Spillage Forecast Models in Hydroelectric Power Plants Using Information from Telemetry Stations and Hydraulic Control

Pedro H. M. Nascimento [ID], Vinícius A. Cabral, Ivo C. Silva Junior *, Frederico F. Panoeiro, Leonardo M. Honório [ID] and André L. M. Marcato [ID]

Electrical Engineering Postgraduate Program, Federal University of Juiz de Fora, Juiz de Fora 36036-900, Brazil; pedro.nascimento@engenharia.ufjf.br (P.H.M.N.); albuquerque.vinicius@engenharia.ufjf.br (V.A.C.); frederico.panoeiro@engenharia.ufjf.br (F.F.P.); leonardo.honorio@ufjf.edu.br (L.M.H.); andre.marcato@ufjf.edu.br (A.L.M.M.)
* Correspondence: ivo.junior@ufjf.edu.br

Abstract: Hydroelectric power plants' operational decisions are associated with several factors, such as generation planning, water availability and dam safety. One major challenge is to control the water spillage from the reservoir. Although this action represents a loss of energy production, it is a powerful strategy to regulate the reservoir level, ensuring the dam's safety. The decision to use this strategy must be made in advance based on level and demand predictions. The present work applies supervised machine learning techniques to predict the operating condition of spillage in a hydroelectric plant for 5 h ahead. The use of this method, in real time, aims to assist the operator so that he can make more assertive and safer decisions, avoiding waste of energy resources and increasing the safety of dams. The Random Forest and Multilayer Perceptron methods were used to define the architecture compared to the forecasting capacity. The proposed methodology was applied to a 902.5 MW Hydroelectric Power Plant located on the Tocantins River, Brazil. The results demonstrate effective assistance to operators in the decision-making, presenting accuracy of up to 99.15% for the spill decision.

Keywords: forecasting; hydroelectric power; machine learning; resources managing; telemetry

1. Introduction

The generation of electric energy by means of Hydroelectric Power Plants (HPPs) has great representativeness in the world. In countries with high water potential, HPPs are commonly the main source of energy generation. This is due to the low cost of energy production with this source, especially when compared to thermoelectric generations [1].

The Brazilian energy matrix consists of 65.7% of energy production from water sources and 19.6% of generation from other renewable sources. In 2019, 70.4% of the energy consumed in Brazil came from water sources and in the first ten months of 2020 this value was 74% [2]. These data show the importance of HPPs for the Brazilian energy scenario.

In Brazil, the National System Operator (ONS) manages the National Interconnected System (SIN), ensuring coordination, control and optimization of the operation in energy generation and transmission.

Planning involving the operation of HPPs is fundamental for the correct management of available water resources and also for the safety of equipment, employees and the population living in the region downstream of the plant. In the specialized literature, it is possible to identify several studies on the operation of HPPs [3,4], which aim to optimize costs, efficiency and resources.

In [5], studies of hydro unit commitment are carried out. In [6], the focus is on the long-term generation scheduling problem. Abritta et al. [7] performed the short-term optimization of HPP considering the indication of spillage by the optimizer. The meta-

heuristic particle swarm optimization and network flow are used in [8] to obtain the optimal solution for reservoir operation rules, through water transfer between basins.

The prediction of natural events that impact the operation of HPPs is the focus of several studies in the literature. Rasouli et al. [9] used three different machine learning models to predict daily flow across a watershed in Canada. The data used in their search are weather forecast data, weather indices and weather data. In the same field of research, the authors of [10] used the single-input sequential adaptive neuro-fuzzy inference system to make river flow predictions. In [11], a water-level fluctuation forecasting model using multilayer perceptron is proposed.

The level of water in the reservoirs is directly associated with the safety of the dam and several studies are carried out in order to guarantee this safety. In [12], the level forecast for HPP is applied and the authors reported that this study contributes to ensuring the safety of the dam, due to maintaining the level at safe limits. One of the ways to control the level of the reservoirs is through pouring. Talib and Hasan [13] presented the implementation of artificial neural networks for the prediction of monthly dam spillage events for an HPP located in northern Malaysia.

As stated in [14], people, in general, expect Artificial Intelligence (AI) to automate routine labor, understand speech or images, make diagnoses in medicine and support scientific research. AI has proved its effectiveness in solving problems that can be written by straight-forward mathematical rules. However, situations that are easy for humans to perform, although difficult to formally describe, such as recognizing a specific person's voice, can be really challenging. Many domains of science had their state-of-the-art improved via machine learning, including in time-series problems [15–17], remote sensing [18] and other applications [19,20].

Multilayer Perceptron (MLP) is a feed-forward ANN with one or more hidden layers [21]. MLP is applied in several fields of study. In medicine, for example, it is widely used in the recognition of diseases through the analysis of images [22,23]. In chemical engineering, it can be used to estimate the molecular weights of chemical compounds [24]. This model of machine learning has also been used in work and research involving water resources. In [25], MLP is used to forecast drought periods in a specific region of Pakistan. The forecast of seasonal rainfall in the Tarim River basin, China, was the objective of Hartmann et al. [26]. MLP can be used in a hybrid way, with other optimization techniques, aiming at a better efficiency of the models under study. Phitakwinai et al. [27] combined MLP with the Cuckoo Research Algorithm to predict the level of the Ping River, Thailand, 7 h in advance. In [28], the optimization algorithm based on the behavior of whales is used in conjunction with the MLP to carry out the annual precipitation forecast in a given region of Senegal.

The Random Forest (RF) method is derived from the Decision Tree, which is based on the hierarchy and the importance of its branches. The RF aggregates a large number of results from the trained decision trees with training subsets and random variables, resulting in a combination of the individual results of the trained trees [29]. This methodology has the capacity to solve problems with different objectives, such as problems of classification, grouping and regression. In [30], the RF technique is used in the prediction of Alzheimer's disease through neuroimaging analysis. In the line of image analysis, Belgiu and Drăguţ [31] presented a review of the application of this methodology in remote sensing. In [32], the noise forecast of wind turbines is presented. In [33], the failure detection in wireless sensor networks with RF classification is presented.

The main objective of this study is the development of a model to predict the need or not for water spills a few hours in advance. The main purpose of this tool is to assist in the decision making of the HPP operation. Through this tool, water resource management and dam safety are optimized, ensuring that high levels in the reservoir are not reached.

