Electrical Power and Energy Systems 157 (2024) 109876

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

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

Hypertuned temporal fusion transformer for multi-horizon time series forecasting of dam level in hydroelectric power plants

Stefano Frizzo Stefenon [a,b,*], Laio Oriel Seman [c], Luiza Scapinello Aquino da Silva [d], Viviana Cocco Mariani [d,e], Leandro dos Santos Coelho [d,f]

[a] Digital Industry Center, Fondazione Bruno Kessler, Trento, Italy
[b] Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy
[c] Department of Automation and Systems, Federal University of Santa Catarina, Florianopolis, Brazil
[d] Department of Electrical Engineering, Federal University of Parana, Curitiba, Brazil
[e] Mechanical Engineering Graduate Program, Pontifical Catholic University of Parana, Curitiba, Brazil
[f] Industrial and Systems Eng. Graduate Program, Pontifical Catholic University of Parana, Curitiba, Brazil

ARTICLE INFO

Dataset link: https://github.com/SFStefenon/AutoTFT

Keywords:
Decision-making
Hydroelectric power plant
Temporal fusion transformer
Time series forecasting

ABSTRACT

This paper addresses the challenge of predicting dam level rise in hydroelectric power plants during floods and proposes a solution using an automatic hyperparameters tuning temporal fusion transformer (AutoTFT) model. Hydroelectric power plants play a critical role in long-term energy planning, and accurate prediction of dam level rise is crucial for maintaining operational safety and optimizing energy generation. The AutoTFT model is applied to analyze time series data representing the water storage capacity of a hydroelectric power plant, providing valuable insights for decision-making in emergency situations. The results demonstrate that the AutoTFT model surpasses other deep learning approaches, achieving high accuracy in predicting dam level rise across different prediction horizons. Having a root mean square error (RMSE) of 2.78×10^{-3} for short-term forecasting and 1.72 considering median-term forecasting, the AutoTFT shows to be promising for time series prediction presented in this paper. The AutoTFT had lower RMSE than the adaptive neuro-fuzzy inference system, long short-term memory, bootstrap aggregation (bagged), sequential learning (boosted), and stacked generalization ensemble learning approaches. The findings underscore the potential of the AutoTFT model for improving operational efficiency, ensuring safety, and optimizing energy generation in hydroelectric power plants during flood events.

1. Introduction

Hydroelectric power plants with reservoirs play a crucial role in electricity generation and regulation, providing flexibility in the management of accumulated water [1]. In Brazil, the electric power system relies heavily on a hydrothermal energy matrix [2], where hydroelectric plants with dams are essential for long-term energy planning since they can store water for drought periods [3]. The operation of a hydroelectric power plant involves optimizing energy generation to minimize costs over time and ensure future energy supply [4]. Plant operators follow prescribed operating procedures set by the national system operator, which include considerations of inflow characteristics, reservoir levels, and energy demand [5].

The sudden rise in river levels during storms, particularly in regions prone to heavy rainfall like southern Brazil, poses a significant challenge for hydroelectric plants [6]. The possibility of abrupt variations in water inflows due to storms and floods represents a risk for these plants and their surroundings [7]. When a plant exceeds its storage capacity and needs to release water, not only is the available energy lost, but there is also a maximum spill limit that must be observed to ensure safety and prevent damage to the dam [8].

Predicting dam level rise accurately and promptly is of utmost importance for the safe and efficient operation of hydroelectric power plants. It can serve as a valuable decision-support indicator for plant operators, helping them anticipate emergency situations caused by rapid water level fluctuations [9]. Effective predictions and evaluation of the hydroelectric power plants can enable proactive measures to be taken, such as adjusting energy generation [10] and reservoir management strategies [11], optimizing spillage [12], and ensuring the stability and reliability of the electrical power system [13]. Considering the vast nomenclature that is used in this field, Table 1 presents the abbreviations used here.

In light of this challenge, this paper proposes the use of an automatic hyperparameters tuning temporal fusion transformer (AutoTFT)

* Corresponding author at: Digital Industry Center, Fondazione Bruno Kessler, Trento, Italy.
E-mail address: sfrizzostefenon@fbk.eu (S.F. Stefenon).

https://doi.org/10.1016/j.ijepes.2024.109876
Received 28 July 2023; Received in revised form 7 November 2023; Accepted 12 February 2024
Available online 21 February 2024

Table 1
Abbreviations and acronyms used.

Name	Abbreviations
Adaptive neuro-fuzzy inference system	ANFIS
Artificial neural networks	ANNs
Automatic hyperparameters tuning TFT	AutoTFT
Empirical wavelet transform	EWT
Exponential linear unit	ELU
Feed-forward network	FFN
Fuzzy c-means	FCM
Gated linear unit	GLU
Gated residual network	GRN
Grid partition	GP
Interquartile range of error	IQR
Long short-term memory	LSTM
Maximum absolute error	MaxAE
Mean absolute error	MeanAE
Mean square error	MSE
Median absolute error	MedianAE
Normalized mean absolute error	NMAE
Number of configurations explored	NCE
Root mean square error	RMSE
Root mean squared percentage error	RMSPE
Root mean squared propagation	RMSprop
Seasonal-trend decomposition based on loess	STL
Sequence-to-sequence	Seq2Seq
Standard deviation of error	Std Dev
Stochastic gradient descent with momentum	SGDM
Temporal fusion	TF
Temporal fusion transformer	TFT
Variance of error	VAR

model for predicting dam level rise during floods. The temporal fusion transformer (TFT) model combines the strengths of recurrent neural networks, autoregressive models, and attention mechanisms, allowing it to capture complex temporal patterns and make accurate predictions [14]. The hyperparameters tuning process optimizes the model's hyperparameters to achieve superior performance and generalization [15].

The evaluation of the proposed model is conducted using a time series dataset representing the water storage capacity of a hydroelectric power plant with a dam. The dataset includes historical information on inflows, water levels, and other relevant factors. The performance of the AutoTFT model is compared against existing forecasting approaches.

The main contributions of this paper are as follows:

- The proposal of an AutoTFT model for predicting dam level rise during floods, leveraging its ability to capture complex temporal patterns and make accurate predictions.
- The evaluation of the proposed model using real-world time series data from a hydroelectric power plant, comparing its performance against existing forecasting approaches.
- The identification of key factors and considerations in dam level prediction, including the analysis of inflow characteristics, reservoir management strategies, and the impact of extreme weather events.

The remainder of this paper is organized as follows: Section 2 provides an overview of related works in time series forecasting for energy systems, highlighting their strengths and limitations. Section 3 describes the methodology, including the architecture and components of the AutoTFT model. Section 4 presents the results and discussion, analyzing the performance of the proposed model and its implications for dam-level prediction. Section 5 concludes the paper, summarizing the key findings and suggesting future research directions in the field of dam-level forecasting for hydroelectric power plants.

2. Related works

In Brazil, the power generation schedule has been studied mainly regarding optimizing the hydrothermal resources for security and economy in using the system [16]. The economy is considering the energy cost over time, and security is regarding the confidence in having the available power supply [17]. Although these concerns rule scheduling problems, when there is a rise in the dam caused by floods, the manager of the power plant becomes responsible for the control of the power plant [9], and this is the focus of the study presented here.

A promising way to evaluate the rise of dam water level is to evaluate the time series of its variation; therefore time series forecasting may be an alternative for manager decision-making in electric power plants with dams. In [8] using a seasonal-trend decomposition based on loess (STL), the authors had a mean square error of 0.017 and 0.019 in two monitoring points using their method called STL-extra-trees long short-term memory (LSTM) for the prediction of dam displacement.

Streamflow predictions are a critical aspect of hydrological research and play a pivotal role in various water resource management and environmental applications. A substantial body of literature exists concerning streamflow prediction methods and models. In the context of streamflow simulation for Brazilian Atlantic Rainforest basins, Vilanova et al. [18] builds upon existing research by exploring the feasibility of employing artificial neural networks (ANNs) with distinct input and output configurations. The findings reveal that daily rainfall up to the third antecedent day exhibits strong correlations with runoff and streamflow, while accumulated rainfall up to the preceding 90 days correlates strongly with baseflow and streamflow.

