scientific reports

Check for updates

OPEN

Convolutional neural networks with transfer learning for natural river flow prediction in ungauged basins

Henrique Echternacht[1], Luciana Campos[2], Alfeu Dias de Martinho[3]✉, Danilo Pinto Moreira de Souza[4], Rodrigo Barbosa de Santis[2], Tiago Silveira Gontijo[5], Matteo Bodini[6], Angela Gorgoglione[7], Camila Martins Saporetti[8] & Leonardo Goliatt[9]

With the increasing integration of artificial intelligence (AI) in several scientific domains, there is a rising demand for advanced AI tools capable of addressing advanced research challenges. A challenge of paramount importance lies in accurately predicting the streamflow within river basins. Effective river flow prediction holds significant relevance, particularly given the substantial societal implications of river usage, encompassing areas such as transportation, agriculture, and power generation. The present study introduces a novel approach to streamflow prediction involving the development of a Deep Learning (DL) model that combines a convolutional neural network with Transfer Learning (TL) techniques to predict streamflow in river systems. With the aim of training the developed DL model, the study employed a time-series dataset containing hydrological data related to two distinct river basins, i.e., Paraíba do Sul, in Brazil, and Zambezi in the state of Mozambique. The developed DL models exhibited the capability to effectively predict the river flow with a one-day horizon, relying on the preceding three or seven days of historical data. To overcome the limited availability of training data and reduce the training time of DL models, TL was leveraged to incorporate two additional distinct time-series datasets, i.e., historical streamflow data from the São Francisco River in Brazil, and climate data from Delhi, India. The application of TL significantly reduced training time, leading only to a minimal decrease in prediction performance. Indeed, in the case DL models were trained on data collected from the Paraíba do Sul River, a substantial reduction in training time was observed - up to 27% - with a modest percentage decrease of 0.31% in test predictive performance (R^2). Similarly, TL induced a significant reduction in training time of up to 48%, while resulting in a modest 2% reduction in test predictive performance (R^2) for the Zambezi dataset. The findings underscore the significance of TL as a strategic and viable approach to improve the efficiency of river flow prediction models in the context of basins with limited hydrological data available.

Reliable river flow data are critical, as they play a fundamental role in managing the environmental, societal, and economic aspects of water resources[1,2]. Indeed, accurate streamflow data enable hydrologists to assess the health of aquatic ecosystems, support flood and drought management, and shape decision-making processes in agriculture, industry, and urban development[3,4]. Additionally, by relying on accurate streamflow data, hydroelectric power facilities can regulate water discharge or retention to effectively face energy demand, thereby influencing river flow dynamics to support residential, agricultural, and industrial needs[5,6].

Streamflow modeling relying on river flow data has been leveraged in several applications, including reservoir operations[7], predicting hydroelectric energy output[8], mitigating flood risks[9], planning hydraulic

[1]Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil. [2]Department of Computer Science, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil. [3]Department of Exact Sciences and Technology, Púnguè University, Campus Universitário de Cambinde, Matundo, Tete, Mozambique. [4]Computational Modeling Program, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil. [5]Campus Centro Oeste, Federal University of São João del-Rei, Divinópolis, MG, Brazil. [6]Dipartimento di Economia, Management e Metodi Quantitativi, Università degli Studi di Milano, Milano, Italy. [7]Department of Fluid Mechanics and Environmental Engineering (IMFIA), School of Engineering, Universidad de la República, Montevideo, Uruguay. [8]Department of Computational Modeling, Polytechnic Institute, Rio de Janeiro State University, Nova Friburgo, RJ, Brazil. [9]Department of Computational and Applied Mechanics, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil. ✉email: alfeu.martinho@unipungue.ac.mz

infrastructure[10], evaluating human-induced impacts on water resources[11], and predicting extreme climatic events[12]. Such applications underscore the importance of reliable streamflow predictions[13]. As a result, the demand for accurate streamflow prediction to enhance water resource management has grown significantly, both on short-term and long-term horizons[13,14]. Nevertheless, ensuring reliability in such remains a significant challenge in the field, and such a challenge arises mainly from the dynamic interactions and high spatial and temporal variability of the factors governing the hydrological cycle, which drives the conversion of precipitation into river flows[15]. Additionally, the inherently nonlinear nature of rainfall-runoff processes further complicates the development of accurate predictive models[16].

The latest advancements in water monitoring technologies have significantly improved the ability to collect comprehensive daily data across entire watersheds, from primary sources to the smallest tributaries[17–19]. As a result, extensive historical data repositories were created, encompassing data representing basic flow-related measurements to more complex data, such as land-use patterns[20] and water pollutant concentrations in water[21]. Within the latter datasets, hydrological data are usually recorded periodically to enable effective monitoring, and the collected observations can be represented as time series, capturing both typical values and anomalies, such as unusually high or low river flow volumes.

The available data supported the development of Machine Learning (ML) based models to predict river behavior[22–24]. However, it must be noted that in past years, streamflow modeling has traditionally relied on Physically Based Models (PBMs) that are capable of representing the fundamental equations governing the hydrological cycle[25,26]. Consequently, implementing PBMs necessitates detailed information about the considered basin, such as climatic conditions, soil properties, and drainage channel characteristics.

The requirement for such extensive hydrometeorological data can make physically based models impractical in poorly monitored basins. While PBMs often achieve high accuracy when sufficient input data are available, they typically cannot function reliably in data-scarce contexts until the necessary parameters are provided[13]. In contrast, data-driven models can remain effective and offer reliable predictions even with limited input data[27], making them attractive alternatives in ungauged or under-monitored basins[28,29].

Regarding the employment of the latter method, it is worth noting that TL was already presented in 1976, with the publication of an article describing its novel application[30]. Since then, it has been proven to represent a viable solution to contemporary challenges related to the development of effective DL models[31], such as the need for large amounts of labeled data[32] and the high computational cost associated with the efficient training of deep neural networks[33]. In particular, the main idea of TL is to address the previously mentioned issues by reusing effective previously trained models and their acquired knowledge to solve similar problems. In this context, only the parameters of specific neural network layers are adjusted to fit the new application context, thereby avoiding the need to retrain the entire network.

In contrast, data-driven methods have recently emerged as powerful tools for streamflow prediction[34–37]. Unlike PBMs, data-driven models leverage patterns in historical data, often yielding accurate results with lower computational costs and without requiring the explicit representation of complex hydrological mechanisms[38]. Indeed, the latter models can identify correlations between input and output data of hydrological systems without considering the involved physical processes[22,39]. In particular, data-driven models can predict the hydrological components of interest by employing mathematical equations that only fit the available training data[40]. In addition, DL models have shown promising performance in capturing intricate land-atmosphere interactions, making them well-suited for river flow predictions, especially in data-rich regions[41,42].

Several traditional ML-based hydrological models, for instance, based on *k*-Nearest Neighbors, Support Vector Machine, Random Forest, Artificial Neural Networks, and DL, have proven to be effective in predicting the streamflow[22,39]. However, despite ML and DL-based models having proven their effectiveness, a persistent challenge in the context of hydrology is represented by the prediction in ungauged basins, where monitoring data are limited, sparse, or even nonexistent[43]. Indeed, nowadays, several world regions still present a low density of hydrometeorological stations compared to other data-rich regions[44]. Moreover, existing station networks suffer from issues related to limited historical data, uneven distribution, data series inconsistencies, and limited measurement periods[36,45]. Such data gaps are partly due to the region's vast size and the inaccessibility of areas where monitoring stations are located, such as national parks and indigenous lands.

Convolutional Neural Networks (CNNs), originally inspired by the concept of local receptive fields[46], have evolved significantly and are now widely applied beyond image analysis. In recent years, CNNs have also demonstrated strong performance in time-series prediction tasks, including hydrological forecasting, due to their ability to capture local temporal patterns and reduce dimensionality through convolution and pooling layers[47,48]. In this context, the development of data-driven models capable of generalizing beyond data-rich environments is of critical importance[49]. Recent studies[36,44,45,50] have demonstrated the effectiveness of machine learning and deep learning approaches for daily streamflow prediction in regions with limited hydrometeorological data, highlighting their strong performance under data-scarce conditions. Furthermore, it was noted that the effectiveness of ML models in streamflow prediction is not only due to their resilience in handling the scarcity of available hydrometeorological information but also in dealing with the frequent non-stationary nature of the considered data[45,51].

In addition to the approaches mentioned above, TL offers a still underexplored but promising approach to address the limited availability of hydrological training data in several world regions: the concept of TL in deep neural networks allows a model to be trained on well-monitored systems, with calibrated parameters subsequently applied to systems where monitoring data are limited or even unavailable[52,53]. Such an approach facilitates knowledge transfer from observed basins to a few measured or unmeasured basins[54–56]. Indeed, hydrologists could transfer learned DL parameters to predict the river flow in basins with limited or no available hydrological data through pre-training DL models on well-monitored basins. Thus, the TL technique enables knowledge transfer across basins with differing hydrological characteristics, allowing for more robust and widespread river

flow predictions[55,56]. As such, TL not only holds potential for advancing the reliability of streamflow prediction, but it also addresses the PBMs drawbacks by supporting model applications across diverse hydrological contexts, potentially improving in a significant manner DL predictive capabilities in the context of hydrology[54–56].

Research motivations

Despite the significant growth in hydrological modeling research in recent years, the Brazilian literature on AI, in particular, related to DL models for streamflow prediction, remains relatively underexplored[36,44,45,50]. Therefore, the present study aims to advance such a field by developing DL models that incorporate TL, which can significantly contribute to the field of hydrology research within the considered state.

In particular, the current study relies on the usage of Convolutional Neural Networks (CNNs) as a specialized kind of DL model, which has been widely adopted in the past literature[22,57,58]. Indeed, it is worth noting that, while CNNs have predominantly been applied in image classification and facial recognition research[59,60], they have also demonstrated remarkable performance in the context of time-series modeling, as evidenced by prior studies[61,62]. Moreover, the present study proposes the utilization of TL techniques[63], where a pre-trained CNN model, trained initially on a specific dataset, is repurposed to predict data from a different context, thus reducing the need for retraining the underlying DL model from scratch.

Recent literature on Transfer Learning (TL) has explored its applications in hydrological modeling and machine learning adaptation. Studies have categorized and reviewed TL methodologies, examining their connections with other machine learning approaches and outlining potential directions for future research[54]. Several works have investigated how pre-trained models can be adapted for use in different geographical regions or data-limited environments. One approach demonstrated the effectiveness of leveraging large-scale datasets to initialize and fine-tune machine learning models for streamflow prediction across diverse basins in Asia, South America, and Europe[64]. Another study proposed a TL framework based on Transformer architectures, enabling accurate flood forecasting in areas with limited data by utilizing models trained on data-rich basins without relying heavily on basin-specific attributes[65]. Additionally, TL techniques have been applied to improve the accuracy of Global Hydrological Models through LSTM-based corrections, particularly in ungauged basins, where results showed significant improvements over uncorrected models[66]. Further research has also examined the applicability of TL in snow-influenced watersheds using LSTM models trained on U.S. watershed data to predict daily streamflow in similar environments[55].

