climate

MDPI

Article

Neuro-Fuzzy Forecasting of Precipitation Occurrence in the Next Hour in Rio Grande, Southern Brazil

Alan Dorneles, André Andrade Longaray *, Jaci Saraiva, Francine Borges and Jeferson Machado

Instituto de Matemática, Estatística e Física (IMEF), Universidade Federal do Rio Grande, Rio Grande 96203-900, Brazil; alanricardo.97@furg.br (A.D.); jaci.saraiva@furg.br (J.S.); francine_sborges@furg.br (F.B.); jeferson.machado@furg.br (J.M.)

* Correspondence: andrelongaray@furg.br

Highlights

What are the main findings?

- An ANFIS with two inputs and nine rules forecast precipitation occurrence in the next hour with an AUC of 0.809 on a continuous, chronologically separated test period (2019–2024), comparable to tuned SVM-RBF and random forest models. However, a simple persistence score based on antecedent precipitation performed better (AUC 0.843).
- Average pressure and relative humidity computed over the preceding 3 h were the only inputs, and the nine rules can be read in physical terms.

What are the implications of the main findings?

- At the natural frequency of precipitation (8%), the precision of the alerts was low (0.17 at the selected threshold), which limits operational use without additional predictors.
- The findings highlight the importance of integrating physical interpretability into machine learning models, and the need for temporally independent evaluation at the natural frequency of precipitation and tuned benchmarks when reporting forecasting skill.

Abstract

This study evaluates an adaptive neuro-fuzzy inference system (ANFIS) to forecast precipitation occurrence in the next hour at a single station in Rio Grande, southern Brazil. Hourly data from INMET automatic station A802 (2002–2024) were organized into consecutive, non-overlapping 3 h windows, each labelled by the occurrence of precipitation in the following hour. The predictors were the window averages of atmospheric pressure and relative humidity. The ANFIS (three bell-shaped membership functions per input, nine rules) was trained on 2002–2018 and evaluated on a continuous test period (2019–2024; 16,326 windows, 8.0% with precipitation), with tuned support vector machine (SVM-RBF) and random forest models as references. ANFIS reached an AUC of 0.809 (95% CI 0.794–0.823) similar to SVM-RBF and random forest (both 0.807) and to a logistic regression (0.808), with the lowest inference cost and an interpretable rule base, but below a persistence score based on antecedent precipitation (0.843); it outperformed persistence only when no precipitation had occurred in the preceding hour (0.739 versus 0.620). At the selected threshold, it detected 81% of the precipitation windows with a precision of 0.17. The results indicate that a compact neuro-fuzzy model

Academic Editor: Charles Jones

Received: 19 August 2026
Revised: 15 November 2026
Accepted: 19 November 2026
Published: 24 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

with two surface variables provides physically interpretable rules but only moderate skill, mostly associated with ongoing precipitation.

Keywords: ANFIS; precipitation forecasting; neuro-fuzzy computing; atmospheric variability; applied machine learning; regional climatology; atmospheric modeling

1. Introduction

Precipitation forecasting remains a major challenge in meteorology, especially in regions characterized by high atmospheric variability and strongly nonlinear atmospheric processes [1–4].

Despite advances in numerical and statistical models, substantial limitations persist in accurately forecasting precipitation occurrence, highlighting the need for alternative approaches capable of capturing the complexity of atmospheric phenomena [5].

In this context, the integration of fuzzy logic and artificial neural networks (ANNs) has emerged as a promising approach [6–11].

Neuro-fuzzy models combine the ability of fuzzy systems to represent uncertain knowledge [12,13] with the learning and adaptive capabilities of ANNs, enabling more effective modeling of nonlinear relationships [11,14]. This complementarity makes these models particularly suitable for applications in complex environmental systems, such as precipitation forecasting.