The proposed methodology was applied to a dataset referring to the region of the Brazilian HPP named as HPP Lajeado. To achieve the proposed objective, ten telemetry

stations (TSs) had their data extracted, analyzed and treated. Such equipment records the flow, precipitation and water level of the river, which are necessary for this study.

Historical information was then applied to two different machine learning approaches for implementing a spill prediction model (SFM): the first approach is based on random forest (RF) and the second is based on artificial neural networks (ANNs). The output of the model is the definition of whether or not to spill over the next 5 h. Thus, the model can be used as an excellent tool to support the decision of the HPP operator.

The operating condition of spillage occurs through the opening of the hydroelectric power plant gates. Thus, the definition of the occurrence of this operative condition is extremely important, for operational and/or security reasons. When the operator does not open the floodgates, the plant can reach critical levels in its reservoir, leading to the need to pour a very large volume of water in a short time, which can cause flooding in the riverside communities downstream of the dam. On the other hand, when the operator opens the floodgates, without being really necessary, there is a waste of water resources for energy production.

The main contributions of this article are the application of the RF and MLP techniques for the prediction of pouring with considerable sensitivity to changes in pouring condition over the forecast hours; a comparison of different strategies for building a training database that impact the performance of the tool; a comparison of the efficiency of the model according to the variation of different parameters of the architecture; and a methodology for forecasting the spillage operating condition that can be applied to other HPPs.

2. Materials and Methods

The spillage forecasting methodology can be characterized by five stages including the analysis of the predictions obtained by the trained model. The methodology flowchart is shown in Figure 1. Block (1) represents the correlation analysis step between the problem data; data adjustments are made in Block (2) to ensure the balance of information; Block (3) represents the training stage of the models; the forecast is applied with the trained models and the consequent treatment of those predictions in Block (4); and the results are analyzed in Block (5). These steps are described in more detail in the next sections.
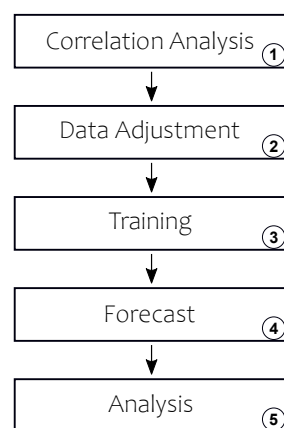

Figure 1. The spillage forecasting methodology. (1) Data correlation analysis step. (2) Data adjustment step. (3) Model development and training stage. (4) Forecasting step and treatment. (5) Model forecast analysis step.

2.1. Data Correlation

The first stage of SFM performs correlation analyzes between the historical measurement data of the TSs and the historical measurement of the HPP. After this analysis, the number of entries in the forecasting model can be reduced by applying only the characteristics most correlated with the HPP's operating conditions.

Pearson's correlation coefficient (ρ) is the index selected to evaluate the correlation between the problem data [34]. This index determines the linear correlation between two scale variables via Equation (1):

$$\rho = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{1}$$

in which x_i and y_i are the variables' measured values, $\bar{x}$ and $\bar{y}$ are the variables' average values and N is the amount of data analyzed. Values between -1 and 1 are attributed to the ρ coefficient, which represent:

- $\rho = 1 \rightarrow$ perfect positive correlation between two variables;
- $\rho = -1 \rightarrow$ perfect negative correlation between two variables, i.e., if one is increased, the other is decreased; and
- $\rho = 0 \rightarrow$ variables are not linearly related to each other.

The polynomial regression technique through the coefficient of determination (R^2) is also applied to check if there is any non-linear correlation in the data. Therefore, the correlation study becomes more robust, since Pearson's coefficient assesses only linear correlations between the data. The expression of the coefficient of determination is presented in Equation (2):

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

where,

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

2.2. Data Adjustment

The training of the model for a spill prediction needs a set of historical data. Analysis of HPP spill history can show that the frequency of occurrence of water spillage is much lower than the frequency of non-spillage. Thus, an application of a dataset without treatment can impact the training stage of the model, because an architecture that will never indicate flow can be obtained.

Thus, a strategy for balancing the dataset was applied, so that the training used as input a set with indications of pouring and non-pouring with similar frequencies. This strategy is based on the repetition of data that indicate pouring (less frequently) until a set is reached with the same amount of data that indicate no need for pouring. Figure 2 illustrates the imbalance observed in a database and the adjustment implemented to improve the training of the proposed model.
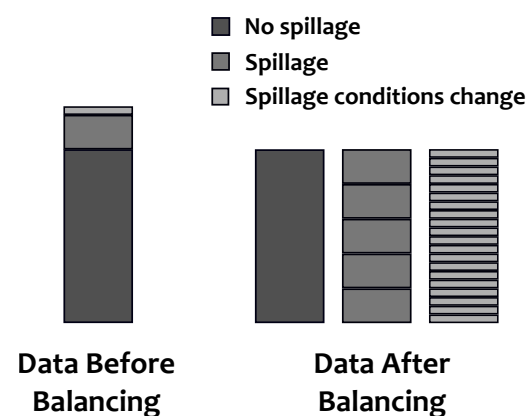

Figure 2. Adjustment performed on the training bench.

The relevance of this treatment based on simple data balancing is analyzed throughout this article. Some models were trained with data modified with this balancing procedure. This treatment occurs only in the training database of the model, that is, after the division of the training and test group. The data balancing is applied exclusively in the database associated with the training of the spill prediction model.

2.3. Training

In this step, a scaling process is applied so that the input variables are normalized between zero and 120% of the highest historical value registered in each input set. In this way, the training makes the forecasting model more robust to future measurements with higher values than those existing in the original database.

The spill prediction models are based on two machine learning methodologies: Random Forest and Multilayer Perceptron (Figure 3). The main motivations for using the machine learning methodologies used here are that both have been used in the specialized literature as classifying tools [35] and presened good performance with data in tabular form [36,37].
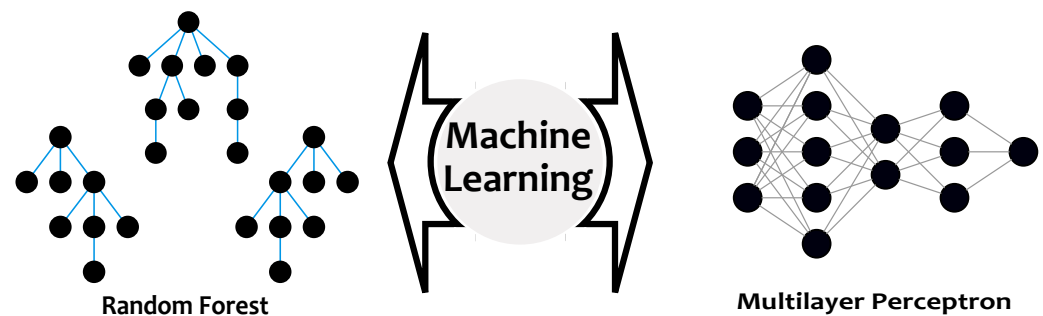

Figure 3. Machine learning structures.

