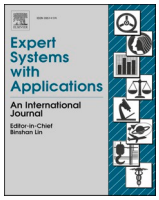
Expert Systems With Applications 256 (2024) 124900

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Robust drought forecasting in Eastern Canada: Leveraging EMD-TVF and ensemble deep RVFL for SPEI index forecasting

Masoud Karbasi [a,b,*], Mumtaz Ali [c], Aitazaz Ahsan Farooque [b,d,*], Mehdi Jamei [b,e,f], Khabat Khosravi [b], Saad Javed Cheema [b], Zaher Mundher Yaseen [g]

[a] Water Engineering Department, Faculty of Agriculture, University of Zanjan, Zanjan, Iran
[b] Canadian Centre for Climate Change and Adaptation, University of Prince Edward Island, St Peters Bay, PE, Canada
[c] UniSQ College, University of Southern Queensland, QLD, 4305, Australia
[d] Faculty of Sustainable Design Engineering, University of Prince Edward Island, Charlottetown, PE, Canada
[e] Department of Civil Engineering, Faculty of Civil Engineering and Architecture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[f] New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Nasiriyah 64001, Iraq
[g] Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Keywords:
Drought
Forecasting
Machine learning
Preprocessing
Feature selection
Atlantic Canada

ABSTRACT

Drought stands as a highly perilous natural catastrophe that impacts numerous facets of human existence. Drought data is nonstationary and noisy, posing challenges for accurate forecasting. This study proposes a novel hybrid framework integrating TVF-EMD preprocessing, LASSO feature selection and Ensemble Deep RVFL modeling for improved multistep ahead drought prediction. Using decomposed $SPEI_{12}$ values, six machine-learning techniques (Support Vector Regression (SVR), Simple RVFL, Ensemble Deep RVFL, and Recurrent Neural Network (RNN), XGBoost, Random Forest (RF)) were applied to forecast the $SPEI_{12}$ drought index. The present study involved forecasting drought in two Canadian stations located in the eastern region (Charlottetown in Prince Edward Island and Fredericton in New Brunswick), where agriculture is rainfed and mostly affected by drought. The statistical period of 1980–2022 was considered for analysis. Following the decomposition of drought data with TVF-EMD, lagged data was generated using the TVF-EMD results. Training time was decreased by utilizing the Lasso regression feature selection algorithm to select effective inputs. Various statistical measures, including the root mean square error (RMSE) and correlation coefficient (R), were employed to assess the precision of the models. The research findings indicated that the TVF-ED-RVFL model achieved the highest level of precision in forecasting multistep ahead (1,3,6 and 12) $SPEI_{12}$ drought index for both Charlottetown and Fredericton stations. During testing, the TVF-ED-RVFL model predicted 1-month $SPEI_{12}$ for Charlottetown ($R = 0.9995$, $RMSE = 0.0352$) and Fredericton ($R = 0.9974$, $RMSE = 0.0560$). For multistep ahead forecasting, the R-values range from 0.9924 for 3-months ahead to 0.9242 for 12-months ahead in Charlottetown and range from 0.9846 for 3-months ahead to 0.8293 for 12-months ahead in Fredericton. By increasing the forecasting horizon, the accuracy of models decreased. The present study's outcomes can contribute to enhancing water management practices during periods of drought.

1. Introduction

Drought is among the most catastrophic and costliest natural hazards worldwide. It is well-known as a creeping phenomenon (i.e., gradually pans out) but covers extensive areas, even on a continental scale (Hao et al., 2014; Maity et al., 2021; Wilhite et al., 2007). Drought negatively affects surface and groundwater resources, soil erosion and degradation, agriculture, food scarcity, and causes economic and social damages. Drought has occurred recently in some parts of the world, even in Europe (i.e., central Europe), known as a green continent, and leads to numerous and irreparable damages (Felsche & Ludwig, 2021).

Canada faced at least ten severe droughts, like other parts of the world. Historical drought occurrences in Canada are 1910–11, 1914–15, 1917–20, 1928–30, 1931–32, 1936–38, 1948–51, 1960–62, 1988–89,

* Corresponding authors.
E-mail addresses: m.karbasi@znu.ac.ir, mkarbasi@upei.ca (M. Karbasi), Mumtaz.Ali@unisq.edu.au (M. Ali), afarooque@upei.ca (A. Ahsan Farooque), jmehdi@upei.ca (M. Jamei), kkhosravi@upei.ca (K. Khosravi), scheema@upei.ca (S. Javed Cheema), z.yaseen@kfupm.edu.sa (Z.M. Yaseen).

https://doi.org/10.1016/j.eswa.2024.124900
Received 31 March 2024; Received in revised form 11 July 2024; Accepted 25 July 2024
Available online 30 July 2024

and 1999–2005 (Bonsal et al., 2011; Maybank et al., 1995). It has been reported by Bonsal et al. (2011) that the drought period of 1999–2005 was extremely severe and extended from British Columbia from the west to Atlantic provinces from the west and led to considerable agricultural, environmental, economic, and societal damages. The drought in the Prairies Provinces from 2001 to 2002 caused $3 Billion in losses in the agricultural sector, resulting in a $6 Billion drop in Canada's GDP (Wheaton et al., 2008). Earth's climate system has recently changed, and changes are speeding up because of climate change. These changes directly impact drought occurrences' frequency, duration, and severity. If this trend continues, it may lead to humanitarian catastrophes, especially for future generations.

One of the primary keys to drought mitigation is drought monitoring and forecasting, leading to the construction of an early warning online system. To accomplish this aim, many drought indices have been created and employed to monitor and evaluate various categories of drought, including agricultural, meteorological, socioeconomic, and hydrological droughts (Eden, 2012; van Hoek, 2016). Palmer Drought Severity Index (PDSI) (Palmer, 1968), Standardized Precipitation Index (SPI) (Mckee et al., 1993), Surface Water Supply Index (SWSI) (Doesken & Garen, 1991), and Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) are among well-known indices of drought assessment. Most of these indices are developed based on readily available meteorological variables, such as SPI, SPEI, and PDSI. SPI and SPEI are among the most widely used methods of drought assessment. The advantage of SPEI over SPI is that SPI only considers precipitation, while SPEI involves both precipitation and evapotranspiration simultaneously (i.e., temperature impact on drought occurrence).

There are four primary groups of drought monitoring and assessment methods: linear statistical methods, dynamical/physical/conceptual methods, remote sensing techniques, and data-driven/machine learning models. Drought occurrences are stochastic, chaotic, and nonlinear; therefore, simple and linear models such as ARMIA cannot capture drought variations accurately. Although dynamical/physical/conceptual methods are able to evaluate spatial and temporal damages caused by drought, vast input data is required, and calibration is time-consuming. Drought monitoring through remote sensing data is popular, as it is able to investigate both spatial and temporal scales, but drought time-series analysis is time-consuming, especially for large areas.

There has been a growing trend in the application of machine learning (ML) and deep learning (DL) methodologies within the domain of hydrology. The advantages of ML/DL can be the non-linearity structure, high flexibility, high robustness, less input variable, reasonable accuracy, insensitivity to missing data, and ability to handle big data with different dimensions (Khosravi et al., 2019; Tao et al., 2024; Yaseen, 2023). Artificial Neural Network (ANN) is one of the oldest ML-based models that is widely used in the field of hydrology and specifically in drought assessment (Belayneh et al., 2014; Tareke & Awoke, 2023; Wable et al., 2023). Because of its low convergence and generalization power, ANN, combined with a fuzzy logic approach and adaptive neuro-fuzzy inference system (ANFIS), is developed. Support vector regression (SVR), extreme learning machine (ELM), and Gene expression programming (GEP) are also among popular ML models that are frequently used in hydrology and water resources management. Mokhtarzad et al. (2017) compared the prediction capability of ANN, ANFIS, and SVR for drought perdition at Bojnourd city station and finally stated that the SVR model leads to higher performance (Mokhtarzad et al., 2017). Wang et al. (2022) implemented ELM, SVR, and their combination with a wavelet model for drought prediction in Dez Dam, Iran (Wang et al., 2022). They finally stated that the wavelet has increased each model's modeling performance, and ELM is a more effective and promising tool for drought prediction. Overall, all these traditional models need the weights of their membership functions set accurately. In addition, some of them are hyperparameter models (i.e., SVR), limiting their application. Yaseen et al. (2021) compared the performance of minimum probability machine regression (MPMR), random forest (RF), M5 Tree (M5tree), online sequential-ELM (OSELM), and extreme learning machine (ELM) for drought prediction in Bangladesh and finally revealed that ELM model leads to the highest performance (Yaseen et al., 2021).

Recently, the application of DL algorithms has attracted the attention of researchers globally due to their higher performance compared to traditional models (Maity et al., 2021). In addition, DL models are developed to solve the problem end to end, while ML models separate the issue into different parts. Some other advantages of DL are feature generation automation, better self-learning capabilities, advanced analytics, and scalability. Panahi et al. (2020) developed a convolutional neural network (CNN) deep learning model for groundwater potential areas and then compared its performance with the SVR models as a benchmark. They finally declared that the groundwater potential map derived by the CNN model has 85 % accuracy, while those of SVR have 75 %. Abbes et al. (2023) developed Long Short-Term Memory (LSTM) combined with a Multi-Resolution Analysis Wavelet Transform (MRA-WT), called MRA-WT-LSTM for drought modeling at prediction in Iran (Abbes et al., 2023). They finally compared their results with ANN, SVR, random forest (RF), and LSTM models. They showed that the proposed hybrid framework of MRA-WT-LSTM with R^2 greater than 0.93 outperforms other models. Kadam et al. (2024) developed and employed RNN, GRU, and LSTM deep-learning models for drought prediction (Kadam et al., 2024). They finally stated that deep learning models are promising for drought modeling and prediction, and all models lead to good performance, but LSTM models have a higher prediction capability than other models.

In addition, ensemble-based or hybridization techniques are among the popular approaches to increase the modeling prediction power of each ML/DL model. Bui et al. (2020) developed several models of random forest (RF), M5P, reduced error pruning tree (REPT), random tree (RT), and their ensemble with Bagging algorithm for water quality index (WQI) prediction (Bui et al., 2020). Finally, they declared that ensemble-based models, due to higher flexibility and benefit from two models simultaneously, have a higher prediction power than standalone models. Kaur and Sood (2020) implemented ANN, SVR, and deep neural network (DNN) for drought assessment in three divisions of the Texas area, USA (Kaur & Sood, 2020). They finally revealed that the DNN model outperformed other models besides just coupling two models, hybridizing the feature selection algorithm and model's hyperparameter tuning reported as a promising technique (Jamei et al., 2023), which rarely addressed in the field of water resources, while has a significant impact on the modeling performance.

RVFL is a particular form of a neural network that utilizes random weights, as suggested in Pao et al., (1994). Randomly generating the weights of the enhancement layer enhances the efficiency of training a neural network. The improvement layer functions as an unsupervised method for extracting features. A different autonomous random neural network was introduced by Schmidt et al., (1992), which is a specific instance of RVFL that does not include the functional link. Several studies have conducted comparisons on the performance of RVFL and its other variants with other machine learning techniques in time series forecasting problems (Cheng et al., 2024; Gao et al., 2021, 2022, 2023; Qiu et al., 2018). Gao et al. (2022) implemented an edRVFL network for electricity load forecasting. They augmented the random enhancement features using empirical wavelet transformation (EWT). The raw load data were decomposed by EWT in a walk-forward manner, ensuring no future data leakage during the decomposition process. Subsequently, all sub-series generated by the EWT, including the raw data, were input into the edRVFL for forecasting. The proposed model was evaluated on sixteen publicly available time series from the Australian Energy Market Operator for the year 2020. The simulation results showed that the proposed model outperformed eleven other forecasting methods across two error metrics and statistical tests in electricity load forecasting tasks. Jamei et al. (2023) utilized a time-varying filter-based empirical mode

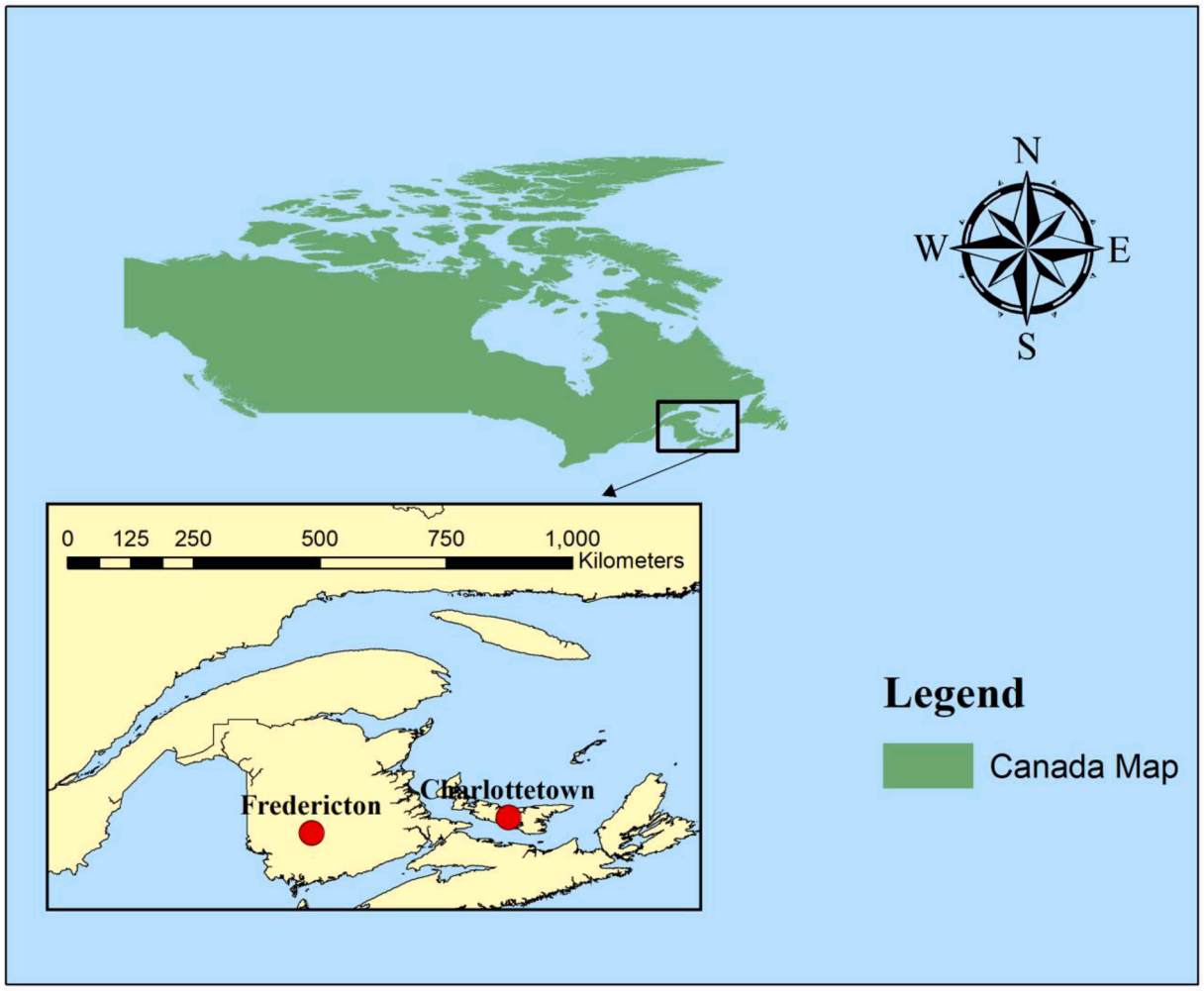


Fig. 1. Location of Stations.

decomposition (TVF-EMD) combined with an ensemble deep Random Vector Functional Link (edRVFL) scheme to forecast PM2.5 and PM10 concentrations 1-hour and 3-hours ahead in two regions of Australia. To validate the TVF-EMD-DRVFL model, they also employed classical RVFL and bidirectional gated recurrent unit neural network (Bi-GRU) models in both hybrid and standalone frameworks. The findings indicate that the TVF-EMD-DRVFL model delivers the highest accuracy for forecasting PM2.5 and PM10 concentrations at t + 1 and t + 3 for the Brisbane and North Parramatta stations.

The novelty and gap in the research lies in developing a hybrid modeling framework that integrates data preprocessing techniques with machine learning algorithms for improved monthly drought forecasting. Non-stationarities and noise in drought time series data can hinder the model if not treated properly. Conventional EMD decomposition might pose mode mixing problems by which the corresponding IMFs might contain high-frequency and low-frequency signals. This is because there is no way to differentiate between intrinsic oscillations and noise in the shifting process. The TVF-EMD method tries to solve this problem by using a time-varying filter when shifting the data instead of the local mean. This filter has the ability to either increase or decrease its cutoff frequency depending on the localized features of the signal. We can use this adaptive, time-varying filter to eliminate high-frequency noises while allowing the intrinsic low-frequency oscillations inherent in the IMFs. This enables each IMF to correspond to the oscillations involving a single characteristic scale and prevents mode coupling. It extracts signals at different time scales more accurately. When performing the TVF-ED analysis on the drought time series beforehand, it is better to decompose the data into a trend, seasonality, and noise and obtain low-frequency patterns relevant to drought.