Previous applications of neuro-fuzzy models have demonstrated their potential for precipitation estimation and forecasting, particularly in representing nonlinear relationships between meteorological variables [5–7,10]. However, these studies differ in their prediction targets, temporal resolutions, input variables, and evaluation procedures, limiting the direct comparison of their reported performance. Furthermore, short-term precipitation studies have investigated rainfall onset detection, precipitation occurrence, and intense rainfall forecasting using different approaches, including threshold-based methods, support vector machines, and neural networks [15–17]. Although these investigations provide valuable evidence, differences in forecasting horizons, precipitation definitions, class distributions, and validation strategies make it difficult to determine how model performance translates across forecasting tasks.

Within this context, an important methodological question concerns the predictive capability of a compact and interpretable ANFIS model based on a restricted set of routinely available surface meteorological observations. In particular, evaluating such a model under chronological separation of training and test periods, while preserving the observed precipitation frequency during testing, provides a clearer assessment of discrimination and false-alarm behavior. These considerations motivate an investigation that combines physically interpretable predictors, temporal validation, comparisons with tuned machine-learning models, and explicit recognition of the limitations associated with class imbalance and single-station observations.

The municipality of Rio Grande, Rio Grande do Sul, Brazil, has characteristics that make precipitation forecasting particularly challenging, including the influence of different meteorological systems, such as cold fronts, cyclones, and mesoscale convective complexes, as well as marked temporal variability in precipitation events [18,19]. These conditions reinforce the importance of developing predictive models capable of adequately representing local atmospheric dynamics.

Earlier studies, such as those by Britto et al. [20], indicate that the average frequency of frontal systems entering the region is approximately 16 per season, reinforcing the idea that local precipitation is predominantly associated with these systems. In a more detailed

analysis, Britto et al. [19] evaluated frontal system activity between 1993 and 1996, identifying 194 cold fronts, which corresponds to an average of four cold fronts per month. These results corroborate the observations of Pegoraro et al. [21], who identified approximately 28 precipitating systems during summer, predominantly convective systems, and 57 during winter, most of which were associated with frontal systems.

Furthermore, more recent analyses, including Silva et al. [18], reinforce this trend by highlighting that cold fronts play a significant role in precipitation formation throughout the year in the Rio Grande region. These systems not only define seasonal rainfall patterns but are also associated with extreme events that contribute to high interannual variability in precipitation.

Against this background, the present study aims to evaluate the predictive performance and interpretability of a compact adaptive neuro-fuzzy inference system (ANFIS) for forecasting precipitation occurrence in the following hour at INMET station A802, located in Rio Grande, southern Brazil. The model uses average atmospheric pressure and average relative humidity calculated over antecedent 3 h observation windows, with three membership functions per input and nine fuzzy rules. Its performance is assessed using a chronological training period (2002–2018) and a subsequent test period (2019–2024), while preserving the observed precipitation frequency within the sampled test windows. Comparisons with tuned SVM-RBF and Random Forest models are conducted to examine discrimination, classification performance, computational efficiency, and interpretability. The investigation also considers the limitations associated with class imbalance, non-overlapping temporal sampling, and observations from a single meteorological station.

2. Materials and Methods

2.1. Study Area

The study was conducted in Rio Grande, located in the southernmost part of Brazil, in the state of Rio Grande do Sul (Figure 1). The area covers approximately 2834 km² and lies in a region characterized by high climate variability and the frequent occurrence of extreme precipitation events [22,23]. These events are associated with synoptic and subsynoptic systems, as well as with factors related to large-scale atmospheric circulation and local circulations over South America [24].