Research objectives

The primary objective of the study is to evaluate the effectiveness of CNNs, enhanced by TL, for reliable river flow predictions, specifically focusing on applications in both data-rich and data-scarce hydrological environments. To achieve the latter aim, the study was guided by the following key objectives:

- Investigating how CNN models, pre-trained on hydrometeorological data collected from well-monitored basins, could improve streamflow predictions in the context of ungauged basins with limited data available.
- Assessing the impact of contextual attributes represented in the pre-training datasets – such as basin-specific hydrological and climatic conditions – on the transferability and resulting performance of CNN models adapted via TL.
- Establishing general guidelines for effectively selecting pre-training datasets and implementing TL methodologies for hydrological modeling in multiple, but even often, data-limited environments.
- Analyzing the efficiency improvements in time and computational resources achieved when employing TL jointly with CNNs.

The forthcoming part of the article is organized as follows: "Material and methods" section describes the leveraged data and applied streamflow predicting methodologies; "Experimental results and discussion" section presents and discusses the obtained results, highlighting both their strengths and limitations; and "Conclusion" section concludes the present study.

Material and methods

Study sites

The first set of streamflow data on which the current work was focused was collected within the state of Brazil. Within such a state, the National Water Agency (Agência Nacional de Águas—ANA)[67] is currently responsible for overseeing the National Hydrometeorological Network (Rede Hidrometeorológica Nacional—RHN), *i.e.*, a comprehensive system comprising 4641 monitoring stations distributed across the whole country. Such stations are divided into distinct categories: several stations are responsible for monitoring hydrological parameters related to river systems, such as water levels, flow rates, water quality metrics, and sediment transport, while other stations primarily measure the evolution of the rainfall.

The available network of the above-mentioned hydrometeorological stations is particularly critical for managing water resources in several key regions of the state of Brazil, such as the Paraíba do Sul hydrographic basin, which is of significant importance, as it spans three of the country's most populous states, *i.e.*, Rio de Janeiro, São Paulo, and Minas Gerais. Additionally, its strategic location places the basin between Brazil's two major industrial hubs: Rio de Janeiro and São Paulo. Moreover, such a basin is home to an extensive industrial complex with over 6,000 manufacturing facilities and 120 hydroelectric power plants, contributing approximately 11% of the national gross domestic product[68].

The second set of considered data, leveraged to carry the present study, was collected from the Zambezi watershed, which is located in Southern Africa and spans eight African countries, *i.e.*, Angola, Zambia, Namibia, Botswana, Zimbabwe, Tanzania, Malawi, and Mozambique[69]. Precipitation phenomena play a crucial role in the

latter region, and its abundance or eventual scarcity may have significant impacts. Consequently, drought is the primary natural disaster affecting such basins. In response to the persisting climatic challenges, the region hosts some of the world's largest dams, including the Kariba Dam, which is the third largest globally, and the Cahora Bassa Dam, *i.e.*, the twelfth largest globally[70].

Figure 1 illustrates the locations of the river basins employed in the present study, along with the hydrological monitoring stations from which the historical streamflow data were collected.

Study scenarios

Moreover, in order to assess the effectiveness of TL methodologies, CNN models were trained using datasets collected from additional and distinct hydrometeorological contexts, relying on two extra sets of data, *i.e.*, river flow time series from the São Francisco river basin, located in the state of Brazil, and data of a different nature, specifically climate time-series data collected from the city of Delhi, India. Relying on the above-described datasets, three different experimental scenarios were developed in the present study:

- Baseline scenario with basin-specific hydrological data: in this first scenario, the CNN model was trained solely with data collected from the investigated basins. In particular, two separate CNN models were trained: the first one was trained to rely on data obtained exclusively from the Paraíba do Sul river basin, and the second CNN employed data uniquely from the Zambezi River basin. Such a baseline scenario allowed for a focused evaluation of the models' performance within a single data environment, thus enabling a direct assessment of the models' performance without employing the TL method.
- TL with analoguous hydrological data: the second scenario aimed to evaluate the effectiveness of TL by pre-training a CNN model on hydrologically similar data, recovered from the well-monitored São Francisco river basin, before adapting it to the target basins considered within the present study. Such a pre-trained CNN model was then fine-tuned to predict river flows within the two target basins on which the present research focused. This approach tested whether knowledge gained from hydrologically similar contexts could enhance the CNN models' predictive performance.
- TL with climate data: in the third scenario, a CNN model was pre-trained relying on climate time-series data recovered from a different geographical and hydrometeorological context. Such a DL model was then adapted to the target basins considered in the current study through TL. In particular, by employing climate data, this scenario investigated whether CNN models could properly reach remarkable predicting performance even when the pre-training dataset differs significantly from the target dataset, providing valuable insights to the employed data-driven model.

The overarching objectives of the above-described three scenarios aim to address the following research questions: could reliable and computationally efficient DL models be derived using TL methodologies in the context of the studied basins? Moreover, to what extent does the contextual disparity between the pre-training dataset and the target application context impact the CNN models' final performance? Such questions have been explored and addressed in detail throughout the remaining part of the article.

Historical data used for transfer learning

The datasets leveraged in the present work were divided into two main groups. In particular, the first group was composed of time series data that served as the focal point of the present research, for which the objective was to generate reliable predictions, *i.e.*:

- River flow data collected from the Paraíba do Sul basin: time series data in this case were composed of 4748 samples, collected between the years 2002 and 2014, representing the daily flow in cubic meters per second, measured at the collection point in the municipality of Campo dos Goytacazes municipality, taken from the ANA system[71] (refer to the Fig. 2).
- River flow of the Zambezi basin: time series composed of 5844 samples, collected from the years 2003 to 2018, representing the daily flow in cubic meters per second, measured at the Cahora Bassa dam, Mozambique[72] (refer to the Fig. 3).
- River flow at the São Francisco Basin: time series with 4018 samples, collected from the years 2002 to 2012, representing the daily flow in cubic meters per second, measured in the Três Marias reservoir and taken from the ONS (Operador Nacional do Sistema Elétrico—Brazilian National Electric System Operator)[73] (refer to the Fig. 4).
- Delhi city temperature time series: time series with 1575 samples collected between the years 2013 and 2017, representing the average daily temperature in degrees Celsius, measured in Delhi in India[74] (refer to the Fig. 5).

Regarding the above-reported time series, it must be noted that the same time interval could not be employed to train the CNN model due to the varying data availability collected from different national operators and environmental agencies. Moreover, such time series are typically non-stationary and exhibit distinct components, particularly trends reflecting long-term directional changes, seasonal components capturing recurring patterns due to periodic influences, and a high degree of stochasticity representing random fluctuations.

To ensure data consistency and mitigate the effects of missing or anomalous values, all time series were subjected to a preprocessing pipeline before training. Missing values were treated using linear interpolation to preserve temporal coherence while avoiding abrupt discontinuities. Although no explicit normalization of seasonality or anomalies was carried out in model inputs, this decomposition allowed us to qualitatively assess patterns and informed the windowing process.

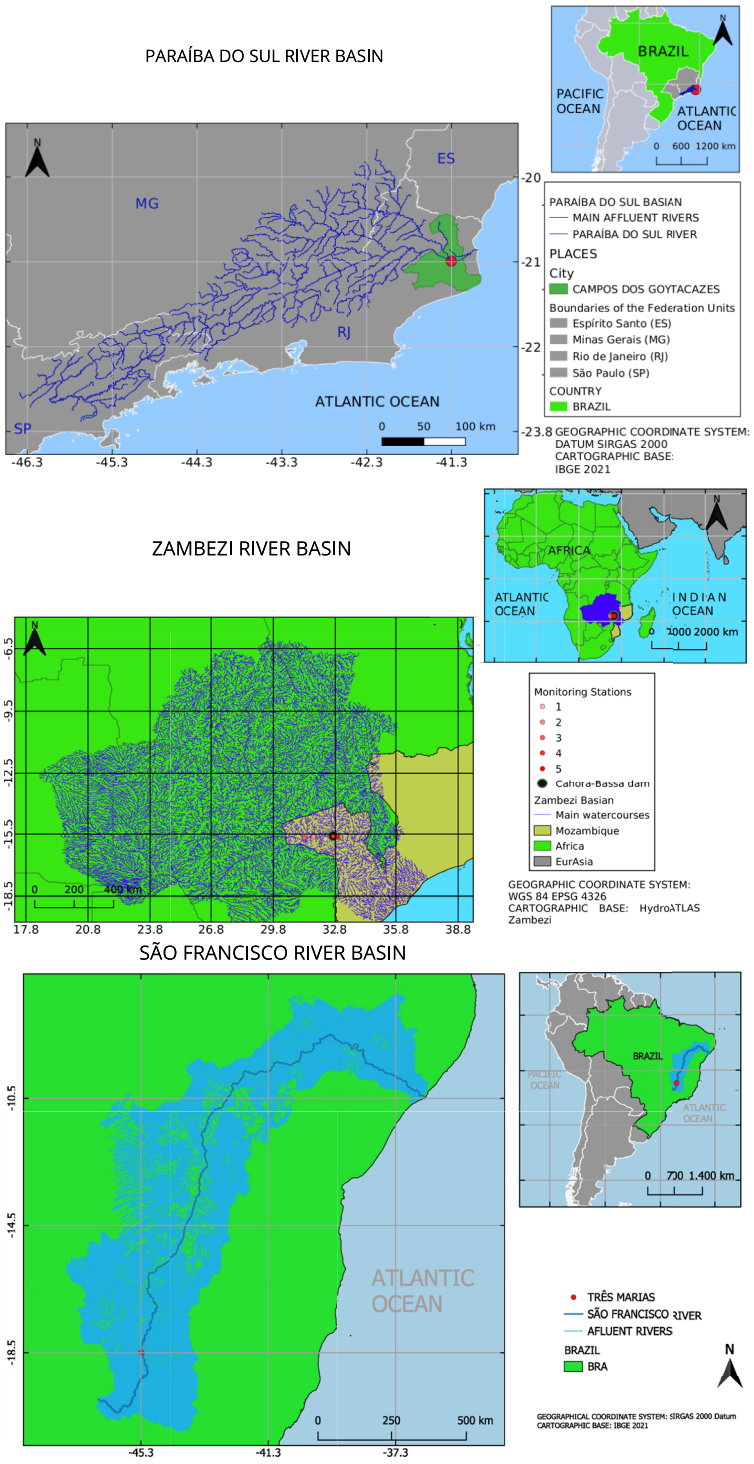

Fig. 1. Geographical locations of the river basins employed in the present study. The stations from which the data were collected are shown in red. The maps depict three river basins' spatial extents and key hydrological features. The top panel shows Brazil's Paraíba do Sul River Basin, including main watercourses, cities, and state boundaries (Espírito Santo, Minas Gerais, Rio de Janeiro, and São Paulo). The middle panel presents the Zambezi River Basin in Africa, highlighting monitoring stations, the Cabora-Bassa dam, and country boundaries, primarily within Mozambique. The bottom panel illustrates the São Francisco River Basin in Brazil, showing main watercourses, tributaries, and the Três Marias Reservoir. Insets in each panel provide a broader geographical context for the respective river basins. The reported maps were prepared by the Authors relying on the QGIS software, version 3.40.0-Bratislava.

