%	-0.2441	-0.8002
	Rel Hum (%)	38.25	69.21	85.89	100	77.09	11.62	15/08%	-0.3379	-0.3641
	Flow(m ³ /s)	0.038	0.12	0.43	3.99	0.3579	0.4	111/7%	3.416	16.69
Winter River	Level(m)	0.963	1.092	1.266	2.358	1.189	0.1453	12/22%	1.67	7.289
	Mean Temp (°C)	-19.7	-1.8	14.8	25.2	6.024	10.18	169/1%	-0.1191	-0.939
	Heat Deg Days (°C)	0	3.2	19.8	37.7	12.39	9.6	77/51%	0.2852	-0.9898
	Total Precip (mm)	0	0	2.8	74	3.266	7.227	221/3%	3.781	19.01
	Dew Point Temp (°C)	-24.88	-5.406	11.09	22.18	2.214	10.27	463/9%	-0.2441	-0.8002
	Rel Hum (%)	38.25	69.21	85.89	100	77.09	11.62	15/08%	-0.3379	-0.3641
	Flow(m ³ /s)	0.083	0.201	0.7965	12.1	0.6633	0.8637	130/2%	5.273	46.4

or even the classical ML models presented in Fig. 1.

Recently, different methodologies have been adopted for Q_{flow} prediction/forecasting based on integrative ML models. Daily scale Q_{flow} was predicted using integrated CatBoost and genetic algorithm for a case study within the Mediterranean region (Kilinc et al., 2023). Hybridized different versions of deep learning models with particle swarm algorithm conducted for modeling daily scale Q_{flow} in Algeria (Zakhrouf et al., 2021). In another work (Deepa et al., 2023), an enhanced variational mode decomposition hybridized with a deep support vector machine for Cahaba river flow in the USA. The feasibility of the Grey Wolf optimization algorithm was tested to tune different ML models for modeling Q_{flow} for multiple-step ahead forecasting at Zambezi River, Zimbabwe (Martinho et al., 2023). Long-term Q_{flow} was forecasted for the Funil reservoir using a newly developed temporal fusion transformer deep learning model (Fayer et al., 2023). Within a semi-arid environment, Q_{flow} was predicted using hybridized Hydrologiska Byråns Vattenbalansavdelning with long short-term memory for Weihe River, Central China (Yu et al., 2023). All the forgoing exhibited research in addition to several others (Ikram et al., 2023; Katipoğlu and Sargöl, 2023; Moazam et al., 2023), confirmed the potential of the hybridized ML algorithms in improving the modeling accuracy for different regions all around the world.

There has been a striking advancement in the research trend on applying ML for Q_{flow} forecasting (Tao et al., 2024). This is due to the need for computer-digitalized technology to allocate Q_{flow} more accurately. Even though the literature showed a variety of ML models for related Q_{flow} concerns, this research field is still experiencing several limitations. Researchers and engineers continue to be primarily interested in exploring robust data pre-processing, feature selection algorithms, or even more reliable prediction models for this field of study. The multivariate variational mode decomposition is a powerful and dependable method of modern data pre-processing technologies (Meng et al., 2022; Yousefi et al., 2023). Its potential for the development of research in a variety of fields has been acknowledged. This pre-processing method can simultaneously and accurately decompose all input characteristics, unlike variational mode decomposition (VMD), which significantly reduces the cost and computation time (Ahmadi et al., 2023). Several other research over the literature showed the capacity of adopted data pre-processing methods as the prior stage for the prediction process using ML models (Ali et al., 2018; Yang et al., 2023; Yang and Li, 2023; Zounemat-Kermani et al., 2021). In¹ addition, as a feature selection approach, the ability of the Boruta approach to capture the high non-linear correlation between feature and target made it a reliable method for filtering and abstracting the important data feature for ML model development. Hence, research on this perspective is always the motive to explore such highly complex non-linear hydrological

issues.

This study investigated a new TOPSIS-based multivariate-multi-temporal intelligent system to forecast the streamflow in the two rivers of PEI, Canada. In order to accurately model and forecast the Q_{flow} a robust Boruta-CART feature selection technique was used to determine the most effective predictors among the available time series data from a couple of previous days (i.e., hydro-meteorological variables, including water level, mean temperature, heat degree days, total precipitation, dew point temperature, and relative humidity). Afterwards, the MVMD decomposition method decomposed the best candidate lagged feature to feed the ML models aiming at streamflow forecasting. Two advanced deep learning approaches, namely CNN-BiGRU and CNN-BiRNN, along with the random vector functional link (RVFL) and kernel extreme learning machine (KELM) models, were constructed for the forecast Q_{flow} . Four hybrid forecasting approaches, namely MVM-CNN-BiGRU, MVM-CNN-BiRNN, MVM-RVFL, and MVM-KELM were validated using several performance metrics, diagnostic analysis, and a TOPSIS scheme for both rivers under this study. The developed intelligent system, owing to remarkable predictive potential, could play a significant role in the decision-making process of flood control in PEI with limited input variables.

2. Methodology and materials

2.1. Area study and data preparation

This study is focused on multivariate-multi-temporal streamflow (Q_{flow}) forecasting using hydro-meteorological signals for two rivers in the Canadian province of Prince Edward Island (PEI), located in the Atlantic region by the Gulf of Saint Lawrence. PEI has a cool, humid climate with moderate winters and relatively warm summers. The annual precipitation of PEI is around 1100 mm, which is evenly distributed throughout the year (Bhatti et al., 2022). The Bear and Winter Rivers in PEI are the two Study Rivers shown in Fig. 2. The station of Bear River at ST Margarets (46° 27' 11" North, 62° 22' 56" West) is situated on the Bear River with a gross drainage area of 14.8 km². The station of Winter River near Suffolk (46°19' 55" North, 63° 03' 55" West) is located on the Winter River with a gross drainage area of 37.5 km² (https://wateroffice.ec.gc.ca/search/historical_e.html). The hydro-meteorological variables, including Level (m), Mean Temperature (°C), Heat Degree Days (°C), Total Precipitation (mm), Dew Point Temperature (°C), Relative Humidity (%) and (Q_{flow}) (m³/s), were used as inputs to develop the streamflow forecasting model. These data were collected from Bear and Winter stations (from 01/01/2015 to 01/01/2020).

Statistical descriptions of all hydro-meteorological variables to develop the forecasting model in two rivers of PEI are shown in Table 1. At the Bear River station, the time series distributions of Mean Temperature (°C), Heat Degree Days (°C), Dew Point Temperature (°C) and

¹ v

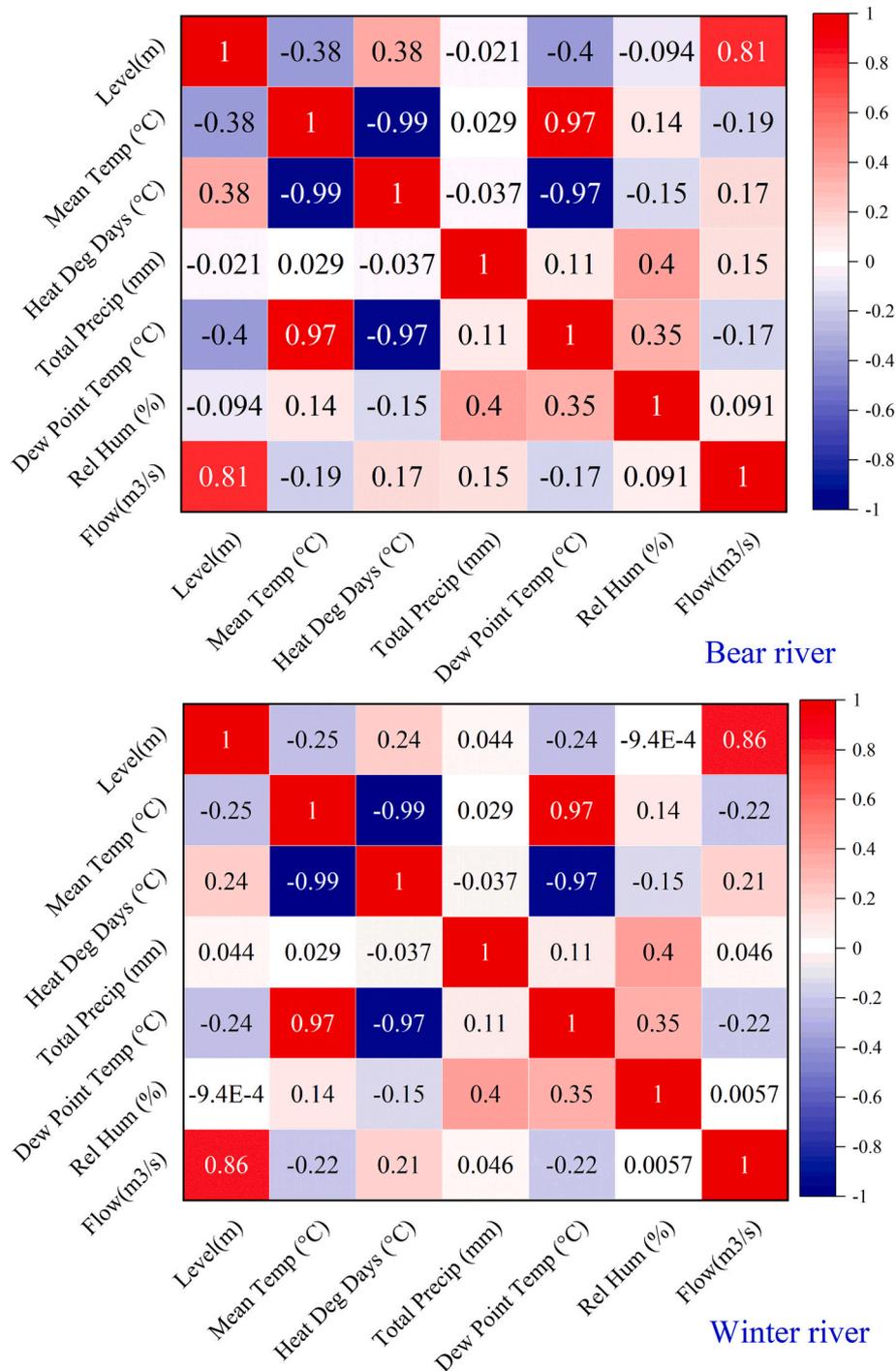


Fig. 3. Correlogram heat map for Bear () and (WINTER) Rivers of PEI to specify the linear correlation between input available signals and streamflow (Q_{flow}).

Relative Humidity (%) are nearly symmetrical and close to normal ($-0.338 < \text{Skewness} < 0.285$); and also with negative Kurtosis values, they have platykurtic distributions. The time series of Level (m), Total Precipitation (mm), and Flow (m^3/s) data at the Bear River station are highly skewed ($\text{Skewness} > 1$), and these data have Leptokurtic distributions with positive kurtosis values ($\text{Kurtosis} > 3$).

At the Winter River station, the time series of Mean Temperature ($^{\circ}C$), Heat Degree Days ($^{\circ}C$), Dew Point Temperature ($^{\circ}C$), and Relative Humidity (%) data have approximately symmetric distributions with skewness values of $[-0.338, 0.285]$. These data with ($\text{Kurtosis} < 3$) have platykurtic distributions. The time series of Level (m), Total Precipitation (mm), and Flow (m^3/s) data at the Winter River station have

Leptokurtic distributions with Kurtosis values (> 3). These data have extremely skewed time series, with skewness values (> 1).

Fig. 3 shows the Pearson correlation coefficient values for the Bear (up) and Winter (bottom) rivers in PEI. This was done so that a rough estimate could be made of how important each input signal was to the streamflow. The correlation coefficient reveals that the water level (Level) for both case studies had the highest influence (Bear|0.8 and Winter|0.86) on streamflow. In contrast, the other signals have no significant linear correlation with the streamflow. However, this technique can only discover the linear interaction between the inputs and the target. Thus, additional accurate exploration is necessary using more advanced feature selection, which has been adopted in the next sections.

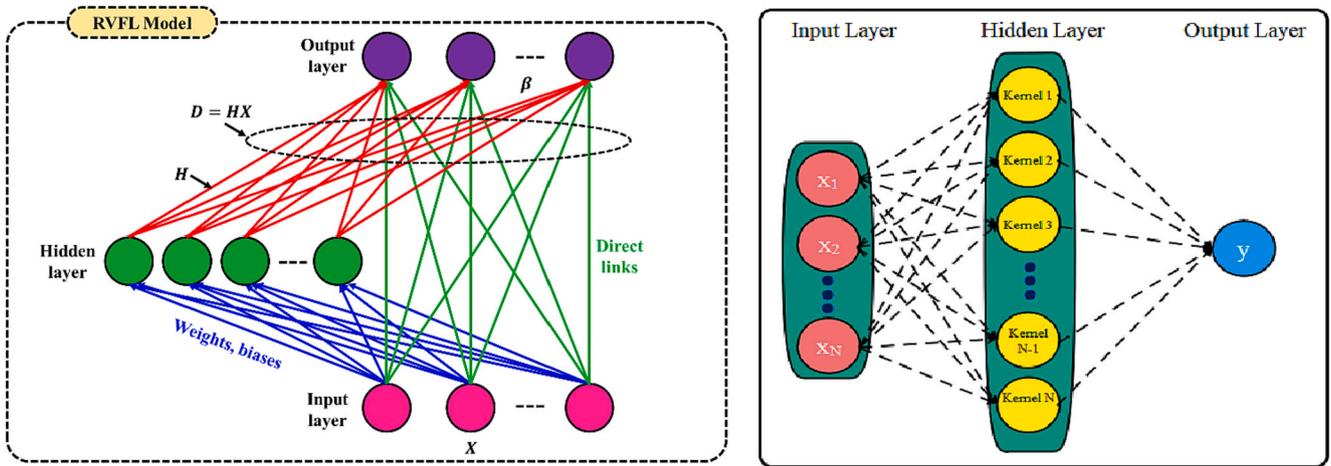


Fig. 4. The network of the RVFL model (left); topology of the KELM model (right).

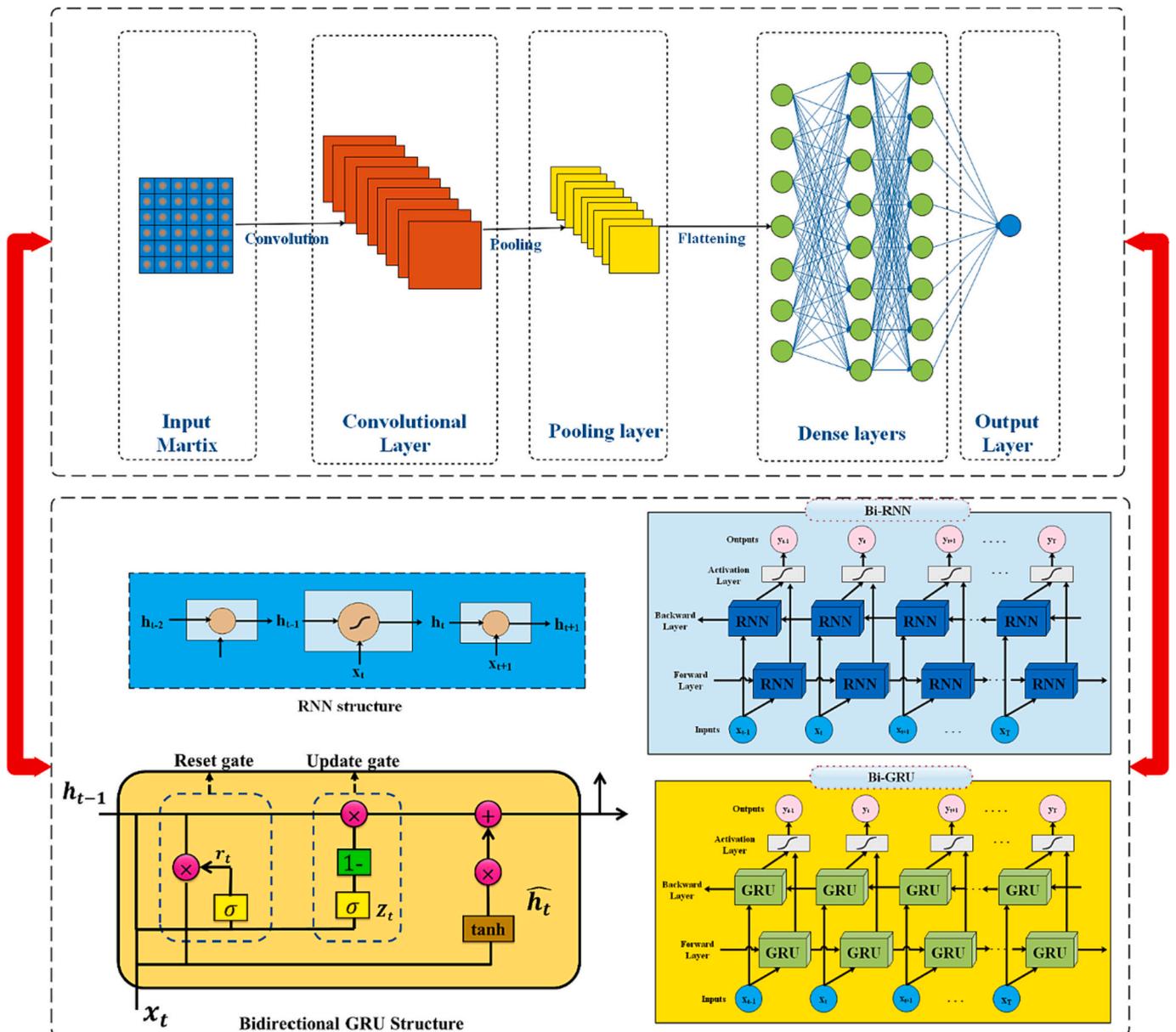


Fig. 5. The main structure of the CNN model (upper panel); Structure of GRU cell, RNN cell, Bi-GRU, and Bi-RNN models.