This study proposes a novel framework for multi-temporal forecasting of the $SPEI_{12}$ drought index using Time-varying filter based Empirical Mode Decomposition (TVF-EMD) for signal decomposition, Least Absolute Shrinkage and Selection Operator (LASSO) regression for feature selection, and Ensemble Deep Random Vector Functional Link (RVFL) machine learning model. The hybridization of data processing steps like TVF-EMD and feature selection aims to reduce complexity and extract meaningful information from time-series drought data. This is integrated with the Ensemble Deep RVFL model, which has not been previously applied for drought forecasting. The framework aims to improve forecasting accuracy compared to standalone machine learning models like Support Vector Regression (SVR), Random Forest (RF), XGBoost, Recurrent Neural Network (RNN), RVFL, and ED-RVFL models. Through comprehensive evaluation and comparison of hybrid and standalone models, the study aims to identify the most effective modeling approach for drought forecasting and assess the lead times for which results are acceptable. This can help address the research gaps in effective monitoring and predicting drought, which is important for sustainable water resource management. The case study focuses on Atlantic Canada, which has high agricultural importance for the region. Drought forecasting is crucial here as most farming is rainfed and thus vulnerable to water shortages. The objectives of the current research can be summarized:

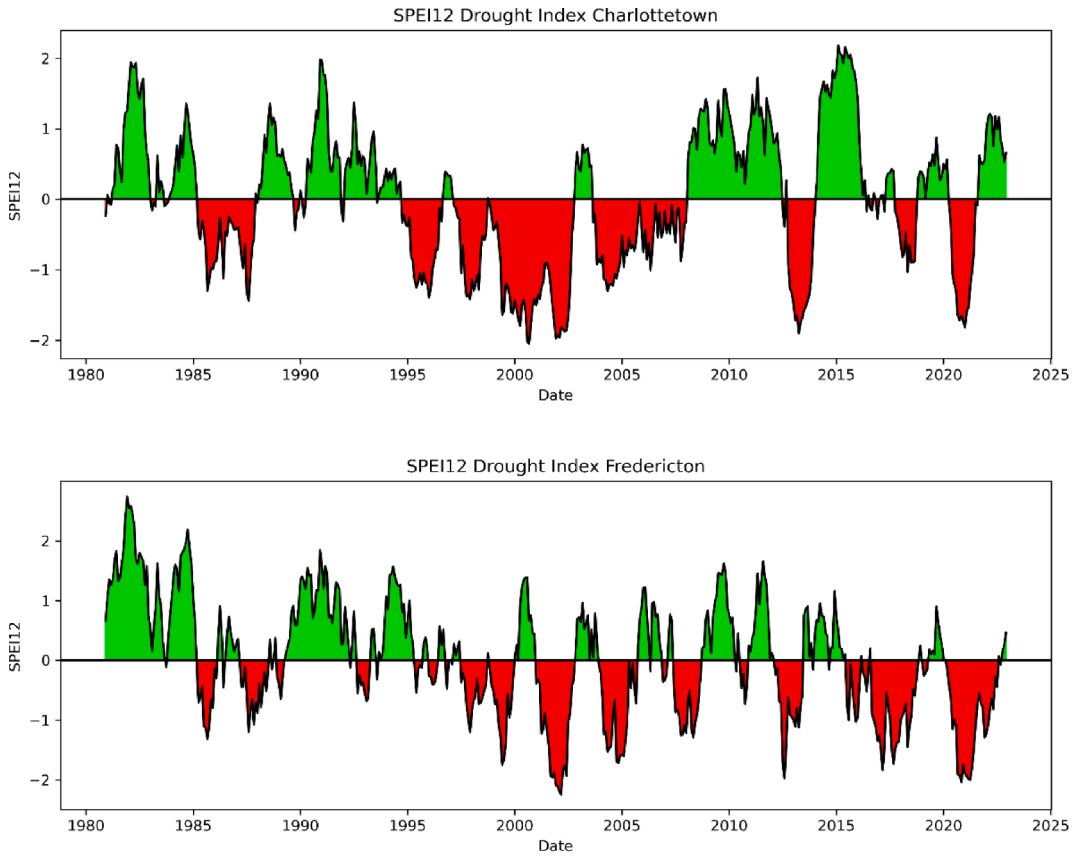


Fig. 2. Time Series of $SPEI_{12}$ drought index at Charlottetown and Fredericton stations.

Table 1
Descriptive statistics of data.

Metric	Charlottetown	Fredericton
Number of Data	504	504
Mean	−0.0017	0.0012
Standard Deviation	0.9840	0.9855
Minimum	−2.0492	−2.2461
Maximum	2.1835	2.7484
Quartile 1	−0.7890	−0.7091
Median	0.0344	0.0030
Quartile 3	0.6778	0.7243
Skewness	−0.0012	0.0688
Kurtosis	−0.7682	−0.4608

i. forecast drought at two stations in Eastern Canada (Charlottetown and Fredericton) for multiple time steps ahead using the SPEI index.
ii. 2-Investigate how data processing techniques (TVF-EMD and feature selection) can improve the results of standalone machine learning models.
iii. Compare the performance of hybrid TVF-based models (TVF-ED-RVFL, TVF-SVR, TVF-RNN, TVF-RF, TVF-XGBoost, TVF-RVFL) vs standalone models (ED-RVFL, RVFL, XGBoost, RF, SVR, RNN).
iv. Evaluate models using extensive statistical metrics like correlation coefficient, RMSE, NSE, KGE, I_A, and U_{95}%. The comprehensive evaluation will provide robust validation of model performance.
v. Analyze multistep ahead scatter plots, time series plots, and residual plots to visually assess model forecasts against observations for different lead times.

2. Material and methods

2.1. Study area and data description

In the current study, two stations in Prince Edward Island (Charlottetown) and New Brunswick (Fredericton) provinces of Canada were selected to forecast agricultural drought. Both provinces have many agricultural lands that produce many crops such as potatoes, wheat, barley etc. Fig. 1 shows the location of Charlottetown and Fredericton stations in Canada. According to the Koppen climate classification system, both stations have a humid continental climate (Dfb) (Beck et al., 2018). The statistical period of 1980–2022 (43 years) was considered to calculate the drought index in both stations. The mean temperature values are 5.8 and 5.7 for Charlottetown and Fredericton, respectively. Annual precipitation and evapotranspiration are (1170 and 716 mm) and (1067 and 849 mm) for Charlottetown and Fredericton, respectively.

To compute the SPEI, the non-exceedance probability of the differences between potential evapotranspiration and precipitation is adjusted using a three-parameter log-logistic distribution that considers frequent negative values (Vicente-Serrano et al., 2010). By applying a logarithmic probability distribution, the water balance is normalized to obtain the SPEI. In this investigation, the Hargreaves and Samani method (Hargreaves & Samani, 1985) was utilized to calculate *PET*. The following equation provides the difference (D_i) between precipitation (P) and *PET* for the given month (i):

$$D_i = P_i - PET_i \tag{1}$$

The summed D values are compiled across various time scales in the subsequent manner:

$$D_n^k = \sum_{j=0}^{k-1} (P_{n-j} - PET_{n-j}) \tag{2}$$

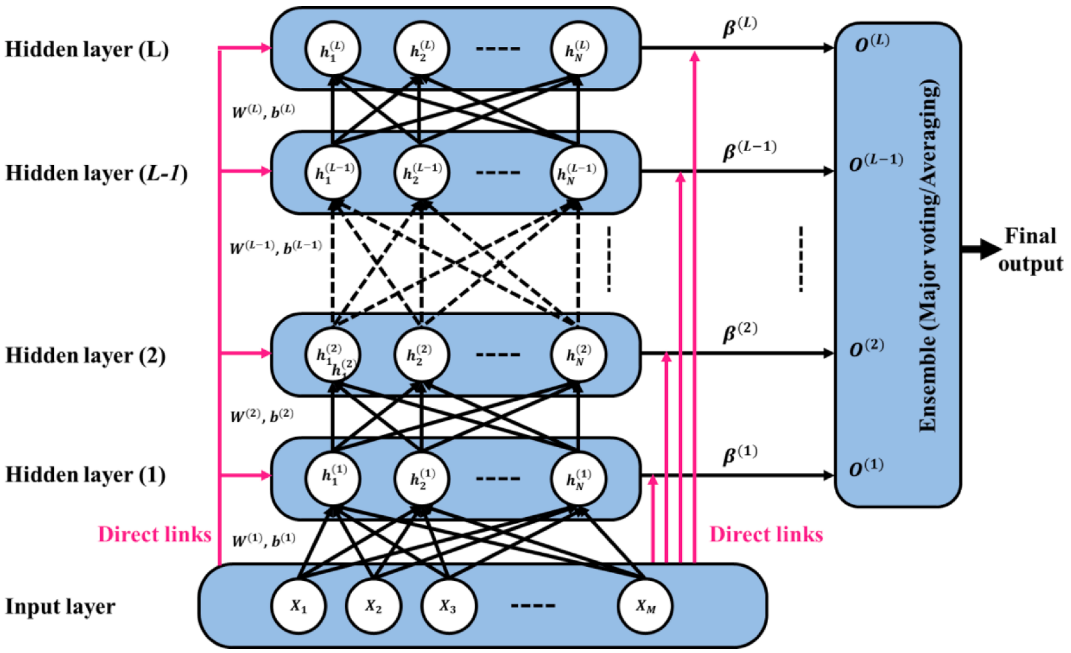


Fig. 3. The topology of the Ensemble Deep RVFL model.

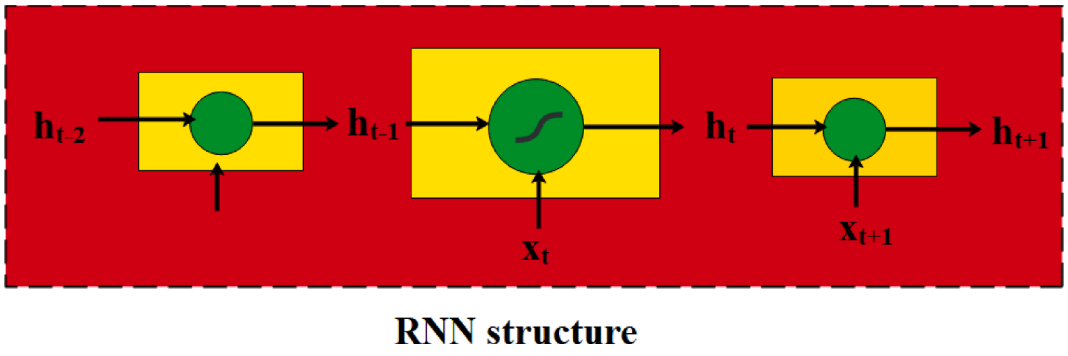


Fig. 4. Schematic diagrams of the RNN Model (Zheng et al., 2023).

Determining the likelihood of an event happening requires three parameters of the log-logistic probability distribution function; this is how SPEI is calculated. In accordance with the log-logistic distribution, one can formulate the probability distribution function of D as follows (Abramowitz & Stegun, 1965):

$$SPEI = \Psi - \frac{\delta_0 + \delta_1\Psi + \delta_2\Psi^2}{1 + d_1\Psi + d_2\Psi^2 + d_3\Psi^3} \quad (3)$$

where

$$\Psi = \sqrt{-2\ln(P)} for P \leq 0.5 \quad (4)$$

$$P = 1 - F(x) for P > 0.5 \quad (5)$$

in which P represents the probability that the D series will surpass and $F(x)$ is the probability distribution of the log-logistic distribution. Readers are directed to the following source (Vicente-Serrano et al., 2010) for additional information regarding the SPEI obtaining procedure and the constant values of δ_0, δ_1, δ_2, d_1, d_1, and d_3. The multi-temporal monthly forecasting of meteorological drought relied on the $SPEI_{12}$ values calculated at the Charlottetown and Fredericton stations. The time series of the $SPEI_{12}$ drought index in both stations is depicted in Fig. 2. According to Fig. 2, most droughts in Charlottetown occurred from 1994 to 2008, and in Fredericton stations, droughts occurred from 2015 to 2022. Table 1 presents the descriptive statistics of the $SPEI_{12}$ drought index in both stations.

2.2. Time-varying filter-based empirical mode decomposition

Li et al. (2017) introduced the concept of the time-varying filter-based empirical mode decomposition (TVF-EMD) method to solve the problems of mode mixing and the final effect in empirical mode decomposition (EMD). They achieved this by using a time-varying filter to perform the shifting process (Li et al., 2017).

When implementing the TVF-EMD approach, it is crucial to carefully consider the appropriate bandwidth threshold and B-spline order as they serve as the main limitations (Li et al., 2017). If these limitations are deemed incorrect, the mode mixing problem cannot be optimized using the TVF-EMD model. Similarly, the local cut-off frequency must be determined to execute time-varying filtering. Further, compared to the decomposition procedures, the TVF-EMD method produces more practical outcomes (Jamei et al., 2023; Wang et al., 2020). Furthermore, in comparison to the decomposition approaches, the TVF-EMD methodology produces more feasible findings (Li et al., 2017):

Step -1: Calculate the local cutoff frequency with a B-spline approximation, which may be explained as:

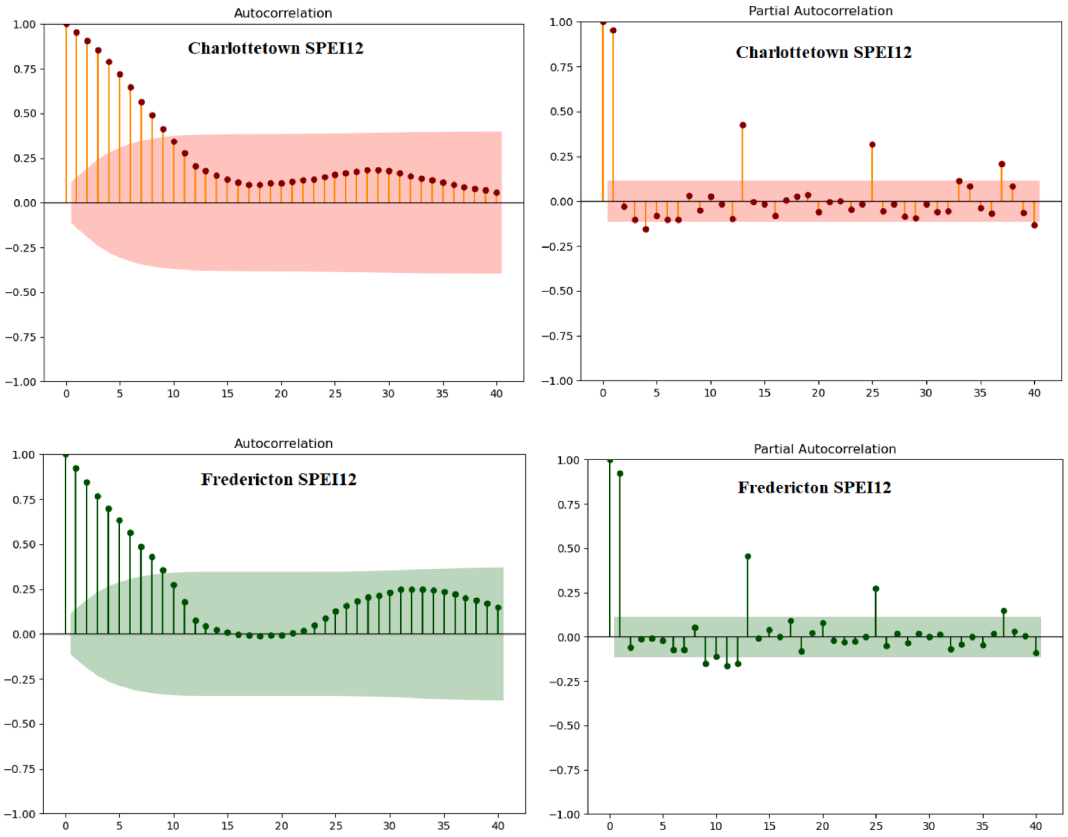


Fig. 5. Determining important lags using ACF and PACF plots.

Table 2
Setting for TVF-EMD for SPEI index decomposition.

Station Name	Stopping criterion	Number of Decomposed IMFs	End_flag	B-spline Order
Charlottetown	0.1	11	0	26
Fredericton	0.1	11	0	26

$$g_m^n(t) = \sum_{k=-\infty}^{\infty} c(k)\beta^n(t/m - k) \tag{1}$$

In this context, $\beta\text{-}n.(t) = \mathrm{B}$ represents the B-spline function, whereas $c(k) = \mathrm{B}$ denotes the B-spline coefficient. By a factor of m the B-spline function is increased in size. n, m, and (k) are the factors that determine the approximation result. When provided with the B-spline order n and knots m, the task of B-spline approximation involves determining the (k) that minimizes the approximation error, and it is computed as:

$$\varepsilon_m^2 = \sum_{t=-\infty}^{+\infty} \left(x(t) - [c]_{\uparrow m} {}^{*} b_m^n(t)\right)^2 \tag{2}$$

In which, $b_m^n(t) = \beta^n(t/m)$, $[\bullet]_{\uparrow\ m}$ = up-sampling operation by m, and * = convolution operator and the solution of $c(k)$ is

$$c(k) = \left[p_m^{n*} x\right]_{\downarrow m}(k) \tag{3}$$

where, $[\bullet]_{\downarrow m}$ = down-sampling operation by m, and p_m^n = pre-filter. Thus, we can write Eq. (1) as:

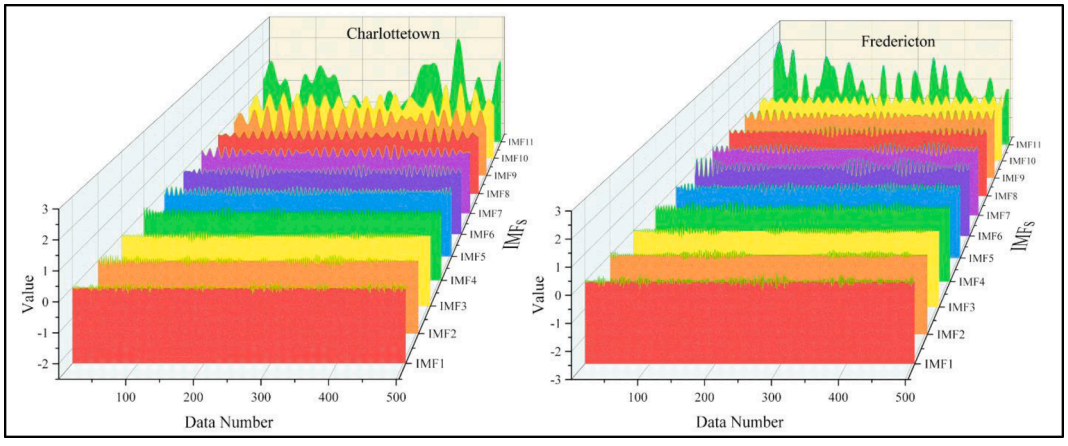


Fig. 6. Drought signal decomposition using TVF-EMD for Charlottetown and Fredericton stations.