The study presented by Saraiva et al. [19] contributes to the ongoing efforts to enhance streamflow forecasting techniques, it conducts a comparative analysis of two machine learning models (ANNs and support vector machines) with wavelet transform and data resampling via the bootstrap method. This analysis was applied to daily streamflow time series data from the Sobradinho Reservoir in northeastern Brazil, covering the period from 1931 to 2015. The findings demonstrate that the ANN outperforms the support vector machines in terms of forecasting accuracy, with the bootstrap, wavelet, and neural network approach emerging as the most effective combination.

Nazari et al. [20] highlight the pressing global issue of flood impacts on communities and the significance of effective reservoir management strategies for flood control and mitigation. By investigating their model performance within a simulated environment, constructed directly from Saint-Venant equations, the research seeks to establish a model that balances high prediction accuracy and scientific consistency, positioning it for real-time applications. By combining aspects of physics and data-driven approaches, it aims to facilitate advancements in hydrological modeling and the development of practical tools for real-time applications.

The work of Siqueira et al. [21] focuses on evaluating the predictive capability of ensemble streamflow forecasts for flood prediction across South America, with a particular emphasis on the continent's major basins. Using medium-range precipitation forecasts from the ensemble prediction model, a comprehensive continental-scale hydrologic-hydrodynamic approach is employed to generate streamflow predictions up to 15 days in advance. The study rigorously assesses these forecasts over 4 years, comparing them against a reference simulation derived from a precipitation dataset that consolidates data from multiple sources.

Agarwal et al. [22] present a comprehensive approach by utilizing integrated modeling techniques, including the wavelet, multilayer perceptron, time-delay neural network, and gamma memory neural network, to forecast hourly river-level fluctuations, incorporating the variable of storage rate change. They showed that the integration of wavelet transforms further augments the predictive power of the models. Poul et al. [23] apply multi-linear regression, ANN, ANFIS, and k-nearest neighbors to forecast monthly flow in the St. Clair River,

bridging the United States and Canada. The findings underscore the significant enhancement in prediction accuracy when lag times of flow, temperature, and precipitation are incorporated into the input variables.

According to Stefenon et al. [24], when performing time series forecasting, it is important to preprocess the data, identify and remove any trends or seasonality, validate the models using appropriate evaluation metrics, and potentially iterate on the models to improve accuracy. They showed that using a seasonal and trend decomposition could have a mean absolute percentage error 18% lower than the baseline method. In the work of Li, Liu, and Tanaka [25], the Hodrick–Prescott filter was applied to denoising the time series and enhance the predictions.

Aiming to reduce abrupt variations in the time series, Branco et al. [26] applied the wavelet transform to preprocess the time series data. Using their wavelet LSTM method, they had a root mean square error (RMSE) of 1.45×10^{-3}, which was lower than the adaptive neuro-fuzzy inference system (ANFIS), group method of data handling, and ensemble learning methods. In [27], the authors had the best result using the stacking ensemble model combined with the wavelet transform, having a coefficient of determination of 0.9982. These works showed that the wavelet transform can yield promising time series denoising results.

The forecasting method depends on the characteristics of the data, the presence of trends or seasonality, the amount of available data, and the specific requirements of the problem at hand. According to Medeiros et al. [28], several models can be applied to forecasting problems, and using the most suitable model is a challenge. Using a recurrent neural network they had an accuracy of up to 97.25% with a mean coefficient of determination of 0.7632 when multiple experiments are evaluated.

Besides using filters for noise reduction, the choice of the model depends on the given problem. Seman et al. [29], compared the blending, bagging, boosting, random subspace, and stacked generalization. They showed that the Hodrick–Prescott filter reduces the RMSE by 2.69 times using the random subspace approach. The proposed method had better results than the original ensemble, standard LSTM, and LSTM with the Hodrick–Prescot filter employed. In general, ensemble models are becoming popular since they have lower computational effort given their properties of combining weaker learners to create a stronger structure [30].

Pereira et al. [31] presented a comparison of several forecasting models applied for time series analysis in smart cities. The prophet was faster to finish the training; however, it did not have the best error results considering the normalized mean absolute error (NMAE). Considering the prediction of American electric power, in their evaluation, the DeepAR had an NMAE of 3.47% been the best model for this application. The DeepAR is a probabilistic forecasting method with autoregressive recurrent networks that presents promising results for time series forecasting [32].

In the work of Nazir et al. [33], the TFT is applied to forecast energy consumption. This topic is important to be evaluated as the management planning of energy use needs to be evaluated over time, given that the time series forecasting of energy can be an extra input for decision making. With a symmetric mean absolute percentage error of 26.46%, the TFT outperforms other deep learning approaches, such as the temporal convolutional network and LSTM.

With an equivalent goal, Wang et al. [34] proposed a variation of the TFT called cross-entity TFT which uses an attention mechanism to inter-entity correlations. In their application, the study is concerned with the scheduling and optimization of the energy system; they show that the presented method had lower errors than other models, even considering different horizons. Several studies cover the scheduling problem considering energy generation [35], this problem can be analyzed considering the hydrothermal unit-commitment [36], where several stochastic programming algorithms can be applied to optimize the use of these resources [37].

According to Ben-Yelun et al. [38], hypertuning is a promising strategy for determining the parameters for the prediction models. Based on this approach tuning the model automatically is an advantage that makes the AutoTFT an interesting strategy for time series forecasting. Sharma et al. [39] applied the hypertuning in the Prophet forecasting model for time series analysis, one advantage of using hypertuning in time series forecasting strategies is that generally, the models take less time to converge than computer vision applications where hypertuning is also employed.

The Optuna proposed by Akiba et al. [40] is becoming popular for hypertuning, especially applied in prediction applications, such as impedance for circuit analysis [41], the critical temperature of superconductors [42], sales estimations [43], and others. As presented by Klaar et al. [44] the use of LSTM combined with empirical wavelet transform (EWT) is a promising approach, even when sequence-to-sequence (Seq2Seq) data is considered. Their method called EWT-Seq2Seq-LSTM which uses the attention mechanism, had a 10.17% lower mean square error (MSE) than the LSTM. Using Optuna, they had a hypertuning that helped the EWT-Seq2Seq-LSTM to achieve better performance.

As presented in [44] the Optuna framework is a that can be applied to tuning the architecture of the network, having an optimized structure for time series forecasting. In their results applying an optimized method, a 10.17% lower MSE was achieved. In the work of Klaar et al. [45], the Optuna using a tree-structured Parzen estimator was employed to optimize the structure of an ensemble learning approach applied for energy price forecasting. Using decomposition methods, the authors achieved an MSE of 3.37×10^{-9}. In their final considerations, Iqbal et al. [46] mentioned that their time series model may be further improved by hypertuning.

Table 2 provides a summary of the most relevant related works in the field of time series forecasting and their key findings. Considering the promising results of the TFT for time series forecasting, this paper employs hyperparameters tuning on this model in a multi-horizon evaluation. Additional information about the structure of the AutoTFT and the application proposed in this paper is presented in the next section.

3. Multi-horizon time series forecasting

Time series forecasting is a statistical approach to predict future values based on historical data points ordered chronologically [47]. Time series forecasting aims to capture the underlying patterns and trends within the data and make predictions about future values [48]. The assumption is that the future values of a variable depend on its past values and potentially other factors [49].

For the time series forecasting considering D samples,

$$x(t-(D-1)\Delta),\dots,x(t-\Delta),x(t) \tag{1}$$

to forecast future value,

$$x(t+P), \tag{2}$$

where Δ is the period of the samples, x is the considered value, t is the time user evaluation, and P are the steps forward. When P is equal to one means one step ahead; therefore for multi-horizon forecasting, the value of P is varied depending on how many steps ahead are evaluated [50].

Every step ahead that is considered takes into account the last prediction, based on this, multi-horizon time series forecasting becomes a problem more difficult to handle since it can have accumulated errors from the previous predictions. Two principal approaches are used to tackle this challenge: the direct method, which involves training separate models for each forecast horizon, and the recursive method, which uses a single model to make a one-step prediction and then feeds this prediction back as input for the subsequent steps. Each approach has implications for the computational efficiency and forecast accuracy

Table 2
Summary of related works in time series forecasting.