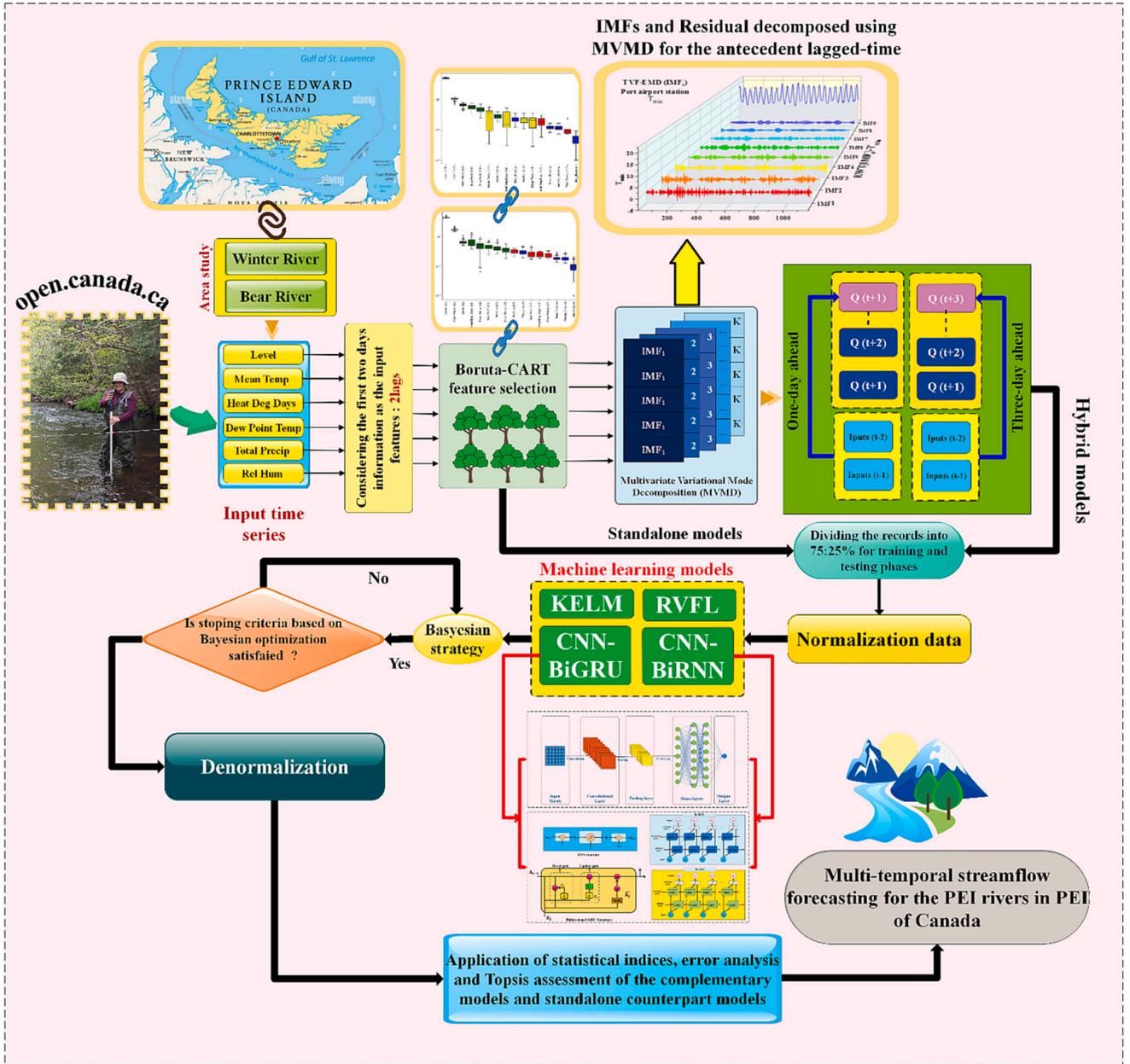


Fig. 6. Workflow of streamflow (Q_{flow}) multi-temporal forecasting of the streamflow values.

2.2. Multivariate variational mode decomposition (MVMD) technique

The MVMD is the extension of the variational mode decomposition (VMD), introduced by Rehman and Aftab (2019). It extracts a predefined K number of multivariate modulated oscillations $u_k(t)$ from input data $x(t)$ containing C number of multichannel signals, and described as (Rehman and Aftab, 2019):

$$x(t) = \sum_{k=1}^K u_k(t), \quad (1)$$

In the above equation, $u_k(t) = [u_1(t), u_2(t), \dots, u_C(t)]$, and $x(t) = [x_1(t), x_2(t), \dots, x_C(t)]$. The implementation of the MVMD decomposition comprises the two conditions (i) minimum the bandwidths sum of the extracted modes, and (ii) the sum of the extracted modes accurately improves the original signal $u_k(t)$ (Gu et al., 2020). Furthermore, to complete that, the vector analytic illustration of $u_k(t)$ designated by $u_+^k(t)$, and utilized the L_2 norm of gradient function of the harmonically

shifted $u_+^k(t)$ to compute the bandwidth of $u_k(t)$. After that, the bandwidth of the modulated multivariate oscillations is projected by shifting the unilateral frequency spectrum of all channels of $u_+^k(t)$ by ω_k and taking the Frobenius norm² of the resulting matrix. The constrained optimization problem for the MVMD is given as (Rehman and Aftab, 2019):

$$\begin{aligned} & \text{minimize} \left\{ \sum_k \sum_c \|\partial_t [u_+^{k,c}(t) e^{-j\omega_k t}] \|^2 \right\}, \\ & \text{Subject to} \sum_k u_{k,c}(t) = x_c(t), c = 1, 2, \dots, C \end{aligned} \quad (2)$$

Here, $u_k(t)$ = analytic modulated signal corresponding to c (channel number) and k (mode number). Next, the unconstrained problem is solved by constructing the augmented Lagrangian (which includes a quadratic term and Lagrangian multipliers λ). Finally, the resulting increased Lagrangian function is given (Gu et al., 2020; Rehman and Aftab, 2019):

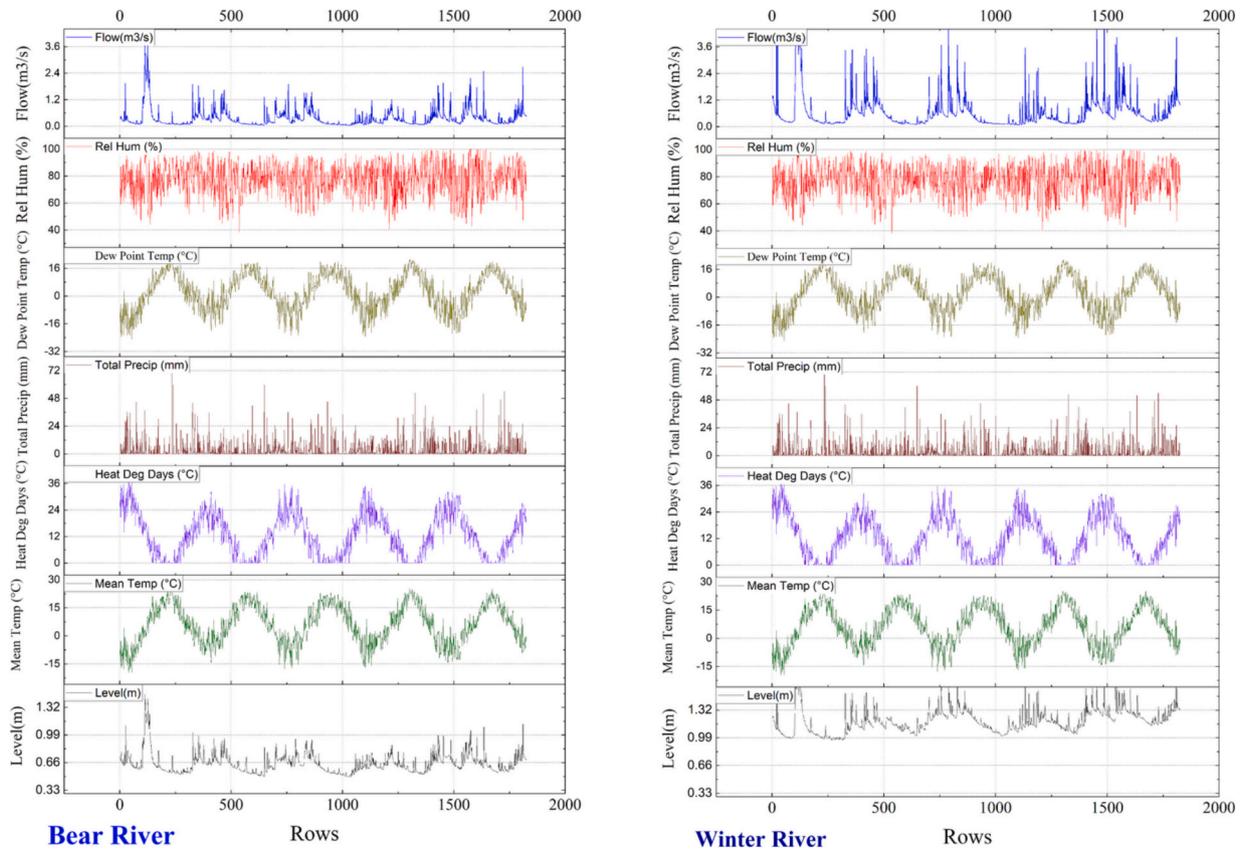


Fig. 7. The original signals of hydro-meteorological predictors to develop the multi-temporal forecasting models.

$$L(\{u_{k,c}\}, \{\omega_k\}, \lambda_c) = \alpha \sum_k \sum_c \|\partial_t [u_{k,c}^k(t) e^{-j\omega_k t}]\|_2^2 + \sum_c \left\| x_c(t) - \sum_k u_{k,c}(t) \right\|_2^2 + \sum_c \langle \lambda_c(t), x_c(t) - \sum_k u_{k,c}(t) \rangle. \quad (3)$$

Eq. (3) is solved using the ADMM (alternative direction technique of multipliers algorithm) method, which includes the mode update and centre frequency update in solving the process of variational problem. For a detailed background of the MVMD, readers refer to Rehman and Aftab (2019).

2.3. Boruta – Classification and regression trees (Boruta-CART)

This technique integrates the Boruta and CART in one topology and is utilized for optimal feature selection in forecasting complex hydrological events (Kursa et al., 2010; Singh et al., 2013). This study explores the application of the Boruta-CART technique in nominating useful inputs for river flow forecasting in the Atlantic regions of Canada. Here, the Boruta algorithm was tuned using the CART algorithm to remove the irrelevant features and point out the relevant inputs (Geurts et al., 2006). In CART, a dataset (either continuous or categorical) is divided into small sub-groups progressively, and then classification and regression trees are constructed (Tsai et al., 2012; Ture et al., 2005). The quality of splits and total error in CART are examined using the Gini Index of Impurity (Smeti et al., 2009). In the Boruta algorithm, the distribution of the Z-score metrics determines the important factors of the predictors. Over the years, this algorithm received effective applications in different fields (Ahmed et al., 2021; Jamei et al., 2022b; Prasad et al., 2019). The working concept of the Boruta algorithm is outlined in the following phases (Kursa et al., 2010; Kursa and Rudnicki, 2010):

1. A randomly ordered duplicate (shadow) variable is generated, x_t^i for the individual input vector x_t . Eliminate the correlations among the shadow predictors and targets, y_t and add the randomness. Forecast the y_t with x_t^i and x_t by using the random forest model.
2. Compute the MDA (mean decrease accuracy) for each x_t and corresponding x_t^i for complete trees as (Hur et al., 2017; Strobl et al., 2008):

$$MDA = \frac{1}{m_{tree}} \sum_{m=1}^{m_{tree}} \frac{\sum_{i \in OOB} I(y_i = f(x_i)) - \sum_{i \in OOB} I(y_i = f(x_t^i))}{|OOB|} \quad (4)$$

In the above Eq. *OOB* represents the out-of-bag (prediction error of each of training samples using bootstrap aggregation), $I()$ defines the indicator function, $(y_t = f(x_t))$ and $(y_t = f(x_t^i))$ states the predicted values before and after permuting.

3. Estimate the Z-scores using Eq. (5) as

$$Z - score = \frac{MDA}{StD} \quad (5)$$

where *StD* indicates the standard deviation of accuracy losses,

4. After that, determine the maximum Z-score among shadow attributes. Next, the inputs Z-scores are equated with matching shadows and inspected through a variable importance distribution. The inputs with Z-scores < maximum Z-score among shadow attributes = insignificant, while inputs with Z-scores > maximum Z-score among shadow attributes = significant.
5. New shadows are formed during each iteration, and end the process, once all features are either established or the iteration threshold is reached to the required level.

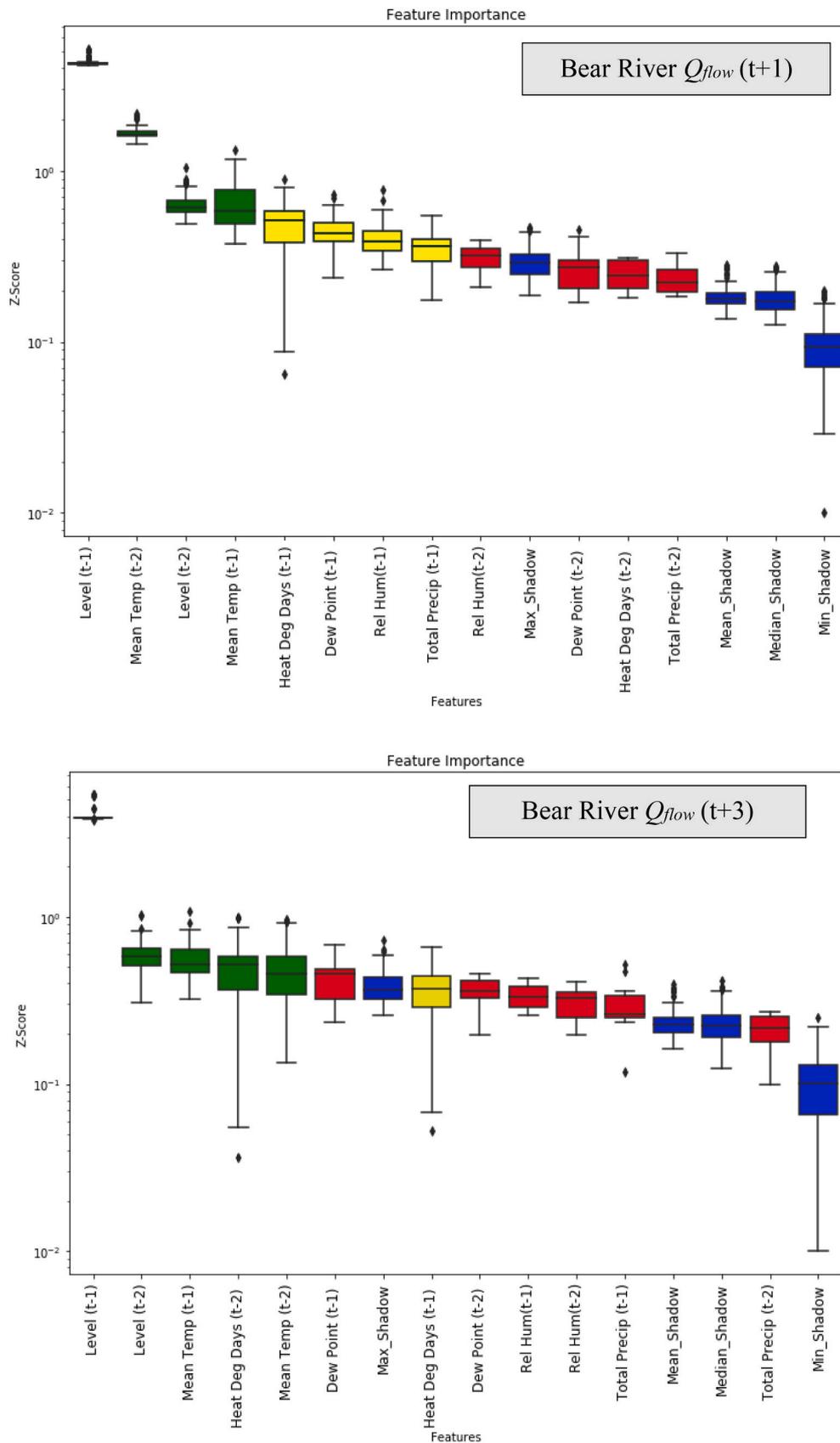


Fig. 8. Box-plot outcomes of the Boruta-CART FS to specify the most critical antecedent components among two first-month information for both understudy horizons (–one and -three days ahead) based on Z-score values in the Bear and Winter Rivers of PEI, Canada. Green colour: Accepted features; Yellow colour: Tentative features; Red colour: Rejected feature; Blue colour: Benchmarked shadows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

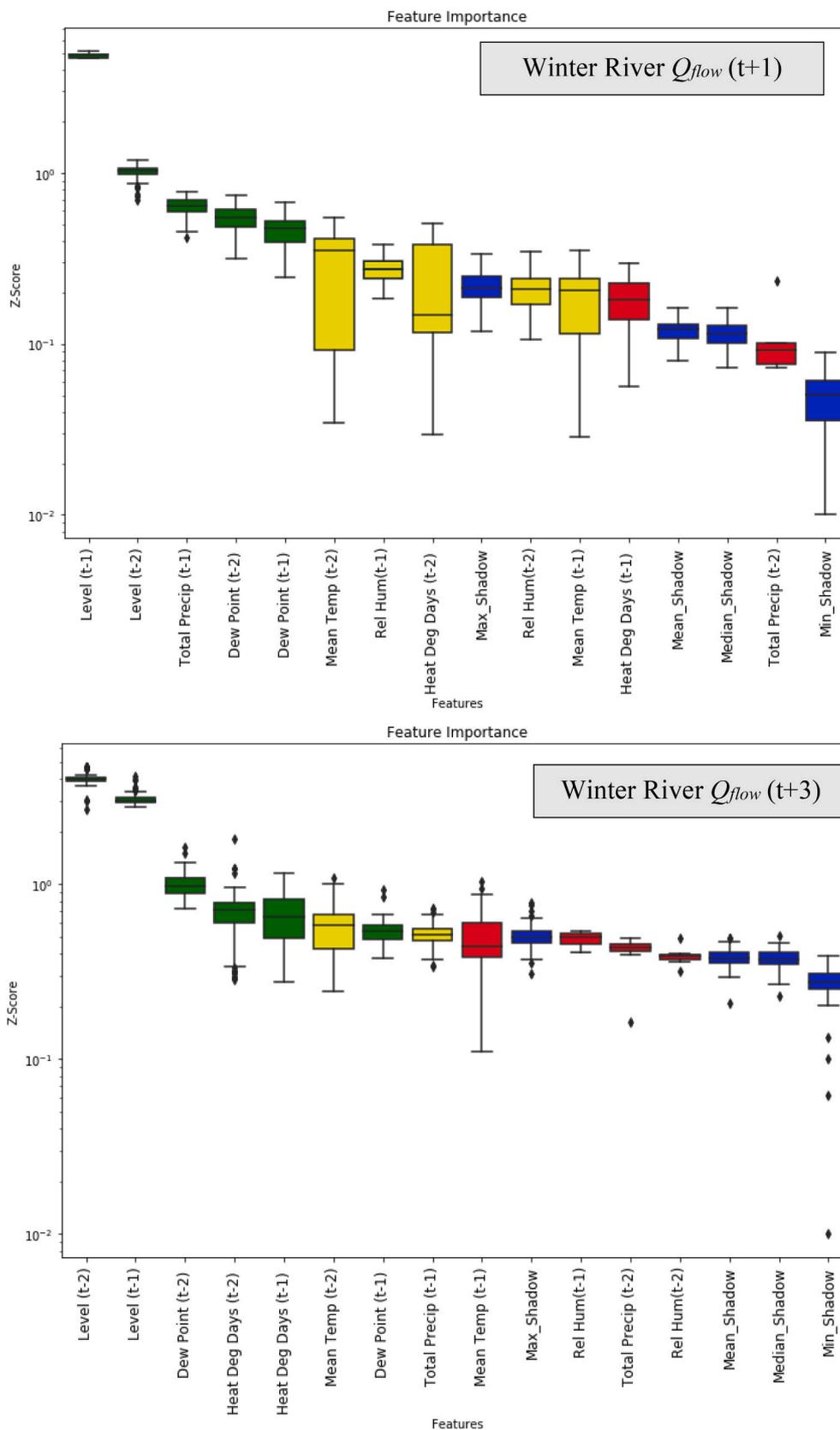


Fig. 8. (continued).