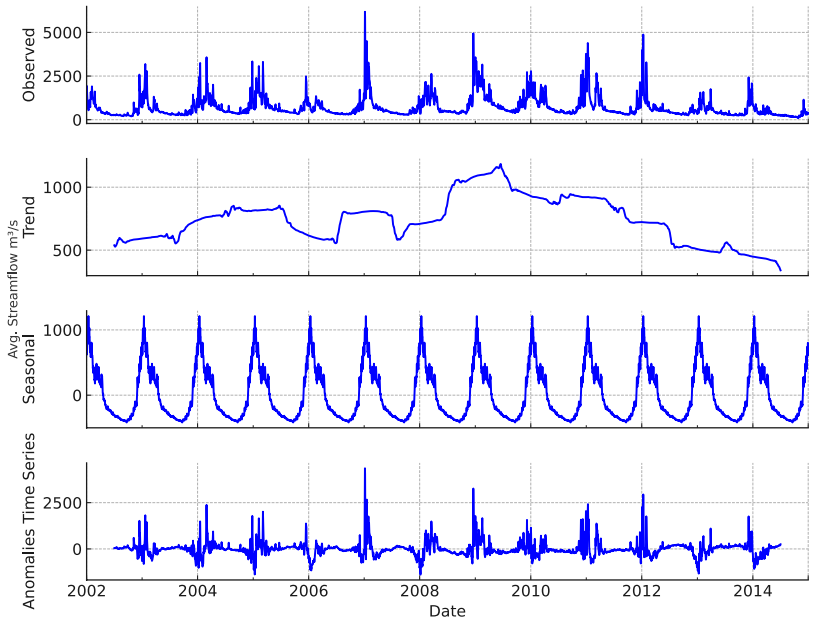

Fig. 2. Time series data of the Paraíba do Sul basin, collected daily from 2002 to 2014, representing river flow in cubic meters per second at the Campo dos Goytacazes monitoring station.

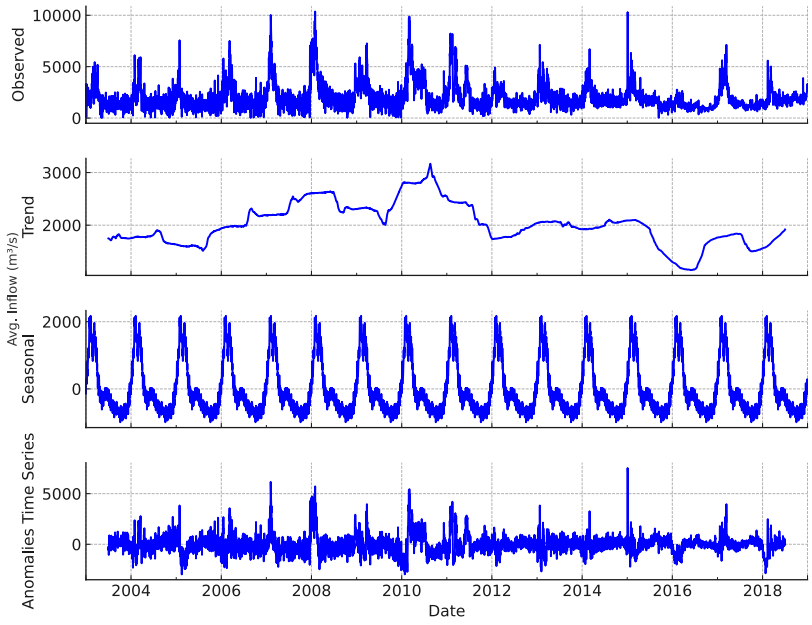

Fig. 3. Time series data of the Zambezi basin, collected daily from 2003 to 2018, representing river flow in cubic meters per second at the Cahora Bassa dam monitoring station.

In addition, a transformation of the time window data was performed relying on a proper function in which the size was defined for the input window M and output window N. Such a process is illustrated in Fig. 6, where the Paraíba do Sul historical data, initially in format (4748, 1), was transformed into format (4745, 3, 1) after applying windowing with $M = 3$ and $N = 1$. In particular, the present study conducted tests with $M = [3, 7]$ and $N = 1$. Additionally, it is worth noticing that, owing to the need for M data points to form the input window and N data points for the output window, the final number of windows was set as $(\text{initial size}) - M - N$.

The designed CNN and TL models

For each of the considered datasets, a specific CNN model was built. CNNs are highly parametric DL architectures in which recursive convolutions are applied to extract patterns from data across multiple computing layers. Regarding the structure of the developed CNNs, the KerasTuner optimizer[75] from the Keras Python library[76]

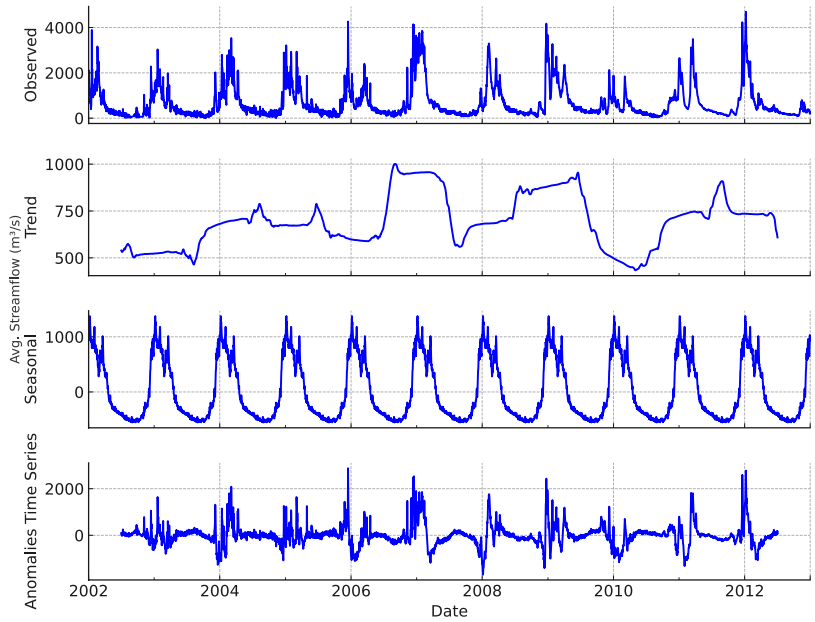

Fig. 4. Time series data of the São Francisco Basin, collected daily from 2002 to 2012, representing river flow in cubic meters per second at the Três Marias reservoir monitoring station.

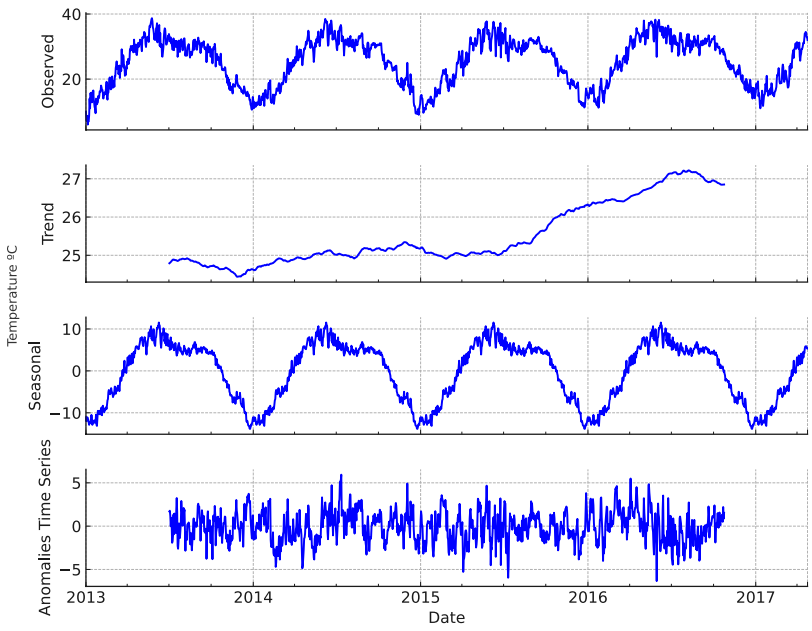

Fig. 5. Time series data of Delhi city, collected daily from 2013 to 2017, representing the average daily temperature in degrees Celsius.

was leveraged to define the number of units in each layer and their corresponding hyperparameters. The Mean Squared Error (MSE) metric was employed as the evaluation criterion to determine the best configurations within KerasTuner. Additionally, the considered datasets were randomly partitioned into training, validation, and test sets, which respectively accounted for 75%, 15%, and 10% of the total available data. The Table 1 specifically presents the search ranges for the architectural details and hyperparameters of each model layer.

In the present study, the development of the DL models was divided into two different stages, as shown in Fig. 7.

In particular:

- During stage 1, the objective was to evaluate the performance of the DL prediction models without employing TL, hence solely training them relying on the time series considered as the focus of this work (refer to section

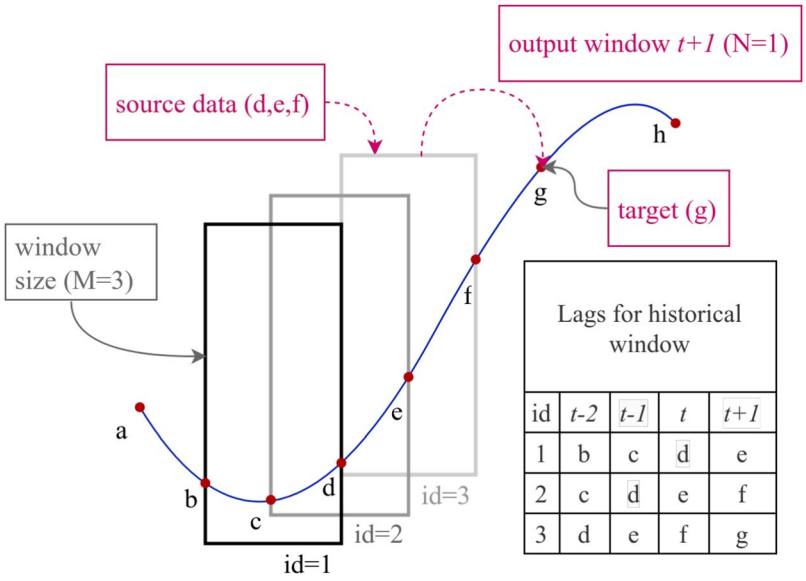

Fig. 6. Windowing process employed on the time series data. The river flow time series are shown in blue, and three different windows are displayed with Ids 1, 2, and 3. In this example, $M = 3$ means a temporal window with three preceding steps. The window with Id=3 has input data as entries (*d*, *e*, *f*); its target is the output *g*, which stands one day ahead.