The Random Forest (RF) method is an ensemble classifier, which uses a large number of decision trees in the search for the objective. It is a methodology that is based on hierarchy and the importance of its branches. Each tree is trained with a dataset grouped randomly and the result of the RF is the combination of the individual results of each tree. There is also the possibility of separating part of the training data so that they are destined for the validation of the trained trees, and the error found during this process is known as out-of-bag error [29].

The MLP method belongs to a category within Artificial Neural Networks. The single-layer perceptron, or just perceptron, is the most basic structure of a neural network. It is able to solve only linearly separable problems and its structure is composed of the artificial neuron with its activation function and adjustable synaptic weights.

Therefore, MLP generalizes the concept of perceptron, enabling the resolution of problems with various levels of complexity. The various types of activation function present in each artificial neuron in the neural network make it possible to use this tool to solve different types of problems, such as classification, regression and others. Synaptic weights can be adjusted by several techniques, but the most common is based on the back-propagation of errors [21].

2.4. Forecasting

The forecast model response consists of five binary values, which represent the condition of the pouring operation for five hours ahead, with a value of 0 indicating the absence of pouring and a value of 1 indicating the need for pouring. Some spillage profiles for 5 h are shown in Figure 4. Figure 4A,B presents common spill configurations, while Figure 4C,D presents configurations that are rarely present in databases in general.

In addition, the two configurations presented in Figure 4C,D are difficult to be achieved by the forecasting model and can hinder the forecasting model training pro-

cess. Thus, to improve training performance and maintain good quality of the forecast type, the 5 h to be forecast are interpreted as a single binary value, in which zero corresponds to the absence of pouring in all 5 h forecasted and one corresponds to the need for pouring in any of the 5 h forecasted.
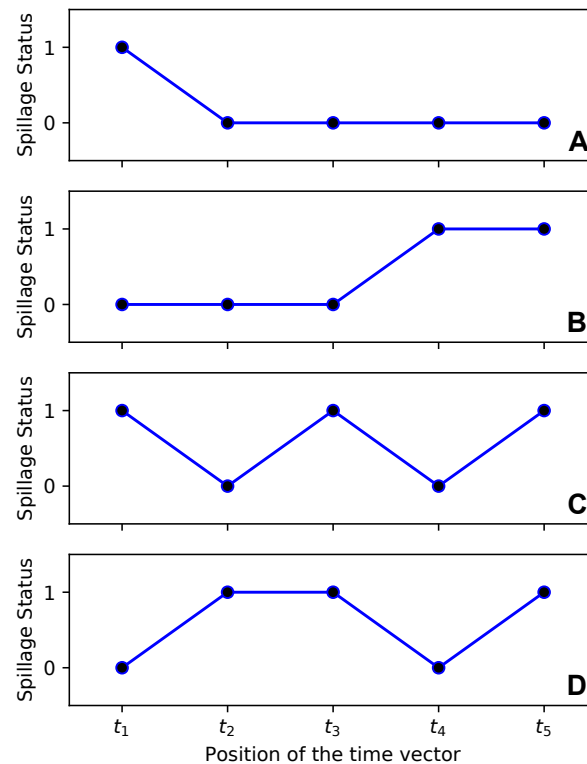

Figure 4. Example of configuring spillage data in 5 h. (**A**) A spillage change in the early hours. (**B**) A spillage change in the final hours. (**C**) Four changes in the spillage. (**D**)Three changes in the spillage.

Throughout the HPP operation, decision making is often associated with the need to change the pouring condition, considered as critical situations for the operation, i.e., when it is necessary to start or stop the pouring of water from the reservoir.

In this study, the occurrences of changes in the state of spillage over the 5-h forecast are also analyzed. Thus, the change in the pouring state is verified in the configurations in which the values obtained for the 5 h of forecast are not all equal to zero (not pouring) or equal to one (pouring).

Since this decision-making is normally carried out by the HPP operator, distinct decisions can be made for scenarios of similar hydrological conditions. Thus, the use of spill prediction models helps in the uniformity of these decisions.

2.5. Analysis

The training analysis of the machine learning models needs to consider the generalizability of the model, that is, the ability to make good quality forecasts regardless of the data used during training. Thus, the database was divided into eight parts, containing 2124 data in each subpackage, as shown in Figure 5. This information corresponds to 88.5 days, equivalent to almost one season.

The four seasons of the year should be considered in the historical dataset, in the training and in the testing of the forecasting models, as the climatic factors have a strong influence on the behavior of the river characteristics and on the measurements of the telemetry stations. Thus, the subpackages were changed to modify the training and test groups, while still presenting characteristics for all seasons. For this, Subpackage A can

only be exchanged with Subpackage E, just as Subpackage B can only be exchanged with Subpackage F and so on. Thus, 16 groups of data were formed for cross-validation, always respecting the condition of maintaining the characteristics of the four seasons both in training and in testing.
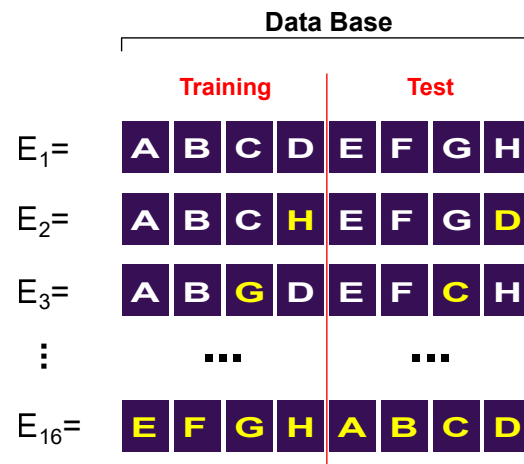

Figure 5. Data split for cross-validation.

The cross-validation predictions are analyzed at each iteration for accuracy, precision, recall, F1-score and training time. With the exception of the latter, all metrics are related to the confusion matrix, as shown in Figure 6.