2.4. Kernel extreme learning machine (KELM)

The extreme learning machine (ELM) (Huang et al., 2006) is a

learning method based on the single-hidden layer feed-forward network. It has the benefits of fast training speed, simple application, and strong generalization ability, which is achieved by randomly giving the input

Table 2

The best characteristics of pre-processing parameter setting, MVMD decomposition information, and optimal antecedent lagged-time components utilizing the Boruta-CART technique to develop the multi-temporal streamflow forecasting (Q_{flow}) model.

River station	Whole components number	Optimal input sub-sequences	MVMD setting
Bear River Horizon: t + 1	40	Level (t-1), Level (t-2), Mean Temp (t-1), Mean Temp (t-2),	k = 10; Tol = 1e-7; DC = 0 Init = 0; τ = 0
Bear River Horizon: t + 3	60	Level (t-1), Level (t-2), Mean Temp (t-1), Mean Temp (t-2) and Heat Deg Days (t-2)	k = 12; Tol = 1e-7; DC = 0 Init = 0; τ = 0
Winter River Horizon: t + 1	55	Level (t-1), Level (t-2), Dew Point Temp (t-1), Dew Point Temp(t-2), and Total Precip (t-1)	k = 11; Tol = 1e-7; DC = 0 Init = 0; τ = 0
Winter River Horizon: t + 3	72	Level (t-1), Level (t-2), Dew Point Temp (t-1), Dew Point Temp(t-2), Heat Deg Days (t-1), and Heat Deg Days (t-2)	k = 12; Tol = 1e-7; DC = 0 Init = 0; τ = 0

DC = A criterion for 0-freq keeping the first mode, Inti= ω distribution type, Tol = Tolerance value for convergence of alternating direction method of multipliers, and τ =The time-step of the dual ascent

weights and hidden layer biases (Gan et al., 2021). The regression model for a given training sample can be described as follows:

$$\hat{y} = f(x) = h(x)\beta = H\beta \quad (6)$$

Where \hat{y} is the output of the ELM model, x is the input, H is the hidden layer output matrix, β is the connection weight between hidden and output layers. To enhance the stability and generalization of the model, a penalty coefficient C is added, which can be solved using the generalized inverse matrix theory.

$$\beta = H^T(I/C + HH^T)^{-1}y \quad (7)$$

Where I is a diagonal matrix, and y is the expected output.

The kernel mapping theory can be used to further improve ELM, making it the kernel extreme learning machine (KELM). This will make ELM less random and improve its ability to generalize (Liu et al., 2020). By defining Ω_{ELM} as a kernel matrix:

$$\Omega_{ELM} = HH^T : \Omega_{ELM_{ij}} = h(x_i)h(x_j) = U(x_i, x_j) \quad (8)$$

Which x_i is the input vector, then:

$$\hat{y} = f(x) = h(x)\beta = h(x)H^T(I/C + HH^T)^{-1}y \quad (9)$$

This paper selects a radial basis kernel $K(x_i, x_j)$ with a broad convergence domain and strong generalizability.

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \quad (10)$$

Where γ is the kernel parameter. It is notable to emphasize that the network training process of KELM requires the optimization of only two hyper-parameters, namely γ and \cdot . This feature makes KELM more convenient compared to other neural network techniques, such as BPNN, which entails the optimization of more than six hyper-parameters. Fig. 4 represents the topology of the KELM model.

2.5. Random vector functional link (RVFL)

Fig. 4 demonstrates the functional structure of the RVFL model, which comprises the input layer neurons that have direct connections to the output layer to avoid the operation of the back-propagation process

(Malik et al., 2023). Pao et al. (1994) invented the idea of RVFL, which is a feed-forward neural network with a single hidden layer. Among the input and hidden (enhancement node) layers, the weights and biases are produced randomly within a domain and kept fixed. The least-square method estimates the output weights during the training process (Katuwal and Suganthan, 2019). Across the world, this technique has received massive applications in the fields of science (Adnan et al., 2021; Ahmadianfar et al., 2019; Bisoi et al., 2019; Mostafa et al., 2023; Rasheed et al., 2020). The inputs of the output layer (D) are a concatenation of randomly, non-linearly transformed features (H) from the hidden layer and original input features (X) from direct links (i.e., $D = HX$). For example, If d = input features and N = number of hidden neurons, then input to output layer cover $d + N$ inputs (Shi et al., 2021). Mathematically, the objective function of the RVFL is defined as (Cheng et al., 2021; Katuwal and Suganthan, 2019):

$$O_{RVFL} = \min_{\beta} \|D\beta - Y\|^2 + \lambda_r \|\beta\|^2 \quad (11)$$

Here, $D = HX$ denotes the concatenation of hidden and original features, Y = target output, λ_r = regularization parameter. β = weight between hidden and output nodes. The direct solution of Eq. (4) results in overfitting. To avoid the overfitting, usually employed two methods i. e., Moore-Penrose pseudoinverse (the solution is given by $\beta = D^+Y$), and regularized least squares (or ridge regression), and the closed form of solution is given by (Suganthan and Katuwal, 2021):

$$\text{Primal Space} : \beta = (D^T D + \lambda_r I)^{-1} D^T Y \quad (12)$$

$$\text{Dual Space} : \beta = D^T (D D^T + \lambda_r I)^{-1} Y \quad (13)$$

The complexity of the inversion matrix is reduced by using the primal or dual solution, depending upon the length of the training samples and the total number of input features.

2.6. CNN-BiGRU

2.6.1. Convolutional Neural Network (CNN)

Feature extraction is a crucial phase for precise prediction by deep learning models. CNN is an effective network for extracting features that were developed in the last decade. The CNN network's fundamental modules for feature extraction are the convolutional and pooling layers (Wang et al., 2023). The 1D-CNN is mainly used for time series data and has a potent feature extraction capacity (Kiranyaz et al., 2021). The 1D-CNN's alternating convolutional and pooling layers can extract non-linear characteristics from unprocessed data, while the fully-connected layer finishes adaptive feature learning (Yao et al., 2021).

Fig. 5 depicts the basic framework of the CNN, which consists of an input layer, multiple convolution layers, multiple pooling layers, a fully connected layer, and an output layer, with the convolutional and pooling layers alternately connected. The CNN feature extraction module is comprised of the input layer, the convolutional layers, and the pooling layers. The output module comprises the connected and output layers (Guo et al., 2023).

CNN's core consists of the convolutional layer, in which the convolutional kernel C_j is utilized to extract features. The precise calculation formula is depicted in Eq. (12).

$$C_j = \text{act} \left(\sum_{i=1}^N A_i \otimes w_i + b_i \right) \quad (14)$$

where A_i is the input of the convolutional layer; \otimes is the convolutional operation; act is the activation function, and the activation function is chosen as a Rectified linear unit (Relu); w_i is the weight matrix and b_i is the bias deviation.

The application of a pooling layer allows data compression and the removal of irrelevant data. In this particular case, the max-pooling layer

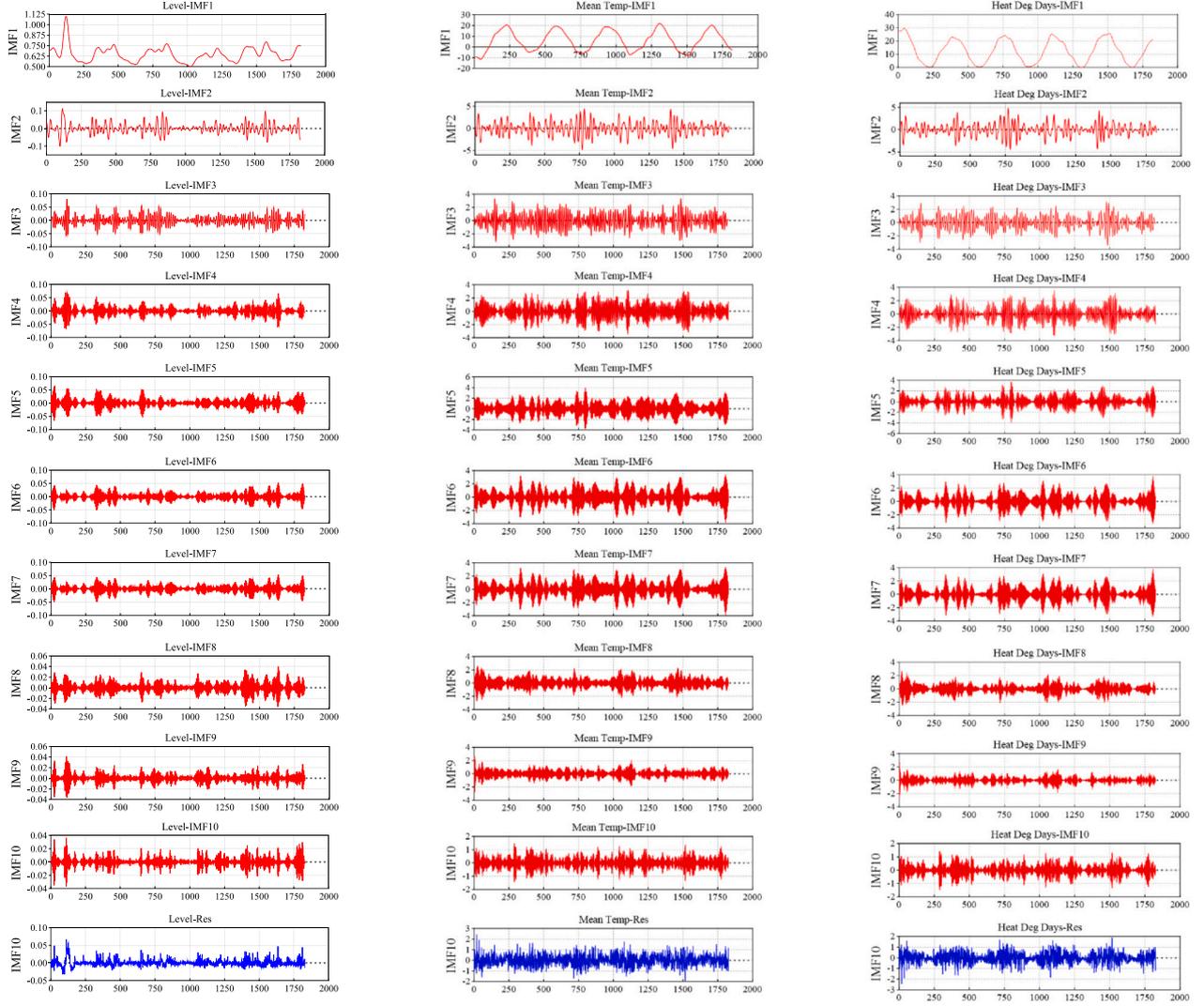


Fig. 9. MVMD-based decomposed sub-components (IMFs and residuals) related to all the climate signals implemented in forecasting of streamflow the Bear River ($Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$) using decomposition level ($k = 10$). (as a sample).

has been chosen. The aggregation of feature information extracted from the convolutional layer occurs in the fully connected layer, creating the final prediction data.

A critical characteristic of the Convolutional Neural Network (CNN) model is its training phase's simplicity, which can be achieved by using a smaller amount of weights compared to the fully connected architecture. In addition, it allows the extraction of key features (Guo et al., 2023).

2.6.2. Bidirectional Gated Recurrent Unit (Bi-GRU)

The Gated Recurrent Unit (GRU) was a sort of RNN that worked analogously to that of the Long Short-Term Memory (LSTM) network. To control gradient disappearance and gradient explosion in conventional RNN models, GRU was first implemented in 2014 (Cho et al., 2014). GRU can mitigate the gradient vanishing issue in RNN, decrease the number of LSTM parameters, and reduce the training period (Niu et al., 2022). The GRU algorithm may be described as follows: The memory cell of GRU primarily consists of two types of gates: reset gates and update gates. The linkage relationship formulas are:

$$z_t = f(W_z X_t + U_z h_{t-1}) \quad (15)$$

$$r_t = f(W_r X_t + U_r h_{t-1}) \quad (16)$$

$$\tilde{h}_t = \tanh(WX_t + U(r_t \odot h_{t-1})) \quad (17)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (18)$$

In this context, X_t represents a vector that indicates the t -th input of X . Meanwhile, r_t and z_t are used to denote the reset and update gates of the t -th memory cell, respectively. The \tilde{h}_t represents the hidden state of the t -th memory cell. The variable h_t represents the t -th output in the Gated Recurrent Unit (GRU). The final output is represented as h_{t-1} . The Hadamard product, denoted as \odot , is a mathematical operation. The activation function f is an example of a sigmoid function. The symbols W and U are used to denote parameters that have been obtained through the process of learning.

A unidirectional neural network configuration always produces output from front to back. Nonetheless, if the current outcome is connected to the conditions of previous and future moments, this hypothesis provides a potent method for extracting complex characteristics from time series. BiGRU, coincidentally, offers the same benefits for establishing this connection. BiGRU is a neural network model comprised of unidirectional and asymmetric GRUs. On the basis of the preceding GRU formulas, BiGRU can be described as:

Table 3

Optimal hyperparameters obtained by the Bayesian optimization scheme associated with the complementary models to -one and -three days ahead forecasting of the streamflow values.

Station	Model	Horizon	Hyperparameters
Bear River	MVMD-CNN-BiGRU	t + 1	CNN Layers = 2, filters = 64, kernel_size = [1,4], BiGRU Layers = 1, BiGRU neurons = 30, Optimizer: Adam, activation function = 'relu', epochs = 43, batch_size = 32, learning_rate = 0.000455, Dense = 100
		t + 3	CNN Layers = 2, filters = 64, kernel_size = [1,4], BiGRU Layers = 1, BiGRU neurons = 30, Optimizer: Adam, activation function = 'relu', epochs = 1000, batch_size = 32, learning_rate = 0.0015, Dense = 100
		t + 1	num_nodes: 20, regular_para = 0.2, weight_random_range = [-1,1], bias_random_range = [0,1], num_nodes: 50, regular_para = 0.9,
		t + 3	weight_random_range = [-1, 1], bias_random_range = [0, 1], Regularization_coefficient = 3E+05,
		t + 1	kernel_parameter =7000
		t + 3	Regularization_coefficient = 9E+06, kernel_parameter =9E+05
	MVMD-CNN-BiRNN	t + 1	CNN Layers = 2, filters = 64, kernel_size = [1,4], BiRNN Layers = 1, BiRNN neurons = 30, Optimizer: Adam, activation function = 'relu', epochs = 200, batch_size = 32, learning_rate = 0.0002, Dense = 100
		t + 3	CNN Layers = 2, filters = 64, kernel_size = [1,4], BiRNN Layers = 1, BiRNN neurons = 30, Optimizer: Adam, activation function = 'relu', epochs = 200, batch_size = 32, learning_rate = 0.0002, Dense = 100
		t + 1	CNN Layers = 2, filters = 64, kernel_size = [1,4], BiGRU Layers = 1, BiGRU neurons = 30, Optimizer: Adam, activation function = 'relu', epochs = 40, batch_size = 32, learning_rate = 0.000455, Dense = 100
		t + 3	CNN Layers = 2, filters = 64, kernel_size = [1,4], BiGRU Layers = 1, BiGRU neurons = 30, Optimizer: Adam, activation function = 'relu', epochs = 1000, batch_size = 32, learning_rate = 0.0015, Dense = 100
		t + 1	num_nodes: 20, regular_para = 0.05, weight_random_range = [-1, 1], bias_random_range = [0, 1], num_nodes: 30, regular_para = 2,
		t + 3	weight_random_range = [-1, 1], bias_random_range = [0, 1], Regularization_coefficient = 9E+05, kernel_parameter =9E+04
Winter River	MVMD-CNN-BiGRU	t + 1	Regularization_coefficient = 1E+06, kernel_parameter =9E+04
		t + 3	kernel_parameter =9E+04
		t + 1	CNN Layers = 2, filters = 64, kernel_size = [1,4], BiRNN Layers = 1, BiRNN neurons = 30, Optimizer: Adam, activation function = 'relu', epochs = 200, batch_size = 32, learning_rate = 0.0001, Dense = 100
	MVMD-CNN-BiRNN	t + 1	CNN Layers = 2, filters = 64, kernel_size = [1,4], BiRNN Layers = 1, BiRNN neurons = 30, Optimizer: Adam, activation function = 'relu', epochs = 200, batch_size = 32, learning_rate = 0.0001, Dense = 100
		t + 3	30, Optimizer: Adam, activation function = 'relu', epochs = 200, batch_size = 32, learning_rate = 0.0001, Dense = 100
		t + 3	30, Optimizer: Adam, activation function = 'relu', epochs = 200, batch_size = 32, learning_rate = 0.0001, Dense = 100

$$BiGRU(X) = function(\overrightarrow{GRU}(X), \overleftarrow{GRU}(X)) \quad (19)$$

where $\overrightarrow{GRU}(X)$ and $\overleftarrow{GRU}(X)$ are the outputs of GRUs that gather information in different ways, and function is a process like sum, concatenation, average, etc. Fig. 5 illustrates the structure of the GRU cell and BiGRU model.