Parameter	Search range
# of convolutional layers	[3–6]
# of pooling layers	[1, 2]
# of dense layers	[1, 3]
# of dropout layers	[0, 2, 5]
# of filters in convolutional layer	[16–256]
Kernel size	[1–3]
Activation function	[Linear-Sigmoid-Relu]
# of neurons in dense layer	[32–256]
Pool-size in pooling layer	[1–3]

Table 1. Search range provided to KerasTuner to define hyperparameters and architecture.

A of Fig. 8). A CNN-based model designed for the Paraíba do Sul river basin and a model for the Zambezi river basin was selected for subsequent comparison with CNN-based prediction models that leveraged TL. In particular, each selected DL model was obtained through a set of ten independent runs, finally considering the best-selected MSE value.

- During stage 2, DL models that relied on TL were selected. Relying on such a method, proper experiments were conducted for each DL model to determine the number of layers that should have been retrained after applying TL to achieve an accurate fit on unseen data. In particular, the performance of the CNN models was evaluated relying on two additional groups of data (refer to the Fig. 8), *i.e.*, datasets collected from the same context as group A in Fig. 8, in this case hydrological flow data which is labeled as group B in the Fig. 8, and datasets recovered from a different context than group A, in this case temperature data labeled as group C in the Fig. 8. Finally, the model selection process mirrored that of the preceding stage, involving a series of ten runs conducted for each considered model.

Within each of the above-described stages, the CNN-based models leveraged M previous days to predict one day ahead for each of the analyzed time series. Furthermore, in each stage, two values of M were evaluated: one model by setting $M = 3$ and another by setting $M = 7$. The set values of M were chosen empirically in the preliminary studies to determine the initial parameters.

Experimental results and discussion

Table 2 reports the values of the performance metrics MSE, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 for the best-recorded predicted results among the ten runs performed for each considered dataset. Such metrics are reported for the CNN model that relied on the previous three days and the model that relied on the last seven days (respectively denoted as $M = 3$ and $M = 7$ in Table 2). The architecture and

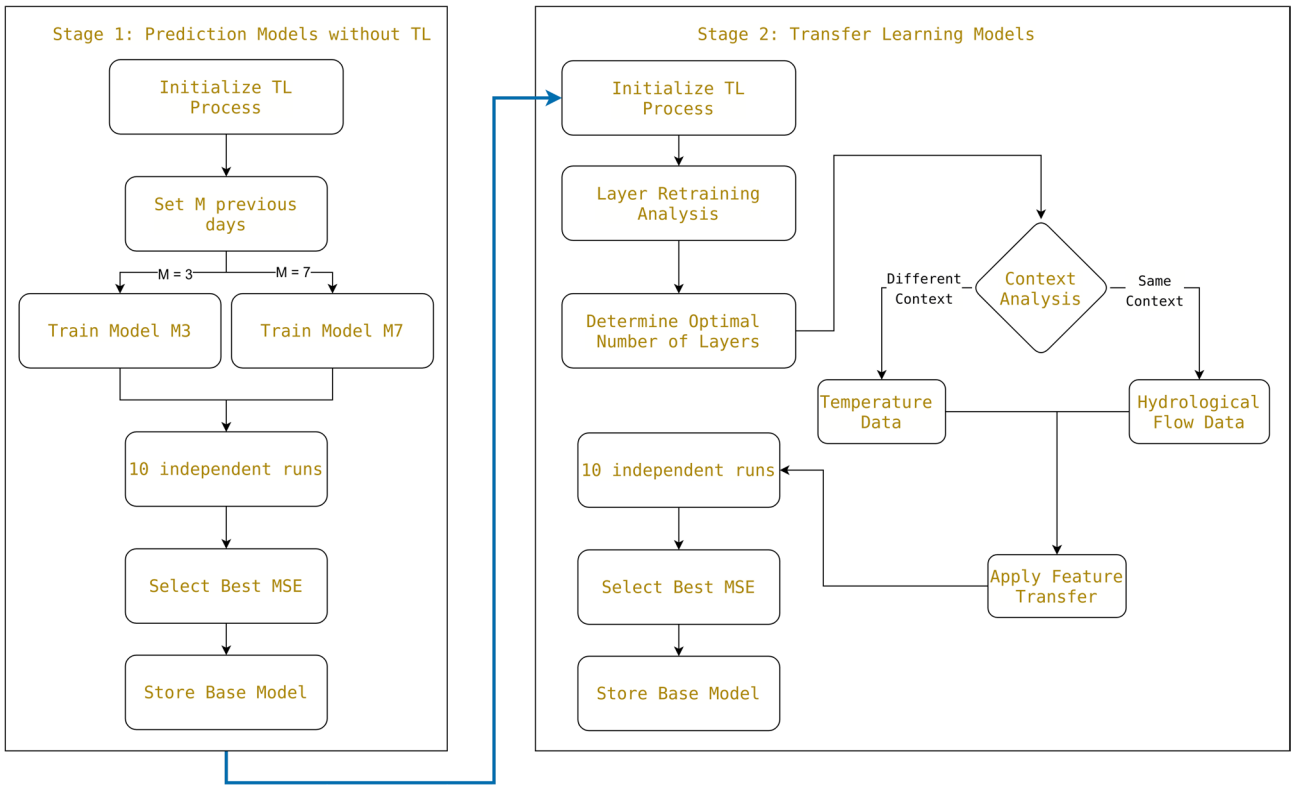

Fig. 7. Processes involved in the stages of prediction and TL models. The diagram illustrates the workflow from the initial data collection and preprocessing to the development and evaluation of CNN-based prediction models, followed by enhancing these models using TL techniques.

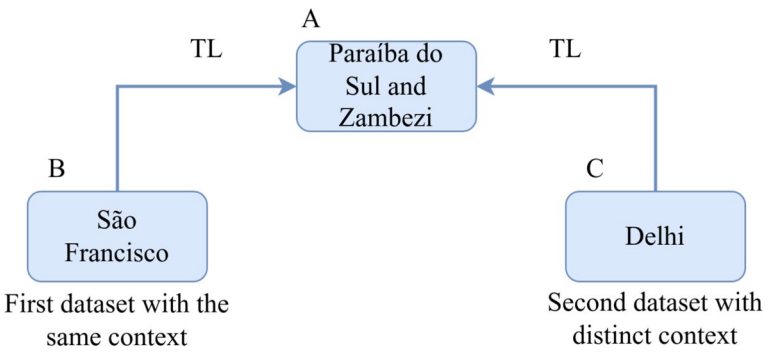

Fig. 8. Historical data used in the first stage. The diagram shows the flow of data from initial collection to model training and validation for CNN-based streamflow prediction. It also illustrates the use of TL with data from different domains.

hyperparameters of the developed CNN models for which performance results were reported in Table 2 were optimized within the specified search ranges outlined in Table 1.

As shown in bold characters within the Table 2, the CNN model with set $M = 3$ achieved the best performance on the data collected from the Paraíba do Sul basin, whereas the model with $M = 7$ yielded the best results when trained on the Zambezi basin data. To analyze the obtained results, it is worth noting that the Paraíba do Sul basin usually exhibits well-defined dry periods far from flood peaks. Thus, a larger input window M may introduce more noise during drought periods, affecting the performance of the DL model. The latter point could explain why the model in which $M = 3$ achieved better results for the basin considered. In contrast, the Zambezi basin usually displays less predictable dry periods; thus, a more significant amount of information could be beneficial for the underlying DL model. As a result, the CNN model that relied on $M = 7$ performed better for the latter basin.

As reported in Fig. 8, two different datasets were used to apply TL in the second stage of the present study, *i.e.*, data collected from the So Francisco basin and climate data from Delhi, India. Table 3 presents the results of

Training dataset		MSE	MAE	RMSE	R^2
Paraíba do Sul	M = 3	**0.00024**	**0.00085**	**0.01547**	**0.91052**
	M = 7	0.00026	0.01017	0.01642	0.89924
Zambezi	M = 3	0.00154	0.02733	0.03925	0.58061
	M = 7	**0.00145**	**0.02722**	**0.03809**	**0.60526**

Table 2. Result of the mean value of the metrics MSE, MAE, RMSE, and R^2 of the ten runs for Paraíba do Sul and Zambezi bases using $M = 3$ and $M = 7$. Top prediction performance results for the trained models are highlighted in bold.

Traning dataset		MSE	MAE	RMSE	R^2
São Francisco	M = 3	0.00201	0.02211	0.04486	0.94741
	M = 7	**0.00170**	**0.02047**	**0.04126**	**0.95593**
Delhi	M = 3	**0.00256**	**0.03977**	**0.05064**	**0.91716**
	M = 7	0.00292	0.04221	0.05406	0.90779

Table 3. Result of the average value of the metrics MSE, MAE, RMSE, and R^2 of the ten runs for base San Francisco and Delhi using $M = 3$ and $M = 7$. Top prediction performance results for the trained models are highlighted in bold.

Dataset	Metric	1 layer retrained	2 layers retrained
São Francisco (M = 7)			
Paraíba do Sul	MSE	0.00164	0.00070
	MAE	0.02781	0.02062
	RMSE	0.02834	0.02653
	R^2	0.69845	0.73700
Zambezi	MSE	0.00172	0.00157
	MAE	0.02994	0.02847
	RMSE	0.04153	0.03961
	R^2	0.53048	0.57300
Delhi (M = 3)			
Paraíba do Sul	MSE	0.00040	0.00025
	MAE	0.01097	0.00871
	RMSE	0.02010	0.01572
	R^2	0.84829	0.90764
Zambezi	MSE	0.00146	0.00154
	MAE	0.02990	0.02846
	RMSE	0.04127	0.03930
	R^2	0.53232	0.57956

Table 4. Performance of transfer learning models with different numbers of retrained layers for São Francisco (M = 7) and Delhi (M = 3) datasets.

the predictive CNN models trained on the São Francisco and Delhi datasets. The possibility of applying TL was confirmed due to the remarkable performance observed on the two different datasets considered. In addition, it is worth noticing (in bold characters within the Table 3) that for the São Francisco basin, the best performing CNN model was the one with $M = 7$, whereas for the Delhi dataset, the best prediction model relied on $M = 3$.