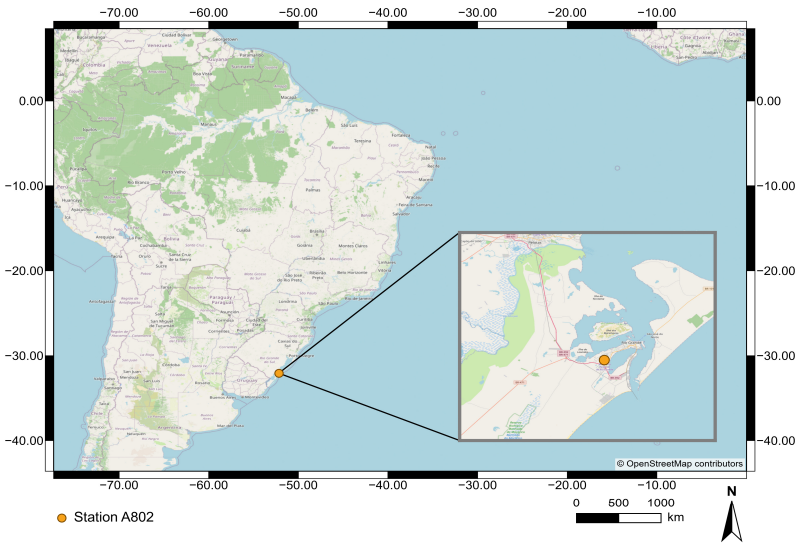


Figure 1. Station A802 on the Brazil Map, located in the municipality of Rio Grande, inside of the Federal University of Rio Grande (FURG) at coordinates 32.07° S and 52.16° W.

The meteorological data used in this study were obtained from automatic station A802 of Brazil's National Institute of Meteorology (INMET), located on the campus of the Federal University of Rio Grande (FURG), at coordinates 32.07°S and 52.16°W, at an altitude of 4.92 m. The station has been in operation since 15 November 2001, and is part of INMET's national network of automatic stations. INMET represents Brazil at the World Meteorological Organization (WMO) and, by delegation from that organization, participates in the international exchange of meteorological messages collected by the South American observation network, within the scope of the World Weather Watch. Therefore, the data used in this study come from an official observational network structured according to operational guidelines and international standards for meteorological observation [25].

Previous climatological studies of Rio Grande, including the study by Krusche et al. [22], reinforce the historical relevance of local meteorological observations for characterizing the region's atmosphere.

2.2. Data Acquisition and Preprocessing

Meteorological data from INMET automatic station A802 were used, covering January 2002 to December 2024 with hourly resolution. The variables considered were atmospheric pressure, relative humidity, precipitation, and date and time records.

An exploratory data analysis was initially performed to identify seasonal patterns, extreme values, and potential inconsistencies in the meteorological time series. Descriptive statistics and graphical visualization techniques were applied to evaluate the distribution of the variables and identify records that were incompatible with the expected physical limits or associated with instrument failure codes. Values such as −9999, as well as physically impossible records for the variables analyzed, were treated as invalid data and removed from the analysis.

The data were then processed to remove missing values and standardize the time series. To ensure temporal consistency and avoid biases associated with imputation, only 3 h windows with complete observations of pressure, relative humidity, and with a valid precipitation record in the target hour, were retained. Windows containing missing values were excluded from the analysis. Thus, the final dataset does not represent the entire original dataset, but rather a subset composed of fully observed atmospheric sequences.

This strategy preserves the physical coherence of the inputs and avoids the introduction of artificial values. However, it reduces the representativeness of the original record, since periods with observational failures were not included in the experiments.

2.3. Dataset Construction and Training/Test Split

Each sample consists of a window of three consecutive hourly observations, from $t-3$ to $t-1$, and its label is the occurrence of precipitation (precipitation > 0 mm) in the following hour, t (Figure 2). Windows are consecutive and non-overlapping (3 h step), so that the label includes both the onset and the continuation of precipitation; no event was excluded because of previous precipitation.

The data were split chronologically. The training period comprises all windows from 1 January 2002 to 31 December 2018 (39,263 windows, of which 3109 (7.92%) were followed by precipitation), and the test period all windows from 1 January 2019 to 31 December 2024 (16,326 windows; 1298 with precipitation, 8.0%). The test set keeps the natural frequency of precipitation. Only the training set was balanced, by retaining all 3109 precipitation windows and randomly selecting the same number of dry windows, yielding 3109 windows per class (6218 windows in total), to prevent the models from being dominated by the dry class.

Variable selection was based on physical knowledge of meteorological processes and on the need to limit the number of fuzzy rules, which grows exponentially with the number of inputs. The antecedent observation-window length (3, 6, 12, or 24 h) was evaluated using five chronological folds with fixed temporal cutoffs, restricted to the 2002–2018 training period. For each fold, class balancing was applied exclusively to the training subset by undersampling non-precipitation windows, and standardization parameters were estimated using only that subset. The subsequent validation subset retained the observed precipitation frequency, providing an evaluation under naturally imbalanced conditions.