2.7. CNN-BiRNN

2.7.1. Bidirectional Recurrent Neural Network (Bi RNN)

An RNN is a type of ANN in which the links between the nodes make a directed graph (where the direction shows the order in time), and the nodes have their memory. An RNN is made by stacking multiple layers on top of each other and putting a chain of values into each RNN cell. The way the RNN layers are set up and the RNN cells are built is essential to how well classification or prediction works. The standard UniRNN layer contains only a forward RNN, ignoring the extraction of features along the reverse time direction. In contrast, a BiRNN layer extracts features using forward and reverse RNNs. The outputs of a BiRNN are then combined using a linear transformation, which can be expressed as

$$y_t = o\left(\overrightarrow{h_t}, \overleftarrow{h_t}\right) = W\left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] + b \quad (20)$$

where, W is the linear transformation weight, $\overrightarrow{h_t}$ is historical information in the positive time direction at time point t , b is linear transformation bias and $\overleftarrow{h_t}$ is historical information in the negative time direction at time point t .

A Bidirectional Recurrent Neural Network (BiRNN) usually performs better than a Unidirectional Recurrent Neural Network (UniRNN) because the BiRNN includes a backward RNN that extracts extra features. Fig. 5 shows the structure of the BiRNN model.

2.8. Technique for order preference by similarity to ideal solution (TOPSIS)

TOPSIS is an analytical multi-criteria decision-making technique that prefers alternatives closest to the positive ideal solution and farthest from the negative perfect solution. TOPSIS is applied to numerical datasets with known criterion importance weights. Furthermore, ranking results are obtained according to the importance weights of the defined criteria. TOPSIS is widely used because its concept is clear and straightforward to understand. It can also measure the relative performance of selected options in a fundamental numerical framework. The mathematical framework of the TOPSIS technique includes the following steps (Ozsahin et al., 2021).

Step 1, Create a decision matrix and determine the importance weights of the criteria. The decision matrix ($X = X_{ij}$) and a weighting vector $W = [w_1, w_2, \dots, w_n]$ are selected, where:

$$X_{ij} \in R, W_j \in R \text{ and } w_1 + \dots + w_n = 1. \quad (21)$$

Step 2, Calculates the normalized values of the decision matrix (n_{ij}) with the eq. (2):

$$n_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^m X_{ij}^2}} \quad (22)$$

Step 3, Computes the weighted normalized values of the decision matrix (v_{ij}) as:

$$v_{ij} = w_j n_{ij} \quad (23)$$

where $i = 1, \dots, m; j = 1, \dots, n$. And w_j is the weight of the j^{th} criteria where $\sum_{j=1}^n w_j = 1$

Step 4, Determine the Positive and negative ideal solutions by identifying the Best Performance (PIS) and Worst Performance (NIS) for each criterion.

The PIS A^+ is calculated with the following equation:

$$A^+ = (v_1^+, v_2^+, \dots, v_n^+) = [[\max_i v_{ij} | j \in I], [\min_i v_{ij} | j \in J]] \quad (24)$$

The NIS A^- is calculated with the equation below:

$$A^- = (v_1^-, v_2^-, \dots, v_n^-) = [[\min_i v_{ij} | j \in I], [\max_i v_{ij} | j \in J]] \quad (25)$$

Table 4

Goodness-of-fit metric to statistically assess the hybrid and standalone models to multi-temporal forecast of the streamflow in Bear River.

	Model	Phase	R	RMSE	MAPE	NSE	U _{95%}	Reliability
$Q_{flow}(t + 1)$	MVMD-CNN-BiGRU	Training	0.985	0.072	19.399	0.970	0.200	67.659
		Testing	0.960	0.098	19.934	0.922	0.272	65.082
	MVMD-RVFL	Training	0.960	0.117	38.599	0.922	0.323	50.117
		Testing	0.949	0.113	29.987	0.896	0.313	49.909
	MVMD-KELM	Training	0.972	0.100	30.743	0.944	0.276	53.171
		Testing	0.930	0.130	30.226	0.863	0.360	46.618
	MVMD-CNN-BiRNN	Training	0.999	0.024	8.328	0.997	0.064	90.603
		Testing	0.955	0.105	20.875	0.911	0.291	59.963
	CNN-BiGRU	Training	0.921	0.164	32.428	0.846	0.455	49.804
		Testing	0.698	0.253	33.691	0.483	0.699	43.693
	RVFL	Training	0.865	0.210	39.152	0.748	0.583	41.973
		Testing	0.662	0.266	38.229	0.429	0.734	37.843
	KELM	Training	0.843	0.225	40.275	0.711	0.625	37.745
		Testing	0.668	0.263	37.304	0.439	0.729	37.477
CNN-BiRNN	Training	0.873	0.205	38.940	0.762	0.567	40.955	
	Testing	0.689	0.256	33.876	0.468	0.709	43.693	
$Q_{flow}(t + 3)$	MVMD-CNN-BiGRU	Training	0.998	0.010	2.903	0.999	0.026	99.922
		Testing	0.933	0.130	22.129	0.863	0.359	58.425
	MVMD-RVFL	Training	0.952	0.128	43.252	0.906	0.355	44.044
		Testing	0.930	0.130	33.073	0.863	0.360	45.238
	MVMD-KELM	Training	0.964	0.112	36.853	0.929	0.310	47.414
		Testing	0.928	0.133	32.872	0.857	0.366	45.788
	MVMD-CNN-BiRNN	Training	0.997	0.024	7.309	0.997	0.063	92.476
		Testing	0.929	0.141	26.103	0.839	0.387	54.029
	CNN-BiGRU	Training	0.792	0.256	48.401	0.628	0.709	33.307
		Testing	0.603	0.284	37.438	0.349	0.784	38.278
	RVFL	Training	0.826	0.236	47.625	0.682	0.655	34.404
		Testing	0.594	0.287	37.775	0.333	0.792	38.278
	KELM	Training	0.783	0.261	47.406	0.612	0.723	29.545
		Testing	0.591	0.287	40.347	0.333	0.795	30.952
CNN-BiRNN	Training	0.801	0.251	47.448	0.641	0.696	33.777	
	Testing	0.598	0.285	36.961	0.342	0.789	38.828	

Table 5

Goodness-of-fit metric to statistically assess the hybrid and standalone models to multi-temporal forecast of the streamflow in Winter River.

	Model	Phase	R	RMSE	MAPE	NSE	U _{95%}	Reliability
$Q_{flow}(t + 1)$	MVMD-CNN-BiGRU	Training	0.985	0.179	49.766	0.962	0.467	35.474
		Testing	0.955	0.247	42.629	0.853	0.687	34.004
	MVMD-RVFL	Training	0.918	0.364	77.362	0.842	1.010	18.481
		Testing	0.944	0.257	42.967	0.871	0.692	42.962
	MVMD-KELM	Training	0.904	0.395	82.576	0.815	1.094	18.324
		Testing	0.928	0.289	42.018	0.837	0.772	36.197
	MVMD-CNN-BiRNN	Training	0.998	0.072	20.441	0.994	0.175	64.135
		Testing	0.943	0.289	41.312	0.837	0.752	39.854
	CNN-BiGRU	Training	0.903	0.398	58.898	0.812	1.099	25.764
		Testing	0.811	0.432	37.402	0.636	1.190	41.316
	RVFL	Training	0.844	0.492	68.399	0.712	1.365	19.890
		Testing	0.783	0.497	51.189	0.520	1.346	30.347
	KELM	Training	0.845	0.491	65.828	0.713	1.362	18.324
		Testing	0.803	0.459	43.394	0.590	1.254	32.358
CNN-BiRNN	Training	0.895	0.410	52.794	0.800	1.135	26.782	
	Testing	0.815	0.429	31.413	0.641	1.183	53.748	
$Q_{flow}(t + 3)$	MVMD-CNN-BiGRU	Training	0.998	0.022	4.034	0.999	0.054	99.216
		Testing	0.924	0.330	35.726	0.788	0.889	39.377
	MVMD-RVFL	Training	0.907	0.386	80.418	0.823	1.071	19.279
		Testing	0.921	0.296	45.012	0.830	0.801	36.630
	MVMD-KELM	Training	0.904	0.397	69.869	0.813	1.100	22.962
		Testing	0.887	0.387	59.413	0.709	0.999	29.853
	MVMD-CNN-BiRNN	Training	0.998	0.070	14.818	0.994	0.187	76.881
		Testing	0.911	0.320	35.165	0.801	0.875	39.560
	CNN-BiGRU	Training	0.778	0.619	125.661	0.544	1.665	15.897
		Testing	0.545	0.706	86.229	0.028	1.877	11.883
	RVFL	Training	0.623	0.717	82.883	0.388	1.989	18.417
		Testing	0.509	0.654	53.859	0.168	1.804	23.443
	KELM	Training	0.682	0.671	77.888	0.464	1.861	23.276
		Testing	0.525	0.639	51.908	0.206	1.768	25.275
CNN-BiRNN	Training	0.839	0.511	70.254	0.689	1.416	24.843	
	Testing	0.496	0.629	42.313	0.229	1.736	36.081	

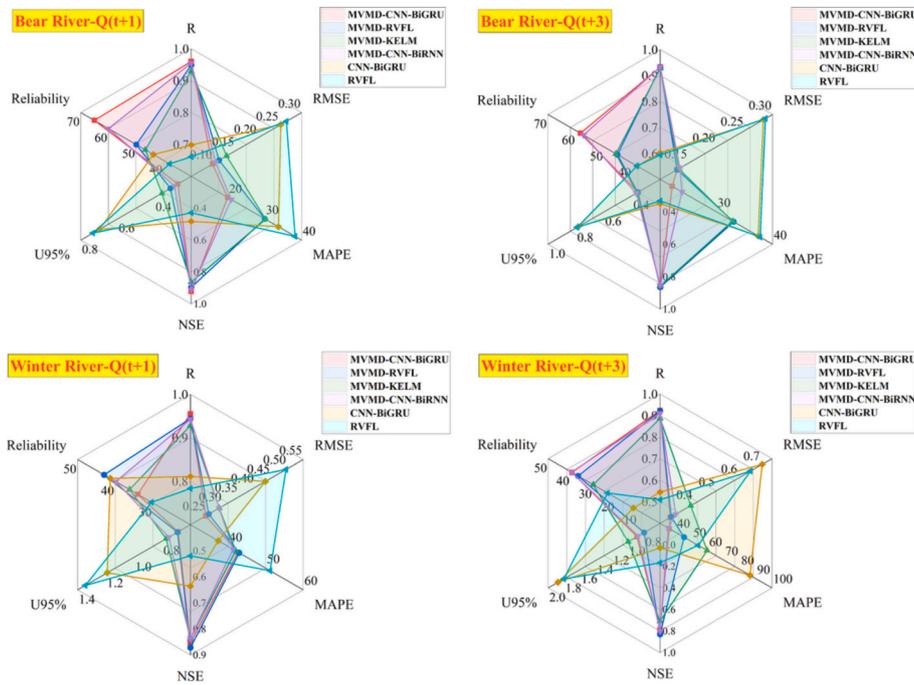


Fig. 10. Spider plots of statistical performance indices to validate the complementary and standalone frameworks aiming the multi-temporal forecasting of streamflow in Bear and Winter Rivers, PEI of Canada.

Where I defined with the benefit criteria, and J determined with the cost criteria,

$$i = 1, \dots, m; j = 1, \dots, n.$$

Step 5, Obtain the separation values of the alternatives and PI and the separation values of the alternatives and NIS. The n-dimensional Euclidean theory is used as the distance metric in this step. Every alternative from the PIS is separated based on formula (6), and every alternative from the NIS is separated according to formula (7) (Ozsahin et al., 2021).

$$d_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2}, i = 1, 2, \dots, m \quad (26)$$

$$d_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2}, i = 1, 2, \dots, m \quad (27)$$

Step 6, Compute the relative closeness to the PIS. The relative closeness of the i^{th} alternative A_i concerning A^+ is as the following equation.;

$$R_i = \frac{d_i^-}{d_i^- + d_i^+} \quad 0 \leq R_i \leq 1, i = 1, 2, \dots, m \quad (28)$$

In the TOPSIS technique, the alternatives are ranked based on their nearness to the PIS. So, the alternative with a higher R_i is the closest alternative to the PIS in eq. (8). Finally, the ranking results can be obtained based on this (Ozsahin et al., 2021).

2.9. Statistical criteria

This section introduces six expert statistical criteria to evaluate the merit of the forecasting model mathematically. The aforementioned metrics include the coefficient (R), root mean square error (RMSE), Kling-Gupta efficiency (KGE) (Gupta et al., 2009), Reliability, Squared Chi-square Distance (SquD), and mean absolute percentage error (MAPE). The following are the mathematical formulations of metrics (Jamei et al., 2021):

$$R = \frac{\sum_{i=1}^N (Q_{obs,i} - \overline{Q_{obs}}) (Q_{for,i} - \overline{Q_{for}})}{\sqrt{\sum_{i=1}^N (Q_{obs,i} - \overline{Q_{obs}})^2 \sum_{i=1}^N (Q_{for,i} - \overline{Q_{for}})^2}} \quad (29)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_{obs,i} - Q_{for,i})^2} \quad (30)$$

$$KGE = 1 - \sqrt{(R - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (31)$$

$$Reliability = \frac{\sum_{i=1}^N K_i}{N} \times 100\% \quad (32)$$

$$K_i = \begin{cases} 1, & \text{if } (RAE_i \leq \delta) \\ 0 & \text{else} \end{cases} \quad (33)$$

$$RAE_i = \frac{|Q_{obs,i} - Q_{for,i}|}{Q_{obs,i}} \times 100\%, RAE \geq 0 \quad (34)$$

$$SquD = \sum_{i=1}^N \frac{(Q_{obs,i} - Q_{for,i})^2}{Q_{obs,i} + Q_{for,i}} \quad (35)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Q_{obs,i} - Q_{for,i}}{Q_{obs,i}} \right| \times 100 \quad (36)$$

where $Q_{obs,i}$ and $Q_{for,i}$ are the values for the observed and forecasted Q values, respectively. $\overline{Q_{obs}}$ and $\overline{Q_{for}}$ are averaged values of observed and forecasted streamflow values. The variable K_i is the number of time intervals where the qualified forecast's threshold value (δ) equals or exceeds the RAE value. According to Chinese law, the δ value has been set at 20%.