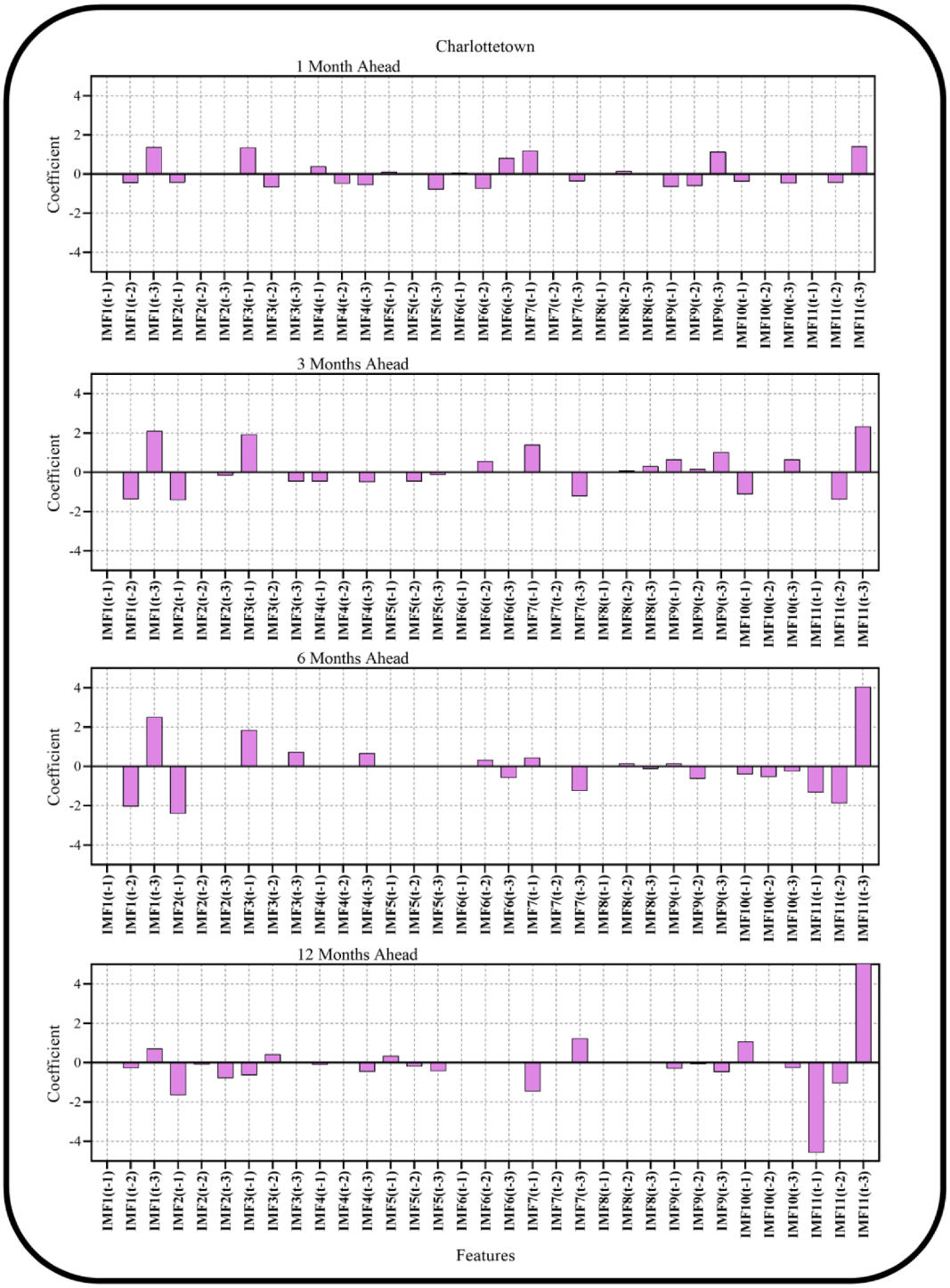


Fig. 7. Determining effective inputs for drought forecasting at Charlottetown station by lasso regression feature selection algorithm.

$$g_m^n(t) = \left[p_m^{n^*}\right]_{\downarrow m} {}^{*}b_m^n(t) \tag{4}$$

The B-spline approximation utilized in the above equation is a specific type of low-pass filtering. Thus, the TVF is generated utilizing the local cut-off frequency that is projected from the input signal. The purpose of this procedure is to acquire the local cut-off frequency, $\varphi'_{\text{bis}}(t) = \varphi'_1(t) + \varphi'_2(t)/2$, here, $\varphi'_1(t)$, and $\varphi'_2(t)$ are slow varying components. Realign the $\varphi'_{\text{bis}}(t)$ to solve the issue of intermittence, i.e., noise, and obtain the final local cut-off frequency by interpolating among the peaks (or remainders).

Step-2: This procedure is executed in order to acquire the local cut-off frequency.

Step-3: Determine if the residual signal satisfies the halting requirements (or meets the criteria for improvement) in the following manner (Jamei et al., 2023; K. Wang et al., 2020):

$$\theta(t) = \frac{B_{\text{Loughlin}}(t)}{\varphi_{\text{avg}}(t)} \tag{5}$$

In which, $B_{\text{Loughtin}}(t)$ and $\varphi_{\text{arg}}(t)$ provides the weighted average instantaneous frequency and Loughlin instantaneous bandwidth of distinct components. The current investigation employed TVF-EMD to anticipate monthly drought at the Charlottetown and Fredericton

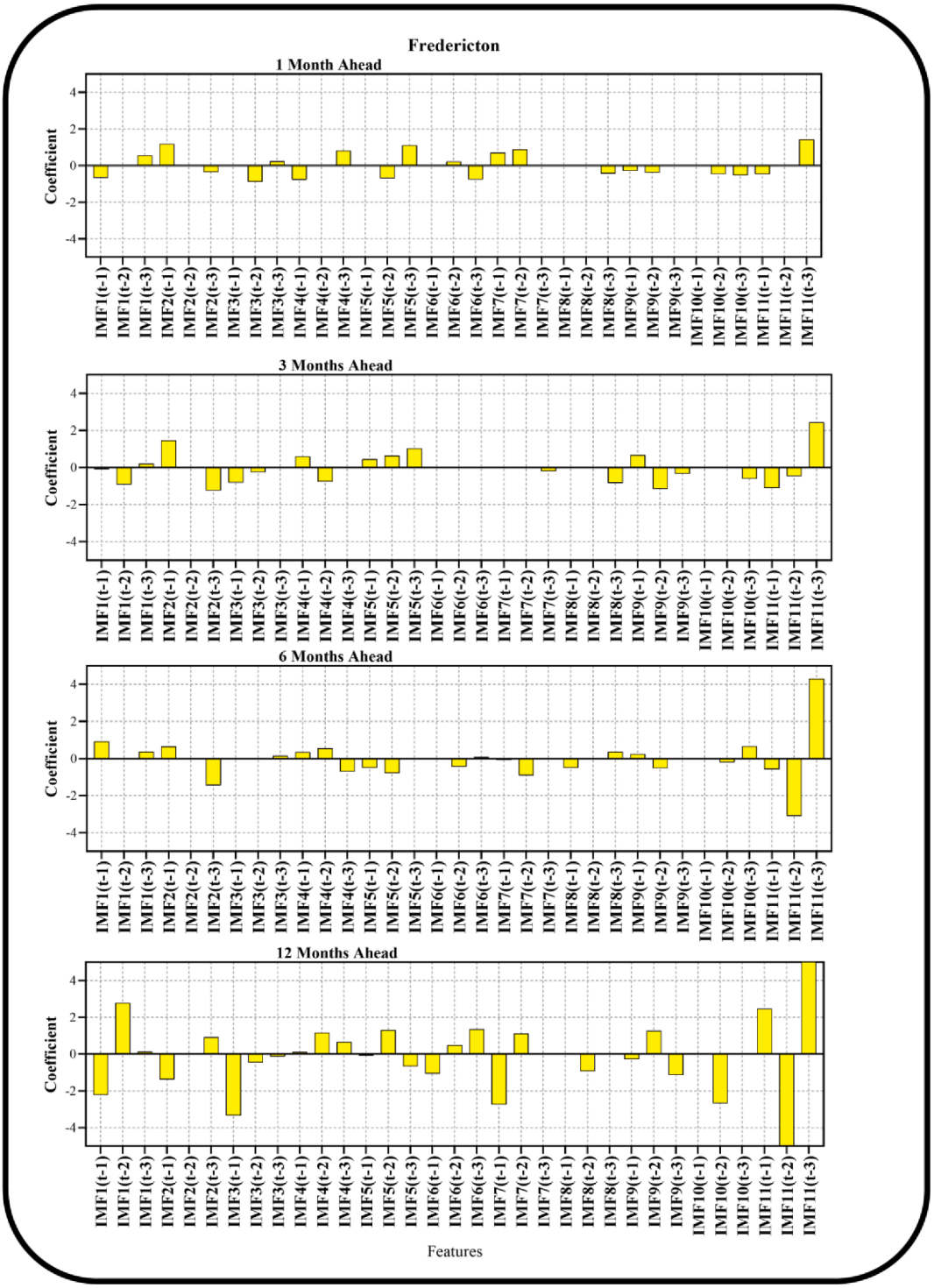


Fig. 8. Determining effective inputs for drought forecasting at Fredericton station by lasso regression feature selection algorithm.

stations. A thorough understanding of the TVF-EMD model can be obtained from the study conducted by Li et al. (2017).

2.3. Least absolute shrinkage and selection operator (LASSO) feature selection

The LASSO is an effective technique for selecting variables and creating regression models. This approach was suggested by Tibshirani (1996) in order to improve the precision of machine-learning model testing. LASSO modifies the model's residua' sum of squares (RSS) by including a penalty component (Tibshirani, 1996). If the RSS surpasses the penalized period, the variable will incur a further penalty; therefore, A regularization procedure will be used to reduce the penalty to zero for a portion of its coefficients. Only features with non-zero coefficients are retained as a consequence. The LASSO method may provide several advantages. At the outset, the regression model's interpretability can be improved by utilizing the LASSO function, which eliminates redundant or least correlated variables. Subsequently, researchers can analyze and deduce the most consequential explanatory variable while excluding extraneous variables that lack correlation.

Table 3
Optimized model adjustment for the forecasting of drought.

Study site	Models	Best parameters
Charlottetown single models	SVR	epsilon = 0.05, kernel = 'rbf', gamma = 0.5, C = 1
	ED-RVFL	num_nodes: 5, regular_para: 0.01, num_layer: 3
	RNN	Learning Rate: 0.00211, Batch Size: 4, Layers: 1, Epochs: 88,.Number of Neurons:15,.Training Algorithm: Adam, Dropout: 0
	XGBoost	'colsample_bytree: 0.8, learning_rate: 0.1, max_depth: 2, min_child_weight: 4, n_estimators: 100, nthread: 3, objective: reg: squarederror, subsample: 0.8
	RF	max_depth: 3, min_samples_leaf: 2, min_samples_split: 18, n_estimators: 10
	RVFL	num_nodes: 10, regular_para: 0.01
Charlottetown TVF based models	TVF-SVR	C = 10, epsilon = 0.05, gamma = 0.05, kernel = 'rbf'
	TVF-ED-RVFL	num_nodes: 5, regular_para: 0.0001, num_layer: 3
	TVF-RNN	Learning Rate: 0.0006, Batch Size: 2, Layers: 1, Epochs: 150,.Number of Neurons:12,. Training Algorithm: Adam, Dropout: 0
		{'colsample_bytree': 0.8, 'learning_rate': 0.1, 'max_depth': 2, 'min_child_weight': 4, 'n_estimators': 200, 'nthread': 3, 'objective': 'reg:squarederror', 'subsample': 0.8}
	TVF-RF	max_depth: 5, min_samples_leaf: 2, min_samples_split: 18, n_estimators: 20
	TVF-RVFL	num_nodes: 10, regular_para: 0.1

Study site	Models	Best parameters
Fredericton single models	SVR	C = 100 epsilon = 0.01, gamma = 0.5, kernel = 'rbf'
	ED-RVFL	num_nodes: 10, regular_para: 0.001, num_layer: 3
	RNN	Learning Rate: 0.0001, Batch Size: 4, Layers: 1, Epochs: 150,.Number of Neurons:20,. Training Algorithm: Adam, Dropout: 0
	XGBoost	{'colsample_bytree': 0.8, 'learning_rate': 0.2, 'max_depth': 2, 'min_child_weight': 4, 'n_estimators': 25, 'nthread': 3, 'objective': 'reg:squarederror', 'subsample': 0.9}
	RF	{'max_depth': 3, 'min_samples_leaf': 2, 'min_samples_split': 12, 'n_estimators': 10}
	RVFL	num_nodes: 5, regular_para: 0.01
Fredericton TVF based models	TVF-SVR	C = 100 epsilon = 0.01, gamma = 0.1, kernel = 'rbf'
	TVF-ED-RVFL	num_nodes: 3, regular_para: 0.01, num_layer: 4
	TVF-RNN	Learning Rate: 0.00225, Batch Size: 4, Layers: 1, Epochs: 250,.Number of Neurons:15,.Training Algorithm: Adam, Dropout: 0
	XGBoost	{'colsample_bytree': 0.8, 'learning_rate': 0.2, 'max_depth': 2, 'min_child_weight': 4, 'n_estimators': 200, 'nthread': 3, 'objective': 'reg:squarederror', 'subsample': 0.9}
	RF	{'max_depth': 7, 'min_samples_leaf': 1, 'min_samples_split': 12, 'n_estimators': 100}
	RVFL	num_nodes: 10, regular_para: 0.1

Additionally, overfitting is minimized by this method. A training model with excessive explanatory variables is prone to overfitting. While the model performs admirably during training, its prediction accuracy on the testing data is subpar. We will ultimately strike a balance between variation and bias. By increasing the penalty term, bias is augmented while variance is reduced, so achieving a balance between variance and bias and simplifying the model. It is assumed that the formula in a linear model is:

$$Y = X\beta + \varepsilon \tag{6}$$

Using explanatory variable X and response variable Y. Xis described as an $X_{n\times k}$ matrix, and Y can be defined as a $Y_{n\times 1}$ matrix in vector form. With an upper constraint of the sum of the absolute values of the model parameters, the LASSO minimizes the RSS. The upper bound is denoted as t. Now, the LASSO formula is represented as Eq. (7).

$$\text{minimize} \sum_{i=1}^{n} \left(y_i - \sum_j x_{ij}\beta_j\right)^2 \text{ subject to } \sum_{j=1}^{k} |\beta_j| < t \tag{7}$$

Thus, the optimization will also be expressed in vector form by:

$$\text{minimized}\left(\frac{|Y - X\beta|_2^2}{n}\right) \text{subject to} \sum_{j=1}^{k} |\beta_j| < t \tag{8}$$

Prior to the sum of the absolute values of the model parameters, a penalty term λ is appended. Greater penalty term magnitude will result in greater model shrinkage. As a result, the estimation of the parameters will be:

$$\widehat{\beta}(\lambda) = \underset{\beta}{\text{argmin}}\left(\frac{|Y - X\beta|_2^2}{n} + \lambda \sum_{j=1}^{k} |\beta_j|\right) \tag{9}$$

Considering the shrinkage, the j^{th} coefficient is assumed in this study as βj(λ) = 0. Subsequently, the model will exclude characteristics with a zero coefficient. Altering the penalty levels enables researchers to modify the quantity of picked characteristics. Hence, the LASSO feature selection technique seems valuable in machine learning for feature selection.