The 3 h configuration achieved the highest mean validation AUC-ROC (0.775), compared with 0.752, 0.711, and 0.732 for the 6-, 12-, and 24 h configurations, respectively. It outperformed these alternatives in 5, 5, and 4 of the 5 folds, respectively, and was therefore selected as the antecedent observation window for the final forecasting framework.

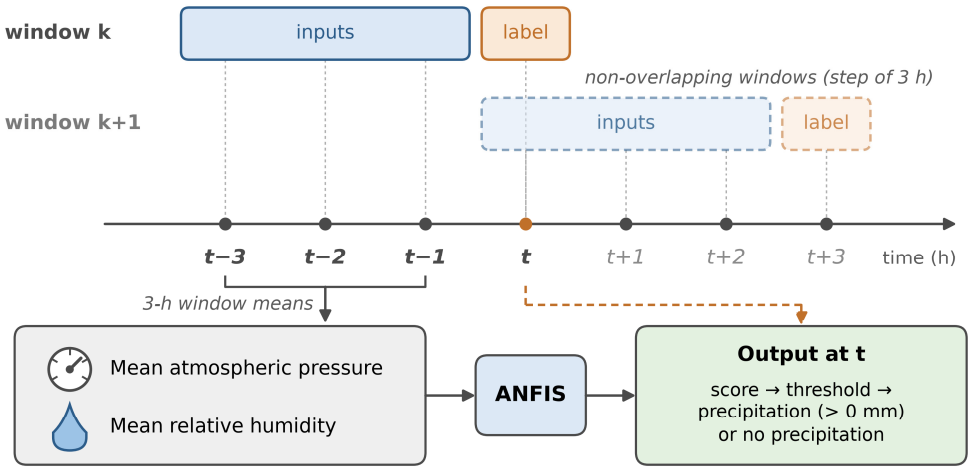


Figure 2. Schematic representation of the 3 h antecedent window (t − 3 to t − 1) used to forecast precipitation occurrence in the following hour (t). The inputs are the means of the window; consecutive windows do not overlap (step of 3 h).

All random procedures (balancing, cross-validation and model initialization) used fixed seeds; the ANFIS training is deterministic because its consequent parameters are initialized at zero. The complete data-processing workflow, from the raw hourly records to the final evaluation, is summarized in Figure 3.

To assess the representativeness of the non-overlapping 3 h sampling procedure, an additional audit was conducted using the original hourly precipitation records. Of the 14,244 h with observed precipitation, 4407 were represented as positive targets in the final dataset (30.9%), primarily reflecting the 3 h sampling interval. The proportion of precipitation onsets among the sampled positive targets was similar to that of the complete precipitation record (37.7% versus 37.5%).

At the event level, the original hourly precipitation record contained 5384 precipitation episodes, defined as sequences of consecutive hours with observed precipitation (>0 mm). Of these, 2900 episodes (53.9%) were represented by at least one positive prediction target in the non-overlapping 3 h sampling procedure. Thus, although only 30.9% of individual precipitation hours were retained as positive targets, the sampling procedure included at least one precipitation hour from more than half of the

identified episodes. Nevertheless, these results do not imply complete temporal coverage of individual episodes or representativeness of every precipitation hour.