TL was then applied to the DL models trained on the Paraíba do Sul and Zambezi river datasets, *i.e.* by retraining the last two layers. It was noted that retraining the previous two layers showed improved prediction performance concerning the top DL models for both the São Francisco and Delhi datasets, as well as for the Paraíba do Sul and Zambezi datasets, as reported in Table 4.

In the context of TL applied to CNNs, it is an established practice to reuse the initial convolutional layers, which capture low-level spatiotemporal features (e.g., edges, textures, periodic patterns) transferable across domains. In contrast, the final Fully Connected (FC) layers learn task-specific representations tied to the source domain's output distribution. As a result, retraining only the last two FC layers is supported by both theoretical principles and empirical evidence[77–79]. The theoretical rationale is that convolutional layers extract generic,

Best architectures hyperparameters	Without TL		With TL	
	Paraiba do Sul	Zambezi	São Francisco	Delhi
M	3	7	7	3
# of Conv. Layers	5	5	6	5
# of pooling layers	1	1	1	1
# of dense layers	3	2	2	1
# of dropout layers	0	0	0	0
M	3	7	7	3
# of filters in Conv. Layer	112	176	48	240
Kernel size	2	2	2	2
Activation function	Relu	Relu	Relu	Relu
# of neurons in dense layer	256	128	224	256
Pool-size in pooling layer	1	1	1	1

Table 5. Best hyperparameters found for each model.

Using M = 3		Training time in seconds			
		With TL		% reduction in training time	
Datasets	Without TL	São Frascisco	Delhi	São Frascisco	Delhi
Paraíba do Sul	661.53	307.06	496.12	54%	25%
Zambezi	504.13	301.62	431.23	40%	14%
Using M = 7		**Training time in seconds**			
		With TL		**% reduction in training time**	
Datasets	**Without TL**	**São Frascisco**	**Delhi**	**São Frascisco**	**Delhi**
Paraíba do Sul	841.60	816.66	811.74	3%	4%
Zambezi	928.36	743.27	765.68	20%	18%

Table 6. Result of the tests performed to present the processing time with and without TL using M = 3 and M = 7.

hierarchical features, while fully connected (FC) layers near the output encode domain-specific mappings. Fine-tuning the latter preserves transferable, low-level knowledge while adapting high-level abstractions to the target task[80]. Empirical studies further validate the latter approach, demonstrating that retraining FC layers effectively balances computational efficiency and accuracy[81,82]. In particular, selectively tuning FC layers maintains the utility of source-domain features while realigning outputs to the target distribution. Last, but not least, techniques for automated optimization of these layers (e.g., via AutoFCL[83]) underscore the effectiveness of this selective fine-tuning strategy.

Table 5 compares the best DL architectures and hyperparameters obtained for the four different datasets considered, *i.e.* Paraíba do Sul, Zambezi, São Francisco, and Delhi. The CNN model architectures reported in the table were trained on such datasets with and without applying TL. Additionally, Table 6 compares the training times of CNN models with and without the application of TL. Such a comparison includes scenarios where TL was implemented using pre-training datasets from similar and different hydrological contexts. Moreover, the reported comparison was based on the top-performing CNN models, selected relying on the top MSE, MAE, RMSE, and R^2 values. The reported tests can be divided into two blocks, *i.e.* one with $M = 3$ and another with $M = 7$. In comparing training times, although both TL applications resulted in a reduction of at least 14% with $M = 3$ and 3% with $M = 7$, the most significant time savings were observed when the pre-training dataset was more similar; regarding the context, to the final dataset. As a final result, TL leads to a training time reduction of 54% with $M = 3$ and 20% with $M = 7$, as shown in Table 6.

The top prediction performance results for the trained models are highlighted in bold in Table 7. As a final result, such performance offers a comparative perspective on the impact of TL on processing efficiency under the considered specific conditions, in particular:

- For the Paraíba do Sul Basin, best results are noted for $M = 3$, which suggests a sensitivity to the basin's characteristic behavior, including well-defined dry periods distant from flood peaks. In hydrological terms, shorter time windows, such as $M = 3$, may be more suitable for capturing the rapid changes in flow due to short-term rainfall events or minor fluctuations in river flow, which are typical during dry periods. Longer windows, like $M = 7$, might capture broader trends and introduce noise during these dry periods, impairing the model's predictive accuracy.

Using M = 3			With TL	
Datasets	**Metrics**	**Without TL**	**São Frascisco**	**Delhi**
Paraíba do Sul	MSE	0.000239	0.00056	**0.00027**
	MAE	0.000859	0.01193	**0.00814**
	RMSE	0.01653	0.02373	**0.015474**
	R^2	0.810523	0.78809	**0.89722**
Zambezi	MSE	0.001541	0.00179	**0.00138**
	MAE	0.027332	0.03135	**0.02582**
	RMSE	0.039259	0.04234	**0.03724**
	R^2	0.580611	0.51146	**0.62209**
Using M = 7			**With TL**	
Datasets	**Metrics**	**Without TL**	**São Frascisco**	**Delhi**
Paraíba do Sul	MSE	**0.00026**	0.00034	0.00036
	MAE	**0.00982**	0.01308	0.01019
	RMSE	**0.01635**	0.01864	0.01912
	R^2	**0.89999**	0.87013	0.86333
Zambezi	MSE	0.00134	**0.00133**	0.00168
	MAE	0.02604	**0.02559**	0.02931
	RMSE	0.03661	**0.03654**	0.04107
	R^2	0.63530	**0.63654**	0.54084

Table 7. Comparison of results with TL and without TL, using M = 3 and M = 7. Top prediction performance results for the trained models are highlighted in bold.

- In contrast, the Zambezi Basin, with its less predictable dry periods, benefits from a larger input window, *i.e.* $M = 7$. Such a longer time window allows the model to capture better seasonal variations and hydrological patterns that span longer periods, including potential flood events and the complex interactions of the basin's tributaries. Hydrologically, this makes sense, as the model can integrate more data points, reflecting a broader understanding of the flow dynamics over time.

Regarding the above two points, within the comparison regarding the employed training time of the developed CNN models, it is possible to observe that the value of *M* directly influenced the performance of the TL model. Indeed, when setting $M = 3$, the pre-training dataset with a different context achieved better results, with a reduction of up to 40% in training time and a 4% improvement in performance, as measured by the R^2 metric, when using the Zambezi dataset. Similarly, when employing the Paraíba do Sul dataset, a 25% reduction in training time was achieved, with only a 2% decrease in performance in terms of the R^2 metric. On the other hand, in the case of $M = 7$, datasets from similar contexts outperformed in almost all cases, with a reduction of up to 20% in training time and a 0.2% improvement in the R^2 metric when using the Zambezi dataset. Relying on the Paraíba do Sul data, a 4% reduction in training time was achieved, with a 2% decrease in performance concerning the R^2 metric.

Furthermore, it is worth noting that for a value of $M = 3$, the context of the TL pre-training datasets had no significant impact on the final performance, with datasets from different contexts yielding the best results. However, for a higher value of $M = 7$, the context of the pre-training dataset directly influenced the final performance, with pre-trained TL models using datasets from the same context outperforming others. Additionally, it is worth noting that the use of TL consistently resulted in a reduction in computational cost, which was particularly significant in some cases.

The findings of this study align with recent advancements demonstrating the effectiveness of deep learning models in hydrological forecasting. One study applied an AutoML framework in conjunction with remote-sensing rainfall data to predict streamflow in the Amazon basin, achieving strong predictive performance, although data scarcity was not specifically addressed[36]. Another study implemented Informer-based TL models for forecasting in data-limited regions, reporting outcomes similar to those of this work in terms of reduced training time and reliable accuracy[52]. In comparison, the approach presented here offers a more streamlined and computationally efficient architecture by utilizing convolutional neural networks (CNNs) with shorter input sequences. Furthermore, unlike previous efforts that relied on LSTM-corrected global hydrological models[66], the CNN-based TL method proposed in this study eliminates the need for large-scale model calibration while maintaining competitive performance, particularly in ungauged basins. These comparisons highlight the potential of CNN-based TL as a practical and effective solution for hydrological forecasting under limited data and computational constraints.

The R^2 values highlight variations in performance across different datasets and model configurations. To better understand these differences, we provide a comprehensive discussion of the key factors influencing the impact of TL. Specifically, we examine why TL enhanced performance in certain cases while leading to a slight reduction in others.

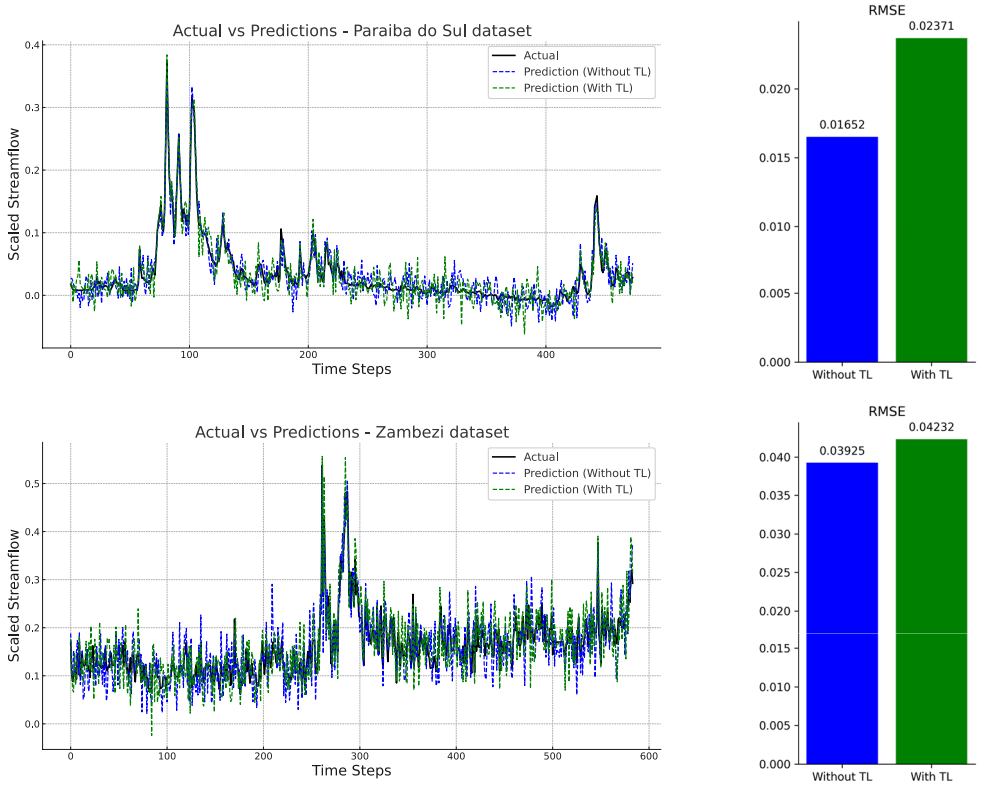

Fig. 9. Comparison of streamflow predictions with and without TL for the Paraíba do Sul and Zambezi River datasets. The left panels show the actual and predicted streamflow, while the right panels present RMSE values for both configurations. The results highlight that, although TL slightly increased the error in these cases, it still provided competitive predictions by reusing pre-trained features, reducing training time and computational effort.