2.4. Machine learning technics

2.4.1. Simple RVFL and ensemble deep RVFL

The deep ensemble-based RVFL model was initially proposed by Cheng et al. (2021). RVFL model stability, robustness, and precision are enhanced with this method of ensemble learning and deep learning integration (Malik et al., 2023; Ren et al., 2016). Fig. 3 demonstrates the deep ensemble RVFL model's design. Similar to the normal RVFL, this model consists of stacks of additional hidden layers following the first hidden layer; its weights and biases are reserved and randomized (Cheng et al., 2021). By means of L hidden nodes, the deep ensemble RVFL model generates a output nodes that have been trained independently. Conversely, every output layer extracts characteristics from its matching intermediate hidden layer and input layer via direct connections. Following the removal of bias words to facilitate understanding, the first layer's output is as follows (i.e., $l = 1$) can be defined as (Cheng et al., 2021):

$$H^{(1)} = f\left(XW^{(1)}\right) \tag{10}$$

Here, $f(.)$ = non-linear activation function, and $W^{(l)} \in \mathbb{R}^{(M_o \times N)}$ = weights between the input layer and the first hidden layer. While for each layer $l > 1$, it is computed as (Shi et al., 2021):

$$H^{(l)} = f\left(\left[H^{(l-1)}X\right]W^{(l)}\right) \tag{11}$$

Here, $W^{(l)} \in \mathbb{R}^{(M_l \times N)}$ = weight between the previous layer and the current layer, and $M_l = M_o + N$. For l^{th} enhancement layer, the loss function is defined as:

$$Loss_l = \|\left[H^l, X\right]\beta_l - Y\|^2 + \lambda\|\beta_l\|^2 \tag{12}$$

Here, β_l is the output vector of the l^{th}. λ is the regularization parameter. Hence, the minimum value of the loss function can be obtained using the closed-form solution (Saunders & Gammerman, 1998):

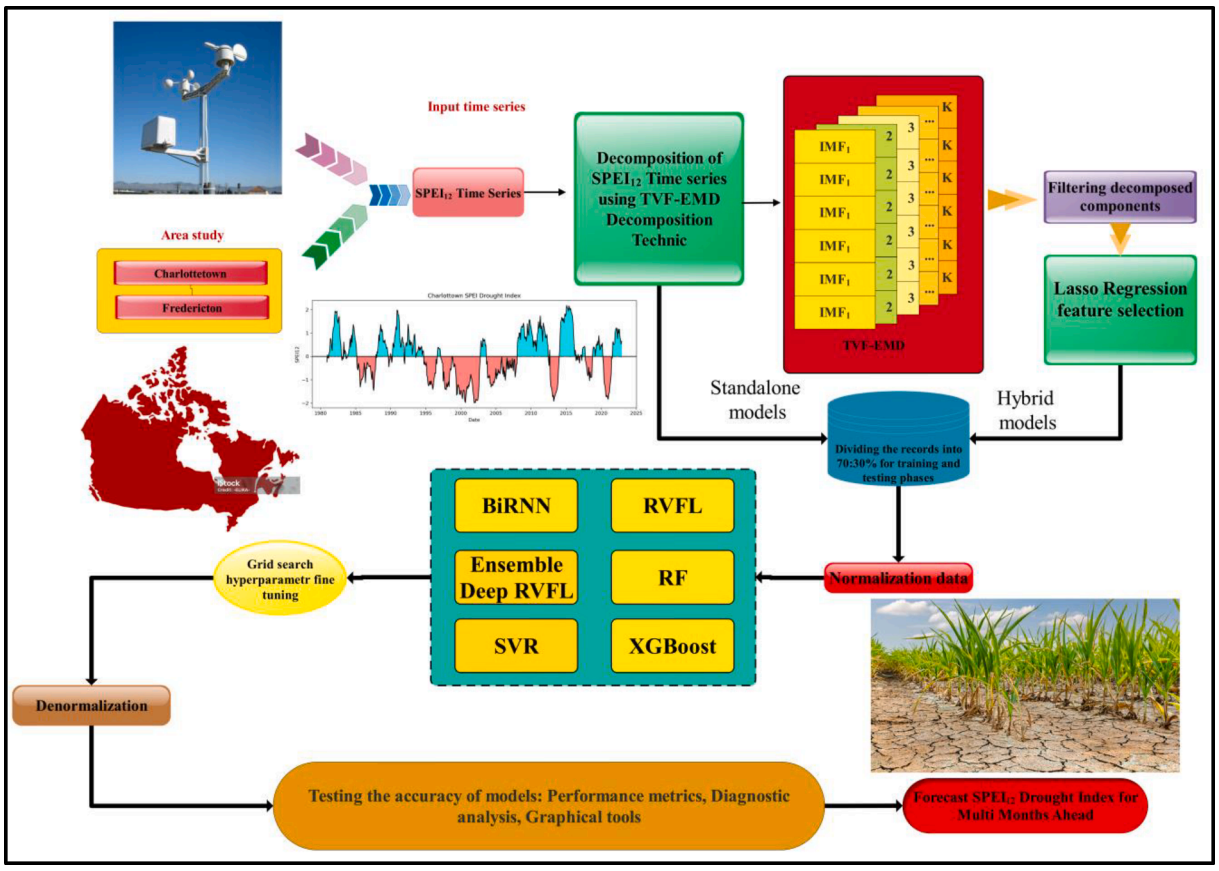


Fig. 9. The flowchart of the current study.

$$\beta_l = \left(\mathrm{D}^{\mathrm{T}} D + \lambda \mathrm{I}\right)^{-1} \mathrm{D}^{T} Y \tag{13}$$

The $\mathrm{D} = \left[\mathrm{H}^{l}, \mathrm{X}\right]$. The L forecast of the deep learning network can be obtained after all the values of the β_l obtained. The ultimate forecast of the deep ensemble RVFL model is thereafter an ensemble of layer outputs, and this may be acquired by majority vote or the average score approach (Du et al., 2022):

$$Y = \left[\mathrm{H}^{l}, \mathrm{X}\right]\beta_l \tag{14}$$

Algorithm 1: Pseudocode Ensemble Deep RVFL model

Input: N, the hidden dimension
L, the number of layers
λ_l, the regularization strength of l th layer
Output:$\beta = [\beta_1, \cdots, \beta_l]$
Initialize the $\mathbf{W}_1, \cdots, \mathbf{W}_L$ randomly
$l = 1$
for $l \leq L$ do
  if $l = 1$ then
    Compute $\mathbf{H}^1$ using $\mathbf{W}_1$ as in Eq. (10)
    Compute β_1 using λ_1 as in Eq. (13)
Else
    Compute $\mathbf{H}^l$ using $\mathbf{W}_l$ as in Eq. (11)
    Compute β_l using λ_l as in Eq. (13)
  End
$l++$
End

2.4.2. Support vector machine regression (SVR)

Support vector regression (SVR) was developed by Smola in 1996 as a variant of the support vector machine (SVM) that includes regression functions. Vapnik (2000) proposed the initial iteration of the Support Vector Machine (SVM) model, which hinged on the structural risk minimization (SRM) principle (Vapnik, 2000). The SVR model was developed in engineering, statistical, and predictive applications to address regression and forecasting issues. The fundamental concept underlying the SVR model is to identify a function dependence f(x) that makes the most minimal feasible use of all data (training x_i, objectives y_i) (Smola, 1996). The expression for the regression function of the SVR model is as follows:

$$f(x) = w \times \phi(x) + b \tag{15}$$

Where ϕ represents the transfer function, f signifies the regression function, and w and b denote the weight and bias, respectively. The regression issue may be expressed as

$$\text{Minimize}: \frac{1}{2}\left\|w\right\|^2 + C\sum\nolimits_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \tag{16}$$

$$\text{Subject to}: \begin{cases} y_i - f(x) \leq \varepsilon + \xi_i \\ f(x) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0,\ i = 1, 2, 3, \cdots, N \end{cases} \tag{17}$$

Where ε represents the boundary value, ξi and ξ_i^* denote the slack variables, and C represents the penalty parameter. Through the application of Lagrange multipliers, the optimization problem undergoes a substantial transformation into quadratic programming yielding a solution for a nonlinear regression function that can be expressed as follows:

$$f(x) = \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right)K(x, x_i) + b \tag{18}$$

The Kernel function is denoted as $K(x, x_i)$, and the dual variables are α_i and α_i^*. Several kernels, including linear, polynomial, sigmoid, and

Table 4
Results of one step ahead drought forecasting in Charlottetown. Note that the best model is boldfaced (black).

Model	Data	R	RMSE	KGE	NSE	$U_{95\%}$	IA
TVF-RNN	Train	0.9752	0.2169	0.0509	0.9469	0.5920	0.9866
	Test	0.9634	0.2988	0.6853	0.9250	0.8246	0.9795
TVF-ED-RVFL	**Train**	**0.9992**	**0.0379**	**0.9988**	**0.9984**	**0.1050**	**0.9996**
	Test	**0.9995**	**0.0352**	**0.9986**	**0.9990**	**0.0977**	**0.9997**
TVF-SVR	Train	0.9983	0.0553	0.9783	0.9965	0.1534	0.9991
	Test	0.9951	0.1289	0.7221	0.9860	0.3485	0.9963
TVF-RF	Train	0.9717	0.2281	0.9412	0.9042	0.6327	0.9838
	Test	0.8421	0.6112	0.6860	0.1370	1.6796	0.8904
TVF-XGBoost	Train	0.9970	0.0743	0.9938	0.9826	0.2062	0.9984
	Test	0.8979	0.5303	0.7636	−0.0087	1.4430	0.9192
TVF-RVFL	Train	0.9991	0.0389	0.9983	0.9987	0.1078	0.9996
	Test	0.9994	0.0386	0.9987	0.9771	0.1069	0.9997
RNN	Train	0.9546	0.2892	0.9146	0.9055	0.8021	0.9766
	Test	0.9572	0.3235	0.8805	0.9120	0.8979	0.9781
ED-RVFL	Train	0.9502	0.2933	0.9291	0.9028	0.8137	0.9739
	Test	0.9617	0.2999	0.9155	0.9244	0.8326	0.9797
SVR	Train	0.9514	0.2899	0.8690	0.9051	0.8039	0.9743
	Test	0.9546	0.3273	0.8113	0.9100	0.9075	0.9753
RF	Train	0.9537	0.2833	0.9094	0.9058	0.7857	0.9755
	Test	0.9559	0.3232	0.9122	0.7523	0.8949	0.9762
XGBoost	Train	0.9634	0.2525	0.9280	0.9343	0.7003	0.9808
	Test	0.9549	0.3257	0.9108	0.7804	0.9024	0.9760
RVFL	Train	0.9505	0.2924	0.9034	0.9299	0.8110	0.9741
	Test	0.9597	0.3080	0.9202	0.8984	0.8553	0.9783

radial basis functions (RBF), comprise the SVR model's structure. The RBF kernel was utilized in the provided study owing to its superior accuracy and robustness in comparison to other kernel functions; it is regarded as the most often employed function in the hydrologic application of SVR (Kalteh, 2013; Khan & Coulibaly, 2006). The RBF can be formulated as

$$K(x, x_i) = \exp\left(-\gamma \|x_i - x\|^2\right) \tag{19}$$

The value "γ," which "represents the kernel function, is the primary determinant of whether the performance of support vector regression is enhanced or diminished.

2.4.3. Recurrent neural network

One type of ANN or artificial neural network that allows ordered consecutive datasets to be occupied successively at the separate time-step onward broadcast is the Recurrent Neural Network (RNN). All sequencing data will share the learnable weight groups. In order for the RNN to be a compatible answer to time-series difficulties, it is assumed that it will absorb arrangement evidence among each signal dataset with its situation (Fayer et al., 2023). Normally, RNN at separately a time-step i can distribute two different mediums, which are called Hidden States h_i and Outcomes y_i. These RNN structures utilizing both are referred to as complete RNNs h_i and y_i interested in scheming of the unseen states h_{i+1} with outcomes y_{i+1} at the following time stage. An RNN attractive individual h_i into the next time step calculation is intended through (Elman, 1990), while Elman network is another name for it. The RNN requires just y_i additionally known as Elman network. A Jordan RNN, which requires individual (Cruse, 1996). Here, Tensorflow, the most popular deep learning framework for recurrent neural networks, was used:

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h) \tag{20}$$

$$y_t = \sigma(W_y h_t + b_y) \tag{21}$$

Where σ is the sigmoid function, W_h, W_y and U_h are weights that can be used to generate the output, transmitted hidden state, and hidden states from the previous time step, where b_h with b_y are bias in related calculations of hidden states and output (Zheng et al., 2023). The characteristics of the recurrent neural networks demonstrate their successful capacity for learning from time series or sequential data, with an improved capacity to discriminate between absolute and relative temporal positions. Fig. 4 shows a schematic structure of the RNN model.

2.4.4. Random forest

Random forest is classified as a supervised machine-learning technique that may be employed for either classification or regression tasks (Breiman, 1999). The random forest method, known for its strong ensemble learning capabilities, holds great potential for a wide range of applications. Ensemble learning mitigates the problem of overfitting in individual decision trees by aggregating the predictions of numerous decision trees, resulting in enhanced accuracy of the overall model. Random forest methods excel in handling nonlinear connections and

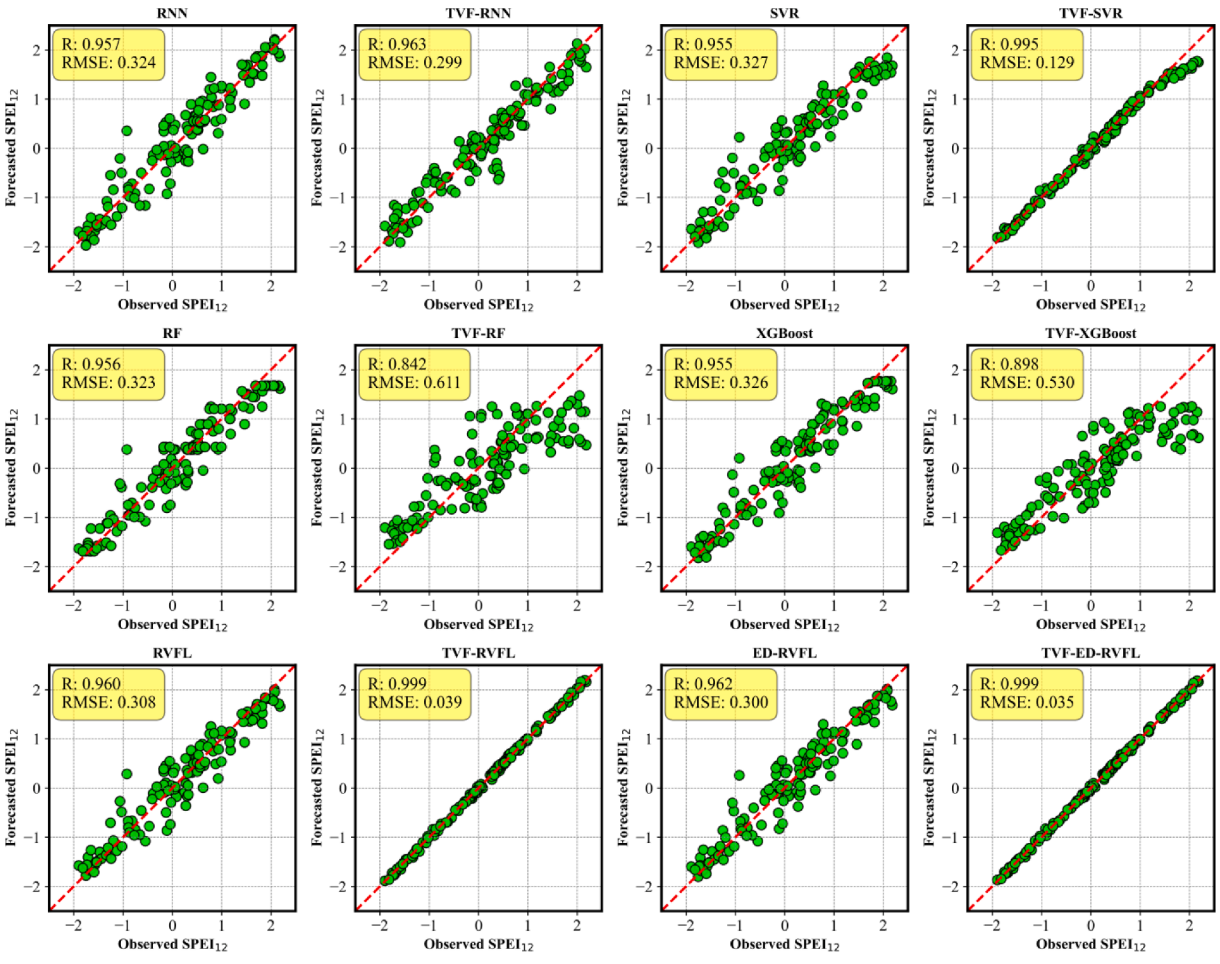


Fig. 10. Scatter plots of observed $SPEI_{12}$ drought index versus Forecasted drought index in Charlottetown station.

complicated data, making them superior to other machine learning models. On the other hand, logistic regression and linear regression models are more appropriate for issues that are linearly separable or need linear fitting (Evans et al., 2011). Furthermore, Random Forests can efficiently manage data with many dimensions, eliminating the need for a time-consuming feature selection procedure. RFs are well-suited for real-world issues that include a large number of features and a huge amount of data (Xue et al., 2024).

2.4.5. XGBoost

XGBoost, short for extreme gradient boosting, is a machine learning technique that is designed to efficiently boost decision trees. Chen and Guestrin created it based on gradient-boosting decision trees (GBDT) (Chen & Guestrin, 2016). XGBoost combines many decision trees, which are considered weak learners, to create a powerful learner. XGBoost stands out for its ability to do multi-threaded parallel and distributed computing. This allows it to execute quickly on datasets with billions of samples or in memory-limited environments (He et al., 2024).

The XGBoost technique enhances running performance by parallelizing the feature selection process for each tree. The samples are selected using a random sampling strategy, which helps to somewhat mitigate the problem of overfitting. The GBDT algorithm follows the principle of training a tree on a training set and sampled labels. This tree is then used to predict the values of the training set samples. The predicted values are subtracted from the actual labels to obtain the residuals. These residuals are then used to train the second tree. Following the training process, the residual of each individual sample is acquired. Subsequently, the nth tree is trained using an analogous approach. XGBoost is a more efficient technique compared to GBDT. It achieves ensemble learning by using gradient boosting to combine numerous CART submodules (Dong et al., 2022).