Study	Methodology	Findings
Li et al. [8]	Applied a combination of STL, extra-trees, and stacked LSTM neural network	Using the proposed STL-extra-trees LSTM they were able to achieve an MSE of 0.017 for the prediction of dam displacement.
Stefenon et al. [9]	Wavelet transform and LSTM with attention for predicting dam levels.	Achieved lower RMSE compared to other methods, highlighting the effectiveness of wavelet transform in denoising time series data.
Pereira et al. [31]	Complete comparison of time series forecasting models for smart cities.	DeepAR model showed promising results with the lowest normalized mean absolute error for the prediction of American electric power.
Salinas et al. [32]	Propose a combination of probabilistic forecasting model and autoregressive recurrent networks.	The proposed DeepAR overcomes several time series foresting models considering different real-world applications.
Nazir et al. [33]	TFT for energy consumption forecasting in smart grids.	TFT outperformed other approaches, including temporal convolutional network and LSTM, in terms of symmetric mean absolute percentage error.
Wang et al. [34]	Cross-entity TFT with attention mechanism for energy system optimization.	Cross-entity TFT demonstrated lower errors compared to other models, particularly in the context of scheduling and optimization of energy systems.
Sharma et al. [39]	Time series analysis using a hypertuned Prophet for the internet of things.	With the hypertuning and other adjustments, the Prophet had an MSE of 0.13 for predicting future energy consumption.
Klaar et al. [44]	EWT combined with LSTM and attention mechanism for time series forecasting.	Demonstrated superior performance of EWT-Seq2Seq-LSTM with attention mechanism in reducing the error compared to the LSTM.
Klaar et al. [45]	Optuna to optimize the structure of an ensemble learning approach.	Showed that a hypertuned model had superior results than the LSTM and the use of attention mechanism and denoising can further improve the predictions.

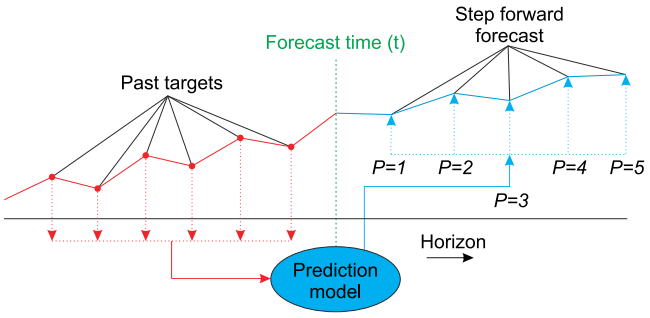

Fig. 1. Concept of prediction for multi-horizon time series.

of the model. An example of this concept is presented in Fig. 1, where each additional step forward P is one extra step ahead in a longer horizon evaluation.

Attention mechanisms have been recently applied to time series forecasting tasks, enhancing the performance of models in capturing long-range dependencies and improving accuracy [51]. In time series forecasting, attention mechanisms allow the model to selectively focus on relevant historical time steps while making predictions for future time steps. This selective attention enables the model to assign different weights or importance to different time steps based on their relevance to the prediction at hand [52].

An approach that incorporates attention mechanisms for time series forecasting is the transformer [53]. Transformers consist of self-attention layers, which allow the model to attend to different parts of the input sequence when making predictions [54]. A model that stands out in this implementation is the TFT.

TFT employs a hierarchical structure that captures both short-term and long-term dependencies in time-series data, making it suitable for a wide range of applications such as forecasting, signal processing, and sequential decision-making [55]. Given the challenge of multi-horizon time series forecasting, Fig. 2 presents how this approach is applied for inflow forecasting in the context of hydroelectric management. The TFT model will be used in this work and is explained in the next subsection.

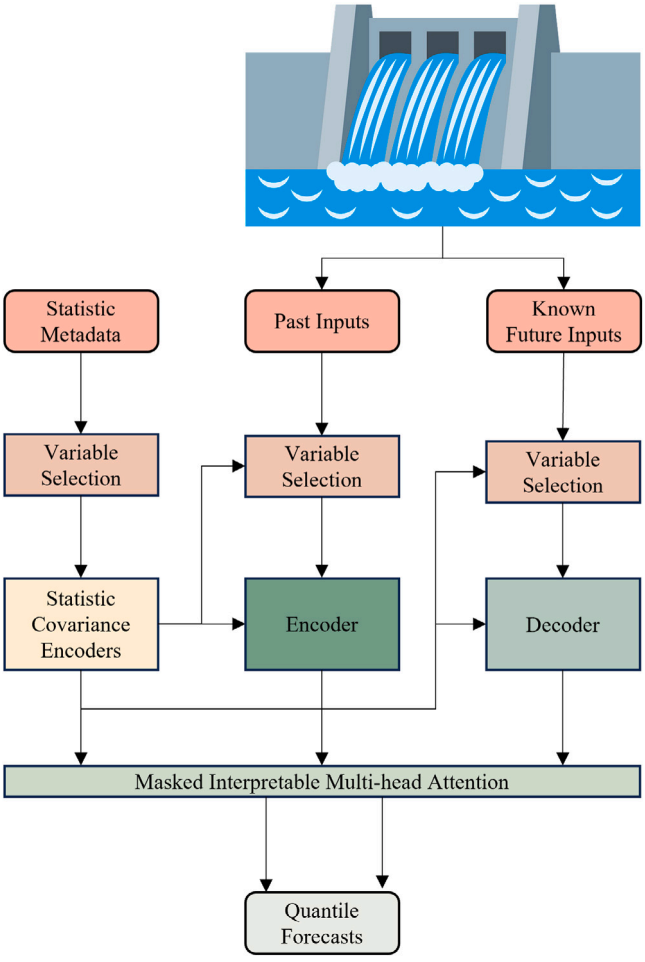

Fig. 2. Inflow forecasting for hydroelectric power plants.

3.1. Temporal fusion transformer

In the TFT architecture, the input time series data is first processed by an encoder consisting of multiple transformer layers. Each transformer layer employs self-attention mechanisms to capture temporal dependencies within the time series [55]. In the transformer encoder the attention mechanism is given by:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimensionality of the key vectors.

Based on that, the multi-head attention [56] is:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{4}$$

and,

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \tag{5}$$

where W_i^Q, W_i^K, and W_i^V are the linear transformation matrices for each attention head.

In this architecture, the feed-forward network (FFN) is:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{6}$$

where x is the input, W_1, W_2 are weight matrices, and b_1, b_2 are bias vectors. Following, the residual connection is:

$$\text{Residual}(x) = \text{LayerNorm}(x + \text{FFN}(x)) \tag{7}$$

where LayerNorm normalizes the input across the last dimension.

The fusion component of TFT comes into play after the encoding stage. It fuses the hidden representations across different time steps to capture temporal patterns and relationships. This fusion process allows the model to learn how the past values of the time series influence future predictions. The fused representations are then fed into a decoder, which predicts the future values of the time series [57].

The temporal fusion (TF) layer is given by:

$$\text{TF}(x) = \text{LayerNorm}(x + \text{FFN}(x)) \tag{8}$$

which is similar to the residual connection in the transformer encoder. The temporal fusion gating is:

$$\text{Gating}(x, g) = \text{sigmoid}(x) \odot g \tag{9}$$

where g is the gating vector. Then, the temporal fusion equation can be described as:

$$\text{TF\_eq}(x_1, \ldots, x_t, g) = \text{TF}(x_1) \oplus \cdots \oplus \text{TF}(x_t) \oplus g \tag{10}$$

where $\oplus$ represents concatenation.

Finally, the output of the encoder (decoder) is a set of hidden representations that encode the temporal and contextual information of the input sequence. The output layer can be represented as:

$$\text{Output}(x) = xW_{\text{out}} + b_{\text{out}} \tag{11}$$

where W_{out} and b_{out} are the output layer's weight matrix and bias vector.

The TFT provides interpretability by incorporating a gating mechanism, enabling it to identify each input variable's importance at different time steps. This allows users to understand the relative significance of different factors in the forecasting process [58]. The ability to handle multivariate data, capture long-range dependencies, and provide interpretability makes the TFT a valuable model for time series analysis and prediction [59].