- Contextual Similarity of Datasets: TL performs well when the pre-training dataset shares hydrological characteristics similar to the target dataset. Our study found that for M = 3, datasets from different hydrological contexts (such as the Zambezi and São Francisco basins) provided better results, likely because the CNN models could adapt to short-term flow dynamics more easily. However, for M = 7, the models performed better when the context of the pre-training datasets was more similar to the target data (such as the Paraíba do Sul and Zambezi basins). This suggests that TL is more effective when the hydrological processes governing the river flows are similar, such as rainfall patterns, seasonal variability, and flow regimes, which can aid the model in extracting relevant features more effectively.
- Model Configuration and Input Window Size (M): The difference in performance between M = 3 and M = 7 is partly attributed to the nature of the basins. For example, the Paraíba do Sul basin, with well-defined dry periods, performed better with M = 3, which helps avoid the inclusion of noise from longer periods. Conversely, with more unpredictable dry periods, the Zambezi basin benefited from the longer time window M = 7, which allowed the model to capture more complex, longer-term hydrological patterns, improving performance.
- Noise and Overfitting: In cases where TL was applied with datasets from dissimilar contexts (e.g., Delhi's climate data), the model could introduce more noise, leading to slight performance reductions. This may be due to the model's attempt to adapt to irrelevant or mismatched features, which could affect its generalization capacity, particularly with the longer input window (M = 7).

Figure 9 compares the time series computed with and without TL for the Paraíba do Sul and Zambezi River datasets. Even as depicted by the latter figure, as a final result, the usage of TL for the São Francisco and Delhi datasets suggests that the trained DL models were capable of leveraging pre-trained weights and features, leading to improved predicting performance, if compared to the DL models trained from scratch.

The findings of this study align with recent advancements demonstrating the effectiveness of deep learning models in hydrological forecasting. One study applied an AutoML framework in conjunction with remote-sensing rainfall data to predict streamflow in the Amazon basin, achieving strong predictive performance, although data scarcity was not specifically addressed[36]. Another study implemented Informer-based TL models for forecasting in data-limited regions, reporting outcomes similar to those of this work in terms of reduced training time and reliable accuracy[52]. In comparison, the approach presented here offers a more streamlined and computationally efficient architecture by utilizing convolutional neural networks (CNNs) with shorter input sequences. Furthermore, unlike previous efforts that relied on LSTM-corrected global hydrological models[66], the CNN-based TL method proposed in this study eliminates the need for large-scale model calibration while

maintaining competitive performance, particularly in ungauged basins. These comparisons highlight the potential of CNN-based TL as a practical and effective solution for hydrological forecasting under limited data and computational constraints.

Model strengths and limitations

The present study proved that TL may be viable for optimizing hydrological models under data constraints, particularly when pre-trained on similar environmental datasets. The method effectively balances model accuracy and computational efficiency, offering practical implications for streamflow prediction in data-scarce world regions.

TL showed remarkable prediction performance when applied to the São Francisco and Delhi datasets, which helped improve predictions for the Paraíba do Sul and Zambezi River datasets. This success can be attributed to TL's ability to leverage previously learned hydrological features from related datasets. Indeed, São Francisco and Delhi likely share common hydrological patterns, such as seasonal variations and extreme events, which the TL models can use to improve predictive performance on new datasets. From a hydrological perspective, such a result indicates that the model can generalize better by learning standard features relevant across different basins.

The effectiveness of TL with datasets from similar and different contexts further emphasizes that the model performs better when pre-trained on datasets from the same hydrological context. Indeed, hydrological processes governing river flows, such as rainfall patterns, evaporation, and seasonal variability, tend to be more similar between datasets from the same region or basin type. Conversely, datasets from different hydrological contexts, such as Delhi's temperature data, introduce additional variability, which explains the reduced performance. Nevertheless, the latter point also highlights the DL models' flexibility in adapting to diverse environmental conditions.

A notable advantage of TL has been noted in experimental results as its ability to reduce the training time, particularly when the pre-training dataset comes from the same hydrological context. This efficiency is attributed to the model's ability to reuse learned features, avoiding the need to start training the DL models from scratch. By adapting faster to new datasets, the learning process becomes more efficient. From a hydrological perspective, the latter result suggests that CNN models can leverage basin-specific dynamics, such as flow regimes and seasonal variability, to expedite training and enhance performance even with limited data.

The reported results also reveal that TL-based models reduce training time and maintain robust performance, even under challenging transfer scenarios. For instance, the proposed approach employed CNNs trained on data-rich regions, including streamflow data from the São Francisco Basin and climate data from Delhi, to enhance model adaptability in new river basins with limited data. These pre-training datasets created distinct and challenging scenarios to evaluate the model's generalization capacity across domains with varying environmental characteristics. The findings demonstrate that the model effectively optimizes its predictive capacity in the target basins, Paraíba do Sul and Zambezi, while maintaining manageable computational demands.

Nonetheless, one significant limitation of TL is the potential for negative transfer, particularly when there is a significant divergence between the pre-training dataset and the target domain[84–86]. Moreover, it must be noted that discrepancies in feature space, temporal dynamics, or underlying climatic and hydrological processes can result in decreased model performance[87–89]. In our study, for example, pre-training using climate data from Delhi led to a modest reduction in accuracy when the model was applied to the Paraíba do Sul basin (Table 7). Furthermore, there remains a lack of well-established strategies for determining what knowledge to transfer and how to transfer it: In practice, these decisions are often made through empirical testing and trial-and-error approaches[52,89,90]. Although recent studies have proposed automated and adaptive transfer mechanisms, these methods are still in the early stages of development and have not yet been broadly implemented[87,89,90]. The previously discussed challenges underscore the importance of:

1. Selecting pre-training datasets with strong hydrological or climatic similarity to the target domain;
2. Applying adaptive transfer techniques to mitigate the risk of negative transfer;
3. Avoiding the use of noisy or corrupted pre-training data, which can lead to overfitting to irrelevant features and degradation in downstream task performance;
4. Ensuring structural compatibility between source and target domains to maximize the effectiveness of feature transfer.

Finally, such hydrological insights underscore that the models' prediction performance is influenced not only by ML techniques but also by the extent to which the hydrological characteristics of each basin are represented in the training data. Understanding the underlying hydrological processes can guide future model development and improve the applicability of machine learning approaches to hydrological modeling. The ability to achieve reduced training times without significant performance loss positions TL-based models as a practical solution for large datasets, where extensive training times might otherwise render their use impractical.

Conclusion

In the present study, DL models were developed relying on TL to analyze their performance regarding training time and predictive performance. The central research questions guiding the carried investigation were: Can accurate DL models with low computational overhead be obtained by implementing TL? To what extent does the context of the pre-training data influence the final DL model prediction performance? To address such questions, this study explored the predictive capabilities of TL models for streamflow prediction within river basins, an issue of substantial environmental, economic, and societal importance.

The DL models were also constructed without using TL as a basis for comparing training time and performance. Thus, it can be asserted that the incorporation of TL resulted in notable reductions in training

time, with a decrease of up to 54% observed while incurring only a slight 2% decrease in the model performance, as indicated by the R^2 metric. Furthermore, it is worth noting that in most cases, pre-trained TL models within contexts aligned with the target domain yielded significantly greater reductions in training time than using pre-training data from unrelated contexts. Such a difference ranged from a minimum of 3 to 54%.

It must be noted that, despite the promising results, it is important to acknowledge the limitations of TL. Indeed, a primary concern is negative transfer, where knowledge from divergent pre-training domains adversely impacts performance in the target domain[84–86]. Such a risk arises from domain mismatches in feature spaces, temporal dynamics, or hydrological processes, thereby degrading model efficacy[87,88]. Compounding this issue, the absence of systematic methods to determine transferable knowledge often requires empirical trial-and-error[52,89]. The latter limitations highlight the critical need for both selecting pre-training datasets with strong hydrological/climatic similarity to target domains and employing adaptive techniques to mitigate domain-mismatch risks[87,88].

While the present study focused on CNNs due to their ability to extract spatial and temporal features effectively in time series data, it is important to acknowledge that other deep learning architectures, such as Long Short-Term Memory (LSTM) networks or hybrid models (e.g., CNN-LSTM), have also demonstrated strong performance in hydrological forecasting. A direct comparison between CNN-TL and these alternative models was not included in this work, as our aim was to evaluate the viability of TL using CNNs under different contextual datasets. Nevertheless, future work should explore such comparative studies to assess the relative advantages of CNN-based TL against other architectures, particularly in terms of training efficiency, prediction accuracy, and robustness in data-scarce environments.

As outlined throughout this study, the relevance of TL is underscored by the notable reductions in the training time it provides while simultaneously minimizing any substantial performance degradation in the models. Nonetheless, several avenues for future research remain open, including but not limited to examining model performance when confronted with incomplete or noisy data, exploring training time efficiencies with larger datasets, and investigating the influence of additional temporal lags on the predictive model.

Data availability

The datasets used and/or analyzed during the current study are available from the Corresponding Author (alfeu.martinho@unipungue.ac.mz) upon reasonable request.

Code availability

Code can be obtained upon request from the Authors.