3. Model configuration

As mentioned above, in this study, a new TOPSIS-based intelligent

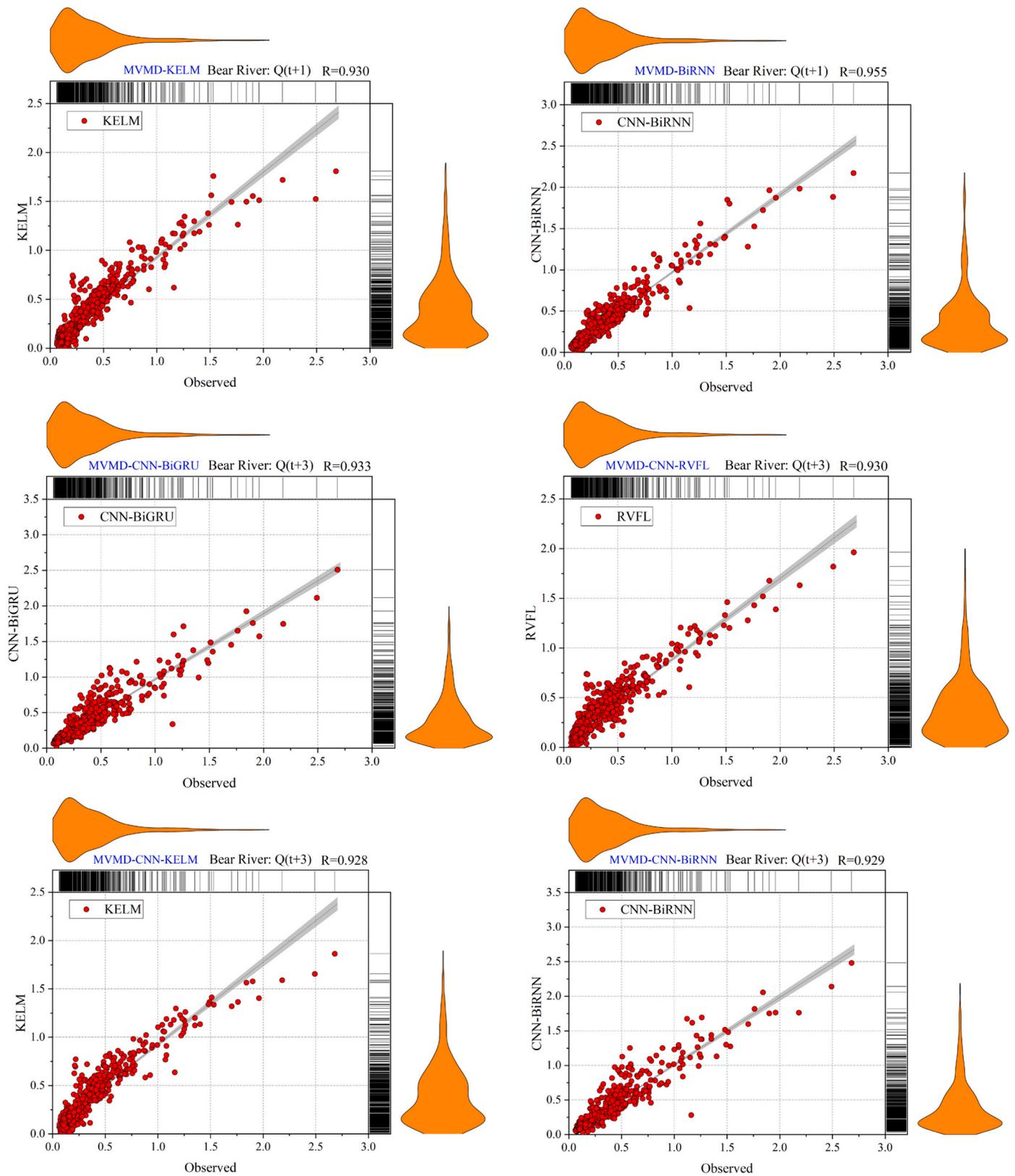


Fig. 11. Comparison between observational and computational values of streamflow over the testing period using the complementary frameworks aiming the multi-temporal forecasting of streamflow in Bear River, PEI of Canada.

system is designed based on the hydro-meteorological variables, including water level (Level), mean temperature (Mean Temp), heat degree days (Heat Deg Days), total precipitation (Total Precip), dew

point temperature (Dew Point Temp), and relative humidity (Rel Hum) which including three segments of Boruta-CART feature selection, MVMD decomposition, and CNN-BiGRU techniques to forecast the daily

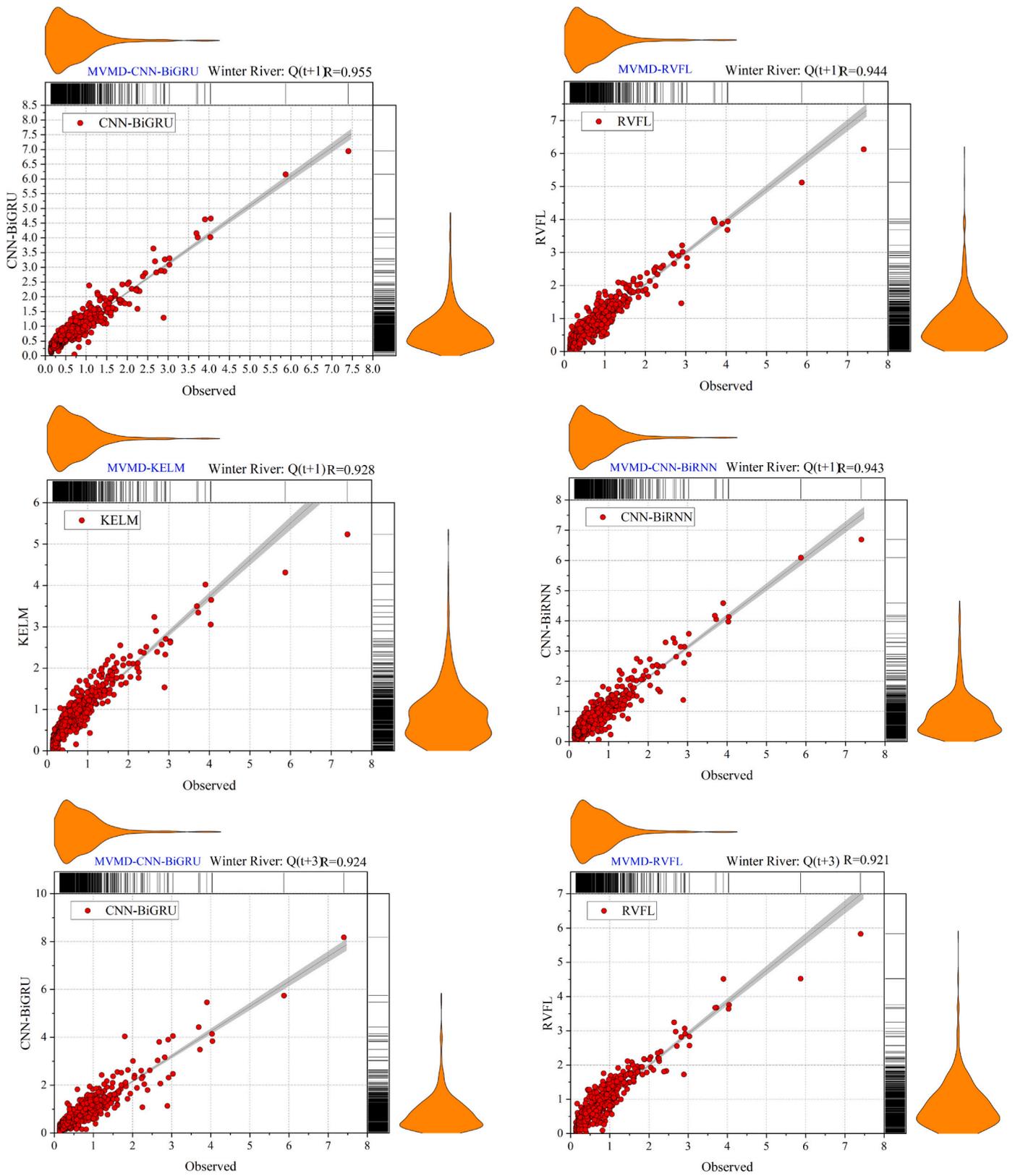


Fig. 11. (continued).

streamflow (Q_{flow}) of two Rivers in the PEI province of Canada. The provided multivariate-multi-temporal framework can specify the significant input signals and describe the sub-sequences role during the training phase. A database of time series related to hydro-climatological drivers from 2015 to 2020 has been used to construct the model.

In this study, the CNN-BiGRU, CNN-BiRNN, RVFL, and KELM ML models have been developed as predictive models. The KELM models and MVMD decomposition techniques were executed in MATLAB 2020a environment. The CNN-BiGRU, CNN-BiRNN, RVFL, and Boruta-CART schemes are provided using the Keras, Tensor Flow, RVFL, Sikit-learn,

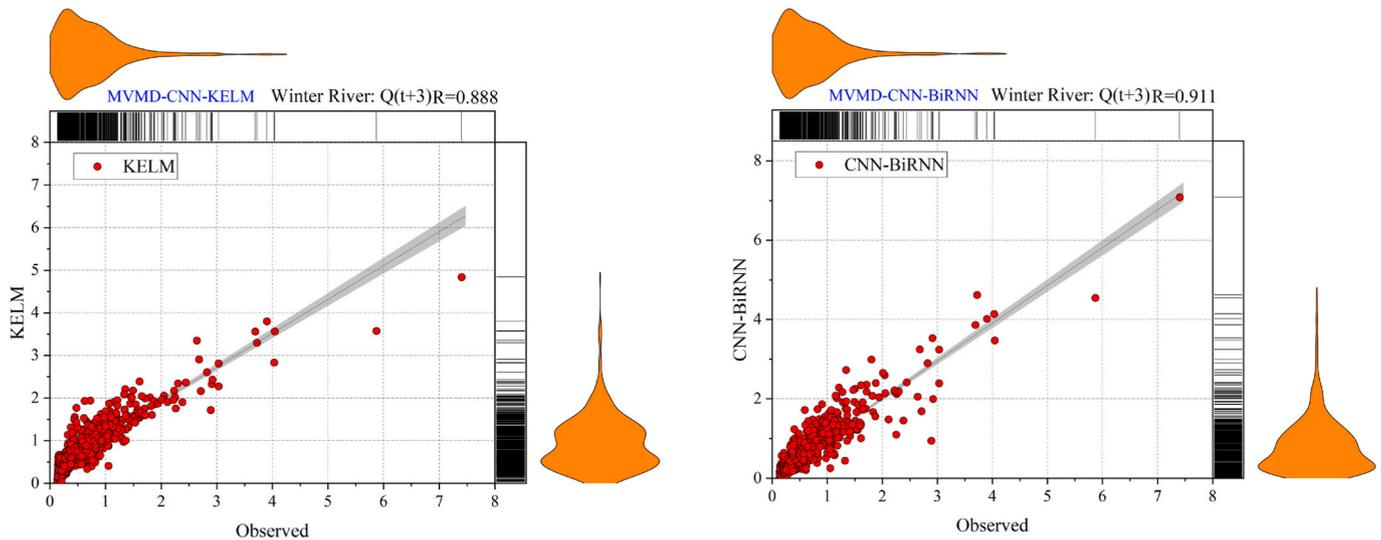


Fig. 11. (continued).

and Boruta open-source libraries (Pedregosa et al., 2011) in Python environment. All models were executed on a system with CPU Intel® Corei7@ 3.60 GHz processor specifications and 128 GB of RAM. Fig. 6 represents the workflow developed for the multi-temporal forecasting of streamflow values in two rivers of PEI, Canada. The development of the streamflow multi-temporal forecasting framework required the following processes:

3.1. Effective signal extraction using Boruta-CART

In this step, antecedent data related to the last two days (the initial two lags) of seven hydro-meteorological variables (as input signals) were defined for the Bear and Winter stations. The dataset from 2015 to 2020 is divided into two sub-intervals. The initial 75% of the data time series is set aside for the training stage, and the remaining 25% is dedicated to the testing stage. To preserve the sequence of time series, the hold-out technique was employed for the data segmentation (Jamei et al., 2023a). Subsequently, the previously mentioned information is considered input features to provide the Boruta-CART feature selection based on the Z-score (factor of feature Importance) to determine the significant signals for daily streamflow forecasting. Fig. 7 exhibits the original input and target time series. The chosen essential signals with a Z-score more effective than the Max_Shadow benchmark score is shown in green colour. The tentative signals are indicated in yellow, and the rejected signals are shown in red. Fig. 8 shows the Z-score values associated with each antecedent information (t-1) and (t-2) of the input signals for both understudy stations. All the hydro-meteorological lagged-time sub-sequences within Fig. 8 are introduced in the nomenclature and Table 2. According to Fig. 8, At the Bear River station, the (Level (t-1), Level (t-2), Mean Temp (t-1), and Mean Temp (t-2)) and (Level (t-1), Level (t-2), Mean Temp (t-1), and Mean Temp (t-2), and Heat Deg Days (t-2)) have the greatest impact on the streamflow (Q_{flow}) for horizon (t + 1) and Horizon (t + 3), respectively. At the winter River station for horizon (t + 1), the (Level (t-1), Level (t-2), Dew Point Temp (t-1), Dew Point Temp(t-2), and Total Precip (t-1)), and for horizon (t + 3), the (Level (t-1), Level (t-2), Dew Point Temp (t-1), Dew Point Temp(t-2), Heat Deg Days (t-1), and Heat Deg Days (t-2)) have the most significant influence on streamflow (see Table 2).

3.2. Signals decomposition via multivariate variational mode decomposition

In the second pre-processing step, the MVMD method simultaneously

decomposes the antecedent sub-components attained by Brouta-CART, consisting of IMFs and residuals for each input time series. One of the primary motivations for using the MVMD technique is the simultaneous decomposition of all the time series predictors to neglect the summing of all the individual forecasts (Malik et al., 2022). A trial-and-error procedure specified the optimal decomposition mode number for each horizon/case study. Table 2 reports the setting parameters of the MVMD (i. e., τ , DC, Init, and k) for each target. At the Bear River station, as a sample, the decomposed components (IMFs and residual, k = 10) for all the hydro-meteorological variables implemented in the construction of the forecasting models are represented in Fig. 9. According to Table 2, the optimal mode numbers of the Bear River for (t + 1) and (t + 3) horizons were 10 and 12, respectively, whereas in the Winter River, the optimal ones were 11 and 12, respectively. Also, according to the lags number, significant lagged-time features, and decomposition mode numbers, all the sub-components to feed the model in Bear and Winter Rivers are equivalent to (t + 1 | 40; t + 3 | 60) and (t + 1 | 55; t + 3 | 72), respectively.

3.3. Machine learning tuning

The last step in configuring the model involves feeding in predictors that the advanced deep learning algorithm feeds by the decomposed components resulting from the previous steps. The primary model is CNN-BiGRU with three comparative ML methods (KELM, RVFL and CNN-BiRNN), which provide four complementary models including MVMD-CNN-BiGRU, MVMD-RVFL, MVMD-KELM and MVMD-CNN-BiRNN. These models are used to create the multi-temporal forecasting model of daily streamflow (Q_{flow}). Setting hyperparameters and their structural architecture is the most important aspect of executing ML-based predictive models (Jamei et al., 2023b). Based on recent research, the main approaches to tuning parameters are algorithms of metaheuristic optimization, schemes of cross-validation schemes (Nested/rolling basis cross-validation) (Huyghues-Beaufond et al., 2020), Bayesian optimization, and more direct methods (such as random search, grid search, and trial and error schemes).

The Bayesian Optimization program was used in this study to optimize hyperparameters related to ML algorithms. Compared to grid search or random search methods, the Bayesian

optimization method can accelerate and improve the optimization of hyperparameters. The Bayesian Optimization package uses Bayesian optimization to explore the hyperparameter space by selecting the next set of hyperparameters to estimate based on past evaluations (Yoshida

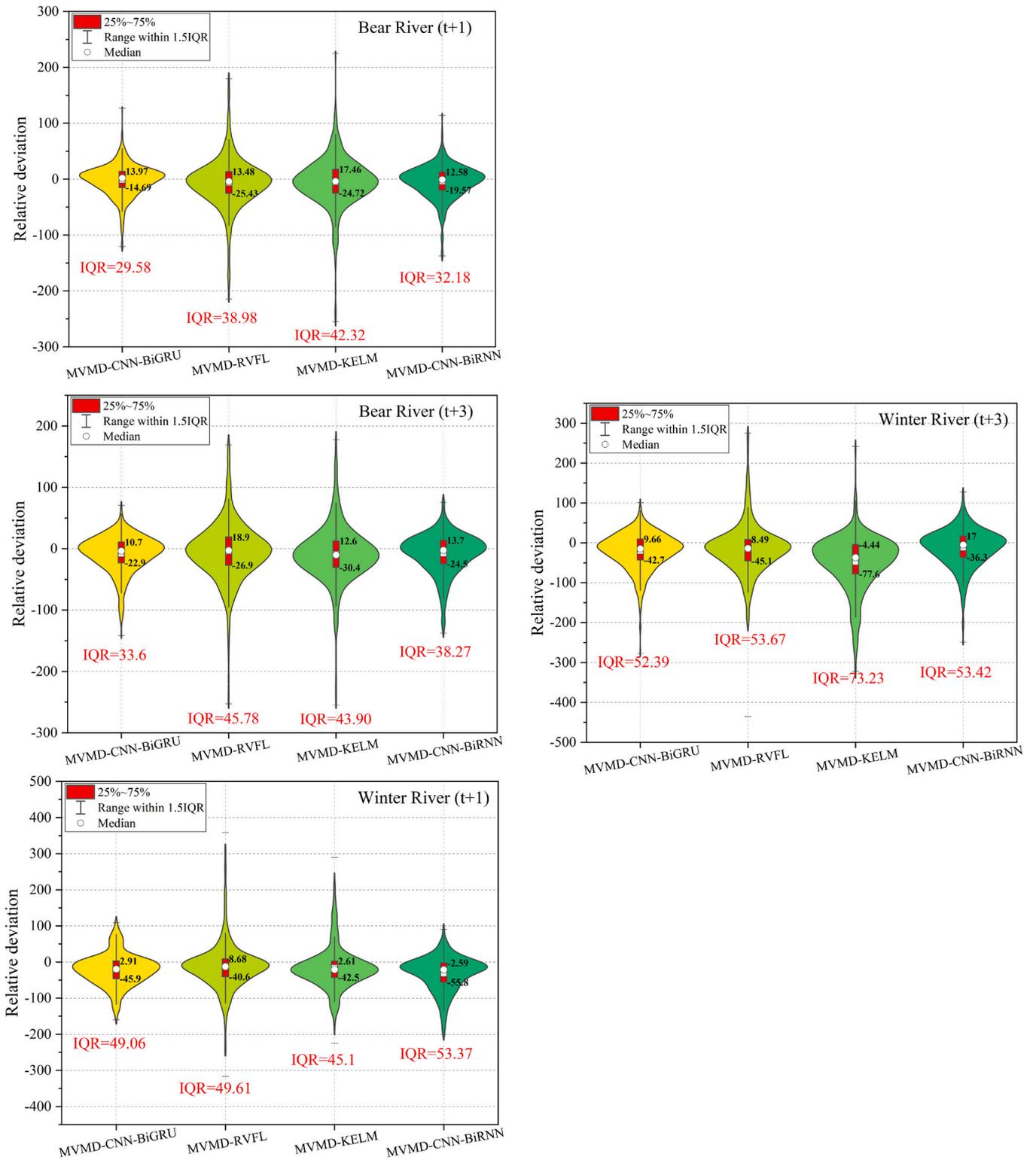


Fig. 12. Diagnostic analysis using the relative deviation in form of the violin plots using the decomposition-based complementary models over the testing period. IQR represents the interquartile range of relation deviation values.

et al., 2023). Table 3 represents the optimal hyperparameters related to the complementary models for -one and -three days ahead forecasting the streamflow in Bear and Winter Rivers of PEI. The important hyperparameters for CNN-BiGRU are listed in Table 3, including the CNN Layers, BiGRU Layers, kernel parameter, Adam, activation function, CNN filters, epochs, and learning_rate. It is worth mentioning that to prevent cumulative errors caused by the multiscale decomposition of the

decomposed signals and convergence improvement, all inputs and target parameters are normalized between 0 and 1 by the following formulation (Jamei et al., 2022a):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{37}$$