The false positive occurs when the forecast indicates the need for pouring, whereas the real historical data indicate the opposite. In this situation, there would be unnecessary losses of water resources that could be used for energy production. The false negative, on the other hand, occurs when the forecast indicates the absence of pouring in future hours, but the data indicate the need. In this situation, there could be an accumulation of water in the reservoir, which would pose a risk to the safety of the dam.

Accuracy is a ratio between the correct predictions and the total predictions made by the model. In some cases, where the data are well distributed, as described in the previous sections, the accuracy may be sufficient to qualify a model. Equation (3) defines the accuracy of the model.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{3}$$

Precision is related to the ratio between the correct positive predictions and the predicted positive predictions. That is, the high precision indicates the reduced occurrence of false positives predicted by the model. Equation (4) presents its formulation.

$$\text{Precision} = \frac{(\text{TP})}{\text{TP} + \text{FP}} \tag{4}$$

The recall rate is defined by the ratio between the correct positive predictions and the sum of all positive observations according to history (FN and TP). Thus, it is possible to know the correct prediction rate for pouring in relation to all times that the model should predict the need for pouring. The recall rate is calculated using Equation (5).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

F1-score is most used in training evaluations with unbalanced databases, as it is calculated by the weighted average of precision and recall. Thus, this index considers both

false positives and false negatives found in the forecasts. The F1-score is calculated through Equation (6).

$$\text{F1} - \text{score} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \tag{6}$$

		Predicted Value	
		Negative	Positive
Actual Value	Negative	**TN** (True Negative)	**FP** (False Positive)
	Positive	**FN** (False Negative)	**TP** (True Positive)

Figure 6. Confusion matrix.

3. Results and Discussion

For the validation of the proposed methodology, data from a Brazilian hydrographic basin were considered in which telemetric stations are installed along the river in locations close to HPP Luis Eduardo Magalhães, also known as HPP Lajeado. This HPP has a generation capacity of 902.5 MW and was inaugurated in 2002. This region is located in the state of Tocantins, within the Tocantins-Araguaia hydrographic basin, as shown in Figure 7.

The historical data necessary for the validation of the tool, which refer to the HPP, are the turbine flow rates, historical flow rates and reservoir levels. Precipitation and river level data, measured by telemetry stations, are also needed. Table 1 lists the telemetry stations in the region and their geographic coordinates.

The information obtained through the TSs was measured in the period from 13 August 2018 to 21 July 2020 with a discretization of 15 min. The data for these stations can be obtained through the National Water Resources Information System managed by National Water Agency (*Agência Nacional de Água*, ANA) [38].

The data extracted from the HPP's operations history are presented with 1 h discretization and were measured in the same period as the data obtained by the TSs. Considering the difference in data discretization, adjustments are made in the set provided by the TSs in order to generate a database also with hourly discretization.
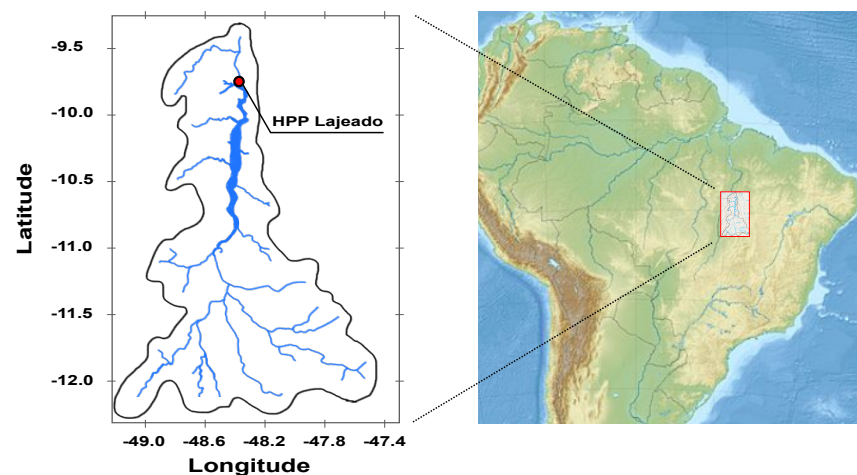

Figure 7. Location of the study region and hydrographic basin area of HPP Lajeado.

Table 1. Location of Telemetry Station.

Telemetry Station	Latitude	Longitude
Jacinto	−11.9817	−48.6569
Jerônimo	−11.7578	−47.8358
Ipueiras	−11.2467	−48.4586
Jurupary	−11.1519	−48.5158
Areias	−10.8950	−48.3483
Mangues	−10.3483	−48.6369
Barramento	−9.7603	−48.3697
Jusante	−9.7442	−48.3625
Lucena	−9.6917	−48.3689
Tocantínia	−9.5669	−48.3792

TSs are installed in open environments and are vulnerable to climate change. Therefore, recurrent corrective and predictive maintenance is necessary. That is, the equipment is occasionally unavailable and, therefore, gaps in the datasets are found, as shown in Figure 8, in which there is a failure of 15 sequential data starting at Hour 200.
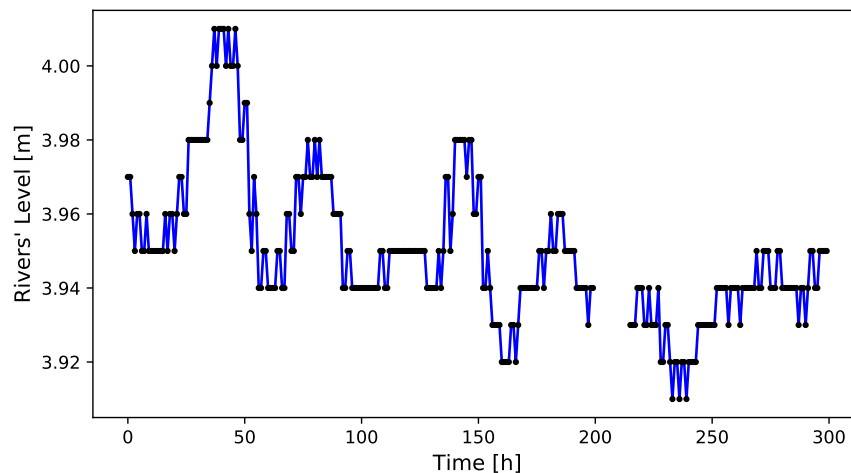
Figure 8. Fault representation of river level data.

Given the significant importance of data quality, it is crucial that no incoherence, error, or inconsistency is present in the datasets since these compromise the performance of the network training process. Therefore, a treatment consisted of linear interpolations was applied to the TSs' data via Equation (7) so that empty intervals could be filled.

$$y = y_0 + (y_1 - y_0)\frac{x - x_0}{x_1 - x_0} \tag{7}$$

in which (x_0, y_0) and (x_1, y_1) are data points.