2.5. Goodness-of-fit indices

Performance metrics, generally referred to as error measures, are fundamental components of evaluation frameworks across multiple disciplines. The performance of ML models was assessed in the present work using six distinct assessment metrics: correlation coefficient (R), root mean square error (RMSE), uncertainty with 95 % confidence level ($U_{95\%}$), Nash–Sutcliffe efficiency index (NSE), Kling-Gupta efficiency (KGE), and I_A (Willmott indicator of agreement) (Willmott, 1982). The mathematical formulations of used metrics are as follows.

$$R = \frac{\sum_{i=1}^{N}\left(SPEI_{o,i} - \overline{SPEI_o}\right)\left(SPEI_{for,i} - \overline{SPEI_{for}}\right)}{\sqrt{\sum_{i=1}^{N}\left(SPEI_{o,i} - \overline{SPEI_o}\right)^2 \sum_{i=1}^{N}\left(SPEI_{for,i} - \overline{SPEI_{for}}\right)^2}} \tag{22}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(SPEI_{o,i} - SPEI_{for,i}\right)^2} \tag{23}$$

$$IA = 1 - \frac{\sum_{i=1}^{N}\left(SPEI_{o,i} - SPEI_{for,i}\right)^2}{\sum_{i=1}^{N}\left(\left|SPEI_{o,i} - \overline{SPEI_o}\right| + \left|SPEI_{o,i} - \overline{SPEI_o}\right|\right)^2} \tag{24}$$

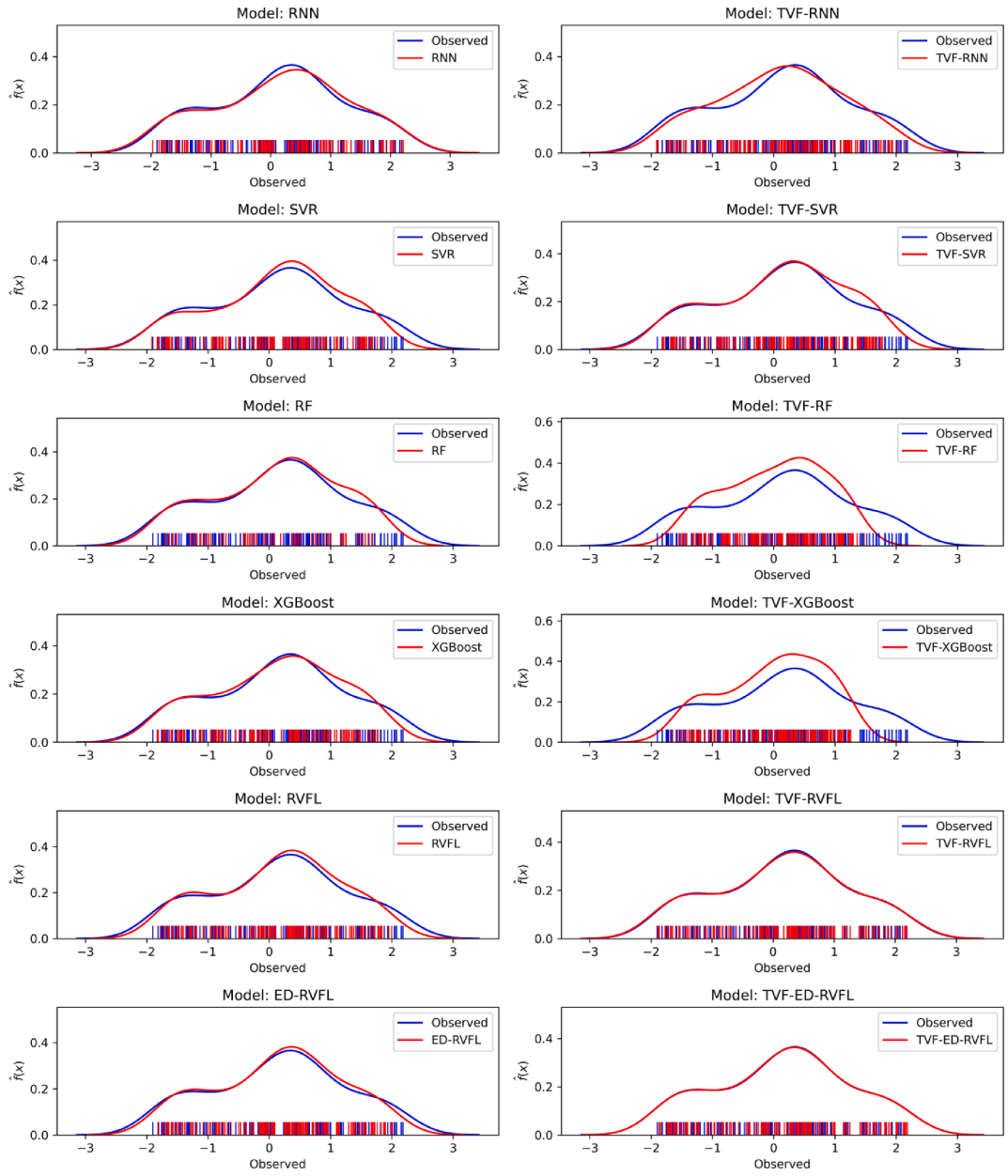


Fig. 11. Rug plots of observed and forecasted values of ML Models at Charlottetown station.

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(SPEI_{o,i} - SPEI_{for,i}\right)^2}{\sum_{i=1}^{N}\left(SPEI_{o,i} - \overline{SPEI_o}\right)^2} \tag{25}$$

$$KGE = 1 - \sqrt{(R-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{26}$$

$$U_{95\%} = 1.96\sqrt{SD_e^2 + RMSE^2}$$

here $SPEI_{o,i}$ and $SPEI_{for,i}$ are observed and predicted SPEI values, respectively. $\overline{SPEI_o}$ and $\overline{SPEI_{for}}$ are averaged data of observed and forecasted SPEI values. Additionally, the comparative deviations between the expected and actual values are denoted by the α, where β is the ratio between the average values projected and observed. When R, NSE, KGE, and I_A are both equal to 1, the model exhibits exceptional performance. The standard deviation of the difference between the predicted and actual values is denoted by SD_e^2.

The correlation coefficient (R) measures the linear relationship between observed and predicted values, indicating the model's ability to capture data trends. RMSE is sensitive to large errors and provides an intuitive measure of prediction accuracy in the same units as the data. Uncertainty with 95 % confidence level ($U_{95\%}$) quantifies the range within which the true predictions lie, offering insights into model reliability. NSE evaluates how well the model predicts compared to the mean of observed data, with values closer to 1 indicating better performance. KGE combines correlation, bias, and variability to measure model accuracy comprehensively. Lastly, I_A assesses the degree of agreement between observed and predicted values, with higher values indicating better performance. Together, these metrics ensure a robust evaluation of the model, confirming its accuracy, reliability, and comprehensive predictive capabilities.

2.6. Model development and figuring out model tuning

In this investigation, a new intelligent framework has been invented for multi-temporal (one, three, six, and twelve months ahead) monthly

Table 5
Results of one step ahead drought forecasting in Fredericton. Note that the best model is boldfaced (black).

Model	Data	R	RMSE	KGE	NSE	$U_{95\%}$	IA
TVF-RNN	Train	0.9941	0.1196	0.7965	0.9852	0.3243	0.9961
	Test	0.9712	0.1910	0.9117	0.9402	0.5304	0.9835
TVF-ED-RVFL	**Train**	**0.9988**	**0.0475**	**0.9983**	**0.9977**	**0.1318**	**0.9994**
	Test	**0.9974**	**0.0560**	**0.9945**	**0.9949**	**0.1556**	**0.9987**
TVF-SVR	Train	0.9998	0.0206	0.9938	0.9996	0.0571	0.9999
	Test	0.9938	0.0886	0.9690	0.9871	0.2453	0.9967
TVF-RF	Train	0.9824	0.1877	0.9636	0.9396	0.9903	0.5206
	Test	0.8704	0.3867	0.7547	0.7910	0.9246	1.0712
TVF-XGBoost	Train	0.9989	0.0470	0.9977	0.9929	0.9994	0.1305
	Test	0.9354	0.2793	0.8721	0.8657	0.9634	0.7732
TVF-RVFL	Train	0.9985	0.0544	0.9969	0.9971	0.9992	0.1509
	Test	0.9972	0.0607	0.9940	0.9701	0.9985	0.1658
RNN	Train	0.9281	0.3663	0.8989	0.8613	1.0159	0.9616
	Test	0.9013	0.3782	0.8325	0.8076	1.0452	0.9467
ED-RVFL	Train	0.9282	0.3659	0.8972	0.8616	1.0149	0.9616
	Test	0.9046	0.3707	0.8408	0.8151	1.0256	0.9483
SVR	Train	0.9354	0.3494	0.8589	0.8738	0.9683	0.9664
	Test	0.8889	0.4145	0.8288	0.7688	1.1458	0.9408
RF	Train	0.9315	0.3587	0.8670	0.8822	0.9622	0.9948
	Test	0.7926	0.5390	0.5236	0.6226	0.8772	1.4621
XGBoost	Train	0.9362	0.3473	0.8753	0.8828	0.9646	0.9633
	Test	0.8943	0.3926	0.7926	0.7966	0.9418	1.0820
RVFL	Train	0.9259	0.3716	0.8572	0.8950	0.9603	1.0308
	Test	0.9024	0.3767	0.8090	0.8196	0.9467	1.0399

forecasting of the $SPEI_{12}$ drought index in two Atlantic provinces of Canada, Prince Edward Island, and New Brunswick. The provided framework is built on the Lasso regression feature selection EMD-TVF and Ensemble Deep RVFL schemes, generated in MATLAB 2023 and Python 3.11 environments. Here, the case study focuses on two critical Atlantic Canada cities that play a significant role in the country's agricultural aspects, like potato production, seed production, and agricultural education. The available datasets for monthly $SPEI_{12}$ are over 43 years (1980–2022). The TVF-EMD preprocessing was performed in MATLAB, whereas the second preprocessing stage associated with Lasso feature selection and all the ML models were performed in Python using Scikit-learn (Pedregosa et al., 2011), Keras (Arnold, 2017), and TensorFlow (Abadi, 2016) libraries platforms. Here, the first 70 % of the whole time series of $SPEI_{12}$ was allocated as the training dataset (30 %), and the rest of the time series was used to validate the implemented frameworks.

First, after filling in the missing datasets, the most influential antecedent information (input time lags) for every horizon of $SPEI_{12}$ was designated using the autocorrelation function (AFC) and partial autocorrelation function (PACF) as the most renowned techniques among hydrological researchers (Jamei et al., 2024). The outcomes of antecedent information related to all the preferred horizons are shown in Fig. 5. According to the ACF and PACF, the first three lags of $SPEI_{12}$ in both cases studied, owing to the higher value of the autocorrelation coefficient measures (more than 75 %) and temporal dependency, are qualified as inputs to participate in the second preprocessing stage (signal decomposition). As the drought indices signal naturally has a lot of non-stationery and complexities, reaching the permissible and promising forecasting accuracy just by applying the standalone ML framework is unlikely and arduous (Jamei et al., 2023). For this reason, emerging signal decomposition techniques with high mathematical potential have been widely considered for integration with the ML model in hydrological modeling. Here, the TVF-EMD, a new decomposition technique, was implemented to reduce the complexity of the $SPEI_{12}$ signals for each case study (Charlottetown and Fredericton). The most crucial setting parameters of TVF-EMD are the number of decomposed IMFs, B-spline order, and stopping criterion, which for all the situations in modeling are listed in Table 2. Using a trial-and-error procedure, the number of decomposed IMFs related to both stations was estimated to be equal to 11. Fig. 6 illustrates the decomposed drought index ($SPEI_{12}$) signal for Charlottetown (left side) and Fredericton (right side).

The third stage of preprocessing before the feed ML model is filtering the redundant sub-sequences from the input feature pool to reduce the computational cost and enhance the accuracy of the $SPEI_{12}$ forecasting process. For this purpose, the Lasso regression feature selection was adopted to acquire the optimal input feature of four scenarios (horizons) associated with the Charlottetown and Fredericton stations. Figs. 7 and 8 depict the effective input features of the $SPEI_{12}$ (t + 1), $SPEI_{12}$ (t + 3), $SPEI_{12}$ (t + 6), and $SPEI_{12}$ (t + 12) scenarios. In the Charlottetown station, 21, 22, 20, and 23 sub-components among all 33 pool members have been merited for the $SPEI_{12}$ (t + 1), $SPEI_{12}$ (t + 3), $SPEI_{12}$ (t + 6), and $SPEI_{12}$ (t + 12) scenarios, respectively. In contrast, in the

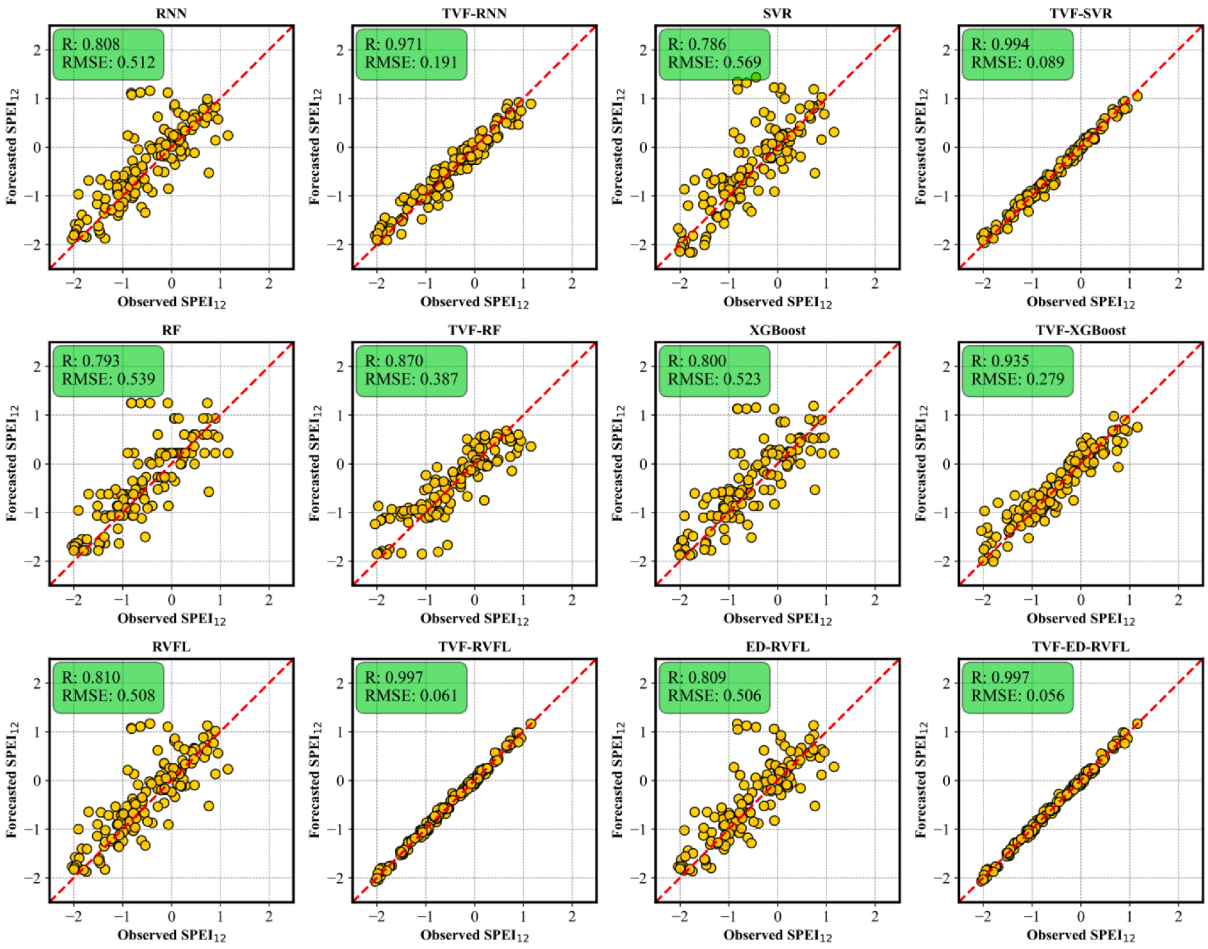


Fig. 12. Scatter plots of observed $SPEI_{12}$ drought index versus Forecasted drought index in Fredericton station.

Fredericton station, 21, 22, 23, and 26 sub-components have deserved for the $SPEI_{12}$ (t + 1), $SPEI_{12}$ (t + 3), $SPEI_{12}$ (t + 6), and $SPEI_{12}$ (t + 12) scenarios, respectively.

Forecasting hydrological problems using complementary intelligent frameworks is intrinsically complicated enough that it cannot overlook the optimal-tuning hyperparameters of ML models. The inaccurate tuning hyperparameters of ML and particularly deep learning models might result in disappointing outcomes and can be seriously tarnished due to applying such tough preprocessing.

Thus, the process above is considered one of the most critical stages of model development. The grid search tuning scheme is a reliable and popular scheme that adequately matches Python models. The grid search tool has been implemented to optimize the SVR, ED-RVFL, and RNN model hyperparameters. The optimized value of all the model settings in complementary and standalone frameworks is reported in Table 3 for each case study. In this regard, the "rbf" kernel function, epsilon, and gamma are the primary hyperparameters for the SVR model. The gamma and epsilon ranges, the SVR's main hyperparameters, are acquired (0.05–0.5) and (0.01–0.05), respectively. Also, the significant hyperparameters related to the RNN are the layer, neuron number, training algorithm, learning rate value, and epoch. Here, the range of learning rate and neuron number are computed (0.0006–0.00225) and (10–20) using the grid search strategy. In contrast, the TVF-ED-RVFL, as the primary ML model, depends on the node number (num_nodes), regulation parameter (regular_para), and the number of layers. According to Table 3, the best optimal ranges of num_nodes and regular_para fall into the (3–10) and (0.0001–0.1), respectively. Fig. 9 shows the whole framework of the current study.

3. Results and discussion

The TVF-ED-RVFL, TVF-RVFL, TVF-RF, TVF-XGBoost, TVF-RNN, TVF-SVR, ED-RVFL, RNN, XGBoost, RF, RVFL, and SVR models were judged and compared against each other in both train and test periods to forecast $SPEI_{12}$ index for Charlottetown and Fredericton stations. A set of statistical goodness-of-fit metrics, namely R, RMSE, NSE, KGE, IA, and $U_{95\%}$, and several types of diagnostic plots were also used to examine the precision of these models.