Received: 15 December 2024; Accepted: 12 June 2025
Published online: 04 July 2025

References

1. Jain, S. K. & Singh, V. P. *Water Resources Systems Planning and Management* (Elsevier, 2023).
2. Brocca, L. et al. River flow prediction in data scarce regions: Soil moisture integrated satellite rainfall products outperform rain gauge observations in West Africa. *Sci. Rep.* **10**, 12517. https://doi.org/10.1038/s41598-020-69343-x (2020).
3. Yaseen, Z. M., El-Shafie, A., Jaafar, O., Afan, H. A. & Sayl, K. N. Artificial intelligence based models for stream-flow forecasting: 2000–2015. *J. Hydrol.* **530**, 829–844. https://doi.org/10.1016/j.jhydrol.2015.10.038 (2015).
4. Leandro, F. R. et al. An alert system for flood forecasting based on multiple seasonal Holt-Winters models: A case study of southeast Brazil. *Sustain. Water Resour. Manag.* **10**, 171 (2024).
5. Kelman, J. Water supply to the two largest Brazilian metropolitan regions. *Aquat. Procedia* **5**, 13–21. https://doi.org/10.1016/j.aqpro.2015.10.004 (2015).
6. Zhang, X. et al. Flow regime changes in the Lancang river, revealed by integrated modeling with multiple earth observation datasets. *Sci. Total Environ.* **862**, 160656. https://doi.org/10.1016/j.scitotenv.2022.160656 (2023).
7. Hurkmans, R. T., Van Den Hurk, B. A. R. T., Schmeits, M., Wetterhall, F. & Pechlivanidis, I. G. Seasonal streamflow forecasting for fresh water reservoir management in the Netherlands: An assessment of multiple prediction systems. *J. Hydrometeorol.* **24**, 1275–1290. https://doi.org/10.1175/JHM-D-22-0107.1 (2023).
8. Liu, Y. et al. Effect of the quality of streamflow forecasts on the operation of cascade hydropower stations using stochastic optimization models. *Energy* **273**, 127298. https://doi.org/10.1016/j.energy.2023.127298 (2023).
9. Bahramian, K., Nathan, R., Western, A. W. & Ryu, D. Probabilistic conditioning and recalibration of an event-based flood forecasting model using real-time streamflow observations. *J. Hydrol. Eng.* **28**, 04023003. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002236 (2023).
10. Eldardiry, H. & Hossain, F. The value of long-term streamflow forecasts in adaptive reservoir operation: The case of the high Aswan dam in the transboundary Nile river basin. *J. Hydrometeorol.* **22**, 1099–1115. https://doi.org/10.1175/JHM-D-20-0241.1 (2021).
11. Park, S.-Y., Moon, H.-T., Kim, J.-S. & Lee, J.-H. Assessing the impact of human-induced and climate change-driven streamflow alterations on freshwater ecosystems. *Ecohydrol. Hydrobiol.* https://doi.org/10.1016/j.ecohyd.2023.09.003 (2023).
12. Tran, V. N., Ivanov, V. Y. & Kim, J. Data reformation–a novel data processing technique enhancing machine learning applicability for predicting streamflow extremes. *Adv. Water Resour.* **182**, 104569. https://doi.org/10.1016/j.advwatres.2023.104569 (2023).
13. Yifru, B. A., Lim, K. J. & Lee, S. Enhancing streamflow prediction physically consistently using process-based modeling and domain knowledge: A review. *Sustainability* **16**, 1376. https://doi.org/10.3390/su16041376 (2024).
14. Tran, T. D. & Kim, J. S. Machine learning modeling structures and framework for short-term forecasting and long-term projection of streamflow. *Stoch. Env. Res. Risk Assess.* **38**, 793–813. https://doi.org/10.1007/s00477-023-02621-y (2024).
15. Belvederesi, C., Zaghloul, M. S., Achari, G., Gupta, A. & Hassan, Q. K. Modelling river flow in cold and ungauged regions: A review of the purposes, methods, and challenges. *Environ. Rev.* **30**, 159–173. https://doi.org/10.1139/er-2021-0043 (2022).
16. Sivakumar, B., Berndtsson, R., Olsson, J. & Jinno, K. Evidence of chaos in the rainfall-runoff process. *Hydrol. Sci. J.* **46**, 131–145. https://doi.org/10.1080/02626660109492805 (2001).
17. Jan, F., Min-Allah, N. & Düştegör, D. Iot based smart water quality monitoring: Recent techniques, trends and challenges for domestic applications. *Water* **13**, 1729. https://doi.org/10.3390/w13131729 (2021).

18. Pham-Duc, B., Prigent, C., Aires, F. & Papa, F. Comparisons of global terrestrial surface water datasets over 15 years. *J. Hydrometeorol.* **18**, 993–1007. https://doi.org/10.1175/JHM-D-16-0206.1 (2017).
19. Paulson, R. W. *Hydrologic Data Collection for River Basin Management* 21–44 (Springer, 1999).
20. Pandey, P. C., Koutsias, N., Petropoulos, G. P., Srivastava, P. K. & Ben Dor, E. Land use/land cover in view of earth observation: Data sources, input dimensions, and classifiers-a review of the state of the art. *Geocarto Int.* **36**, 957–988. https://doi.org/10.1080/10106049.2019.1629647 (2021).
21. Zhu, M. et al. A review of the application of machine learning in water quality evaluation. *Eco-Environ. Health* **1**, 107–116. https://doi.org/10.1016/j.eehl.2022.06.001 (2022).
22. Ng, K. et al. A review of hybrid deep learning applications for streamflow forecasting. *J. Hydrol.* **625**, 130141. https://doi.org/10.1016/j.jhydrol.2023.130141 (2023).
23. Troin, M., Arsenault, R., Wood, A. W., Brissette, F. & Martel, J.-L. Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years. *Water Resour. Res.* **57**, e2020WR028392. https://doi.org/10.1029/2020WR028392 (2021).
24. Herschy, R. W. Chapter 10. weirs and flumes. In: *Streamflow Measurement*, vol. 3, CRC Press (2014).
25. Aires, U. R. V. et al. Modeling of surface sediment concentration in the Doce river basin using satellite remote sensing. *J. Environ. Manage.* **323**, 116207. https://doi.org/10.1016/j.jenvman.2022.116207 (2022).
26. Jaiswal, R., Ali, S. & Bharti, B. Comparative evaluation of conceptual and physical rainfall-runoff models. *Appl. Water Sci.* **10**, 48. https://doi.org/10.1007/s13201-019-1122-6 (2020).
27. Saravani, M. J. et al. Predicting chlorophyll-a concentrations in the world's largest lakes using Kolmogorov-Arnold networks. *Environ. Sci. Technol.* **59**, 1801–1810. https://doi.org/10.1021/acs.est.4c11113 (2025).
28. Shen, C. et al. Hess opinions: Incubating deep-learning-powered hydrologic science advances as a community. *Hydrol. Earth Syst. Sci.* **22**, 5639–5656. https://doi.org/10.5194/hess-22-5639-2018 (2018).
29. Balthazar, L. D. et al. Long-term natural streamflow forecasting under drought scenarios using data-intelligence modeling. *Water Cycle* **5**, 266–277. https://doi.org/10.1016/j.watcyc.2024.07.001 (2024).
30. Bozinovski S. Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica* (2020). https://doi.org/10.31449/inf.v44i3.2828.
31. Shao, L., Zhu, F. & Li, X. Transfer learning for visual categorization: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **26**, 1019–1034. https://doi.org/10.1109/TNNLS.2014.2330900 (2015).
32. Tan, C. et al. A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018* (eds Kůrková, V. et al.) 270–279 (Springer International Publishing, 2018). https://doi.org/10.1007/978-3-030-01424-7_27.
33. Zhang, Q. et al. Waste image classification based on transfer learning and convolutional neural network. *Waste Manage.* **135**, 150–157. https://doi.org/10.1016/j.wasman.2021.08.038 (2021).
34. Vilaseca, F., Castro, A., Chreties, C. & Gorgoglione, A. Assessing influential rainfall-runoff variables to simulate daily streamflow using random forest. *Hydrol. Sci. J.* **68**, 1738–1753. https://doi.org/10.1080/02626667.2023.2232356 (2023).
35. Gorodetskaya, Y., Silva, R. O., de Melo Ribeiro, C. B. & Goliatt, L. Enhancing short-term streamflow forecasting of extreme events: A wavelet-artificial neural network hybrid approach. *Water Cycle* **5**, 297–312. https://doi.org/10.1016/j.watcyc.2024.09.001 (2024).
36. Bodini, M. Daily streamflow forecasting using AutoML and remote-sensing-estimated rainfall datasets in the amazon biomes. *Signals* **5**, 659–689. https://doi.org/10.3390/signals5040037 (2024).
37. de Sousa Jr, M. F. et al. Streamflow prediction based on machine learning models and rainfall estimated by remote sensing in the Brazilian Savanna and Amazon biomes transition. *Model. Earth Syst. Environ.* **10**, 1191–1202 (2024).
38. Gorgoglione, A., Russo, C., Gioia, A., Iacobellis, V. & Castro, A. First flush occurrence prediction and ranking of its influential variables in urban watersheds: Evaluation of XGBoost and SHAP techniques. In Gervasi, O., Murgante, B., Misra, S., Rocha, A. M. A. C. & Garau, C. (eds.) *Computational Science and Its Applications—ICCSA 2022 Workshops*, 423–434 (Springer International Publishing, Cham, 2022). https://doi.org/10.1007/978-3-031-10545-6_29.
39. Kumar, V., Kedam, N., Sharma, K., Mehta, D. & Caloiero, T. Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. *Water* **15**, 2572. https://doi.org/10.3390/w15142572 (2023).
40. Souza, D. P., Carlos, G. E., da Christo, S. E., Martinho, A. D. & Goliatt, L. Multi-criteria polynomial neural networks for hydrological time series modeling. *Earth Sci. Inform.* **18**, 1–19 (2025).
41. Shen, C., Chen, X. & Laloy, E. Broadening the use of machine learning in hydrology. *Front. Water* https://doi.org/10.3389/frwa.2021.681023 (2021).
42. Tao, H. et al. Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: A comprehensive review, assessment, and possible future research directions. *Eng. Appl. Artif. Intell.* **129**, 107559. https://doi.org/10.1016/j.engappai.2023.107559 (2024).
43. Sivapalan, M. et al. IAHS decade on predictions in ungauged basins (pub), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* **48**, 857–880. https://doi.org/10.1623/hysj.48.6.857.51421 (2003).
44. Filho, H. et al. Nowcast flood predictions in the Amazon watershed based on the remotely sensed rainfall product PDIRnow and artificial neural networks. *Environ. Monit. Assess.* **196**, 245. https://doi.org/10.1007/s10661-024-12396-6 (2023).
45. De Sousa, J. et al. Streamflow prediction based on machine learning models and rainfall estimated by remote sensing in the Brazilian Savanna and Amazon biomes transition. *Model. Earth Syst. Environ.* **10**, 1191–1202. https://doi.org/10.1007/s40808-023-01837-9 (2023).
46. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**, 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455 (1968).
47. Zhao, B., Lu, H., Chen, S., Liu, J. & Wu, D. Convolutional neural networks for time series classification. *J. Syst. Eng. Electron.* **28**, 162–169. https://doi.org/10.21629/JSEE.2017.01.18 (2017).
48. Gao, J. et al. Robusttad: Robust time series anomaly detection via decomposition and convolutional neural networks. CoRR. arXiv:abs/2002.09545 (2020).
49. Narbondo, S., Gorgoglione, A., Crisci, M. & Chreties, C. Enhancing physical similarity approach to predict runoff in ungauged watersheds in sub-tropical regions. *Water* **12**, 528. https://doi.org/10.3390/w12020528 (2020).
50. Silva, E., Coutinho, A., Cardoso, J. & Bezerra, S. Jucazinho dam streamflow prediction: A comparative analysis of machine learning techniques. *Hydrology* **11**, 97. https://doi.org/10.3390/hydrology11070097 (2024).
51. Xie, T. et al. Hybrid forecasting model for non-stationary daily runoff series: A case study in the Han River Basin, China. *J. Hydrol.* **577**, 123915. https://doi.org/10.1016/j.jhydrol.2019.123915 (2019).
52. Ghobadi, F., Yaseen, Z. M. & Kang, D. Long-term streamflow forecasting in data-scarce regions: Insightful investigation for leveraging satellite-derived data, informer architecture, and concurrent fine-tuning transfer learning. *J. Hydrol.* **631**, 130772. https://doi.org/10.1016/j.jhydrol.2024.130772 (2024).
53. Muhammad, A. U. & Abba, S. I. Transfer learning for streamflow forecasting using unguaged MOPEX basins data set. *Earth Sci. Inf.* **16**, 1241–1264. https://doi.org/10.1007/s12145-023-00952-6 (2023).
54. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**, 1345–1359. https://doi.org/10.1109/TKDE.2009.191 (2010).
55. Khoshkalam, Y., Rousseau, A. N., Rahmani, F., Shen, C. & Abbasnezhadi, K. Applying transfer learning techniques to enhance the accuracy of streamflow prediction produced by long short-term memory networks with data integration. *J. Hydrol.* **622**, 129682. https://doi.org/10.1016/j.jhydrol.2023.129682 (2023).