The gated residual network (GRN) is a residual gated mechanism that allows the network to skip the non-linear transformation of an input a in a context c. The primary function of the gating mechanism within GRNs is to modulate the contribution of the input and the residual connection, thereby enhancing the model's capacity to learn complex functions and mitigating the vanishing gradient problem often encountered in deep networks. The residual block is given by:

$$\eta_1 = \text{ELU}(\mathbf{W}_1\mathbf{a} + \mathbf{W}_2\mathbf{c} + \mathbf{b}_1) \tag{12}$$

$$\eta_2 = \mathbf{W}_2\eta_1 + b_2 \tag{13}$$

$$\text{GRN}(\mathbf{a}, \mathbf{c}) = \text{LayerNorm}(a + \text{GRU}(\eta_2)) \tag{14}$$

where ELU is the exponential linear unit, GLU represents the gated linear unit, and provides the flexibility of suppressing unnecessary parts of the GRN. Consider GRN's output γ then GLU transformation is defined by:

$$\text{GLU}(\gamma) = \sigma(\mathbf{W}_4\gamma + b_4) \odot (\mathbf{W}_5\gamma + b_5) \tag{15}$$

Table 3
Hyperparameters evaluated in the hypertuning procedure (adjusted model).

Hyperparameter	Way to set	Range
`hidden_size`	Choice	[8, 32]
`n_head`	Choice	[2, 8]
`learning_rate`	Log uniform	[1e−4, 1e−1]
`scaler_type`	Choice	['robust', 'standard']
`max_steps`	Choice	[500, 1000]
`batch_size`	Choice	[8, 32]
`random_seed`	rand int	[1, 20]

3.2. Quantile outputs

TFT can be extended to provide quantile outputs, allowing the model to predict not just the expected value, but also the uncertainty of the prediction. This can be particularly useful in applications where it is essential to have a more complete understanding of the prediction, such as in weather or economic forecasting [60].

Quantile regression is a type of regression that predicts not only the mean of the target variable but also the distribution of the target variable. In TFT, this can be achieved by modifying the final layer of the network to output quantiles of the prediction instead of just the mean. Given a set of quantiles, τ, the quantile regression loss is defined as:

$$L(\hat{y}, y) = \sum_{i=1}^{n} \rho_{\tau_i}(y_i - \hat{y}_i) \tag{16}$$

where n is the number of data points, $\hat{y}$ is the predicted value, y is the actual value, and ρ_τ is the quantile loss function defined as:

$$\rho_\tau(z) = \begin{cases} \tau z & z \geq 0 \\ (\tau - 1)z & z < 0 \end{cases} \tag{17}$$

The quantile loss function measures the deviation between the predicted and actual values at the specified quantiles. By optimizing this loss function, the TFT model can learn to predict the quantiles of the target variable, allowing it to capture not only the mean of the target but also its distribution [61].

3.3. Hypertuning

To perform a complete evaluation of the AutoTFT model, this model is evaluated using default and adjusted hyperparameters. The adjusted hyperparameters being evaluated are presented in Table 3. These hyperparameters are the hidden size, number of heads (`n_head`), learning rate, scaler type, maximum steps (`max_steps`), batch size, and random seed.

During the hypertuning process combinations of hyperparameters are tried and evaluated automatically to define the best structure for the model, therefore evaluations of the hyperparameters individually are not necessary. The complete architecture of the AutoTFT model is presented in Fig. 3.

The original model is compared to the hypertuned one in a statistical assessment where the Wilcoxon test is applied. The Wilcoxon test is

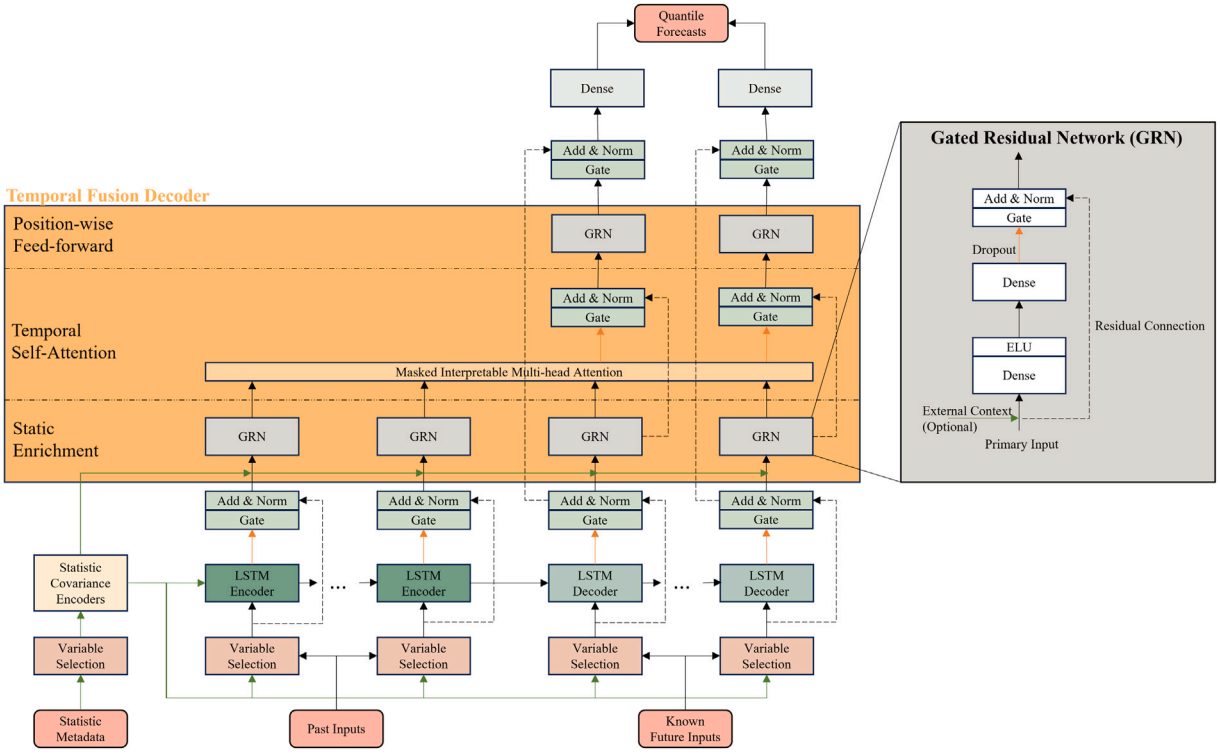

Fig. 3. AutoTFT architecture.

a nonparametric hypothesis test to assess whether the population mean ranks differ. In this paper, the Wilcoxon test was applied considering the predicted and observed values (for the short and median-term). The null hypothesis is that the distributions of the two vectors are the same. The alternative hypothesis is that the true location shift is not equal to zero [62].

Considering that there is a discussion regarding the use of shallow architectures and deep learning. For the final evaluation of the AutoTFT, a benchmarking with the ANFIS [63], bootstrap aggregation (bagged) [64], sequential learning (boosted) [65], stacked generalization [66] ensemble learning approaches [67], and LSTM [68] are presented. Variations of the structure of these models are presented for a complete assessment. The summary flowchart of the proposed approach and evaluation used in this paper is presented in Fig. 4.

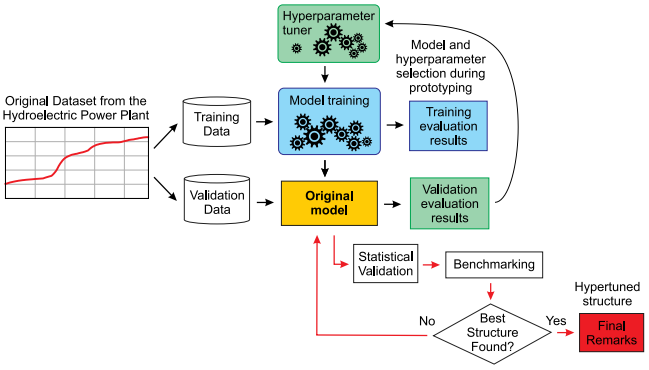

Fig. 4. Summary flowchart of the proposed approach.

4. Results and discussion

The TFT algorithm was written in Python. The experiments were computed on Google Colab using an NVIDIA Tesla T4 graphics processing unit of 16 GB with allocated 12 GB of random access memory. In the analysis presented in this paper, the RMSE, root mean squared percentage error (RMSPE), maximum absolute error (MaxAE), mean absolute error (MeanAE), and median absolute error (MedianAE) are evaluated [69].