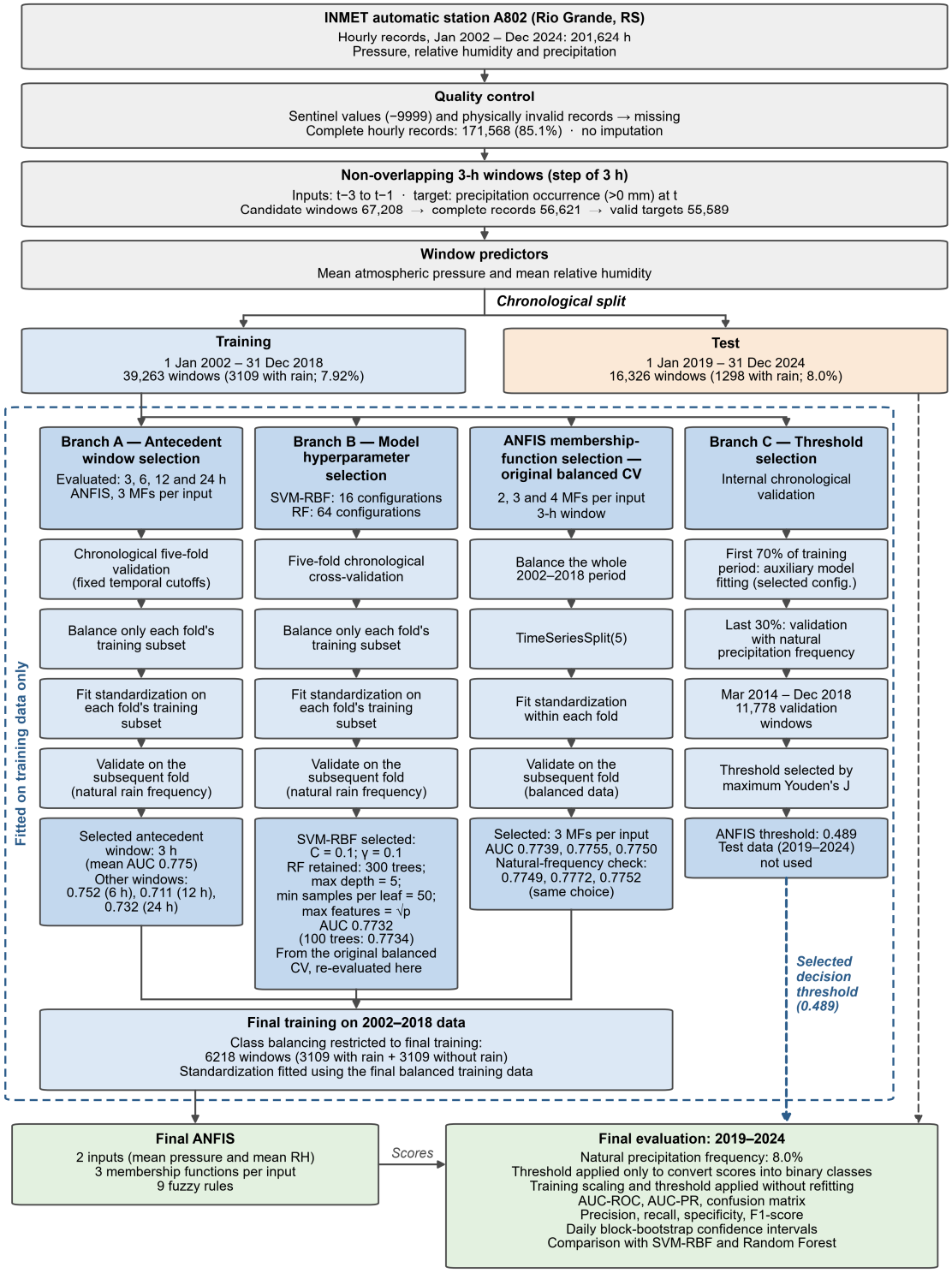


Figure 3. Data-processing and model-development workflow, from the hourly meteorological observations at INMET station A802 to the final evaluation. The diagram distinguishes antecedent window selection, hyperparameter optimization, ANFIS membership-function selection, and internal decision-threshold selection. All model-development procedures were restricted to the 2002–2018 training period. The final models and selected decision thresholds were evaluated on the chronologically subsequent 2019–2024 test period without refitting.

To quantify how strongly an antecedent dry-period criterion would restrict the sample, a 24 h dry-period requirement, commonly used to isolate rainfall onsets, was applied retrospectively to the present sampling. Only 510 of the 2900 represented episodes would satisfy it, so that 2390 episodes (82.4%) would be excluded and the restricted sampling would represent only about 9.5% of the 5384 identified episodes. Of the excluded episodes, 1920 began less than 24 h after the end of the preceding episode (median interval, 5 h). Over the complete record, the median interval between consecutive precipitation episodes was 10 h (90th percentile, 91 h), and the excluded episodes had a median duration of 3 h (mean, about 4 h; maximum, 43 h). Restricting the analysis to events preceded by long dry periods would therefore remove most of the precipitation episodes observed at the station.