Table 4 displays the goodness-of-fit values to inspect the standalone and hybrid TVF-based frameworks to forecast $SPIE_{12}$ in Charlottetown and Fredericton stations. The TVF-ED-RVFL model expressed highest accuracy by means of (R = 0.9992, RMSE = 0.0379, NSE = 0.9984, KGE = 0.9988, IA = 0.9996, $U_{95\%}$ = 0.1050) and (R = 0.9995, RMSE = 0.0352, NSE = 0.9990, KGE = 0.9986, IA = 0.9997, $U_{95\%}$ = 0.0977), following with TVF-RVFL (R = 0.9991, RMSE = 0.0389, NSE = 0.9987, KGE = 0.9983, IA = 0.9996, $U_{95\%}$ = 0.1078), and (R = 0.9994, RMSE = 0.0386, NSE = 0.9771, KGE = 0.9987, IA = 0.9997, $U_{95\%}$ = 0.1069) in training and testing periods to forecast $SPEI_{12}$. The hybrid TVF-RVFL and TVF-RNN are reasonably good at forecasting SPEI12 but could not surpass the TVF-ED-RVFL model. ED-RVFL is better than the SVR, RF, XGBoost, RVFL, and RNN models in the standalone version of the models. By comparing the TVF-based hybrid and standalone models, it is noted that the TVF-based hybrid version of the models, except TVF-RF and TVF-XGBoost, outperformed the standalone counterparts in forecasting $SPEI_{12}$ in Charlottetown station.

The scatter plots in Fig. 10 inspect the capability of machine learning models between the forecasted and observed $SPEI_{12}$ for Charlottetown.

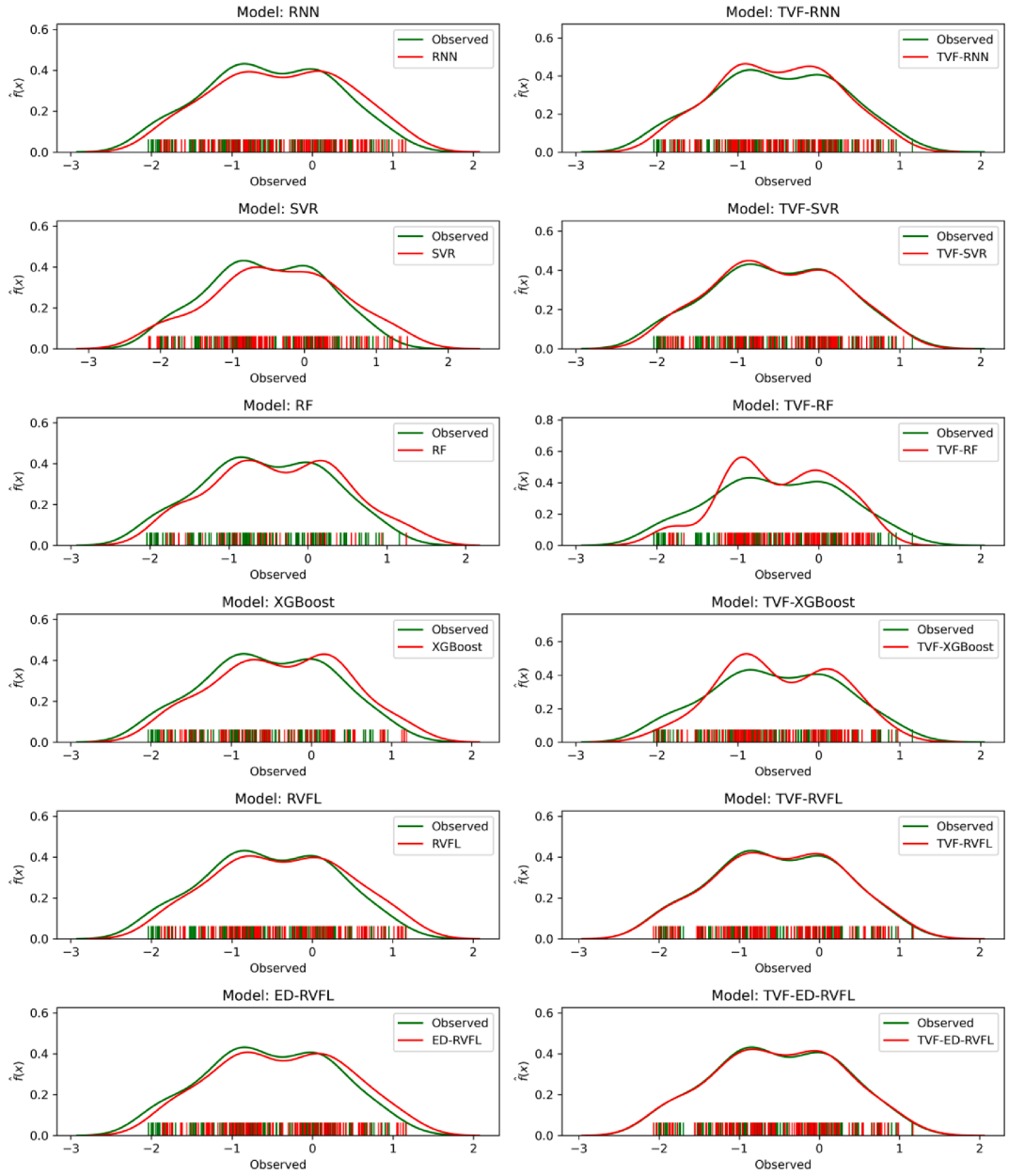


Fig. 13. Rug plots of observed and forecasted values of ML Models at Fredericton station.

The scatter diagrams supplement the models' forecasting competence by embedding the R and RMSE metrics. By inspecting, the TVF-ED-RVFL model reached the highest accuracy level by gaining R = 0.9995, RMSE = 0.0352, followed by TVF-RFL, TVF-SVR, and TVF-RNN to forecast $SPEI_{12}$. Moreover, the hybrid TVF-based version of the models demonstrates better precision in terms of scatter plots along with R and RMSE metrics over the standalone models. The accuracy improvement can be seen up to 4 % in hybrid models (i.e., specifically in the case of TVF-ED-RVFL) against standalone counterparts. Hence, Fig. 10 verified that the TVF-ED-RVFL model is the most accurate $SPEI_{12}$ forecasting model at Charlottetown.

Fig. 11 exhibits the Rug-distribution of the predicted $SPEI_{12}$ generated by the hybrid and standalone versions of the models vs. the observed $SPEI_{12}$. The density distribution of the TVF-ED-RVFL and TVF-RVFL models for the Charlottetown station appeared to be precisely overlying with the observed $SPEI_{12}$ as compared to other models. The ED-RVFL and RVFL methods portray a relatively better density distribution against observed $SPEI_{12}$ than the SVR, RF, XGBoost, and RNN models in the standalone group of models. Fig. 11 further established the better predictability of the TVF-ED-RVFL in terms of Rug-distribution plots to forecast $SPEI_{12}$.

The TVF-ED-RVFL model in Table 5 also accomplished the maximum accuracy values to forecast $SPEI_{12}$ at Fredericton station compared to other machine learning models. The numerical values obtained by the TVF-ED-RVFL model are displayed as R = 0.9988, 0.9974; RMSE = 0.0475, 0.0560; NSE = 0.9977, 0.9949; KGE = 0.9983, 0.9945; IA = 0.9994, 0.9987; and $U_{95\%}$ = 0.1318, 0.1556 in both training and testing periods to forecast $SPEI_{12}$. Moreover, the TVF-RVFL and TVF-SVR are also reasonably good at acquiring a second position, while TVF-RNN comes in fourth position in forecasting $SPEI_{12}$. TVF-XGBoost and TVF-RF ranked fifth and sixth, respectively. It is again true that the hybrid version of the models surpasses the standalone models for better accuracy. However, the TVF-ED-RVFL model generally proved to be a precise $SPEI_{12}$ performer for Fredericton station.

The scatter plots for Fredericton station in Fig. 12 discovered the model evaluation based on the hybrid and standalone models between

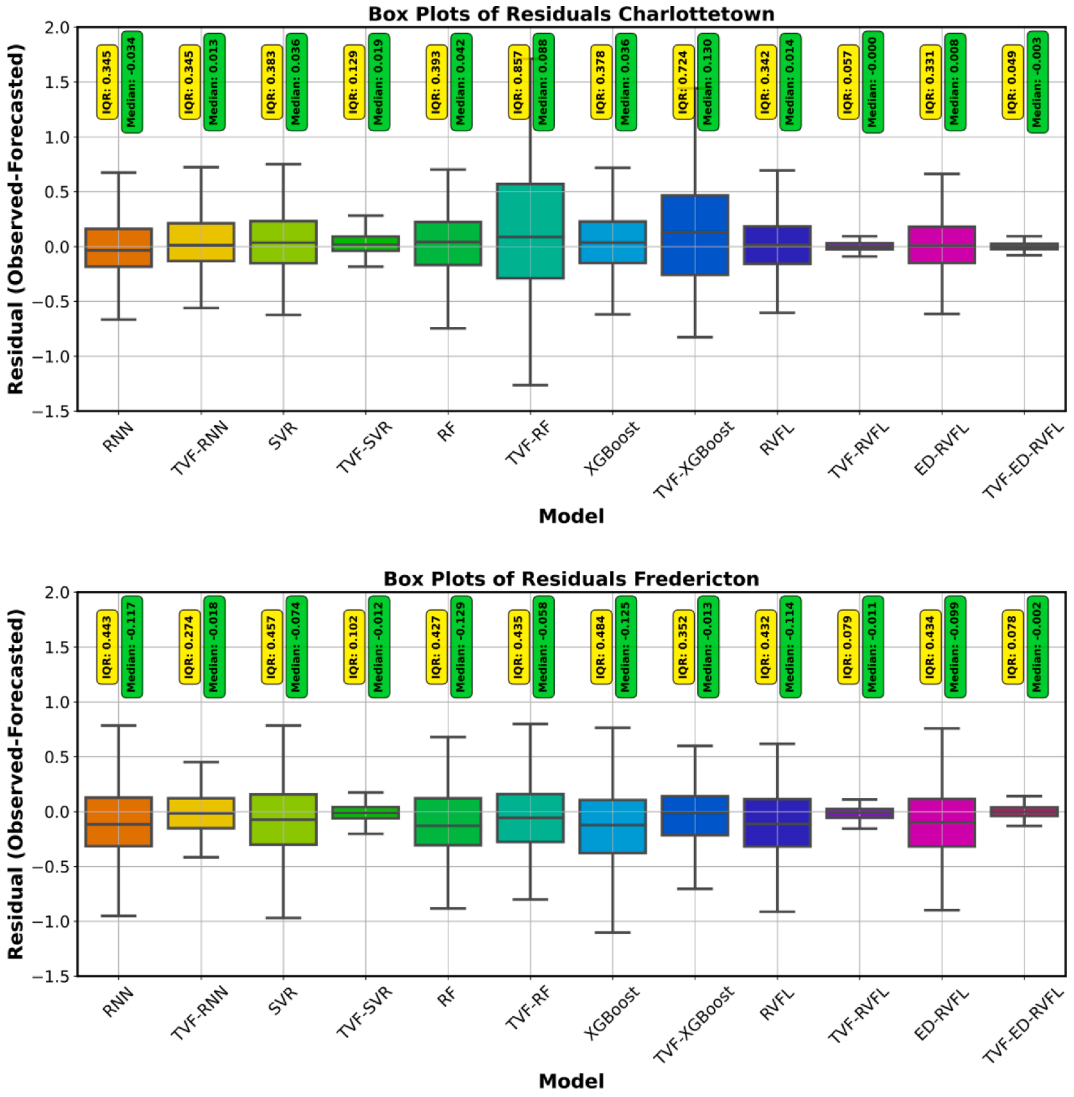


Fig. 14. Box plots of Residual values for Charlottetown and Fredericton stations.

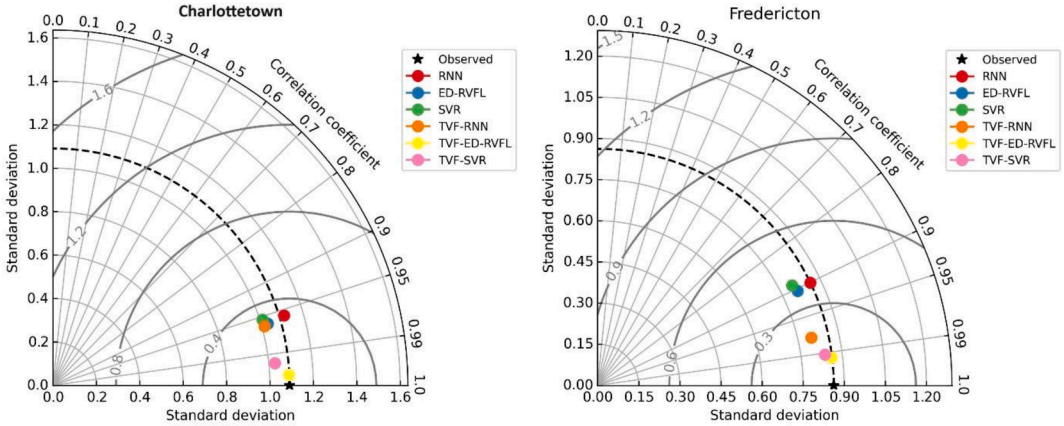


Fig. 15. Taylor diagrams for Charlottetown and Fredericton stations.

the observed and forecasted $SPEI_{12}$ in combination with R and RMSE values. The TVF-ED-RVFL expressed higher accuracy to forecast $SPEI_{12}$ with R = 0.9974, and RMSE = 0.0560 as compared to TVF-RVFL (R = 0.9972, RMSE = 0.0607), TVF-SVR (R = 0.9938, RMSE = 0.0886) and TVF-RNN (R = 0.9712, RMSE = 0.1910). The standalone models are relatively low in forecasting $SPEI_{12}$ at Fredericton station as compared to their hybrid versions. The TVF-ED-RVFL yet again offered improved accuracy in forecasting $SPEI_{12}$ compared to other models.

The Rug-Histogram plot in Fig. 13 exhibits the forecasting $SPEI_{12}$ for Fredericton station generated by the TVF-ED-RVFL, TVF-SVR, TVF-RNN,

Table 6
Results of Multi steps ahead drought forecasting for Charlottetown station.

Model	Steps Ahead	Data	R	RMSE	NSE	KGE	IA	U95%
Single	1	Train	0.9502	0.2933	0.9028	0.9291	0.9739	0.8137
		Test	0.9617	0.2999	0.9244	0.9155	0.9797	0.8326
	3	Train	0.8633	0.4749	0.7453	0.8040	0.9223	1.3172
		Test	0.8545	0.5721	0.7282	0.6666	0.9150	1.5869
	6	Train	0.7109	0.6647	0.5053	0.5839	0.8122	1.8436
		Test	0.6083	0.8710	0.3622	0.3051	0.7481	2.4148
	12	Train	0.3908	0.8615	0.1526	0.1301	0.5015	2.3895
		Test	−0.2229	1.2405	−0.2935	−0.6235	0.1687	3.4354
TVF-based	1	Train	0.9992	0.0379	0.9984	0.9988	0.9996	0.1050
		Test	0.9995	0.0352	0.9990	0.9986	0.9997	0.0977
	3	Train	0.9926	0.1147	0.9852	0.9894	0.9963	0.3182
		Test	0.9924	0.1485	0.9815	0.9315	0.9956	0.4123
	6	Train	0.9854	0.1608	0.9710	0.9768	0.9926	0.4460
		Test	0.9833	0.2028	0.9654	0.9696	0.9915	0.5633
	12	Train	0.9578	0.2699	0.9168	0.9209	0.9773	0.7486
		Test	0.9242	0.4184	0.8529	0.8441	0.9603	1.1612

Table 7
Results of multi steps ahead drought forecasting for Fredericton station.

Model	Steps Ahead	Data	R	RMSE	NSE	KGE	IA	U95%
Single	1	Train	0.9282	0.3659	0.8616	0.8972	0.9616	1.0149
		Test	0.9046	0.3707	0.8151	0.8408	0.9483	1.0256
	3	Train	0.7763	0.6185	0.6026	0.6821	0.8638	1.7155
		Test	0.6379	0.6374	0.3338	0.4772	0.7729	1.7392
	6	Train	0.5649	0.8064	0.3191	0.3815	0.6845	2.2368
		Test	0.4462	0.7895	−0.0220	0.0363	0.6028	2.0814
	12	Train	0.2010	0.9253	0.0393	−0.1531	0.2310	2.5666
		Test	−0.0024	1.0361	−0.7605	−0.7933	0.4424	2.5690
TVF-based	1	Train	0.9988	0.0475	0.9977	0.9983	0.9994	0.1318
		Test	0.9974	0.0560	0.9949	0.9945	0.9987	0.1556
	3	Train	0.9917	0.1258	0.9835	0.9876	0.9958	0.3491
		Test	0.9846	0.1369	0.9693	0.9681	0.9921	0.3801
	6	Train	0.9784	0.2020	0.9573	0.9655	0.9889	0.5603
		Test	0.9569	0.2321	0.9117	0.8811	0.9753	0.6413
	12	Train	0.8285	0.5294	0.6856	0.7369	0.8968	1.4684
		Test	0.8293	0.5118	0.5705	0.4143	0.8389	1.3483

TVF-RVFL, TVF-XGBoost, TVF-RF, ED-RVFL, RVFL, XGBoost, RF, SVR, and RNN models. The distribution of the TVF-ED-RVFL and TVF-RVFL models appeared to be accurate, with the observed $SPEI_{12}$ overlapping with the density compared to other benchmarking models. Again, the hybrid models are better in relation to the standalone models by producing the same density distributions. Thus, Fig. 13 confirms that the TVF-ED-RVFL performs better for the Fredericton station than other benchmarking models.

The residual box plots in Fig. 14 deliver a more concrete and conclusive comparison of the models' forecasting ability for both Charlottetown and Fredericton stations along with the interquartile range (IQR) using the hybrid TVF-ED-RVFL, TVF-SVR, TVF-RNN, ED-RVFL, SVR and RNN models to forecast $SPEI_{12}$. It is evident that the TVF-ED-RVFL model offered more accurate forecasts with lower residual box plot distribution and IQR = 0.049 (Charlottetown station) and 0.078 (Fredericton stations) to forecast $SPEI_{12}$, followed by TVF-RVFL model. The other models are relatively lower in forecasting $SPEI_{12}$ for both stations based on residual box plots and IQR. Thus, TVF-ED-RVFL models attain precise $SPEI_{12}$ forecasting at both Charlottetown and Fredericton stations.