4.1. Dataset

The considered data is based on measurements from the automatic hydraulic control of a hydraulic power plant during a flood. In 20 days and 1 hour, the level of the dam increased from 20.46% to 86.27%. This situation raises the problem of emergency situations in power plants with dams, and how is necessary for the manager of the power plant to make decisions when it happens.

Given this task, the recorded data from a hydraulic power plant located in Pelotas River, Santa Catarina state, Brazil, is used to perform a time series forecasting to have the information in advance, and do not face emergency situations. The considered measurements were recorded in July 2020, considering a time interval of 1 h, corresponding to 481 records.[1] The recorded data is presented in Fig. 5.

4.2. Model evaluation

The first evaluation is regarding the time step predicted horizon (previously defined as P) and the number of configurations explored (NCE). The results are presented in Table 4.

The errors were higher when the analysis is considering the median term compared to the short-term forecasting. Using a higher NCE the model needs more time to be computed, given that not all the results were better when more configurations were explored, the maximum of NCE equal to 10 is evaluated. For the median-term forecasting with a horizon equal to 30, the best result was found using an NCE equal to

[1] https://github.com/SFStefenon/AutoTFT.

Table 4
Median and short-term evaluation considering different horizons and configurations explored.

NCE	Horizon	RMSE	RMSPE	MaxAE	MeanAE	MedianAE	Time (s)
3	1	3.03×10^{-2}	3.51×10^{-3}	3.03×10^{-2}	3.03×10^{-2}	3.03×10^{-2}	123.94
	2	7.87×10^{-2}	9.12×10^{-3}	8.55×10^{-2}	7.84×10^{-2}	7.84×10^{-2}	101.83
	3	7.23×10^{-1}	8.38×10^{-2}	9.20×10^{-1}	7.04×10^{-1}	6.68×10^{-1}	122.09
	10	2.70×10^{-1}	3.13×10^{-2}	3.19×10^{-1}	2.66×10^{-1}	2.80×10^{-1}	219.21
	20	5.82×10^{-1}	6.77×10^{-2}	7.88×10^{-1}	5.76×10^{-1}	5.43×10^{-1}	608.96
	30	3.41	3.96×10^{-1}	6.15	2.96	2.41	898.52
5	1	3.33×10^{-2}	3.87×10^{-3}	3.33×10^{-2}	3.33×10^{-2}	3.33×10^{-2}	143.70
	2	9.22×10^{-3}	1.07×10^{-3}	1.27×10^{-2}	7.88×10^{-3}	7.88×10^{-3}	178.28
	3	7.94×10^{-3}	9.20×10^{-4}	1.13×10^{-2}	7.05×10^{-3}	7.52×10^{-3}	195.32
	10	1.26×10^{-1}	1.47×10^{-2}	2.86×10^{-1}	9.96×10^{-2}	6.17×10^{-2}	754.31
	20	1.73	2.01×10^{-1}	2.76	1.49	1.12	420.98
	30	8.93×10^{-1}	1.04×10^{-1}	1.48	8.15×10^{-1}	6.60×10^{-1}	1779.95
10	1	2.26×10^{-3}	2.62×10^{-4}	2.26×10^{-3}	2.26×10^{-3}	2.26×10^{-3}	371.92
	2	6.64×10^{-3}	7.70×10^{-4}	7.44×10^{-3}	6.58×10^{-3}	6.58×10^{-3}	481.35
	3	1.02×10^{-2}	1.18×10^{-3}	1.23×10^{-2}	9.66×10^{-3}	1.15×10^{-2}	316.18
	10	1.68×10^{-1}	1.94×10^{-2}	2.18×10^{-1}	1.63×10^{-1}	1.67×10^{-1}	901.76
	20	1.90	2.20×10^{-1}	2.72	1.81	1.84	3330.00
	30	1.32	1.54×10^{-1}	2.48	1.22	1.25	2381.50

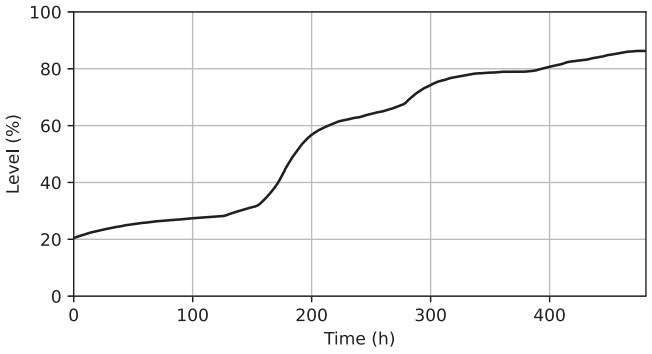

Fig. 5. Original signal: Variation of the level of the reservoir of the power hydraulic plant.

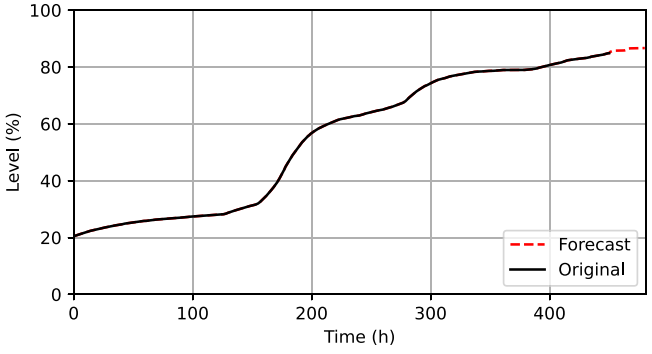

Fig. 6. Original and forecasted signals.

5, using this configuration the forecasting and the original signal are presented in Fig. 6.

The horizon equal to 30 means that the manager would have the information of a possible emergency situation 30 h before it occurs, which would be appropriate to alert the team to perform the precautions given this possible scenario. The comparison of the observed and the predicted results for a horizon of 30 h are presented in Fig. 7.

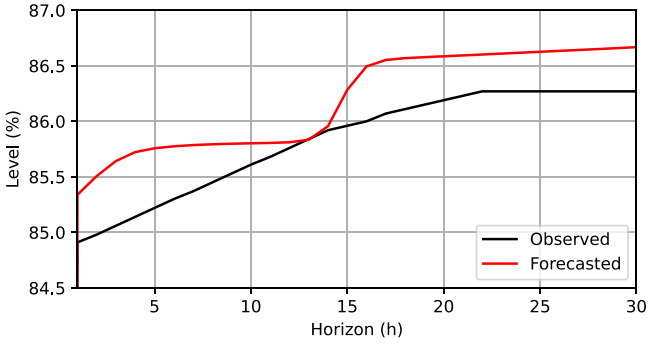

Fig. 7. Observed and forecasted signals.

Even considering a measurement of the level of the dam during a flood, the variation of the level was not abrupt which makes possible medium-term forecasting, if abrupt variations happen short-term forecasting would be more adequate where the model forecasts considering a shorter horizon. Short-term forecasting has a more dynamic adjustment since the has fewer steps ahead to predict.

First evaluations were computed using default values of hyperparameters (pre-defined) for the AutoTFT. Given that complete hypertuning with adjusted setup values would possibly enhance the performance of the model, the adjusted hyperparameters were evaluated. After the hypertuning, the model using the updated values is called adjusted. The setup for this evaluation is explained in Section 3.3.

Considering the hyperparameter search space, statistical analysis with the evaluation of the variance of error (VAR), the standard deviation of error (Std Dev), and the interquartile range of error (IQR) are also presented in Table 5. The use of hyperparameters given a range of possible variations was shown to be more promising than the use of default values in this evaluation. The RMSE of the adjusted AutoTFT was 80 times lower than the model using its default setup for short-term forecasting. Considering the p-value the null hypothesis is rejected for all models, meaning that the distribution of the two populations, from which the samples are taken, has a median distinct.

The AutoTFT model had a promising outcome for the time series forecasting considering a signal with low noise, not having high frequencies. When high noising signals are considered the application of this method is less recommended since the focus of the predictions of the proposed method is the signal trend. When high frequencies can be disregarded, the combination of denoising techniques and trend-decomposing methods may be a solution to deal with complex signals having high nonlinearities.