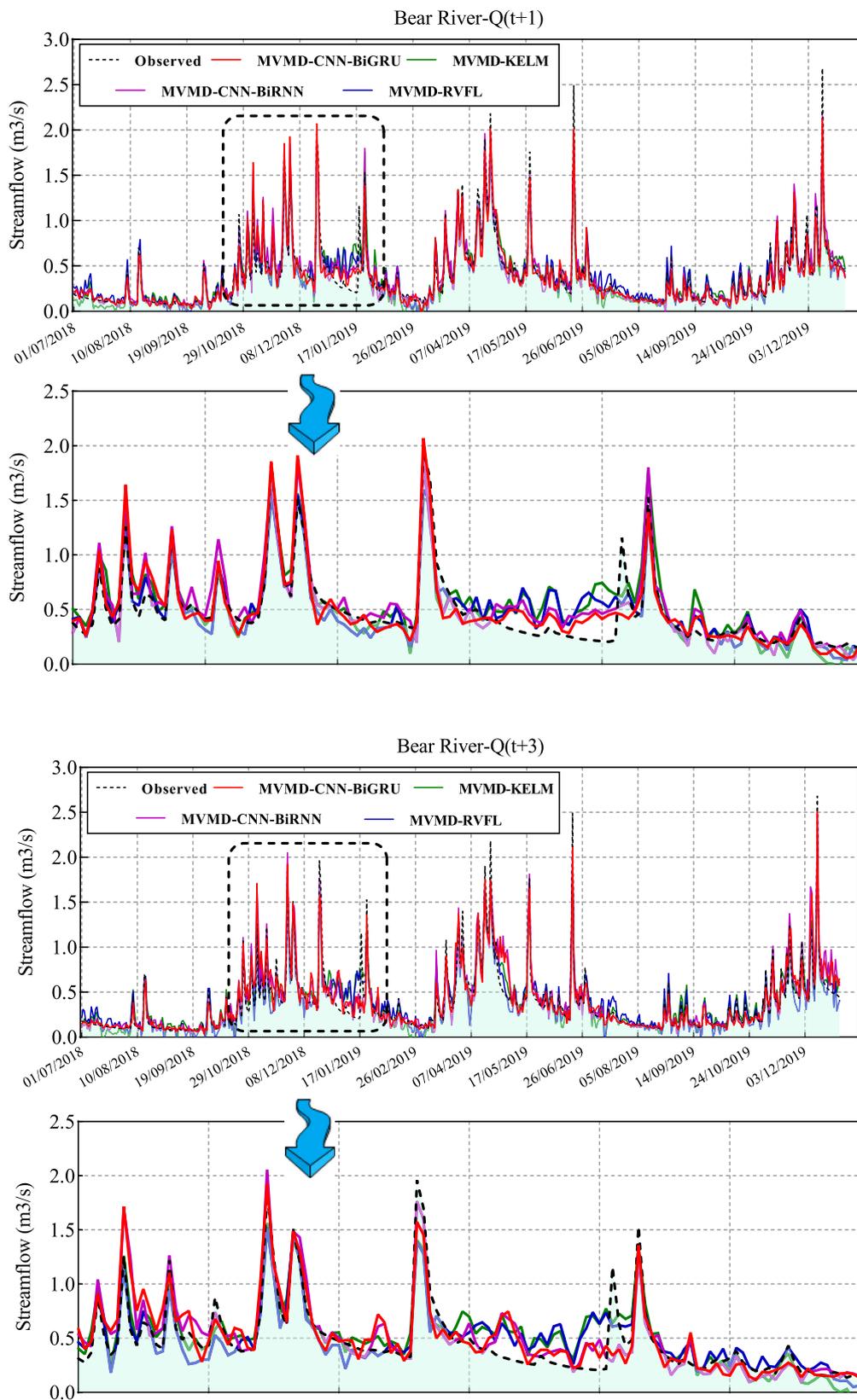


Fig. 13. Trend plot of observational and computational streamflow values over the testing period using the hybrid intelligent frameworks in two understudy stations.

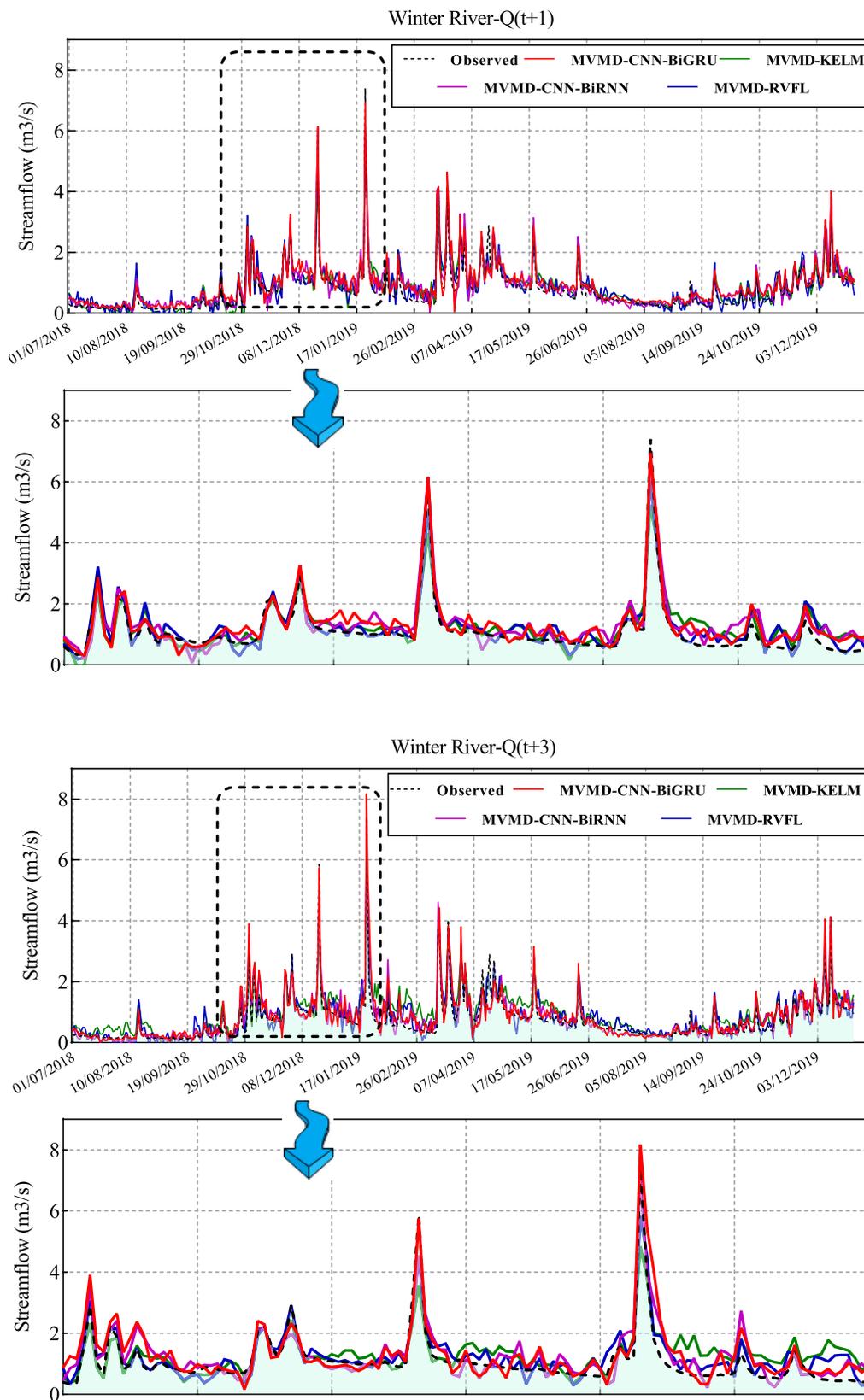


Fig. 13. (continued).

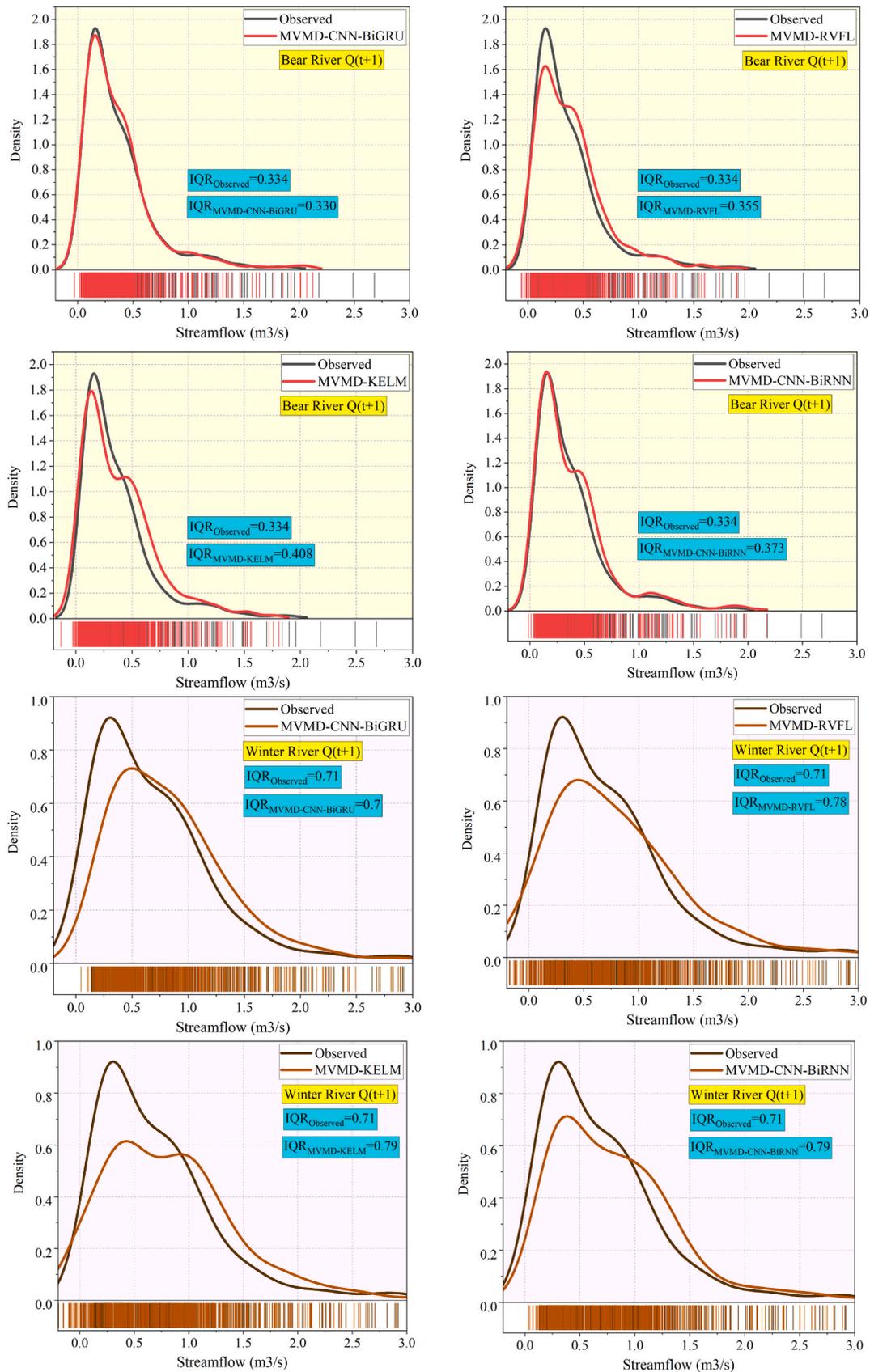


Fig. 14. Smooth kernel-based Rug histograms to assess the forecasting compatibility of each hybrid model vs comparison with the observational streamflow over the testing period (just for -one day ahead).

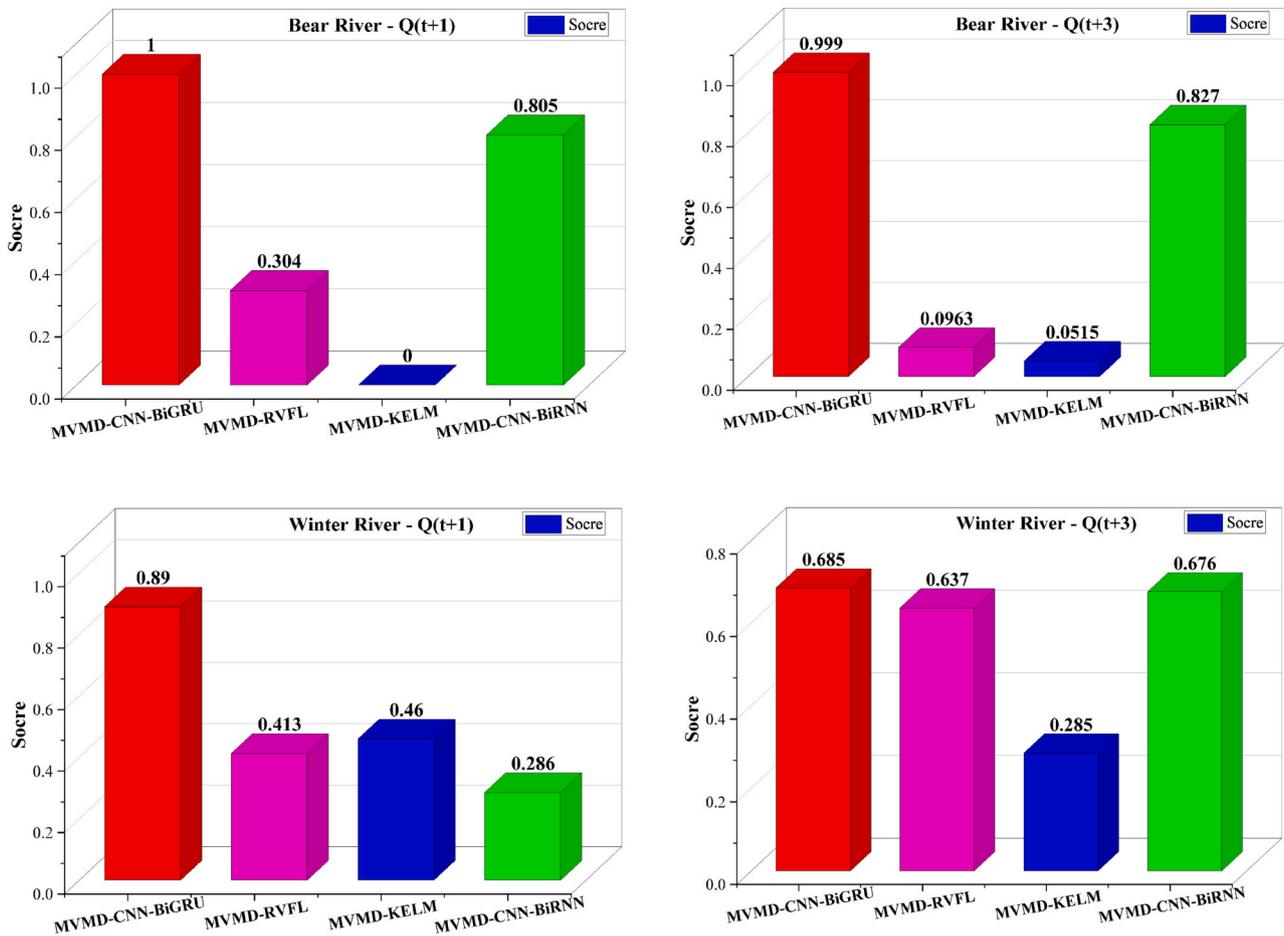


Fig. 15. Robustness assessment of the hybrid forecasting models over the testing period based on the TOPSIS score factor.

where X_{norm} denotes the normalized value of the X parameter; X is the original value; X_{min} and X_{max} are the maximum and minimum values of the original parameter (X), respectively.

4. Results assessment

The goodness-of-fit metrics to statistically assess the hybrid and standalone models to forecast multi-temporal streamflow in Bear River and Winter River Canada in training and testing periods in Tables 1 and 2, along with the diagnostic plots. The standalone CNN-BiGRU, RVFL, KELM, CNN-BiRNN and their hybrid versions MVMD-CNN-BiGRU, MVMD-RVFL, MVMD-KELM, and MVMD-CNN-BiRNN models were designed and constructed to forecast streamflow at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ multi-temporal forecasting horizons.

Table 4 illustrates the multi-temporal streamflow forecast for Bear River to assess the performance accuracy of the MVMD-CNN-BiGRU, MVMD-RVFL, MVMD-KELM, and MVMD-CNN-BiRNN, along with the CNN-BiGRU, RVFL, KELM, CNN-BiRNN models to at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$. The hybrid MVMD-CNN-BiGRU seemed to be the most precise model in terms of ($R = 0.985$, $RMSE = 0.072$, $MAPE = 19.399$, $NSE = 0.970$, $U_{95\%} = 0.200$, $Reliability = 67.659$)-training phase and ($R = 0.960$, $RMSE = 0.098$, $MAPE = 19.934$, $NSE = 0.922$, $U_{95\%} = 0.272$, $Reliability = 65.082$)-testing phase to forecast streamflow at $Q_{flow}(t + 1)$, following by MVMD-CNN-BiRNN, MVMD-RVFL, MVMD-KELM, CNN-BiGRU, CNN-BiRNN, RVFL, and KELM models. For forecasting streamflow at $Q_{flow}(t + 3)$, again MVMD-CNN-BiGRU model expressed better accuracy than other hybrid and standalone versions of the models. The MVMD-CNN-BiGRU report accurate values in the training and testing phases as compared to other models. For both multi-temporal

forecasting horizons, the MVMD-CNN-BiGRU model accomplished consistent accuracy to consolidate the better performance over the comparing model (Table 4) for Bear River. Additionally, it can be noted that the hybrid models overall revealed better accuracy than their standalone counterpart models.