Fig. 15 presented the Taylor diagram to assess the efficiency based on the observed and forecasted $SPEI_{12}$ using TVF-ED-RVFL, TVF-SVR, TVF-RNN, ED-RVFL, SVR, and RNN for both Charlottetown and Fredericton stations. Taylor diagrams are very useful for comprehensively evaluating the models' comparability, centered on standard deviation and correlation coefficient. For the Charlottetown station, the TVF-ED-RVFL model positioned closely to the observed $SPEI_{12}$ by registering a correlation coefficient between 0.99 and 1 with a standard deviation (of 1.0 to 1.2). This gives the TVF-ED-RVFL model the first ranking in the Taylor diagram, and TVF-SVR achieved the second-ranking. The hybrid models are rationally satisfactory as compared to the standalone models, but their forecasting accuracy cannot be exceeded. The standalone Ed-RVFL, SVR, and RNN models are positioned far away from the observed $SPEI_{12}$. The TVF-ED-RVFL model again falls in close proximity to observed $SPEI_{12}$ for Fredericton station in relation to other comparing models to forecast $SPEI_{12}$. Therefore, the TVF-ED-RVFL model outperformed against comparing models at both stations to forecast $SPEI_{12}$.

3.1. *Multistep ahead forecasting*

Because of the better performance of TVF-ED-RVFL and ED-RVFL models in single step ahead drought forecasting in both stations, these models were used to forecast multistep ahead drought forecasting. Tables 6 and 7 demonstrate the multistep ahead, i.e., 1-, 3-, 6-, and 12-months $SPEI_{12}$ forecasting in Charlottetown and Fredericton stations to measure the performance of the TVF-ED-RVFL and ED-RVFL models. For Charlottetown station, the TVF-ED-RVFL seemed to be the highest precise model for all forecasting horizons in terms of R, RMSE, NSE, KGE, IA, and $U_{95\%}$ metrics when comparing against the single/standalone ED-RVFL model (Table 6). A significant accuracy improvement can be noted by the TVF-ED-RVFL model for all 1-, 3-, 6-, and 12-month $SPEI_{12}$ forecasting in Charlottetown station, which confirms the

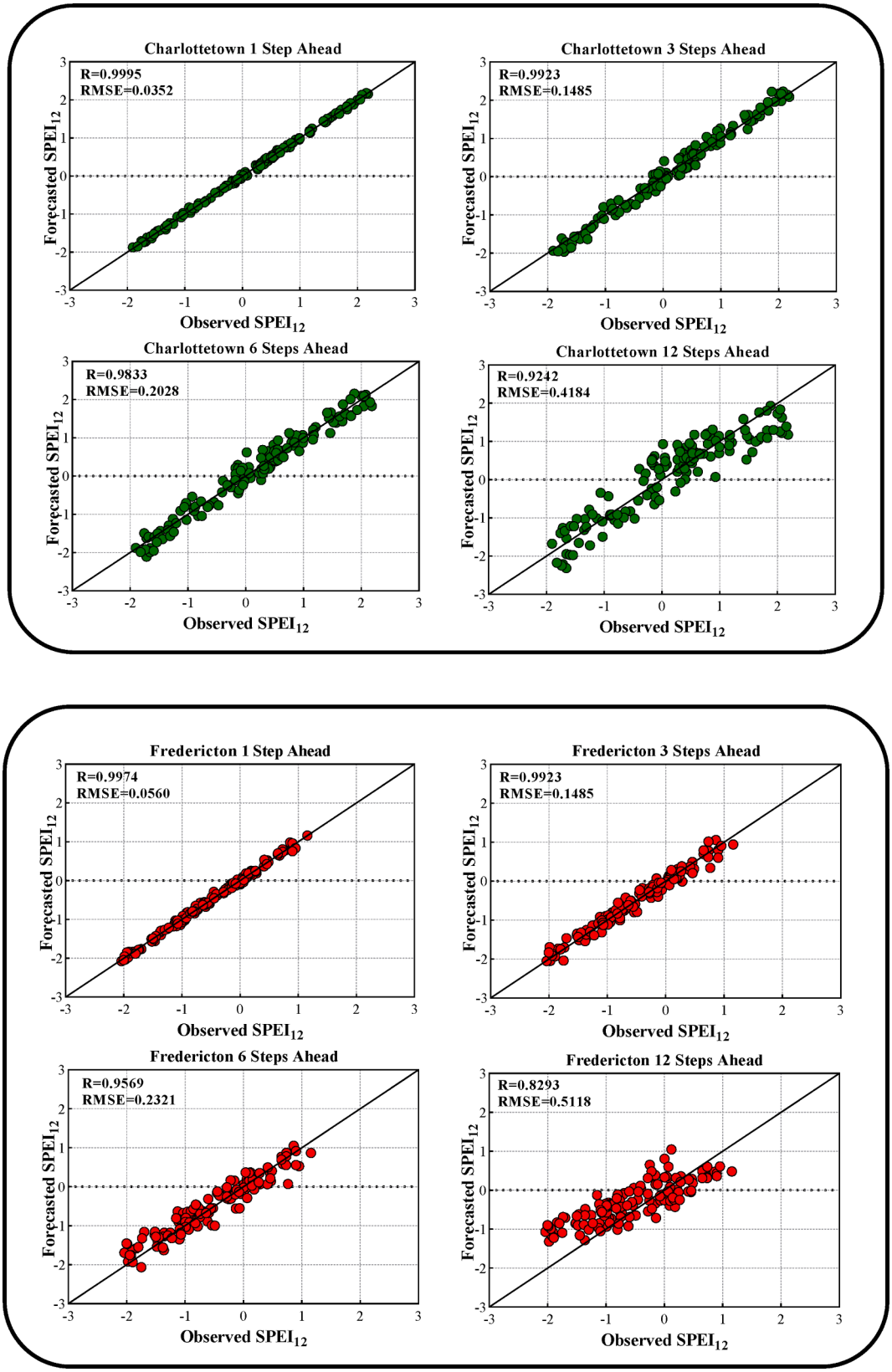


Fig. 16. Scatter plots of Multistep Ahead Forecasting of $SPEI_{12}$ for Charlottetown and Fredericton stations.

suitability of the TVF decomposition method with ED-RVFL (Table 6). Similarly, the TVF-ED-RVFL displays better accuracy in Table 7 to forecast 1-, 3-, 6-, and 12-month $SPEI_{12}$ for Fredericton station as compared to the ED-RVFL model. Here, it is to be noted that the TVF-ED-RVFL model performance is getting lower in the long term (i.e., 12 months) ahead of $SPEI_{12}$ forecast as compared to the 1-, 3-, and 6-months ahead for Fredericton station. But overall, the TVF-ED-RVFL achieves better accuracy as compared to the ED-RVFL to forecast 1-,

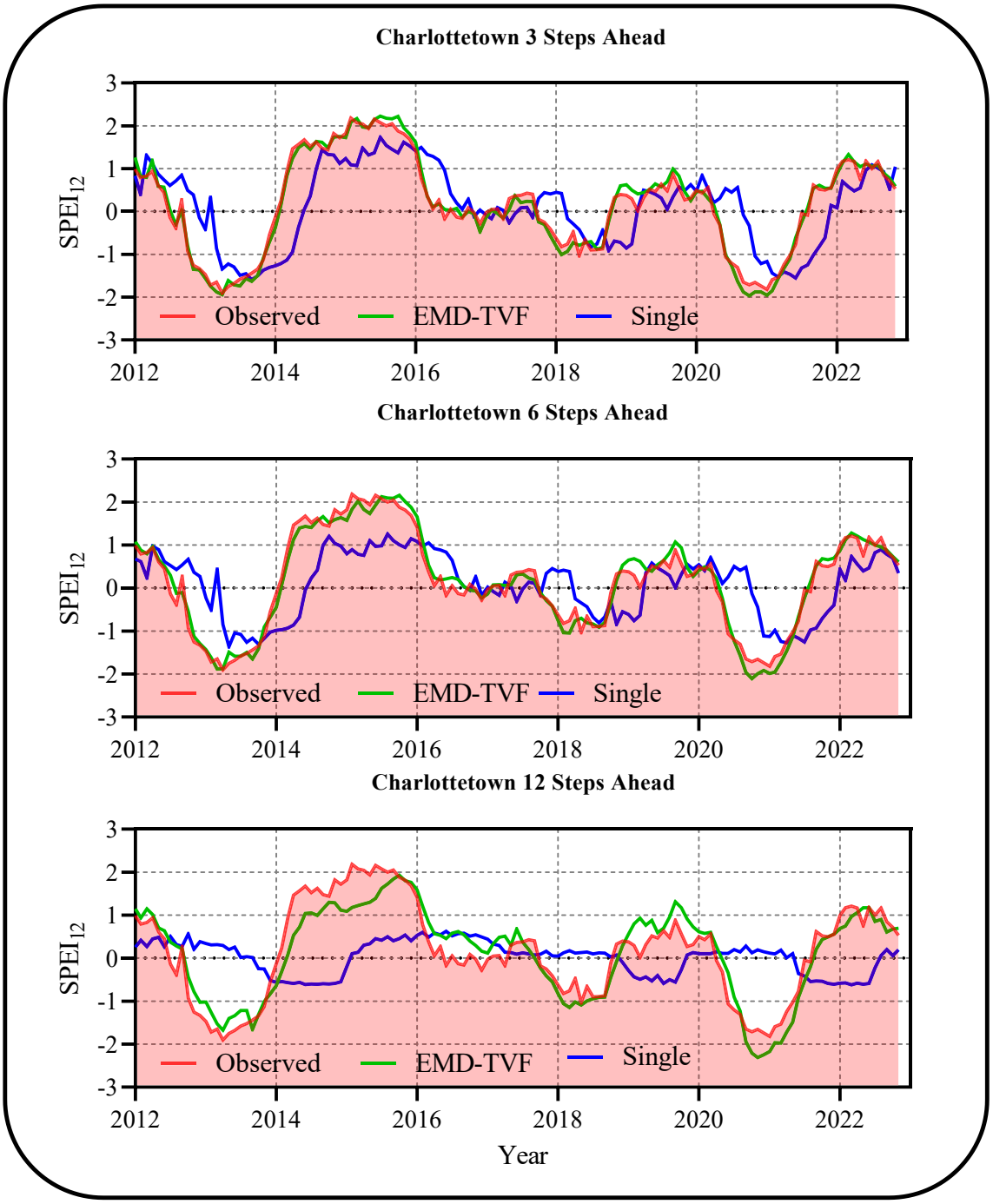


Fig. 17. Time series of observed and multistep ahead Forecasted values of $SPEI_{12}$ for Charlottetown and Fredericton stations.

3-, 6-, and 12-months ahead $SPEI_{12}$ for both Charlottetown and Fredericton stations (Tables 6 and 7).

The scatter plots in Fig. 16 characterize the observed and forecasted $SPEI_{12}$ of 1-, 3-, 6-, and 12-months ahead generated by the TVF-ED-RVFL model for Charlottetown and Fredericton stations. Besides, the R grouping with the RMSE metric was also displayed in Fig. 16. The TVF-ED-RVFL model at 1-, 3-, 6-, and 12 months ahead to forecast $SPEI_{12}$ exhibited good accuracy between the observed and forecasted values along with higher R and lower RMSE magnitudes in relation to Charlottetown station. The outcomes are also similar in the case of Fredericton station, which forecasts 1-, 3-, 6-, and 12 months ahead of $SPEI_{12}$. The TVF-ED-RVFL model forecast is slightly decreased for 12 months ahead of $SPEI_{12}$ in the scatter diagram as compared to 1-, 3-, and 6-months for both stations, and this also coincides with Tables 6 and 7. Thus, Fig. 16 established that TVF-ED-RVFL model is better for forecasting multistep ahead $SPEI_{12}$.

The time-series plots in Fig. 17 compared the annual trends from 2012 to 2022 using TVF-ED-RVFL and ED-RVFL and observed 3-, 6-, and 12-months ahead $SPEI_{12}$ for Charlottetown and Fredericton stations. The TVF-ED-RVFL model clearly attained better accuracy in terms of parallel and consistent trends against the observed SPEI in 3-, 6-, and 12-months ahead forecast horizons as compared to the ED-RVFL model for Charlottetown station. The TVF-ED-RVFL produced more consistent trends plot at 3-months, followed by 6-months, and 12-months ahead $SPEI_{12}$

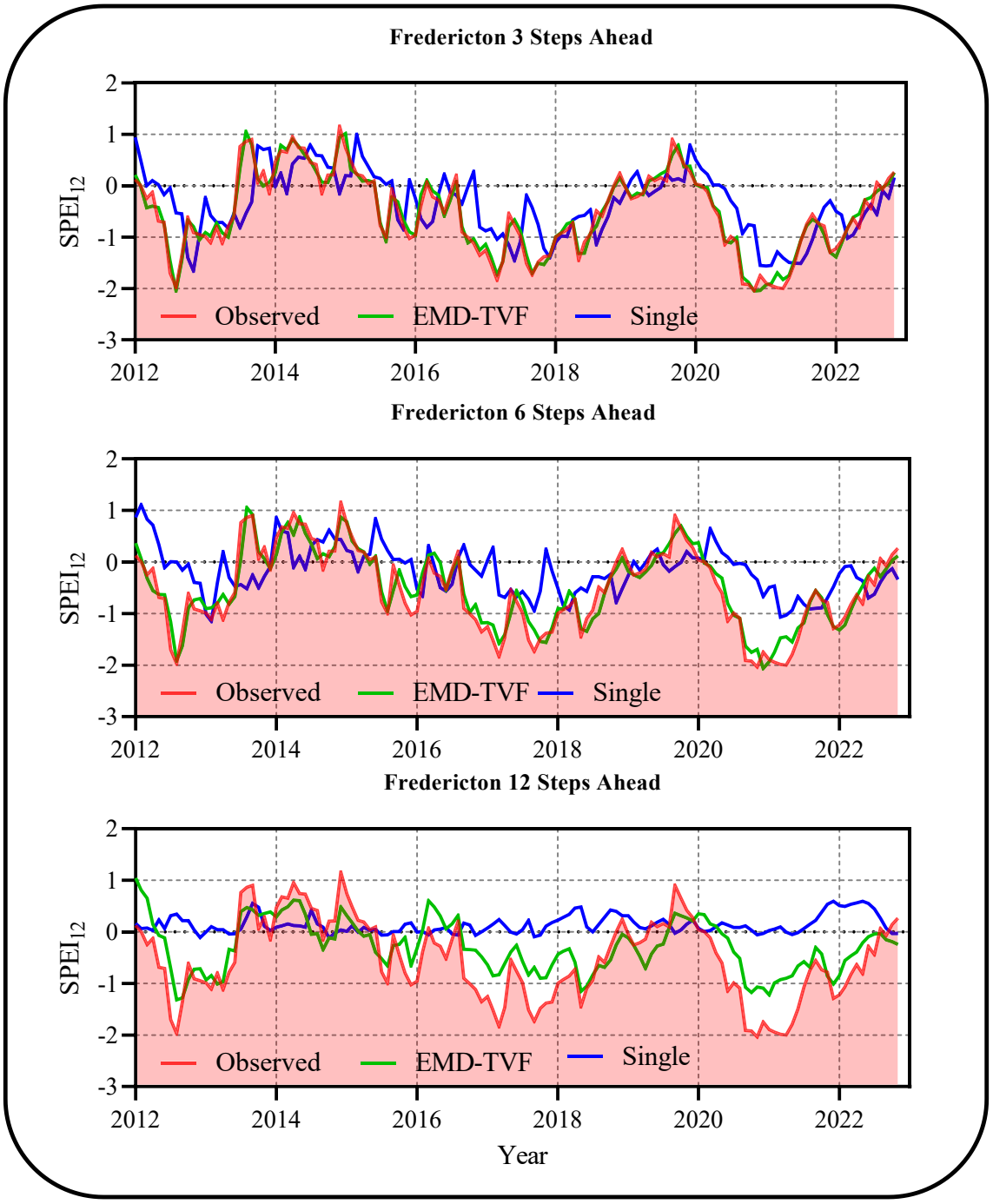


Fig. 17. (*continued*).

forecasting. This further established that the forecasting accuracy was higher for the short and medium-term (3-, and 6-months) $SPEI_{12}$ as compared to 12-months ahead. Similarly, the TVF-ED-RVFL model shows more consistency in trends with observed $SPEI_{12}$ for the Fredericton station.

3.2. Discussion

The current paper shows improved precision compared to conventional approaches in multistep-ahead drought forecasting in Eastern Canada using TVF-EMD, LASSO feature selection, and Ensemble Deep RVFL. Our work includes innovative preprocessing and modeling methods, setting it apart from existing research on drought prediction using ML and DL.

The decomposition methods are widely used in time series analysis especially for the nonstationary and noisy data, for instance drought indices. TVF-EMD application in this research was central since it facilitated the breakdown of the $SPEI_{12}$ values and helped in pattern and trend identification from the collected information. This step significantly improves the next machine learning models to be used. For example, Belayneh et al. (2014) and Özger et al., (2020) applied wavelet transform and EMD for drought forecasting. Also, Karbasi et al. (2023) have used discrete wavelet transform and empirical wavelet transform (EWT) to forecast drought in Iran. Their results showed better performance of EWT-based models.

There are many applications of EMD and its extensions, including

TVF-EMD, in a large number of fields in which time series are analyzed. For instance, Jamei et al. (2023) adopted TVF-EMD to improve the exactness of the forecast of the PM2. 5 and the levels of PM 10 they prove the effectiveness of the decomposition techniques dealing with the elaboration of the detailed analysis of the actual datasets. Karbasi et al. (2023) used a combination of TVF-EMD and bidirectional RNN to forecast reference evapotranspiration, and their results showed that TVF-based models had higher accuracy.