Data acquisition, implementation and testing were performed using Python. For training, validation and testing of the models, the libraries Scikit-Learn, Tensorflow and Keras were used. All developments were carried out on a machine with a Windows 10 64-bit operating system, with an Intel Core i5 processor and 1.6 GHz frequency, 6 GB of RAM and 240 GB solid state disk (SSD).

The first step that must be performed is the definition of which telemetric stations will be used for training the model. The main criteria used for the selection of the TS is the position and correlation between the data. The Jusante, Lucena and Tocantínia stations are located downstream from the HPP, so they were not considered in this study. In addition, the Barramento station represents HPP data and was also not considered in this step.

The river level data, captured by the TS, were subjected to a correlation with the level and flow measured at the HPP. Linear correlation analyzes with Pearson's method and non-linear correlations with second and third degree polynomial regression were performed. The results of the correlations of the TSs with the level of the dam are shown in Figure 9A, in which it is observed that the TSs Jurupary and Areias maintain all the correlation values above 0.7 whereas the other TSs do not exceed the value of 0.4.
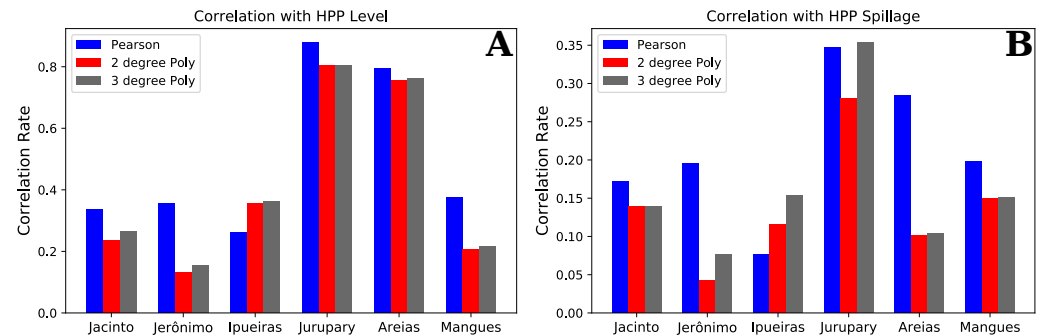

Figure 9. TSs correlation results with the HPP. (**A**) Correlation values with HPP level data. (**B**) Correlation values with HPP spillage data.

Figure 9B shows the correlation of the measures of the TSs with the spill data of the HPP. It is possible to verify that the TS Jurupary presented the three correlation values superior to the other TSs. TS Areias also showed a good linear correlation. Hence, the Jurupary and Areias TSs are the only stations used as input for the spillage forecast model.

The structure of SFM inputs and outputs is represented in Figure 10. The inputs are composed of data from turbine flow, spillway flow, reservoir level, river level and precipitation measured at Jurupary and Areias stations. At each training stage, the data from these stations refer to the measurements of 10 h prior to the moment of the forecast, and this value was obtained empirically.

The model also has as input data the turbine flow programming data of five hours ahead. These data are provided by an optimization model that is used to define the operation schedule based on HPP's generation goals. The model provides five binary values referring to the forecast of the following hours as output data.
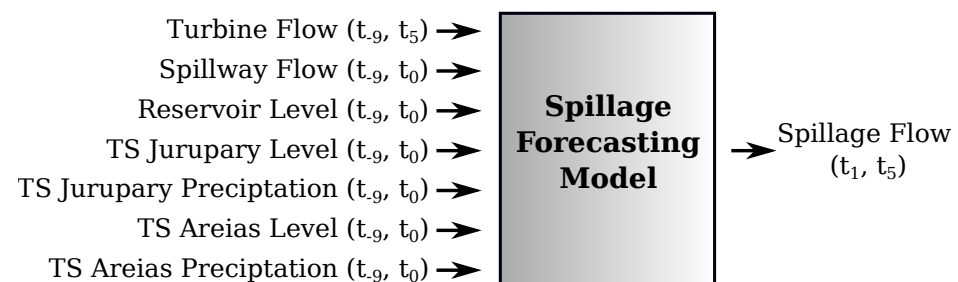

Figure 10. SFM's structure.

A sensitivity analysis was performed for both machine learning methods. The models were subjected to training and tests with five of the 16 groups presented in Section 2.5, chosen at random. For the assessment, accuracy was assessed in two different situations, named Accuracy I and Accuracy II. Accuracy I corresponds to the evaluation of results for all data used in the test phase. Accuracy II assesses only the cases in which changes in the pouring state occur during any of the five hours of forecast, defined as critical situations for the operation. In this sensitivity analysis, the training was carried out with the data adjusted so that the different types of output were equally distributed in the set used.

The RF model is a methodology that explores ensemble learning in which the sensitivity analysis are performed regarding the number of trees/estimators that form the forest.

For the development and training of this model, the Random Forest Classifier from the Scikit-Learn library was used. All the characteristics of the RF model, with the exception of the number of estimators, were kept constant throughout this work. Each tree created is subjected to training with three quarters of all data intended for training the model. The Gini Impurity function is used to measure the quality of the divisions, which occur until the leaves are pure. Out-of-bag assessment is not applied to the models. The rest of the model settings are defined as the standard features of this library. Thus, six quantities of estimators to be evaluated are defined: 10, 50, 100, 500, 1000 and 1500 estimators.

Table 2 shows the average of all metrics calculated for the five training and tests performed with each model. The comparison of the results shows that the differences between the results are small. However, the model with 1000 estimators showed good results in all metrics and the best result for Accuracy II. Thus, this RF configuration with 1000 estimators was chosen to be used in training.

Table 2. Average of the results of five different trainings for each configuration of the RF-model.