The MVMD-CNN-BiGRU model also acquired the highest accuracy metrics to forecast multi-temporal streamflow at both forecast horizon $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ for Winter River, as revealed in Table 5, against comparing models. The MVMD-CNN-BiGRU model exhibited the highest values of $R = 0.985$, 0.955 ; $RMSE = 0.179$, 0.247 ; $MAPE = 49.766$, 42.629 ; $NSE = 0.962$, 0.853 ; $U_{95\%} = 0.467$, 0.687 ; and $Reliability = 35.474$ in both training and testing phases respectively to forecast streamflow at $Q_{flow}(t + 1)$. Similarly, the MVMD-CNN-BiGRU model accomplished the highest degree of accuracy using these goodness-of-fit metrics, i.e., R , $RMSE$, $MAPE$, NSE , $U_{95\%}$, and $Reliability$ to forecast streamflow for Bear River at $Q_{flow}(t + 3)$ as compared to MVMD-CNN-BiRNN, MVMD-RVFL, MVMD-KELM, CNN-BiGRU, CNN-BiRNN, RVFL, and KELM models (Table 5). Once again, the hybrid version of the models surpasses the standalone models in forecasting multi-temporal streamflow at both forecast horizons. But overall, the MVMD-CNN-BiGRU model displayed performance dominance over all other models compared to forecast multi-temporal streamflow for both stations.

Fig. 10 plots the R , $RMSE$, $MAPE$, NSE , $U_{95\%}$, and $Reliability$ metrics in terms of a spider diagram to assess the performance consistency of the constructed models. Based on these criteria for both stations at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$, the accuracy of the MVMD-CNN-BiGRU model against MVMD-CNN-BiRNN, MVMD-RVFL, MVMD-KELM, CNN-BiGRU, CNN-BiRNN, RVFL, and KELM evidenced to be outstanding. For

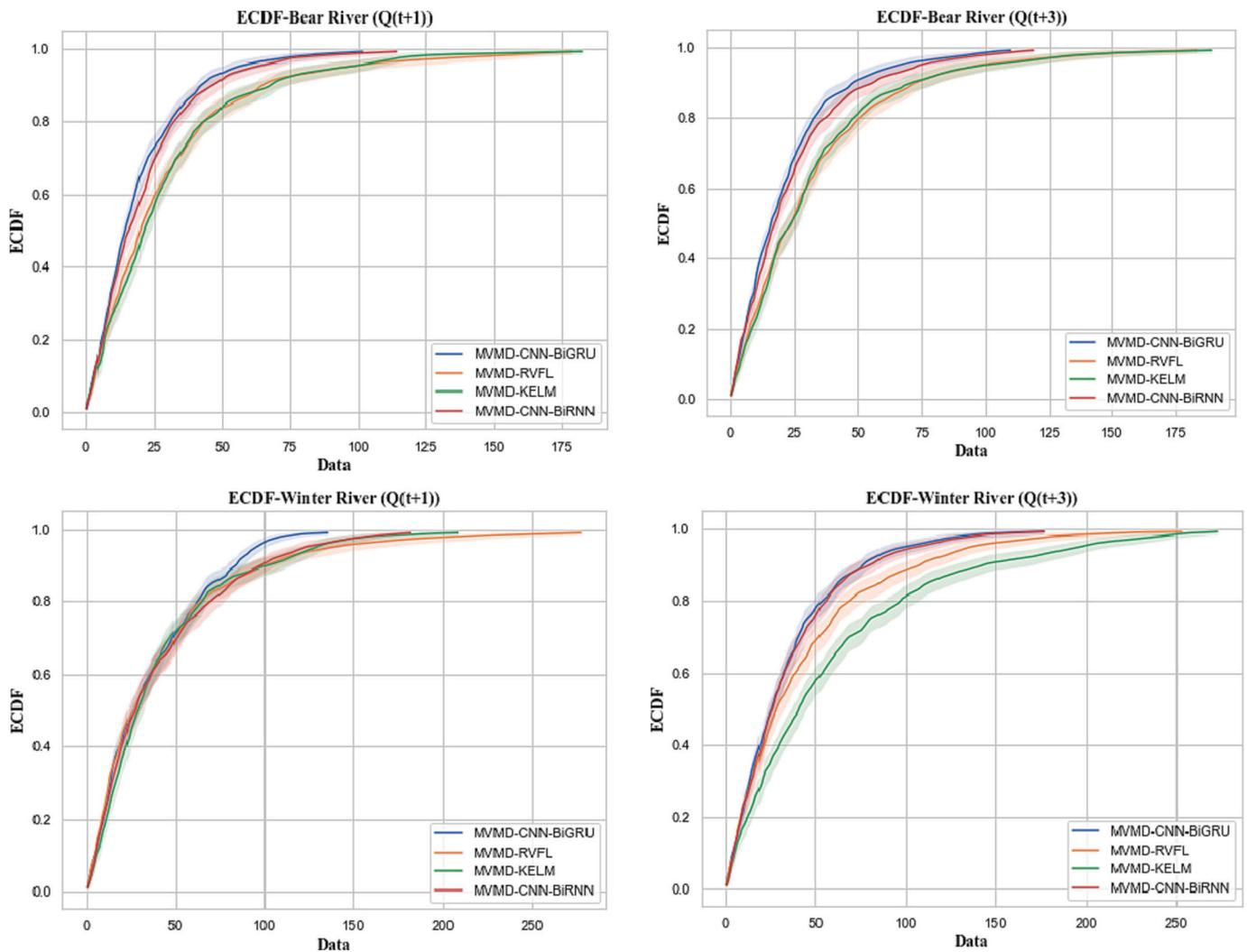


Fig. 16. The ECDF values relative deviation for all the hybrid frameworks to multi-temporal (-one and -three ahead horizons) forecast the streamflow in Bear and Winter Rivers during the testing period.

example, the webs of spider plots are more stretched towards the centre for RMSE and MAPE by acquiring lower magnitudes of these error metrics. This spiderweb spread outwards in the case of R, NSE, $U_{95\%}$, and Reliability in terms of higher values MVMD-CNN-BiGRU model at both Bear River and Winter River at $Q_{flow}(t + 1)$ against comparing models. Similarly, the MVMD-CNN-BiGRU appeared to be the most precise model in terms of spider plots to forecast streamflow at $Q_{flow}(t + 3)$ for both rivers. These spider plots in Fig. 10 are also confirmed in Table 5, which ensures the highest accuracy of the proposed MVMD-CNN-BiGRU model based on these assessment metrics.

Fig. 11 examines the comparison between observed and forecasted values of streamflow over the testing period using MVMD-CNN-BiGRU against MVMD-CNN-BiRNN, and MVMD-RVFL, MVMD-KELM models for Bear River and Winter River efficiency in terms of scatter plots. Moreover, the violin distribution in combination with R metrics was also incorporated in Fig. 11. The scatter plots provide a supplementary assessment of the forecasting ability between the observed and forecasted streamflow in both stations at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$. The MVMD-CNN-BiGRU model at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ to forecast streamflow displayed the highest precision with $R = 0.960$ and 0.933 , followed by MVMD-CNN-BiRNN, MVMD-RVFL, and MVMD-KELM models for Winter River. Also, the violin distributions of forecast streamflow were almost identical for the MVMD-CNN-BiGRU model compared to the observed streamflow at both forecast horizons. For Bear

River, the MVMD-CNN-BiGRU again presented better accuracy to forecast streamflow at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ compared to other models. The scatter diagrams in Fig. 11 confirmed that MVMD-CNN-BiGRU better forecasts multi-temporal streamflow for both stations.

In Fig. 12, the violin plots provide a diagnostic analysis using the relative deviation along with IQR values as well as the interquartile range using only the hybrid MVMD-CNN-BiGRU vs MVMD-CNN-BiRNN, MVMD-RVFL, and MVMD-KELM models to forecast multi-temporal streamflow for Bear and Winter Rivers. The boxplot (red colour) lies between 25% to 75%, the whiskers range within 1.51 IQR values, and the white circle shows the median. By observing, it is apparent that the MVMD-CNN-BiGRU displayed a more accurate relative deviation distribution with the lowermost IQR = 29.5, 49.1 $Q_{flow}(t + 1)$ and 33.6 and 52.32 $Q_{flow}(t + 3)$ to forecast streamflow at Bear and Winter River as compared to MVMD-CNN-BiRNN, MVMD-RVFL, and MVMD-KELM models. Thus, MVMD-CNN-BiGRU models accomplish accurate streamflow forecasting based on Violin plots for rivers at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$.

Fig. 13 demonstrates the comparisons Trend plot between the observed and forecasted streamflow values using MVMD-CNN-BiGRU vs. MVMD-CNN-BiRNN, MVMD-RVFL, and MVMD-KELM models at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ for Bear and Winter Rivers. This diagram delivers the trends of the observed and forecasted streamflow trends produced by these models from 01/07/2018 to 03/12/2019. For Bear

River, the MVMD-CNN-BiGRU model portrays similar trends concerning the observed streamflow, which can be further confirmed in the specific four months interval (i.e., 29/10/2018–26/02/2019) at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ as compared to the MVMD-CNN-BiRNN, MVMD-RVFL, and MVMD-KELM models. Similarly, the MVMD-CNN-BiGRU model again reported identical trends for Winter River to forecast multi-temporal streamflow. Consequently, the MVMD-CNN-BiGRU model delivers accurate forecasting for both rivers.

Fig. 14 displays the smooth kernel-based Rug histograms to assess the forecasting compatibility of MVMD-CNN-BiGRU vs MVMD-CNN-BiRNN, MVMD-RVFL, and MVMD-KELM models together with the IQR values to forecast streamflow at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ for both rivers. For Bear River, the MVMD-CNN-BiGRU is ranked the best and the most precise model generating identical density with IQR values to forecast streamflow, flowing by MVMD-RVFL, MVMD-CNN-BiRNN, and MVMD-KELM at both forecasting horizons. The MVMD-CNN-BiGRU is again the top model in terms of smooth kernel-based Rug histograms and IQR against comparing models to forecast streamflow at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ for Winter River. Thus, Fig. 14 confirmed that the MVMD-CNN-BiGRU model has better streamflow forecasting ability.

The bar graphs in Fig. 15 signify the TOPSIS score factor achieved by the MVMD-CNN-BiGRU vs. MVMD-CNN-BiRNN, MVMD-RVFL, and MVMD-KELM models to forecast multi-temporal streamflow at $Q_{flow}(t + 1)$, and $Q_{flow}(t + 3)$. It is visible that the MVMD-CNN-BiGRU model accomplished the highest TOPSIS score = 1.00 and 0.99 for Bear River and 0.89 and 0.685 for Winter River at both $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ compared to other benchmarking models (see; Fig. 15). The TOPSIS score bar graphs in Fig. 15 established that the MVMD-CNN-BiGRU exhibits higher accuracy to forecast streamflow at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ compared to other models. The TOPSIS score suggests that the forecasting of MVMD-CNN-BiGRU is better and more accurate for Bear River than the Winter River.

The forecasted streamflow of Bear River and Winter River are plotted in Fig. 16 using the empirical cumulative distribution function (ECDF) to depict a more tangible view of the models. For both Rivers, the ECDF of the MVMD-CNN-BiGRU showed a very close profile at both $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ forecasting horizons against MVMD-CNN-BiRNN, MVMD-RVFL, and MVMD-KELM models. Hence, Fig. 16 further determines that the MVMD-CNN-BiGRU is a better model for forecasting multi-temporal streamflow at Bear and Winter Rivers at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$.

5. Discussion

The goal of this work is to design a novel data decomposition-based multivariate-multi-temporal Boruta-CART integrated with CNN-BiGRU algorithm to forecast streamflow using hydro-meteorological drivers at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ for Bear and Winter rivers. The MVMD technique demarcates the hydro-meteorological drivers into IMFs-based signals, and then the most significant lags (i.e., IMFs signals) were identified by the Boruta-CART model at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$. The Boruta-CART-based selected signals were lastly employed into the CNN-BiGRU which resulted in the MVMD-CNN-BiGRU model to forecast streamflow. The MVMD helps to overcome the non-stationarity and non-linearity issues in the hydro-meteorological drivers whereas the Boruta-CART model plays a vital role in optimizing the forecasting accuracy through selecting the most influential signals. The comparison was performed against the standalone CNN-BiGRU, RVFL, KELM, CNN-BiRNN and their hybrid versions MVMD-CNN-BiGRU, MVMD-RVFL, MVMD-KELM, and MVMD-CNN-BiRNN models to forecast streamflow. The MVMD-CNN-BiGRU model reported the highest precision against all these benchmarks comparing models for both rivers using a set of goodness-of-fit metrics. But further recommendations are suggested here for future work.

According to the Boruta-CART FS outcome, it can simply explain the effectiveness of each input signal during the model for both case studies

and horizons. In other words, the interpretability outcomes related to each hydro-meteorological signal can be accurately described during the pre-processing stage. In this regard, confidently, it can be concluded that the lags of the water level parameter (especially Level (t-1)) in all the horizons and both case studies is recognized as the most influential feature due to the highest importance factor (Z-score > 12). Also, the antecedent information of the mean temperature in forecasting the streamflow horizons of the Bear River was identified as the second most important feature, whereas in the Winter River, the dew point temperature (Dew Point Temp) and heat degree days (Heat Deg Days) were the second and third most important features, respectively. It is noteworthy that the relative humidity (Rel Hum) has been selected for none of the target, and the total precipitation (Total Precip) just participated in forecasting the Q_{flow} for horizon (t + 1) in the Winter River. Thus, Rel Hum had the least important factor in forecasting the streamflow values.

Deep learning appeared to be very effective in accurate forecasts, however, there are a few limitations and restrictions exist which confine their capability and lack the aspect of the model's prediction explainability. Therefore, the combination of explainable AI models for example Local Interpretable Model-Agnostic Explanations (LIME) (Mishra et al., 2017) and Shapley Additive explanations (SHAP) (Shapley, 1953) with MVMD-CNN-BiGRU can be a possible emerging area for future research. Moreover, the physics-based models with MVMD-CNN-BiGRU can be considered another direction to enlighten the physical viewpoint. The underlying model uncertainties during the learning process can be handled by the integration of the Bayesian Model Averaging (Slougher et al., 2010) and bootstrapping (Tiwari and Chatterjee, 2011) methods with MVMD-CNN-BiGRU to improve the forecasting skills.

The outcomes based on the achieved forecasting accuracy established that the MVMD-CNN-BiGRU is an efficient streamflow forecasting model using hydro-meteorological drivers at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ as compared to other models. To enhance the scope of this work, the MVMD-CNN-BiGRU model can be adopted in other hydrological issues such as water level, flooding, and water resource management. Potentially, the MVMD-CNN-BiGRU model can be tested in renewable and sustainable energy, agriculture, and other climate change sectors.

6. Conclusion

A novel data decomposition-based multivariate-multi-temporal Boruta-CART integrated with CNN-BiGRU deep learning algorithm for daily streamflow forecasting was designed for the first time based on hydro-meteorological drivers. The main novelty encompasses the MVMD, Boruta-CART, and CNN-BiGRU into one topology to forecast monthly multi-temporal streamflow at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$ for Bear and Winter rivers. The first stage introduced the MVMD to decompose the hydro-meteorological drivers into signals. Next step, the most influential lags (i.e., signals) were determined by the Boruta-CART feature selection at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$. Finally, the selected signals were incorporated into the CNN-BiGRU to design the MVMD-CNN-BiGRU model to forecast streamflow at $Q_{flow}(t + 1)$ and $Q_{flow}(t + 3)$. Extensive results and analysis showed that the MVMD-CNN-BiGRU exhibited higher accuracy in forecasting streamflow when compared with the other models. The MVMD-CNN-BiGRU model outperformed the other models in forecasting the stream flow for Bear River [$R = 0.960$, $RMSE = 0.098$, $MAPE = 19.934$, $NSE = 0.922$, $U_{95\%} = 0.272$, $Reliability = 65.082$]-($t + 1$); [$R = 0.933$; $RMSE = 0.130$; $MAPE = 22.129$; $NSE = 0.863$, $U_{95\%} = 0.359$, $Reliability = 58.425$]-($t + 3$). Likewise, the MVMD-CNN-BiGRU model resulted in higher accuracy for streamflow forecasting at Winter River. Overall, the results suggested that the hybrid version of the models achieved higher accuracy than the standalone models. The developed expert systems could be used to develop an early warning system for flood control and management to reduce the damage in flood-vulnerable areas. This information could be beneficial for both in-season and flash flooding, in terms of control measures and

analyzing the river patterns in case of uncertain water fluctuations in the face of climate change. The proposed novel MVMD-CNN-BiGRU model can be applied to other emerging sectors, such as renewable and sustainable energy, climate change, agriculture, and environmental areas, to solve the current challenges.

Funding

No funding.

Ethics approval

Not applicable

CRediT authorship contribution statement

Mozhdeh Jamei: Conceptualization, Data curation, Validation, and Methodology. **Mehdi Jamei:** Conceptualization, Data curation, Methodology, Software. **Mumtaz Ali:** Conceptualization, Investigation, Validation, Writing – original draft. **Masoud Karbasi:** Methodology, Visualization, Writing – original draft, Writing – review & editing. **Aitazaz A. Farouque:** Supervision, Writing – review & editing. **Anurag Malik:** Methodology, Writing – original draft. **Saad Javed Cheem:** Writing – original draft, Writing – review & editing. **Travis J. Esau:** Writing – review & editing. **Zaher Mundher Yaseen:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no competing interests.

Data availability

Data will be made available on request.

Acknowledgement

Authors would like to thank Natural Science and Engineering Council of Canada and Atlantic Canada Opportunities Agency for providing funding. Authors are also thankful to the University of Prince Edward Island for providing resources.