Table 5
Statistical evaluation of the errors given default and adjusted values of hyperparameters.

	Short-term $P = 3$		Median-term $P = 30$	
	Default	Adjusted	Default	Adjusted
RMSE	2.23×10^{-2}	2.78×10^{-3}	2.15	1.72
RMSPE	2.58×10^{-3}	3.23×10^{-4}	2.50×10^{-1}	2.00×10^{-1}
MaxAE	2.79×10^{-2}	4.34×10^{-3}	2.57	2.08
MeanAE	2.09×10^{-2}	2.38×10^{-3}	2.09	1.67
MedianAE	2.49×10^{-2}	1.90×10^{-3}	2.26	1.79
VAR	6.21×10^{-5}	4.61×10^{-6}	2.40×10^{-1}	1.65×10^{-1}
Std Dev	7.88×10^{-3}	2.15×10^{-3}	4.90×10^{-1}	4.06×10^{-1}
IQR	1.80×10^{-2}	5.26×10^{-3}	4.66×10^{-1}	7.42×10^{-1}
P-value[a]	0.375	0.125	9.31×10^{-10}	9.31×10^{-10}

[a] Results of Wilcoxon test.

Table 6
Benchmarking (one step ahead).

Model	Structure	RMSE	RMSPE	MaxAE	MeanAE	MedianAE	Time (s)
ANFIS	GP	1.25×10^{-1}	1.50×10^{-2}	2.75×10^{-1}	1.04×10^{-1}	9.93×10^{-2}	10.22
	FCM	8.45×10^{-2}	1.01×10^{-2}	1.63×10^{-1}	6.95×10^{-2}	6.89×10^{-2}	**3.84**
	Subtractive	1.29×10^{-1}	1.53×10^{-2}	2.32×10^{-1}	1.05×10^{-1}	1.10×10^{-1}	4.20
Ensemble	Stacked	4.28	5.45×10^{-1}	6.01	4.10	3.99	26.75
	Bagged	3.04	3.83×10^{-1}	3.48	3.02	3.09	19.40
	Boosted	2.76	3.48×10^{-1}	3.18	2.75	2.82	188.55
LSTM	Adam	1.81	2.24×10^{-1}	2.18	1.79	1.76	22.21
	SGDM	1.38	1.68×10^{-1}	2.09	1.27	1.26	19.00
	RMSprop	3.29×10^{-1}	4.11×10^{-2}	6.59×10^{-1}	2.79×10^{-1}	2.44×10^{-1}	22.96
AutoTFT	Default	2.14×10^{-1}	2.48×10^{-2}	2.14×10^{-1}	2.14×10^{-1}	2.14×10^{-1}	437.14
	Adjusted	$\mathbf{4.12} \times 10^{-2}$	4.77×10^{-3}	$\mathbf{4.12} \times 10^{-2}$	$\mathbf{4.12} \times 10^{-2}$	$\mathbf{4.12} \times 10^{-2}$	495.05

The underlined values are the best results for each method and the bold values are the overall best results.

To highlight that the applied hypertuned AutoTFT (adjusted) is appropriate to perform the time series forecasting presented in this paper, Table 6 presents a comparative analysis for one step ahead to other models. The ANFIS, ensemble learning approaches, LSTM, and AutoTFT using default settings are compared to the proposed strategy.

The long-term forecasting horizon helps the operator to have the information in advance, however, when the horizon becomes bigger the predictions may have higher errors. Considering that one step ahead means one hour in this study, one-step-ahead forecasting is acceptable for ab action in the power plant decision-making, and therefore, is used for the final benchmarking presented here.

Considering the ANFIS model, for a complete evaluation, different structures were analyzed, the fuzzy c-means (FCM), grid partition (GP), and subtractive clustering are considered. For the ensemble approach, the stacked, bagged, and boosted are evaluated. For the LSTM, the results of the use of the Adam, stochastic gradient descent with momentum (SGDM), and root mean squared propagation (RMSprop) optimizers are presented.

The AutoTFT with default setup had results slightly better than the LSTM with RMSProp optimizer. If this model version is compared to the ANFIS, the model becomes inferior and inadequate. Considering a defined hyperparameter search space (adjusted model), the results of the AutoTFT were considerably better than all other models. These results show that proper model tuning may enhance its forecasting capability, being the best strategy for the task discussed here.

5. Final remarks

The time series forecasting of the level of dams in hydroelectric power plants shows to be a promising way to improve the security of the power plant giving extra information for decision-making during emergency situations. Additionally, the information may be incorporated into the stochastic optimization problem of power generation scheduling. Given those possibilities, the application of the proposed method showed to be practicable.

The results present in this work showed that wider hypertuning is better than the use of standard (pre-defined) tuning values. Considering an adjusted model, it was possible to evaluate the forecasting considering different horizons. The evaluation showed that it becomes harder to predict values considering longer terms.

With an RMSE of 4.12×10^{-2} the adjusted AutoTFT outperforms well-established models such as ANFIS, LSTM, bootstrap aggregation, sequential learning, and stacked generalization ensemble learning approaches for one step ahead. A disadvantage of the AutoTFT is its higher requested computational effort, which results in more time for training, in this study this was not a restriction since the evaluation was performed offline, however, for application in embedded systems the AutoTFT may not be the most suitable approach.

In future research, it would be interesting to evaluate the performance of this AutoTFT model considering signals with higher noise, which was not the case in the measurement of the level of the hydroelectric dam presented here. Given high nonlinearities, the use of a hybrid method, such as some discussed in the literature review would be interesting.

CRediT authorship contribution statement

Stefano Frizzo Stefenon: Writing – original draft. **Laio Oriel Seman:** Methodology, Software, Formal analysis. **Luiza Scapinello Aquino da Silva:** Writing – review & editing. **Viviana Cocco Mariani:** Writing – review & editing. **Leandro dos Santos Coelho:** 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

The dataset is available at: https://github.com/SFStefenon/AutoTFT.

References

[1] Santos K, Finardi E. Piecewise linear approximations for hydropower production function applied on the hydrothermal unit commitment problem. Int J Electr Power Energy Syst 2022;135:107464. http://dx.doi.org/10.1016/j.ijepes.2021.107464.

[2] Larroyd PV, Pedrini R, Beltrán F, Teixeira G, Finardi EC, Picarelli LB. Dealing with negative inflows in the long-term hydrothermal scheduling problem. Energies 2022;15(3):1115. http://dx.doi.org/10.3390/en15031115.

[3] Gomes e Souza H, Finardi EC, Brito BH, Takigawa FYK. Partitioning approach based on convex hull and multiple choice for solving hydro unit-commitment problems. Electr Power Syst Res 2022;211:108285. http://dx.doi.org/10.1016/j.epsr.2022.108285.

[4] Colonetti B, Finardi E, Brito S, Zavala V. Parallel dual dynamic integer programming for large-scale hydrothermal unit-commitment. IEEE Trans Power Syst 2023;38(3):2926–38. http://dx.doi.org/10.1109/TPWRS.2022.3187059.

[5] Brito BH, Finardi EC, Takigawa FYK, Pereira AI, Gosmann RP, Weiss LA, Fernandes A, Morais DTSd. Exploring symmetry in a short-term hydro scheduling problem: The case of the santo antônio hydro plant. J Water Resour Plan Manage 2022;148(1):05021026. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0001495.

[6] Araújo PVN, Amaro VE, Silva RM, Lopes AB. Delimitation of flood areas based on a calibrated a DEM and geoprocessing: Case study on the Uruguay River, Itaqui, southern Brazil. Nat Hazards Earth Syst Sci 2019;19(1):237–50. http://dx.doi.org/10.5194/nhess-19-237-2019.

[7] Duan H-F, Gao X. Flooding control and hydro-energy assessment for urban stormwater drainage systems under climate change: framework development and case study. Water Resour Manag 2019;33:3523–45. http://dx.doi.org/10.1007/s11269-019-02314-8.

[8] Li Y, Bao T, Gong J, Shu X, Zhang K. The prediction of dam displacement time series using STL, extra-trees, and stacked LSTM neural network. IEEE Access 2020;8:94440–52. http://dx.doi.org/10.1109/ACCESS.2020.2995592.