Estimators	Accuracy I	Accuracy II	F1-Score	Precision	Recall	Time (s)
10	0.9639	0.4360	0.8990	0.8628	0.9391	1
50	0.9685	0.4452	0.9111	0.8755	0.9515	7
100	0.9667	0.4611	0.9069	0.8662	0.9526	14
500	0.9684	0.4531	0.9109	0.8733	0.9536	70
1000	0.9683	0.4733	0.9108	0.8709	0.9559	138
1500	0.9683	0.4724	0.9109	0.8701	0.9570	207

In MLP, sensitivity analysis is performed to determine a good configuration of the hidden layers and their respective numbers of neurons. The Sequential model of Tensorflow with densely connected layers was used for the creation of the MLP model and the consequent training. With the exception of the number of hidden layers and their respective number of neurons, all other characteristics of the model were maintained in all training sessions. The Rectified Linear Unit function (ReLU) was chosen as the neuron activation function due to the excellent results of other models based on ANN that also apply it [39,40]. The training of the internal parameters of the network was carried out through the Adam Optimizer, which is a stochastic gradient descent method existing in the Keras library. The results are validated with 20% of the training data separated, automatically and immediately, before the training starts. The training aims to minimize the mean square error (MSE) between observed and predicted data. The training also considers the input data referring to the measurements 10 h before the forecast time and the number of training epochs is limited to 50.

Two levels of trainable parameters were defined for this analysis of the MLP models: 27,000 and 42,500. Thus, using models with 1–4 hidden layers, neuron quantities per hidden layer were defined in order to keep the trainable parameters close to the determined levels. Table 3 presents the characteristics and results of the five training sessions carried out with the same groups of data used in the training of the RF models. The configurations with three and four hidden layers presented better average results of Accuracy I, F1-score and precision. The recall rate for all settings is similar. Analyzing Accuracy II, two configurations presented better results, but the configuration with one hidden layer and 526 neurons did not present good results in the other metrics. Therefore, the MLP model with 82 neurons in each of its four hidden layers was defined as the model to be used in the continuation of training, as it presents a better set of results.

After the sensitivity analysis described, the models are retrained. In this stage, the 16 training sessions are carried out with the different partitioning of the database, presented in Section 2.5. In addition, the database used exclusively to train the model can be adjusted so that the following three conditions are equally distributed in the dataset: spillage condition, non-spillage condition and change in spillage condition, and these conditions refer to the 5-h forecast.

As a result of this adjustment of the training data, the amount of data used in this step increased significantly. This training stage with adjustment of the dataset was compared with training in which this adjustment does not occur in order to evaluate the improvements provided by this adjustment. Finally, the training of the model was also carried out considering the original triple dataset for assessing the performance of the model with the increase of the available dataset. Then, three different forms of training database were used for both RF and MLP techniques. Table 4 shows the settings and the nomenclature assigned to each training stage.

Table 3. Average of the results of five different trainings for each configuration of the RF-model.

Hidden L. / Neur.	Trainable P.	Accuracy I	Accuracy II	F1-Score	Precision	Recall	Time (s)
1/332	26,894	0.7971	0.5669	0.6078	0.4602	0.9755	111
2/128	26,855	0.8941	0.5353	0.7488	0.6178	0.9689	109
3/97	26,874	0.9275	0.5345	0.8049	0.6980	0.9586	117
4/82	27,065	0.9274	0.6129	0.8137	0.7284	0.9467	119
1/526	42,611	0.7826	0.6206	0.6019	0.4457	0.9758	114
2/169	42,424	0.8542	0.5154	0.7037	0.5644	0.9657	112
3/127	42,804	0.9314	0.4823	0.8195	0.7386	0.9453	118
4/106	42,617	0.9316	0.5620	0.8096	0.7182	0.9514	127

Table 4. Description of the treatment performed in the SFM training database.

Model Name	Type	Training Database Modification
M_1	RF	Data balancing
M_2	RF	No modification
M_3	RF	Data triplication
M_4	MLP	Data balancing
M_5	MLP	No modification
M_6	MLP	Data triplication

Figure 11 presents the boxplot representation of the metrics obtained for all 16 different training sessions. Regarding Accuracy I, shown in Figure 11A, the RF-based models performed better with respect to the median of the values and less dispersion of results, when compared to MLP models. The Accuracy II of the RF models showed low variation, regardless of the treatment data of the training data, as shown in Figure 11B. MLP with equally distributed data is the technique that obtained the best results, although it has shown greater dispersion. The values of Accuracy I are usually higher than the values of Accuracy II, since Accuracy II refers only to the moments when a change in the operating condition of the spill occurs, that is, in the most difficult moments to make the forecast and with less number of samples for training.

The M4, M5 and M6 models show lower precision values compared to the other models and greater dispersion of the results, as shown in Figure 11C. These results demonstrate that these trained models have many false positives. That is, these models define the need for pouring when it should not occur in some simulations. In addition, the results also demonstrate that the precision of the M2 and M5 models showed the smallest dispersions when compared with the other methods.
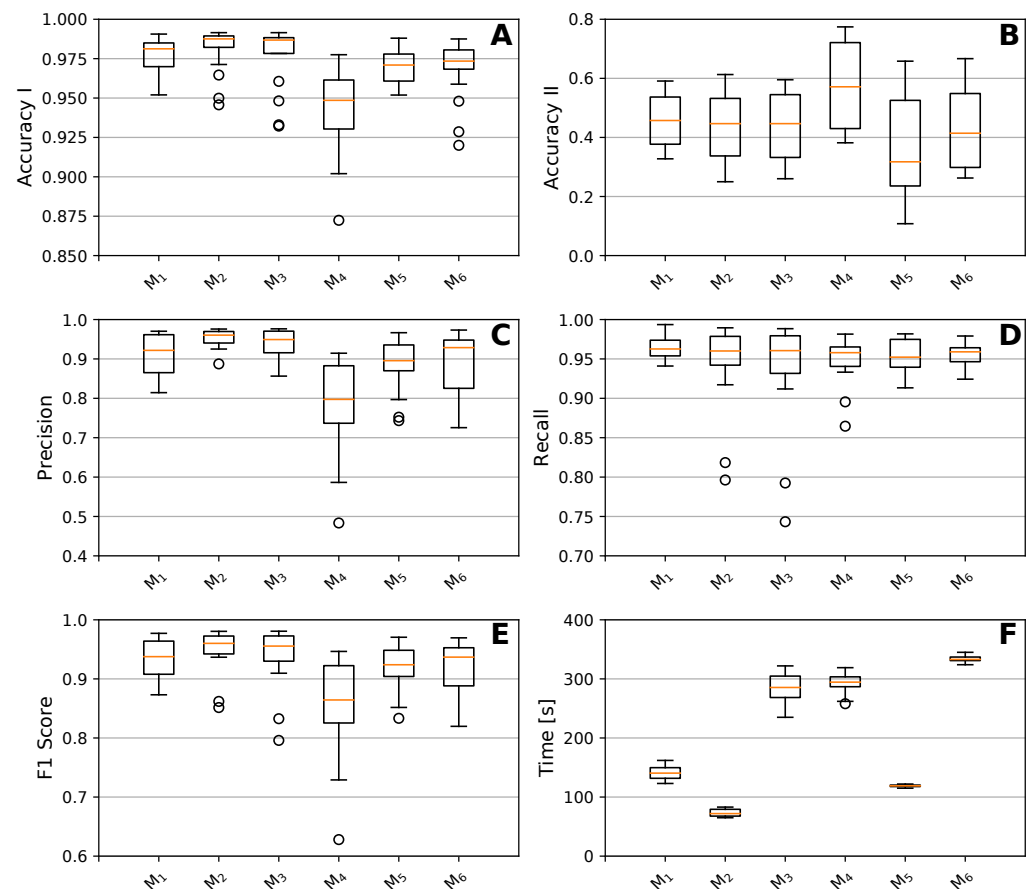
Figure 11. The boxplot diagram with the values of the metrics resulting from training with 16 different training and test databases. (**A**) Accuracy I rate results. (**B**) Accuracy II rate results. (**C**) Precision rate results. (**D**) Recall rate results. (**E**) F1 score rate results. (**F**) Time results