References

- Adnan, R.M., Mostafa, R.R., Islam, A.R.M.T., Gorgij, A.D., Kuriqi, A., Kisi, O., 2021. Improving drought modeling using hybrid random vector functional link methods. *Water* 13, 3379. <https://doi.org/10.3390/w13233379>.
- Ahmadi, F., Tohidi, M., Sadriazade, M., 2023. Streamflow prediction using a hybrid methodology based on variational mode decomposition (VMD) and machine learning approaches. *Appl Water Sci* 13, 135.
- Ahmadianfar, I., Jamei, M., Chu, X., 2019. Prediction of local scour around circular piles under waves using a novel artificial intelligence approach. *Mar. Georesources Geotechnol.* 0, 1–12. <https://doi.org/10.1080/1064119X.2019.1676335>.
- Ahmed, A.A.M., Deo, R.C., Ghahramani, A., Raj, N., Feng, Q., Yin, Z., Yang, L., 2021. LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios. *Stoch. Env. Res. Risk A.* <https://doi.org/10.1007/s00477-021-01969-3>.
- Al-Areeq, A.M., Abba, S.I., Yassin, M.A., Benaaf, M., Ghaleb, M., Aljundi, I.H., 2022. Computational machine learning approach for flood susceptibility assessment integrated with remote sensing and GIS techniques from Jeddah, Saudi Arabia. *Remote Sens.* <https://doi.org/10.3390/rs14215515>.
- Ali, M., Deo, R.C., Downs, N.J., Maraseni, T., 2018. Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting. *Comput. Electron. Agric.* 152, 149–165. <https://doi.org/10.1016/j.compag.2018.07.013>.
- Allawi, M.F., Salih, S.Q., Kassim, M., Ramal, M.M., Mohammed, A.S., Yaseen, Z.M., 2022. Application of computational model based probabilistic neural network for surface water quality prediction. *Mathematics* 10, 3960.
- Bakshi Ostadkalayeh, F., Moradi, S., Asadi, A., Moghaddam Nia, A., Taheri, S., 2023. Performance improvement of LSTM-based deep learning model for streamflow forecasting using Kalman filtering. *Water Resour. Manag.* 1–17.

- Beyaztas, U., Shang, H.L., Yaseen, Z.M., 2021. A functional autoregressive model based on exogenous hydrometeorological variables for river flow prediction. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2021.126380>.
- Bhatti, A.Z., Farouque, A.A., Krouglicof, N., Peters, W., Li, Q., Acharya, B., 2022. Prospective climates, and water availabilities under different projections of environmental changes in Prince Edward Island, Canada. *Water (Switzerland)* 14. <https://doi.org/10.3390/w14050740>.
- Bisoi, R., Dash, P.K., Mishra, S.P., 2019. Modes decomposition method in fusion with robust random vector functional link network for crude oil price forecasting. *Appl. Soft Comput.* 80, 475–493. <https://doi.org/10.1016/j.asoc.2019.04.026>.
- Cheng, W.X., Suganthan, P.N., Katuwal, R., 2021. Time series classification using diversified ensemble deep random vector functional link and resnet features. *Appl. Soft Comput.* 112, 107826.
- 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. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1724–1734. <https://doi.org/10.3115/v1/D14-1179>.
- Deepa, S.N., Natarajan, N., Berlin, M., 2023. Enhanced variational mode decomposition with deep learning SVM kernels for river streamflow forecasting. *Environ. Earth Sci.* 82, 544.
- Fan, J., Zhao, G., Mu, X., Lu, A., Tian, P., Gao, P., Sun, W., 2023. Effects of cascading reservoirs on streamflow and sediment load with machine learning reconstructed time series in the upper Yellow River basin. *Catena* 225, 107008.
- Fayer, G., Lima, L., Miranda, F., Santos, J., Campos, R., Bignoto, V., Andrade, M., Moraes, M., Ribeiro, C., Capriles, P., 2023. A temporal fusion transformer deep learning model for long-term streamflow forecasting: a case study in the Funil reservoir, Southeast Brazil. *Knowledge-Based Eng. Sci.* 4, 73–88.
- Gan, L., Zhao, X., Wu, H., Zhong, Z., 2021. Estimation of remaining fatigue life under two-step loading based on kernel-extreme learning machine. *Int. J. Fatigue* 148, 106190.
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach. Learn.* 63, 3–42.
- Gu, C., Qiao, X., Jin, Y., Liu, Y., 2020. A novel fault diagnosis method for diesel engine based on MVMD and band energy. *Shock. Vib.* 2020, 1–17. <https://doi.org/10.1155/2020/8247194>.
- Guo, Z., Yang, C., Wang, D., Liu, H., 2023. A novel deep learning model integrating CNN and GRU to predict particulate matter concentrations. *Process. Saf. Environ. Prot.* 173, 604–613.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol.* 377, 80–91.
- Hamzah, F.B., MohdHamzah, F., Mohd Razali, S.F., Jaafar, O., AbdulJamil, N., 2020. Imputation methods for recovering streamflow observation: a methodological review. *Cogent. Environ. Sci.* <https://doi.org/10.1080/23311843.2020.1745133>.
- Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: theory and applications. *Neurocomputing* 70, 489–501.
- Hur, J.-H., Ihm, S.-Y., Park, Y.-H., 2017. A variable impacts measurement in random Forest for Mobile cloud computing. *Wirel. Commun. Mob. Comput.* 2017, 1–13. <https://doi.org/10.1155/2017/6817627>.
- Huyghues-Beaufond, N., Tindemans, S., Falugi, P., Sun, M., Strbac, G., 2020. Robust and automatic data cleansing method for short-term load forecasting of distribution feeders. *Appl. Energy* 261, 114405.
- Ibrahim, K.S.M.H., Huang, Y.F., Ahmed, A.N., Koo, C.H., El-Shafie, A., 2022. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alexandria Eng. J.* 61, 279–303. <https://doi.org/10.1016/j.aej.2021.04.100>.
- Ikram, R.M.A., Hazarika, B.B., Gupta, D., Heddham, S., Kisi, O., 2023. Streamflow prediction in mountainous region using new machine learning and data preprocessing methods: a case study. *Neural Comput. & Applic.* 35, 9053–9070. <https://doi.org/10.1007/S00521-022-08163-8/METRICS>.
- Jamei, M., Ahmadianfar, I., Chu, X., Yaseen, Z.M., 2021. Estimation of triangular side orifice discharge coefficient under a free flow condition using data-driven models. *Flow Meas. Instrum.* 77, 101878 <https://doi.org/10.1016/j.flowmeasinst.2020.101878>.
- Jamei, M., Ali, M., Karbasi, M., Xiang, Y., Ahmadianfar, I., Yaseen, Z.M., 2022a. Designing a multi-stage expert system for daily ocean wave energy forecasting: a multivariate data decomposition-based approach. *Appl. Energy* 326, 119925. <https://doi.org/10.1016/j.apenergy.2022.119925>.
- Jamei, M., Maroufpoor, S., Aminpour, Y., Karbasi, M., Malik, A., Karimi, B., 2022b. Developing hybrid data-intelligent method using Boruta-random forest optimizer for simulation of nitrate distribution pattern. *Agric. Water Manag.* 270, 107715 <https://doi.org/10.1016/J.AGWAT.2022.107715>.
- Jamei, M., Ahmadianfar, I., Karbasi, M., Malik, A., Kisi, O., Yaseen, Z.M., 2023a. Development of wavelet-based Kalman online sequential extreme learning machine optimized with Boruta-Random Forest for drought index forecasting. *Eng. Appl. Artif. Intell.* 117, 105545.
- Jamei, M., Ali, M., Malik, A., Karbasi, M., Rai, P., Yaseen, Z.M., 2023b. Development of a TVF-EMD-based multi-decomposition technique integrated with encoder-decoder-bidirectional-LSTM for monthly rainfall forecasting. *J. Hydrol.* 129105.
- Katipoğlu, O.M., Sarıgöl, M., 2023. Application of boosted tree algorithm with new data preprocessing techniques in the forecasting one day ahead streamflow values in the Tigris basin, Türkiye. *J. Hydro-Environ. Res.* 50, 13–25.

- Katuwal, R., Suganthan, P.N., 2019. Stacked autoencoder based deep random vector functional link neural network for classification. *Appl. Soft Comput.* 85, 105854 <https://doi.org/10.1016/j.asoc.2019.105854>.
- Kilinc, H.C., Ahmadianfar, I., Demir, V., Heddam, S., Al-Areeq, A.M., Abba, S.I., Tan, M. L., Halder, B., Marhoon, H.A., Yaseen, Z.M., 2023. Daily scale river flow forecasting using hybrid gradient boosting model with genetic algorithm optimization. *Water Resour. Manag.* 1–16.
- Kiranyaz, S., Avcı, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J., 2021. 1D convolutional neural networks and applications: a survey. *Mech. Syst. Signal Process.* 151, 107398.
- Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the boruta package. *J. Stat. Softw.* 36 <https://doi.org/10.18637/jss.v036.i11>.
- Kursa, M.B., Jankowski, A., Rudnicki, W.R., 2010. Boruta – a system for feature selection. *Fundam. Informat.* 101, 271–285. <https://doi.org/10.3233/FI-2010-288>.
- Liu, H., Zhang, Y., Zhang, H., 2020. Prediction of effluent quality in papermaking wastewater treatment processes using dynamic kernel-based extreme learning machine. *Process Biochem.* 97, 72–79. <https://doi.org/10.1016/j.procbio.2020.06.020>.
- Liu, Y., Hou, G., Huang, F., Qin, H., Wang, B., Yi, L., 2022. Directed graph deep neural network for multi-step daily streamflow forecasting. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2022.127515>.
- Malik, A., Jamei, M., Ali, M., Prasad, R., Karbasi, M., Yaseen, Z.M., 2022. Multi-step daily forecasting of reference evapotranspiration for different climates of India: a modern multivariate complementary technique reinforced with ridge regression feature selection. *Agric. Water Manag.* 272, 107812 <https://doi.org/10.1016/j.agwat.2022.107812>.
- Malik, A.K., Gao, R., Ganaie, M.A., Tanveer, M., Suganthan, P.N., 2023. Random vector functional link network: recent developments, applications, and future directions. *Appl. Soft Comput.* 110377 <https://doi.org/10.1016/j.asoc.2023.110377>.
- Martinho, A.D., Saporetto, C.M., Goliatt, L., 2023. Approaches for the short-term prediction of natural daily streamflows using hybrid machine learning enhanced with grey wolf optimization. *Hydrol. Sci. J.* <https://doi.org/10.1080/02626667.2022.2141121>.
- Meng, A., Zhu, Z., Deng, W., Ou, Z., Lin, S., Wang, C., Xu, X., Wang, X., Yin, H., Luo, J., 2022. A novel wind power prediction approach using multivariate variational mode decomposition and multi-objective crisscross optimization based deep extreme learning machine. *Energy* 260, 124957. <https://doi.org/10.1016/j.energy.2022.124957>.
- Mishra, S., Sturm, B.L., Dixon, S., 2017. Local interpretable model-agnostic explanations for music content analysis. In: *ISMIR*, pp. 537–543.
- Moazam, H.M.Z.H., Dehghani, Adnan, Mortazavizadeh, F., Ranjbar, V., Mirzaei, M., Hin, L.S., Ng, J.L., Dehghani, Amin, 2023. Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches. *Ecol. Inform.* 102119.
- Mosaffa, H., Sadeghi, M., Mallakpour, I., Jahromi, M.N., Pourghasemi, H.R., 2022. Application of machine learning algorithms in hydrology. In: *Computers in Earth and Environmental Sciences*. Elsevier, pp. 585–591.
- Mostafa, R.R., Kisi, O., Adnan, R.M., Sadeghifar, T., Kuriqi, A., 2023. Modeling potential evapotranspiration by improved machine learning methods using limited climatic data. *Water* 15, 486. <https://doi.org/10.3390/w15030486>.
- Niu, D., Yu, M., Sun, L., Gao, T., Wang, K., 2022. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. *Appl. Energy* 313, 118801.
- Nourani, V., Baghanam, A.H., Adamowski, J., Kisi, O., 2014. Applications of hybrid Wavelet-Artificial Intelligence models in hydrology. A review. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2014.03.057>.
- Ozsahin, D.U., Gökçekuş, H., Uzun, B., LaMoreaux, James, 2021. Application of Multi-Criteria Decision Analysis in Environmental and Civil Engineering, Book.
- Pao, Y.-H., Park, G.-H., Sobajic, D.J., 1994. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing* 6, 163–180.
- Papacharalampous, G., Tyrallis, H., 2022. A review of machine learning concepts and methods for addressing challenges in probabilistic hydrological post-processing and forecasting. *Front. Water.* <https://doi.org/10.3389/frwa.2022.961954>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2019. Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *Catena* 177, 149–166. <https://doi.org/10.1016/j.catena.2019.02.012>.
- Rasheed, A., Adebisi, A.T., Veluvolu, K.C., 2020. Respiratory motion prediction with random vector functional link (RVFL) based neural networks. *J. Phys. Conf. Ser.* 1626, 012022 <https://doi.org/10.1088/1742-6596/1626/1/012022>.
- Rehman, N. ur, Aftab, H., 2019. Multivariate variational mode decomposition. *IEEE Trans. Signal Process.* 67, 6039–6052. <https://doi.org/10.1109/tsp.2019.2951223>.
- Rozos, E., Dimitriadis, P., Bellos, V., 2022. Machine learning in assessing the performance of hydrological models. *Hydrology* 9, 5.
- Shapley, L.S., 1953. A value for n-person games. In: *Contributions to the Theory of Games (AM-28)*, Vol. II. Princeton University Press, pp. 307–318. <https://doi.org/10.1515/9781400881970-018>.
- Shi, Q., Katuwal, R., Suganthan, P.N., Tanveer, M., 2021. Random vector functional link neural network based ensemble deep learning. *Pattern Recogn.* 117, 107978.
- Singh, K.P., Gupta, S., Rai, P., 2013. Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmos. Environ.* 80, 426–437. <https://doi.org/10.1016/j.atmosenv.2013.08.023>.
- Slougher, J.M., Gneiting, T., Raftery, A.E., 2010. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Am. Stat. Assoc.* 105, 25–35.
- Smeti, E.M., Thanasoulas, N.C., Lytras, E.S., Tzoumerkas, P.C., Goulinopoulos, S.K., 2009. Treated water quality assurance and description of distribution networks by multivariate chemometrics. *Water Res.* 43, 4676–4684. <https://doi.org/10.1016/j.watres.2009.07.023>.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC Bioinform.* 9, 307. <https://doi.org/10.1186/1471-2105-9-307>.
- Suganthan, P.N., Katuwal, R., 2021. On the origins of randomization-based feedforward neural networks. *Appl. Soft Comput.* 105, 107239.
- Tao, H., Abba, S.I., Al-Areeq, A.M., Tangang, F., Samantaray, S., Sahoo, A., Siqueira, H. V., Maroufpoor, S., Demir, V., Dhanraj Bokde, N., Goliatt, L., Jamei, M., Ahmadianfar, I., Bhagat, S.K., Halder, B., Guo, T., Helman, D.S., Ali, M., Sattar, S., Al-Khafaji, Z., Shahid, S., Yaseen, Z.M., 2024. Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: a comprehensive review, assessment, and possible future research directions. *Eng. Appl. Artif. Intell.* 129, 107559 <https://doi.org/10.1016/j.engappai.2023.107559>.
- Tiwari, M.K., Chatterjee, C., 2011. A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting. *J. Hydroinf.* 13, 500–519.
- Tsai, C.C., Lu, M.C., Wei, C.C., 2012. Decision tree-based classifier combined with neural-based predictor for water-stage forecasts in a river basin during typhoons: a case study in Taiwan. *Environ. Eng. Sci.* <https://doi.org/10.1089/ees.2011.0210>.
- Ture, M., Kurt, I., Turhan Kurum, A., Ozdamar, K., 2005. Comparing classification techniques for predicting essential hypertension. *Expert Syst. Appl.* <https://doi.org/10.1016/j.eswa.2005.04.014>.
- Wang, J., Wang, P., Tian, H., Tansey, K., Liu, J., Quan, W., 2023. A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. *Comput. Electron. Agric.* 206, 107705.
- Yang, H., Li, W., 2023. Data decomposition, seasonal adjustment method and machine learning combined for runoff prediction: a case study. *Water Resour. Manag.* 37, 557–581.
- Yang, L., Yu, H., Feng, Q., Barzegar, R., Adamowski, J.F., Wen, X., 2023. Ensemble Learning of Decomposition-Based Machine Learning and Deep Learning Models for Multi-Time Step Ahead Streamflow Forecasting in an Arid Region.
- Yao, D., Li, B., Liu, H., Yang, J., Jia, L., 2021. Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit. *Measurement* 175, 109166.
- Yaseen, Z.M., Awadh, S.M., Sharafati, A., Shahid, S., 2018. Complementary data-intelligence model for river flow simulation. *J. Hydrol.* 567, 180–190. <https://doi.org/10.1016/j.jhydrol.2018.10.020>.
- Yoshida, K., Watanabe, K., Chiou, T.-Y., Konishi, M., 2023. High throughput optimization of medium composition for Escherichia coli protein expression using deep learning and Bayesian optimization. *J. Biosci. Bioeng.* 135, 127–133.
- Yousefi, M., Wang, J., Fandrem Hoivik, Ø., Rajasekharan, J., Hubert Wierling, A., Farahmand, H., Arghandeh, R., 2023. Short-term inflow forecasting in a dam-regulated river in Southwest Norway using causal variational mode decomposition. *Sci. Rep.* 13, 7016.
- Yu, Q., Jiang, L., Wang, Y., Liu, J., 2023. Enhancing streamflow simulation using hybridized machine learning models in a semi-arid basin of the Chinese loess plateau. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2023.129115>.
- Zakhrouf, M., Hamid, B., Kim, S., Madani, S., 2021. Novel insights for streamflow forecasting based on deep learning models combined the evolutionary optimization algorithm. *Phys. Geogr.* 1–24.
- Zhang, S., Gan, T.Y., Bush, A.B.G., Zhang, G., 2023. Evaluation of the impact of climate change on the streamflow of major pan-Arctic river basins through machine learning models. *J. Hydrol.* 619, 129295.
- Zounemat-Kermani, M., Batelaan, O., Fadaee, M., Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: a review. *J. Hydrol.* 598 <https://doi.org/10.1016/j.jhydrol.2021.126266>.