[9] Stefenon SF, Seman LO, Aquino LS, dos Santos Coelho L. Wavelet-Seq2Seq-LSTM with attention for time series forecasting of level of dams in hydroelectric power plants. Energy 2023;274:127350. http://dx.doi.org/10.1016/j.energy.2023.127350.

[10] Ribeiro MHDM, da Silva RG, Moreno SR, Mariani VC, dos Santos Coelho L. Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. Int J Electr Power Energy Syst 2022;136:107712. http://dx.doi.org/10.1016/j.ijepes.2021.107712.

[11] Ren S, Zhang B, Wang W-J, Yuan Y, Guo C. Sedimentation and its response to management strategies of the Three Gorges reservoir, Yangtze river, China. Catena 2021;199:105096. http://dx.doi.org/10.1016/j.catena.2020.105096.

[12] Abritta R, Panoeiro F, Honório L, Silva Junior I, Marcato A, Guimarães A. Hydroelectric operation optimization and unexpected spillage indications. Energies 2020;13(20):5368. http://dx.doi.org/10.3390/en13205368.

[13] Stefenon SF, Oliveira JR, Coelho AS, Meyer LH. Diagnostic of insulators of conventional grid through LabVIEW analysis of FFT signal generated from ultrasound detector. IEEE Latin Am Trans 2017;15(5):884–9. http://dx.doi.org/10.1109/TLA.2017.7910202.

[14] López Santos M, García-Santiago X, Echevarría Camarero F, Blázquez Gil G, Carrasco Ortega P. Application of temporal fusion transformer for day-ahead PV power forecasting. Energies 2022;15(14):5232. http://dx.doi.org/10.3390/en15145232.

[15] Lim B, Arık SO, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. Int J Forecast 2021;37(4):1748–64. http://dx.doi.org/10.1016/j.ijforecast.2021.03.012.

[16] Colonetti B, Finardi E, Larroyd P, Beltran F. A novel cooperative multi-search benders decomposition for solving the hydrothermal unit-commitment problem. Int J Electr Power Energy Syst 2022;134:107390. http://dx.doi.org/10.1016/j.ijepes.2021.107390.

[17] dos Santos KV, Colonetti B, Finardi EC, Zavala VM. Accelerated dual dynamic integer programming applied to short-term power generation scheduling. Int J Electr Power Energy Syst 2023;145:108689. http://dx.doi.org/10.1016/j.ijepes.2022.108689.

[18] Vilanova RS, Zanetti SS, Cecílio RA. Assessing combinations of artificial neural networks input/output parameters to better simulate daily streamflow: Case of Brazilian atlantic rainforest watersheds. Comput Electron Agric 2019;167:105080. http://dx.doi.org/10.1016/j.compag.2019.105080.

[19] Saraiva SV, de Oliveira Carvalho F, Santos CAG, Barreto LC, Freire PKdM. Daily streamflow forecasting in sobradinho reservoir using machine learning models coupled with wavelet transform and bootstrapping. Appl Soft Comput 2021;102:107081. http://dx.doi.org/10.1016/j.asoc.2021.107081.

[20] Nazari LF, Camponogara E, Seman LO. Physics-informed neural networks for modeling water flows in a river channel. IEEE Trans Artif Intell 2022;1–15. http://dx.doi.org/10.1109/TAI.2022.3200028.

[21] Siqueira VA, Fan FM, de Paiva RCD, Ramos M-H, Collischonn W. Potential skill of continental-scale, medium-range ensemble streamflow forecasts for flood prediction in South America. J Hydrol 2020;590:125430. http://dx.doi.org/10.1016/j.jhydrol.2020.125430.

[22] Agarwal S, Roy P, Choudhury P, Debbarma N. Comparative study on stream flow prediction using the GMNN and wavelet-based GMNN. J Water Clim Change 2022;13(9):3323–37. http://dx.doi.org/10.2166/wcc.2022.226.

[23] Khazaee Poul A, Shourian M, Ebrahimi H. A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. Water Resour Manage 2019;33:2907–23. http://dx.doi.org/10.1007/s11269-019-02273-0.

[24] Stefenon SF, Seman LO, Mariani VC, Coelho LS. Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices. Energies 2023;16(3):1371. http://dx.doi.org/10.3390/en16031371.

[25] Li Z, Liu Y, Tanaka G. Multi-reservoir echo state networks with Hodrick–Prescott filter for nonlinear time-series prediction. Appl Soft Comput 2023;135:110021. http://dx.doi.org/10.1016/j.asoc.2023.110021.

[26] Branco NW, Cavalca MSM, Stefenon SF, Leithardt VRQ. Wavelet LSTM for fault forecasting in electrical power grids. Sensors 2022;22(21):8323. http://dx.doi.org/10.3390/s22218323.

[27] Stefenon SF, Ribeiro MHDM, Nied A, Mariani VC, Coelho LS, Leithardt VRQ, Silva LA, Seman LO. Hybrid wavelet stacking ensemble model for insulators contamination forecasting. IEEE Access 2021;9:66387–97. http://dx.doi.org/10.1109/ACCESS.2021.3076410.

[28] Medeiros A, Sartori A, Stefenon SF, Meyer LH, Nied A. Comparison of artificial intelligence techniques to failure prediction in contaminated insulators based on leakage current. J Intell Fuzzy Systems 2022;42(4):3285–98. http://dx.doi.org/10.3233/JIFS-211126.

[29] Seman LO, Stefenon SF, Mariani VC, dos Santos Coelho L. Ensemble learning methods using the Hodrick–Prescott filter for fault forecasting in insulators of the electrical power grids. Int J Electr Power Energy Syst 2023;152:109269. http://dx.doi.org/10.1016/j.ijepes.2023.109269.

[30] Yamasaki M, Freire RZ, Seman LO, Stefenon SF, Mariani VC, dos Santos Coelho L. Optimized hybrid ensemble learning approaches applied to very short-term load forecasting. Int J Electr Power Energy Syst 2024;155:109579. http://dx.doi.org/10.1016/j.ijepes.2023.109579.

[31] Pereira PJ, Costa N, Barros M, Cortez P, Durães D, Silva A, Machado J. A comparison of automated time series forecasting tools for smart cities. In: Progress in artificial intelligence, vol. 21, Lisbon, Portugal: Springer; 2022, p. 551–62. http://dx.doi.org/10.1007/978-3-031-16474-3_45.

[32] Salinas D, Flunkert V, Gasthaus J, Januschowski T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. Int J Forecast 2020;36(3):1181–91. http://dx.doi.org/10.1016/j.ijforecast.2019.07.001.

[33] Nazir A, Shaikh AK, Shah AS, Khalil A. Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (TFT). Results Eng 2023;17:100888. http://dx.doi.org/10.1016/j.rineng.2023.100888.

[34] Wang Z, Zhu Z, Xiao G, Bai B, Zhang Y. A transformer-based multi-entity load forecasting method for integrated energy systems. Front Energy Res 2022;10:952420. http://dx.doi.org/10.3389/fenrg.2022.952420.

[35] dos Santos KV, Colonetti B, Finardi EC, Zavala VM. Accelerated dual dynamic integer programming applied to short-term power generation scheduling. Int J Electr Power Energy Syst 2023;145:108689. http://dx.doi.org/10.1016/j.ijepes.2022.108689.

[36] Beltrán F, Finardi E, de Oliveira W. Two-stage and multi-stage decompositions for the medium-term hydrothermal scheduling problem: A computational comparison of solution techniques. Int J Electr Power Energy Syst 2021;127:106659. http://dx.doi.org/10.1016/j.ijepes.2020.106659.

[37] Beltrán F, Finardi EC, Fredo GM, de Oliveira W. Improving the performance of the stochastic dual dynamic programming algorithm using Chebyshev centers. Optim Eng 2020;23:147–68. http://dx.doi.org/10.1007/s11081-020-09558-z.

[38] Ben-Yelun I, Díaz-Lago M, Saucedo-Mora L, Sanz MÁ, Callado R, Montáns FJ. Self-learning locally-optimal hypertuning using maximum entropy, and comparison of machine learning approaches for estimating fatigue life in composite materials of the aerospace industry. Eng Struct 2023;283:115829. http://dx.doi.org/10.1016/j.engstruct.2023.115829.