The recall metric was the one that showed the least variability of the results due to the variation of the models. Furthermore, the results show that the medians indicate correct answers above 95%, as shown in Figure 11D. It is very important that the trained model has high values for this metric, as it indicates the success rate of the pouring condition. That is, the higher is the value of this metric, the lower is the predictions of false negatives.

F1-score values represent the weighted average of the precision and recall rate. Due to the unbalanced occurrence of the data, this is a rate of great importance in the evaluation of the models. The M2 model, characterized by RF without modification of the training bench, presented the best results, as shown in Figure 11E.

Figure 11F shows the distribution of training times for each model and the results are in line with expectations. The M2 and M5 models had the lowest computational costs due to the lack of adjustment of the training bench. The M1 and M3 models showed great dispersion of training times, due to the greater amount of training data and the random nature of the divisions of the RF branches. The M6 model showed less dispersion than the M4 model, as there is greater variability in the training dataset in the latter model.

The best training and test result for each model is shown in Table 5. The RF model without any adjustment from the training bench (M2) obtained the best performance in relation to Accuracy I. However, this model presented a low Accuracy II, due to errors in forecasting the transition of the operating condition of the pouring. Regarding this metric, the MLP model with an equally distributed training database (M4) obtained the best result: 77.42%. In addition, the Accuracy I result of this model is good and very close to the results of the other models, which shows that it is the best model obtained for predicting the pouring condition of HPP Lajeado.

Table 5. Results of the best training for each model.

Model	Accuracy I	Accuracy II	Precision	Recall	F1-Score
M_1	0.9789	0.5909	0.8640	0.9936	0.9243
M_2	0.9915	0.6129	0.9693	0.9895	0.9793
M_3	0.9908	0.5952	0.9759	0.9856	0.9807
M_4	0.9775	0.7742	0.9145	0.9808	0.9465
M_5	0.9729	0.6581	0.7968	0.9490	0.8663
M_6	0.9875	0.6667	0.9601	0.9791	0.9695

4. Conclusions

Operating decisions at an HPP are associated with several factors such as generation capacity, availability of water resources and dam safety. This work proposes a tool capable of predicting the need for spillage or not for the next 5 h, with the objective of assisting the HPP operator in the decision in real time. The validation of this model was done for the HPP Lajeado on the Tocantins River, Brazil, but it can be used for other It was sent to other HPPs, requiring access to real-time operational measurement information and data from telemetry stations upstream of the dam.

The training and validation of the models were carried out through the historical data of operation of the UHE and the data of the telemetry stations. The analyses performed show that the model based on the RF technique performed better in the forecasting stage when the treatment is performed with a dataset treated in such a way that there is a balanced distribution between the model's response possibilities. However, the application of the MLP technique generated a model that had less chance of false positives and better performance when there are changes in the pouring operating conditions.

The correlation study presented demonstrated an ability to select the data available for the HPP, which provided a reduction in the number of inputs from machine learning models. The fact that these telemetry stations are associated with flaws in their measurement data was overcome by an interpolation that kept the application of the tool viable, since the indices demonstrate a high efficiency in the forecasting capacity.

Although the MLP technique has shown the best result, both techniques have a good ability to predict spillage. Thus, the results demonstrate that the proposed methodology is capable of providing good assertiveness to assist, in real time, HPP operators in decision making regarding the spillage operating conditions.

Abbreviations

HPP	Hydroelectric Power Plant
ANN	Artificial Neural Networks
MLP	Multilayer Perceptron
RF	Random Forest
TS	Telemetry Station
SFM	Spillage Forecast Model
ONS	National System Operator
SIN	National Interconnected System
ANA	National Water Agency
AI	Artificial Intelligence
MSE	Mean square error
TP	True positive
TN	True negative
FP	False positive
FN	False negative
ρ	Pearson's correlation coefficient
R^2	Coefficient of determination
x_i and y_i	Variables measured
$\overline{x}, \overline{y}$	Variables' average values
$\hat{y}$	Predicted value
N	Amount of data analyzed
h	hours
m	meters
s	seconds