Furthermore, the investigations carried out by Abbes et al. (2023) & Kadam et al. (2024) hypothesize that DL models, including LSTM and GRU are more effective compared to other ML methods in drought forecasting. Our results are consistent with this trend since the tested deep learning model, namely the Ensemble Deep RVFL model, was found to be more accurate in terms of forecasting than the selected benchmark models, which include SVR, Simple RVFL, RNN, XGBoost, and Random Forest.

For future studies, the following suggestions are recommended.

- The comparison with more advanced signal decomposition techniques can be adopted to solve the non-stationarity and non-linearity issues of mode mixing. Therefore, multivariate variational mode decomposition (MVMD) (ur Rehman & Aftab, 2019) and multivariate empirical mode decomposition (MEMD) (Rehman & Mandic, 2010) can be the potential approaches against the EMD-TVF method.
- The black-box representation of deep learning limits and confines their ability in terms of explainability. To overcome this, the explainable AI techniques such as Local Interpretable Model-Agnostic Explanations (LIME) (Mishra et al., 2017) and Shapley Additive explanations (SHAP) (Shapley, 1953) can be integrated into the modeling framework for better understanding.
- Another aspect that enriches the modeling strategy is the hybridization of the physics-based models to illustrate the physical overview.
- The Bayesian Model Averaging (Sloughter et al., 2010) and bootstrapping (Tiwari & Chatterjee, 2011) techniques are widely considered to solve the underlying uncertainties of the model and can be adopted in future work.

4. Conclusion

This paper proposed a novel modeling framework to forecast short-, medium- and long-term drought index (i.e., $SPEI_{12}$) using EMD-TVF and Ensemble Deep RVFL to create EMD-TVF-ED-RVFL for Charlottetown and Fredericton stations in Canada. The EMD-TVF method decomposes the 1-, 3-, 6-, and 12-months $SPEI_{12}$ data into IMF signals. Next, the most significant IMFs were determined by the lasso regression feature selection algorithm. Finally, the selected IMFs were incorporated as inputs into the TVF-ED-RVFL model to forecast 1-, 3-, 6-, and 12-months ahead of $SPEI_{12}$. The results confirmed that the EMD-TVF-ED-RVFL model exhibits higher precision to forecast multistep ahead $SPEI_{12}$ compared to the comparing models. The hybrid EMD-TVF-based models were higher in accuracy than the standalone/single version of the models.

The EMD-TVF-ED-RVFL model acquired the highest precision to forecast 1-, 3-, 6-, and 12-months ahead of $SPEI_{12}$ for both Charlottetown and Fredericton stations. For example, to forecast $SPEI_{12}$ (i.e., 1-month), the EMD-TVF-ED-RVFL model in testing period reported the goodness-of-fit metrics ($R = 0.9995$, $RMSE = 0.0352$, $NSE = 0.9990$, $KGE = 0.9986$, $I_A = 0.9997$, $U_{95\%} = 0.0977$)-Charlottetown station and [$R = 0.9974$; $RMSE = 0.0560$; $NSE = 0.9949$; $KGE = 0.9945$; $I_A = 0.9987$; and $U_{95\%} = 0.1556$]- Fredericton station.

As a future recommendation, satellite-derived input data can be used in the proposed EMD-TVF-ED-RVFL model to forecast the SPEI index, further enhancing accuracy. Moreover, the satellite-derived inputs provide an alternative to the ground-based data, which is sometimes hard to attain due to complexity and financial burden. Due to the recent changes in climates, the impacts of droughts are frequent and severe; therefore, synoptic-scale climate mode indices can be used to provide in-depth analysis and enhance the model's accuracy. Moreover, other meteorological and atmospheric inputs (i.e., temperature, precipitation, humidity, vapor pressure, solar radiations, etc.) can be used in the proposed EMD-TVF-ED-RVFL model to forecast the SPEI index. The EMD-TVF-ED-RVFL model can be applied to agriculture, the environmental and hydrology sectors, and renewable and sustainable energy areas to broaden the scope in the future.

CRediT authorship contribution statement

Masoud Karbasi: Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Visualization, Investigation, Validation, Writing – original draft, Writing – review & editing. **Mumtaz Ali:** Investigation, Investigation, Supervision, Validation. **Aitazaz Ahsan Farooque:** Investigation, Supervision, Validation. **Mehdi Jamei:** Methodology, Validation, Writing – original draft, Writing – review & editing. **Khabat Khosravi:** Writing – original draft, Writing – review & editing. **Saad Javed Cheema:** Writing – review & editing. **Zaher Mundher Yaseen:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada, the Department of Environment, Energy and Climate Action, the Government of Prince Edward Island, and the Atlantic Canada Opportunities Agency for funding the project. Special thanks go to the Research Group at the Centre of Excellence in Food Security and Sustainability, University of Prince Edward Island, for their lab and field experimentation assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2024.124900.

References

Abadi, M. (2016). TensorFlow: learning functions at scale. In *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming*. ACM. https://doi.org/10.1145/2951913.2976746.

Abbes, A. B., Inoubli, R., Rhif, M., & Farah, I. R. (2023). Combining deep learning methods and multi-resolution analysis for drought forecasting modeling. *Earth Science Informatics, 16*(2), 1811–1820.

Abramowitz, M., & Stegun, I. A. (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables* (Vol. 55). Courier Corporation.

Arnold, T. B. (2017). kerasR: R Interface to the Keras Deep Learning Library. *Journal of Open Source Software, 2*(14), 296.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data 2018 5:1*, *5*(1), 1–12. https://doi.org/10.1038/sdata.2018.214.

Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural networks and wavelet support vector regression models. *Journal of Hydrology, 508*, 418–429. https://doi.org/10.1016/j.jhydrol.2013.10.052

Bonsal, B. R., Wheaton, E. E., Chipanshi, A. C., Lin, C., Sauchyn, D. J., & Wen, L. (2011). Drought research in Canada: A review. *Atmosphere-Ocean, 49*(4), 303–319.

Breiman, L. (1999). Randon Forests. *Machinelearning202.Pbworks.Com*, 1–35.

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2020.137612

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.

Cheng, R., Gao, R., & Yuen, K. F. (2024). Ship order book forecasting by an ensemble deep parsimonious random vector functional link network. *Engineering Applications of Artificial Intelligence, 133*, Article 108139.

Cheng, W. X., Suganthan, P. N., & Katuwal, R. (2021). Time series classification using diversified ensemble deep random vector functional link and resnet features. *Applied Soft Computing, 112*, Article 107826.

Cruse, H. (1996). *Neural networks as cybernetic systems*. Thieme Stuttgart.

Doesken, N. J., & Garen, D. (1991). Drought monitoring in the Western United States using a surface water supply index. *Proceedings of the 7th Conference on Applied Climatology, Salt Lake City, UT, USA*, 10–13.

Dong, J., Chen, Y., Yao, B., Zhang, X., & Zeng, N. (2022). A neural network boosting regression model based on XGBoost. *Applied Soft Computing, 125*, Article 109067.

Du, L., Gao, R., Suganthan, P. N., & Wang, D. Z. W. (2022). Graph ensemble deep random vector functional link network for traffic forecasting. *Applied Soft Computing, 131*, Article 109809. https://doi.org/10.1016/j.asoc.2022.109809

Eden, U. (2012). *Drought assessment by evapotranspiration mapping in Twente, the Netherlands*. University of Twente.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211. https://doi.org/10.1016/0364-0213(90)90002-E

Evans, J. S., Murphy, M. A., Holden, Z. A., & Cushman, S. A. (2011). Modeling species distribution and change using random forest. In *Predictive species and habitat modeling in landscape ecology* (pp. 139–159). Springer.

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 Engineering and Sciences, 4*(2), 73–88.

Felsche, E., & Ludwig, R. (2021). Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations. *Natural Hazards and Earth System Sciences, 21*(12), 3679–3691.

Gao, R., Du, L., Suganthan, P. N., Zhou, Q., & Yuen, K. F. (2022). Random vector functional link neural network based ensemble deep learning for short-term load forecasting. *Expert Systems with Applications, 206*, Article 117784.

Gao, R., Du, L., Yuen, K. F., & Suganthan, P. N. (2021). Walk-forward empirical wavelet random vector functional link for time series forecasting. *Applied Soft Computing, 108*, Article 107450.

Gao, R., Li, R., Hu, M., Suganthan, P. N., & Yuen, K. F. (2023). Online dynamic ensemble deep random vector functional link neural network for forecasting. *Neural Networks, 166*, 51–69.

Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture, 1*(2), 96–99. https://doi.org/10.13031/2013.26773.

Hao, Z., AghaKouchak, A., Nakhjiri, N., & Farahmand, A. (2014). Global integrated drought monitoring and prediction system. *Scientific Data, 1*(1), 1–10.

He, B., Armaghani, D. J., Tsoukalas, M. Z., Qi, C., Bhatawdekar, R. M., & Asteris, P. G. (2024). A case study of resilient modulus prediction leveraging an explainable metaheuristic-based XGBoost. *Transportation Geotechnics*, Article 101216.

Jamei, M., Ali, M., Karbasi, M., Karimi, B., Jahannemaei, N., Farooque, A. A., & Yaseen, Z. M. (2024). Monthly sodium adsorption ratio forecasting in rivers using a dual interpretable glass-box complementary intelligent system: Hybridization of ensemble TVF-EMD-VMD, Boruta-SHAP, and eXplainable GPR. *Expert Systems with Applications, 237*, Article 121512.

Jamei, M., Ali, M., Malik, A., Karbasi, M., Rai, P., & Yaseen, Z. M. (2023). Development of a TVF-EMD-based multi-decomposition technique integrated with encoder-decoder-bidirectional-LSTM for monthly rainfall forecasting. *Journal of Hydrology*, Article 129105.

Kadam, C. M., Bhosle, U. V., & Holambe, R. S. (2024). Deep learning-driven regional drought assessment: An optimized perspective. *Earth Science Informatics*, 1–15.

Kalteh, A. M. (2013). Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Computers & Geosciences, 54*, 1–8. https://doi.org/10.1016/j.cageo.2012.11.015

Karbasi, M., Jamei, M., Ali, M., Malik, A., Chu, X., Farooque, A. A., & Yaseen, Z. M. (2023). Development of an enhanced bidirectional recurrent neural network combined with time-varying filter-based empirical mode decomposition to forecast weekly reference evapotranspiration. *Agricultural Water Management, 290*, Article 108604.

Karbasi, M., Jamei, M., Malik, A., Kisi, O., & Yaseen, Z. M. (2023). Multi-steps drought forecasting in arid and humid climate environments: Development of integrative machine learning model. *Agricultural Water Management*. https://doi.org/10.1016/j.agwat.2023.108210

Kaur, A., & Sood, S. K. (2020). Deep learning based drought assessment and prediction framework. *Ecological Informatics, 57*, Article 101067. https://doi.org/10.1016/j.ecoinf.2020.101067

Khan, M. S., & Coulibaly, P. (2006). Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering*. https://doi.org/10.1061/(asce)1084-0699(2006)11:3(199)

Khosravi, K., Daggupati, P., Alami, M. T., Awadh, S. M., Ghareb, M. I., Panahi, M., Pham, B. T., Rezaie, F., Qi, C., & Yaseen, Z. M. (2019). Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. *Computers and Electronics in Agriculture, 167*, Article 105041.

Li, H., Li, Z., & Mo, W. (2017). A time varying filter approach for empirical mode decomposition. *Signal Processing, 138*, 146–158. https://doi.org/10.1016/j.sigpro.2017.03.019

Maity, R., Khan, M. I., Sarkar, S., Dutta, R., Maity, S. S., Pal, M., & Chanda, K. (2021). Potential of Deep Learning in drought assessment by extracting information from hydrometeorological precursors. *Journal of Water and Climate Change, 12*(6), 2774–2796.

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. *Applied Soft Computing*. , Article 110377. https://doi.org/10.1016/j.asoc.2023.110377

Maybank, J., Bonsai, B., Jones, K., Lawford, R., O'brien, E. G., Ripley, E. A., & Wheaton, E. (1995). Drought as a natural disaster. *Atmosphere-Ocean, 33*(2), 195–222.

Mckee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *AMS 8th Conference on Applied Climatology* (pp. 179–184).

Mishra, S., Sturm, B. L., & Dixon, S. (2017). Local interpretable model-agnostic explanations for music content analysis. *ISMIR, 53*, 537–543.

Mokhtarzad, M., Eskandari, F., Jamshidi Vanjani, N., & Arabasadi, A. (2017). Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environmental Earth Sciences*. https://doi.org/10.1007/s12665-017-7064-0

Özger, M., Başakın, E. E., Ekmekcioğlu, Ö., & Hacısüleyman, V. (2020). Comparison of wavelet and empirical mode decomposition hybrid models in drought prediction. *Computers and Electronics in Agriculture, 179*. https://doi.org/10.1016/j.compag.2020.105851

Palmer, W. C. (1968). *Keeping track of crop moisture conditions, nationwide: the new crop moisture index*.

Pao, Y.-H., Park, G.-H., & Sobajic, D. J. (1994). Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing, 6*(2), 163–180. https://doi.org/10.1016/0925-2312(94)90053-1

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. *The Journal of Machine Learning Research, 12*, 2825–2830.

Qiu, X., Suganthan, P. N., & Amaratunga, G. A. J. (2018). Ensemble incremental learning random vector functional link network for short-term electric load forecasting. *Knowledge-Based Systems, 145*, 182–196.

Rehman, N., & Mandic, D. P. (2010). Multivariate empirical mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. https://doi.org/10.1098/rspa.2009.0502

Ren, Y., Zhang, L., & Suganthan, P. N. (2016). Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational Intelligence Magazine, 11*(1), 41–53. https://doi.org/10.1109/MCI.2015.2471235

Saunders, C., & Gammerman, A. (1998). Ridge regression learning algorithm in dual variables. *15th International Conference on Machine Learning (ICML '98) (01/01/98)*.

Schmidt, W. F., Kraaijveld, M. A., & Duin, R. P. W. (1992). Feed forward neural networks with random weights. *International Conference on Pattern Recognition*, 1.

L.S. Shapley Shapley, L. S. (1953). 17. A value for n-person games. In *Contributions to the theory of games (AM-28), Volume II* (pp. 307–318). Princeton University Press. 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 Recognition, 117*, Article 107978.

Sloughter, J. M., Gneiting, T., & Raftery, A. E. (2010). Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association, 105*(489), 25–35.

Smola, A. J. (1996). *Regression estimation with support vector learning machines*. Technische Universität München. Master's thesis.

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., … 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. *Engineering Applications of Artificial Intelligence*, *129*(November 2023), 107559. https://doi.org/10.1016/j.engappai.2023.107559.

Tareke, K. A., & Awoke, A. G. (2023). Hydrological drought forecasting and monitoring system development using artificial neural network (ANN) in Ethiopia. *Heliyon, 9*(2), Article e13287. https://doi.org/10.1016/j.heliyon.2023.e13287.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267–288. https://doi.org/10.1111/j.1467-9868.2011.00771.x

Tiwari, M. K., & Chatterjee, C. (2011). A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. *Journal of Hydroinformatics, 13*(3), 500. https://doi.org/10.2166/hydro.2010.142

ur Rehman, H., & Aftab, H. (2019). Multivariate variational mode decomposition. *IEEE Transactions on Signal Processing, 67*(23), 6039–6052.

van Hoek, M. (2016). *Drought monitoring from space: A focus on indicators, early detection and development of a web-based integrated portal*. Beijing: University of Chinese Academy of Sciences.

Vapnik, V. N. (2000). *The nature of statistical learning theory* (second ed.). New York: Springer.

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate, 23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Wable, P. S., Jha, M. K., Adamala, S., Tiwari, M. K., & Biswal, S. (2023). Application of hybrid ANN techniques for drought forecasting in the semi-arid region of India. *Environmental Monitoring and Assessment, 195*(9), 1090.

Wang, G. C., Zhang, Q., Band, S. S., Dehghani, M., Chau, K., Wing, Tho, Q. T., Zhu, S., Samadianfard, S., & Mosavi, A. (2022). Monthly and seasonal hydrological drought forecasting using multiple extreme learning machine models. *Engineering Applications of Computational Fluid Mechanics, 16*(1), 1364–1381.

Wang, K., Fu, W., Chen, T., Zhang, B., Xiong, D., & Fang, P. (2020). A compound framework for wind speed forecasting based on comprehensive feature selection, quantile regression incorporated into convolutional simplified long short-term memory network and residual error correction. *Energy Conversion and Management, 222*(May), Article 113234. https://doi.org/10.1016/j.enconman.2020.113234

Wilhite, D. A., Svoboda, M. D., & Hayes, M. J. (2007). Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water Resources Management*. https://doi.org/10.1007/s11269-006-9076-5

Willmott, C. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society, 63*(11), 1309–1313. https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

Xue, Z., Yi, X., Feng, W., Kong, L., & Wu, M. (2024). Prediction and mapping of soil thickness in alpine canyon regions based on whale optimization algorithm optimized random forest: A case study of Baihetan Reservoir area in China. *Computers & Geosciences*, Article 105667.

Yaseen, Z. M. (2023). A new benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. *Knowledge-Based Engineering and Sciences, 4*(3), 65–103.

Yaseen, Z. M., Ali, M., Sharafati, A., Al-Ansari, N., & Shahid, S. (2021). Forecasting standardized precipitation index using data intelligence models: Regional investigation of Bangladesh. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-82977-9

Zheng, Z., Ali, M., Jamei, M., Xiang, Y., Karbasi, M., Yaseen, Z. M., & Farooque, A. A. (2023). Design data decomposition-based reference evapotranspiration forecasting model: A soft feature filter based deep learning driven approach. *Engineering Applications of Artificial Intelligence, 121*, Article 105984.