[39] Sharma DK, Jain V, Dhingra B, Datta Gupta K, Ghosh U, Al-Numay W. A novel hypertuned prophet based power saving approach for IoT enabled smart homes. Trans Emerg Telecommun Technol 2022;e4597. http://dx.doi.org/10.1002/ett.4597.

[40] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: International conference on knowledge discovery & data mining. Vol. 25, New York, NY, USA: ACM SIGKDD; 2019, p. 2623–31. http://dx.doi.org/10.1145/3292500.3330701.

[41] Lai J-P, Lin Y-L, Lin H-C, Shih C-Y, Wang Y-P, Pai P-F. Tree-based machine learning models with optuna in predicting impedance values for circuit analysis. Micromachines 2023;14(2):265. http://dx.doi.org/10.3390/mi14020265.

[42] Yu J, Zhao Y, Pan R, Zhou X, Wei Z. Prediction of the critical temperature of superconductors based on two-layer feature selection and the optuna-stacking ensemble learning model. ACS Omega 2023;8(3):3078–90. http://dx.doi.org/10.1021/acsomega.2c06324.

[43] Arai K, Fujikawa I, Nakagawa Y, Momozaki T, Ogawa S. Modified prophet+ optuna prediction method for sales estimations. Int J Adv Comput Sci Appl 2022;13(8). http://dx.doi.org/10.14569/IJACSA.2022.0130809.

[44] Klaar ACR, Stefenon SF, Seman LO, Mariani VC, Coelho LS. Optimized EWT-Seq2Seq-LSTM with attention mechanism to insulators fault prediction. Sensors 2023;23(6):3202. http://dx.doi.org/10.3390/s23063202.

[45] Klaar ACR, Stefenon SF, Seman LO, Mariani VC, Coelho LdS. Structure optimization of ensemble learning methods and seasonal decomposition approaches to energy price forecasting in latin America: A case study about Mexico. Energies 2023;16(7):3184. http://dx.doi.org/10.3390/en16073184.

[46] Iqbal M, Iqbal M, Jaskani F, Iqbal K, Hassan A. Time-series prediction of cryptocurrency market using machine learning techniques. EAI Endors Trans Creat Technol 2021;8(28). http://dx.doi.org/10.4108/eai.7-7-2021.170286.

[47] Sauer J, Mariani VC, dos Santos Coelho L, Ribeiro MHDM, Rampazzo M. Extreme gradient boosting model based on improved jaya optimizer applied to forecasting energy consumption in residential buildings. Evol Syst 2022;13:577–88. http://dx.doi.org/10.1007/s12530-021-09404-2.

[48] Sopelsa Neto NF, Stefenon SF, Meyer LH, Ovejero RG, Leithardt VRQ. Fault prediction based on leakage current in contaminated insulators using enhanced time series forecasting models. Sensors 2022;22(16):6121. http://dx.doi.org/10.3390/s22166121.

[49] Ribeiro MHDM, da Silva RG, Ribeiro GT, Mariani VC, dos Santos Coelho L. Cooperative ensemble learning model improves electric short-term load forecasting. Chaos Solitons Fractals 2023;166:112982. http://dx.doi.org/10.1016/j.chaos.2022.112982.

[50] da Silva RG, Moreno SR, Ribeiro MHDM, Larcher JHK, Mariani VC, dos Santos Coelho L. Multi-step short-term wind speed forecasting based on multi-stage decomposition coupled with stacking-ensemble learning approach. Int J Electr Power Energy Syst 2022;143:108504. http://dx.doi.org/10.1016/j.ijepes.2022.108504.

[51] Shih S-Y, Sun F-K, Lee H-y. Temporal pattern attention for multivariate time series forecasting. Mach Learn 2019;108:1421–41. http://dx.doi.org/10.1007/s10994-019-05815-0.

[52] Hu J, Zheng W. Multistage attention network for multivariate time series prediction. Neurocomputing 2020;383:122–37. http://dx.doi.org/10.1016/j.neucom.2019.11.060.

[53] Shen L, Wang Y. TCCT: Tightly-coupled convolutional transformer on time series forecasting. Neurocomputing 2022;480:131–45. http://dx.doi.org/10.1016/j.neucom.2022.01.039.

[54] Huang L, Mao F, Zhang K, Li Z. Spatial-temporal convolutional transformer network for multivariate time series forecasting. Sensors 2022;22(3):841. http://dx.doi.org/10.3390/s22030841.

[55] Zhang H, Zou Y, Yang X, Yang H. A temporal fusion transformer for short-term freeway traffic speed multistep prediction. Neurocomputing 2022;500:329–40. http://dx.doi.org/10.1016/j.neucom.2022.05.083.

[56] Zeng P, Hu G, Zhou X, Li S, Liu P, Liu S. Muformer: A long sequence time-series forecasting model based on modified multi-head attention. Knowl-Based Syst 2022;254:109584. http://dx.doi.org/10.1016/j.knosys.2022.109584.

[57] Li D, Tan Y, Zhang Y, Miao S, He S. Probabilistic forecasting method for mid-term hourly load time series based on an improved temporal fusion transformer model. Int J Electr Power Energy Syst 2023;146:108743. http://dx.doi.org/10.1016/j.ijepes.2022.108743.

[58] Wu B, Wang L, Zeng Y-R. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. Energy 2022;252:123990. http://dx.doi.org/10.1016/j.energy.2022.123990.

[59] Huy PC, Minh NQ, Tien ND, Anh TTQ. Short-term electricity load forecasting based on temporal fusion transformer model. IEEE Access 2022;10:106296–304. http://dx.doi.org/10.1109/ACCESS.2022.3211941.

[60] Han Y, Tian Y, Yu L, Gao Y. Economic system forecasting based on temporal fusion transformers: Multi-dimensional evaluation and cross-model comparative analysis. Neurocomputing 2023;126500.

[61] Sappl J, Harders M, Rauch W. Machine learning for quantile regression of biogas production rates in anaerobic digesters. Sci Total Environ 2023;872:161923. http://dx.doi.org/10.1016/j.scitotenv.2023.161923.

[62] Meléndez R, Giraldo R, Leiva V. Sign, wilcoxon and mann-whitney tests for functional data: An approach based on random projections. Mathematics 2020;9(1):44. http://dx.doi.org/10.3390/math9010044.

[63] Stefenon SF, Kasburg C, Freire RZ, Silva Ferreira FC, Bertol DW, Nied A. Photovoltaic power forecasting using wavelet neuro-fuzzy for active solar trackers. J Intell Fuzzy Systems 2021;40(1):1083–96. http://dx.doi.org/10.3233/JIFS-201279.

[64] Kim D, Baek J-G. Bagging ensemble-based novel data generation method for univariate time series forecasting. Expert Syst Appl 2022;203:117366. http://dx.doi.org/10.1016/j.eswa.2022.117366.

[65] Agrawal RK, Muchahary F, Tripathi MM. Ensemble of relevance vector machines and boosted trees for electricity price forecasting. Appl Energy 2019;250:540–8. http://dx.doi.org/10.1016/j.apenergy.2019.05.062.

[66] Massaoudi M, Refaat SS, Chihi I, Trabelsi M, Oueslati FS, Abu-Rub H. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for short-term load forecasting. Energy 2021;214:118874. http://dx.doi.org/10.1016/j.energy.2020.118874.

[67] Li J, Hao J, Feng Q, Sun X, Liu M. Optimal selection of heterogeneous ensemble strategies of time series forecasting with multi-objective programming. Expert Syst Appl 2021;166:114091. http://dx.doi.org/10.1016/j.eswa.2020.114091.

[68] Borré A, Seman LO, Camponogara E, Stefenon SF, Mariani VC, Coelho LS. Machine fault detection using a hybrid CNN-LSTM attention-based model. Sensors 2023;23(9):4512. http://dx.doi.org/10.3390/s23094512.

[69] Perktold J, Prescott P. Source code for statsmodels.tools.eval_measures: Some measures for evaluation of prediction, tests and model selection. 2023, URL https://www.statsmodels.org/dev/_modules/statsmodels/tools/eval_measures.html.