References

1. IEA. Data & Statistics. 2020. Available online: https://www.iea.org/ (accessed on 21 October 2020).
2. ONS. O Sistema em Números & Geração de Energia, 2020. Available online: http://www.ons.org.br/ (accessed on 1 October 2020).
3. Singh, V.K.; Singal, S. Operation of hydro power plants—A review. *Renew. Sustain. Energy Rev.* **2017**, *69*, 610–619. [CrossRef]
4. Nazari-Heris, M.; Mohammadi-Ivatloo, B.; Gharehpetian, G. Short-term scheduling of hydro-based power plants considering application of heuristic algorithms: A comprehensive review. *Renew. Sustain. Energy Rev.* **2017**, *74*, 116–129. [CrossRef]
5. Finardi, E.; Takigawa, F.; Brito, B. Assessing solution quality and computational performance in the hydro unit commitment problem considering different mathematical programming approaches. *Electr. Power Syst. Res.* **2016**, *136*, 212–222. [CrossRef]
6. Fredo, G.L.M.; Finardi, E.C.; de Matos, V.L. Assessing solution quality and computational performance in the long-term generation scheduling problem considering different hydro production function approaches. *Renew. Energy* **2019**, *131*, 45–54. [CrossRef]
7. Abritta, R.; Panoeiro, F.; Honório, L.; Silva Junior, I.; Marcato, A.; Guimar aes, A. Hydroelectric Operation Optimization and Unexpected Spillage Indications. *Energies* **2020**, *13*, 5368. [CrossRef]
8. Passos de Arag ao, A.; Teixeira Leite Asano, P.; de Andrade Lira Rabêlo, R. A Reservoir Operation Policy Using Inter-Basin Water Transfer for Maximizing Hydroelectric Benefits in Brazil. *Energies* **2020**, *13*, 2564. [CrossRef]
9. Rasouli, K.; Hsieh, W.W.; Cannon, A.J. Daily streamflow forecasting by machine learning methods with weather and climate inputs. *J. Hydrol.* **2012**, *414*, 284–293. [CrossRef]
10. Belvederesi, C.; Dominic, J.A.; Hassan, Q.K.; Gupta, A.; Achari, G. Predicting River Flow Using an AI-Based Sequential Adaptive Neuro-Fuzzy Inference System. *Water* **2020**, *12*, 1622. [CrossRef]
11. Zhou, T.; Jiang, Z.; Liu, X.; Tan, K. Research on the Long-term and Short-term Forecasts of Navigable River's Water-Level Fluctuation Based on the Adaptive Multilayer Perceptron. *J. Hydrol.* **2020**, *591*, 125285. [CrossRef]
12. Shang, Y.; Xu, Y.; Shang, L.; Fan, Q.; Wang, Y.; Liu, Z. A method of direct, real-time forecasting of downstream water levels via hydropower station reregulation: A case study from Gezhouba Hydropower Plant, China. *J. Hydrol.* **2019**, *573*, 895–907. [CrossRef]
13. Talib, A.; Hasan, Y.A. The Application of Artificial Neural Network for Forecasting Dam Spillage Events. In Proceedings of the 5th International Congress on Enviromental Modelling and Software, Ottawa, ON, Canada, 5–8 July 2010; pp. 1–9.
14. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
15. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [CrossRef]
16. Yao, S.; Hu, S.; Zhao, Y.; Zhang, A.; Abdelzaher, T. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 351–360.

17. Chen, J.; Zeng, G.Q.; Zhou, W.; Du, W.; Lu, K.D. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Convers. Manag.* **2018**, *165*, 681–695. [CrossRef]

18. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]

19. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]

20. Mosavi, A.; Ozturk, P.; Chau, K.w. Flood prediction using machine learning models: Literature review. *Water* **2018**, *10*, 1536. [CrossRef]

21. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice Hall PTR: Lebanon, IN, USA, 1994.

22. Hosseinzadeh, M.; Ahmed, O.H.; Ghafour, M.Y.; Safara, F.; Ali, S.; Vo, B.; Chiang, H.S. A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things. *J. Supercomput.* **2020**, *65*, 1–22. [CrossRef]

23. Abdar, M.; Yen, N.Y.; Hung, J.C.S. Improving the diagnosis of liver disease using multilayer perceptron neural network and boosted decision trees. *J. Med. Biol. Eng.* **2018**, *38*, 953–965. [CrossRef]

24. Hemmati-Sarapardeh, A.; Ameli, F.; Varamesh, A.; Shamshirband, S.; Mohammadi, A.H.; Dabir, B. Toward generalized models for estimating molecular weights and acentric factors of pure chemical compounds. *Int. J. Hydrogen Energy* **2018**, *43*, 2699–2717. [CrossRef]

25. Ali, Z.; Hussain, I.; Faisal, M.; Nazir, H.M.; Hussain, T.; Shad, M.Y.; Mohamd Shoukry, A.; Hussain Gani, S. Forecasting drought using multilayer perceptron artificial neural network model. *Adv. Meteorol.* **2017**, *2017*. [CrossRef]

26. Hartmann, H.; Snow, J.A.; Su, B.; Jiang, T. Seasonal predictions of precipitation in the Aksu-Tarim River basin for improved water resources management. *Glob. Planet. Chang.* **2016**, *147*, 86–96. [CrossRef]

27. Phitakwinai, S.; Auephanwiriyakul, S.; Theera-Umpon, N. Multilayer perceptron with Cuckoo search in water level prediction for flood forecasting. In Proceedings of the IEEE 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 519–524.

28. Diop, L.; Samadianfard, S.; Bodian, A.; Yaseen, Z.M.; Ghorbani, M.A.; Salimi, H. Annual Rainfall Forecasting Using Hybrid Artificial Intelligence Model: Integration of Multilayer Perceptron with Whale Optimization Algorithm. *Water Resour. Manag.* **2020**, *34*, 733–746. [CrossRef]

29. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

30. Sarica, A.; Cerasa, A.; Quattrone, A. Random Forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Front. Aging Neurosci.* **2017**, *9*, 329. [CrossRef] [PubMed]

31. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

32. Iannace, G.; Ciaburro, G.; Trematerra, A. Wind turbine noise prediction using random forest regression. *Machines* **2019**, *7*, 69. [CrossRef]

33. Noshad, Z.; Javaid, N.; Saba, T.; Wadud, Z.; Saleem, M.Q.; Alzahrani, M.E.; Sheta, O.E. Fault detection in wireless sensor networks through the random forest classifier. *Sensors* **2019**, *19*, 1568. [CrossRef]

34. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.

35. Awad, M.; Khanna, R. *EfficienT Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Springer Nature: Berlin/Heidelberg, Germany, 2015.

36. Yeung, D.S.; Cloete, I.; Shi, D.; wY Ng, W. *Sensitivity Analysis for Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2010.

37. Cerrada, M.; Zurita, G.; Cabrera, D.; Sánchez, R.V.; Artés, M.; Li, C. Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mech. Syst. Signal Process.* **2016**, *70*, 87–103. [CrossRef]

38. SNIRH. Rede Hidrometeorológica Nacional, 2020. Available online: https://www.ana.gov.br/ANA/ (accessed on 22 July 2020).

39. Jiang, W.; He, G.; Long, T.; Ni, Y.; Liu, H.; Peng, Y.; Lv, K.; Wang, G. Multilayer perceptron neural network for surface water extraction in Landsat 8 OLI satellite images. *Remote Sens.* **2018**, *10*, 755. [CrossRef]

40. Wang, S.H.; Phillips, P.; Sui, Y.; Liu, B.; Yang, M.; Cheng, H. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J. Med. Syst.* **2018**, *42*, 85. [CrossRef]