56. Yao, Y. et al. Can transfer learning improve hydrological predictions in the alpine regions?. *J. Hydrol.* **625**, 130038. https://doi.org/10.1016/j.jhydrol.2023.130038 (2023).
57. Le, X.-H., Nguyen, D.-H., Jung, S., Yeon, M. & Lee, G. Comparison of deep learning techniques for river streamflow forecasting. *IEEE Access* **9**, 71805–71820. https://doi.org/10.1109/ACCESS.2021.3077703 (2021).
58. Zhao, X. et al. A comprehensive review of methods for hydrological forecasting based on deep learning. *Water* **16**, 1407. https://doi.org/10.3390/w16101407 (2024).
59. Bodini, M. A review of facial landmark extraction in 2D images and videos using deep learning. *Big Data Cogn. Comput.* **3**, 14. https://doi.org/10.3390/bdcc3010014 (2019).
60. Bodini, M., D'Amelio, A., Grossi, G., Lanzarotti, R. & Lin, J. Single sample face recognition by sparse recovery of deep-learned LDA features. In *Advanced Concepts for Intelligent Vision Systems* (eds Blanc-Talon, J. et al.) 297–308 (Springer International Publishing, 2018). https://doi.org/10.1007/978-3-030-01449-0_25.
61. Shen, F., Liu, J. & Wu, K. Multivariate time series forecasting based on elastic net and high-order fuzzy cognitive maps: A case study on human action prediction through EEG signals. *IEEE Trans. Fuzzy Syst.* **29**, 2336–2348. https://doi.org/10.1109/TFUZZ.2020.2998513 (2021).
62. Han, Z., Zhao, J., Leung, H., Ma, K. F. & Wang, W. A review of deep learning models for time series prediction. *IEEE Sens. J.* **21**, 7833–7848. https://doi.org/10.1109/JSEN.2019.2923982 (2021).
63. Wen, T. & Keyes, R. Time series anomaly detection using convolutional neural networks and transfer learning (2019). arXiv:1905.13628.
64. Ma, K. et al. Transferring hydrologic data across continents-leveraging data-rich regions to improve hydrologic prediction in data-sparse regions. *Water Resour. Res.* **57**, e2020WR028600. https://doi.org/10.1029/2020WR028600 (2021).
65. Xu, Y. et al. Deep transfer learning based on transformer for flood forecasting in data-sparse basins. *J. Hydrol.* **625**, 129956. https://doi.org/10.1016/j.jhydrol.2023.129956 (2023).
66. Tang, S. et al. Optimal postprocessing strategies with LSTM for global streamflow prediction in ungauged basins. *Water Resour. Res.* **59**, e2022WR034352. https://doi.org/10.1029/2022WR034352 (2023).
67. ANA. Hidroweb. sistema de informações hidrológicas (2020).
68. Ioris, A. A. R. Os limites políticos de uma reforma incompleta: A implementação da lei dos recursos hídricos na bacia do paraíba do sul. *Revista Brasileira de Estudos Urbanos e Regionais* **10**, 61. https://doi.org/10.22296/2317-1529.2008v10n1p61 (2008).
69. Martinho, A. D., Ribeiro, C. B., Gorodetskaya, Y., Fonseca, T. L. & Goliatt, L. Extreme learning machine with evolutionary parameter tuning applied to forecast the daily natural flow at Cahora Bassa dam, Mozambique. In: *International Conference on Bioinspired Methods and Their Applications*, 255–267 (Springer, 2020).
70. Souza, D. P. M., Martinho, A. D., Rocha, C. C., da Christo, S. E. & Goliatt, L. Hybrid particle swarm optimization and group method of data handling for short-term prediction of natural daily streamflows. *Model. Earth Syst. Environ.* **8**, 5743–5759. https://doi.org/10.1007/s40808-022-01466-8 (2022).
71. ANA. Sistema de Hidrometria. http://www.snirh.gov.br/hidrotelemetria/Mapa.aspx (2022). [Online; Accessed 17 Dec 2022].
72. de Cahora Bassa, C. H. Company's official web site. https://www.hcb.co.mz/ (2022). [Online; Accessed 17 Dec 2022].
73. ONS. Operador nacional do sistema elétrico. https://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/dados_hidrologicos_vazoes.aspx (2022). [Online; Accessed 17 Dec 2022].
74. PES. PES University Web Site. https://pes.edu/ (2022). [Online; Accessed 17 Dec 2022].
75. O'Malley, T. et al. Kerastuner. https://github.com/keras-team/keras-tuner (2019).
76. Chollet, F. et al. Keras. https://github.com/fchollet/keras (2015).
77. Mahmoud, K. A. et al. Transfer learning by fine-tuning pre-trained convolutional neural network architectures for switchgear fault detection using thermal imaging. *Alex. Eng. J.* **103**, 327–342. https://doi.org/10.1016/j.aej.2024.05.102 (2024).
78. Ashraf, A., Naz, S., Shirazi, S. H., Razzak, I. & Parsad, M. Deep transfer learning for alzheimer neurological disorder detection. *Multimed. Tools Appl.* https://doi.org/10.1007/s11042-020-10331-8 (2021).
79. Di Maggio, L. G. Intelligent fault diagnosis of industrial bearings using transfer learning and CNNs pre-trained for audio classification. *Sensors* **23**, 211. https://doi.org/10.3390/s23010211 (2022).
80. Garcia-Gasulla, D. et al. On the behavior of convolutional nets for feature extraction. *J. Artif. Intell. Res.* **61**, 563–592. https://doi.org/10.1613/jair.5756 (2018).
81. Zhang, W., Yang, Y., Akilan, T., Wu, Q. J. & Liu, T. Fast transfer learning method using random layer freezing and feature refinement strategy. *IEEE Trans. Cybern.* https://doi.org/10.1109/TCYB.2024.3483068 (2024).
82. Davila, A., Colan, J. & Hasegawa, Y. Comparison of fine-tuning strategies for transfer learning in medical image classification. *Image Vis. Comput.* **146**, 105012. https://doi.org/10.1016/j.imavis.2024.105012 (2024).
83. Basha, S. S., Vinakota, S. K., Dubey, S. R., Pulabaigari, V. & Mukherjee, S. Autofcl: Automatically tuning fully connected layers for handling small dataset. *Neural Comput. Appl.* **33**, 8055–8065. https://doi.org/10.1007/s00521-020-05549-4 (2021).
84. Wang, Z., Dai, Z., Póczos, B. & Carbonell, J. Characterizing and avoiding negative transfer. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11285–11294 (2019). https://doi.org/10.1109/CVPR.2019.01155.
85. Hosna, A. et al. Transfer learning: A friendly introduction. *J. Big Data* **9**, 102. https://doi.org/10.1186/s40537-022-00652-w (2022).
86. Zhang, W., Deng, L., Zhang, L. & Wu, D. A survey on negative transfer. *IEEE/CAA J. Autom. Sin.* **10**, 305–329. https://doi.org/10.1109/JAS.2022.106004 (2023).
87. Chen, S. et al. Enhancing the performance of runoff prediction in data-scarce hydrological domains using advanced transfer learning. *Resour., Environ. Sustain.* **18**, 100177. https://doi.org/10.1016/j.resenv.2024.100177 (2024).
88. Ma, K., Shen, C., Xu, Z. & He, D. Transfer learning framework for streamflow prediction in large-scale transboundary catchments: Sensitivity analysis and applicability in data-scarce basins. *J. Geog. Sci.* **34**, 963–984. https://doi.org/10.1007/s11442-024-2235-x (2024).
89. Ouyang, W. et al. Dive into transfer-learning for daily rainfall-runoff modeling in data-limited basins. *J. Hydrol.* **657**, 133063. https://doi.org/10.1016/j.jhydrol.2025.133063 (2025).
90. Waqas, M. & Humphries, U. W. A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX* **13**, 102946. https://doi.org/10.1016/j.mex.2024.102946 (2024).

Author contributions

Henrique Echternacht: Software, Investigation, Methodology, Writing—original draft. Luciana Campos: Resources, Writing—original draft. Alfeu Dias Martinho: Data curation, Investigation. Danilo Pinto Moreira de Souza: Data curation, Visualization. Rodrigo Barbosa de Santis: Investigation, Methodology, Software. Tiago Silveira Gontijo: Funding acquisition, Investigation, Writing—review & editing. Matteo Bodini: Funding acquisition, Formal analysis, Writing—review & editing. Camila Saporetti: Funding acquisition, Methodology, Writing—review & editing. Angela Gorgoglione: Formal analysis, Validation, Writing—review & editing. Leonardo Goliatt: Conceptualization, Funding acquisition, Supervision, Writing—review & editing.

Funding
The authors acknowledge the financial support provided by the funding agencies CNPq (grants 401796/2021-3, 307688/2022-4, and 409433/2022-5), Fapemig (grants APQ-04458-23 and BPD-00083-22), and Capes (Finance Code 001).

Angela Gorgoglione and Leonardo Goliatt acknowledge the National Research and Innovation Agency (ANII), grant number VCT-1-2024-2-184149, for financial support.

Declarations

Competing interest
The corresponding author states that there is no conflict of interest on behalf of all authors.

Additional information
Correspondence and requests for materials should be addressed to A.D.M.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025