Furthermore, sensitivity to the starting phase of the 3 h windows was evaluated using three alternative alignments (t − 3, t − 2, and t − 1). The resulting AUC values ranged from 0.808 to 0.810, indicating that discrimination performance was relatively stable across the evaluated sampling phases. However, the selected decision thresholds varied from 0.455 to 0.489, with corresponding sensitivities between 0.81 and 0.84. These findings support the stability of the model's discrimination performance while acknowledging that the non-overlapping sampling procedure does not evaluate every precipitation hour.

2.4. ANFIS Model Structure and Adopted Configuration

The model used in this study was the adaptive neuro-fuzzy inference system (ANFIS), proposed by Jang [14], which combines fuzzy inference systems with artificial neural networks. This type of model can represent nonlinear relationships through Takagi-Sugeno-type fuzzy rules, whose parameters are adjusted during training.

The model is divided into five computational layers, as proposed by Jang [14]. In the conceptual representation, square nodes indicate adjustable parameters associated with the antecedents and consequents of the fuzzy rules, while circular nodes represent fixed operations. Inputs x and y are, in this study, the average atmospheric pressure and the average relative humidity of the window. In the implemented model, each input has three membership functions, so the rule index i runs from 1 to 9 and the sums in Equations (3) and (5) extend over the nine rules.

The mathematical formulation presented below follows the conventional five-layer ANFIS architecture described by Jang [14], using a generic rule-based notation to illustrate the inference process. In Equations (1) and (2), the index i identifies a fuzzy rule, while A_i and B_i denote the antecedent fuzzy sets associated with that rule, which may be shared across multiple rules. Therefore, the number of rules does not correspond to the number of distinct membership functions. In the implemented configuration, each of the two meteorological inputs is represented by three membership functions, and their pairwise combinations generate nine fuzzy rules ($R = 9$). Figure 4 provides a simplified conceptual illustration of the inference architecture rather than an exhaustive representation of the implemented network.

The first layer consists of nodes associated with the membership functions.

$$O_{1,i} = \mu_{A_i}(x), \qquad i = 1,2,\ldots,R \tag{1}$$

This layer calculates the firing strength of each rule, representing the degree to which the rule's conditions are satisfied by the given inputs.

$$O_{2,i} = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y), \qquad i = 1,2,\ldots,R \tag{2}$$

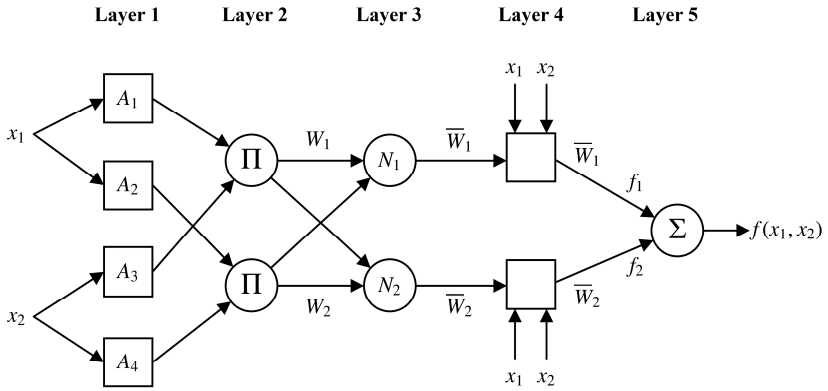


Figure 4. Conceptual illustration of the conventional five-layer ANFIS architecture [10,14,26] showing the fuzzification, rule activation, normalization, consequent evaluation, and output aggregation stages. The diagram illustrates a simplified two-input, two-rule configuration; the ANFIS implemented in this study uses two inputs, three membership functions per input, and nine fuzzy rules.

The third layer is responsible for normalizing the firing strengths of the rules, ensuring that each rule contributes proportionally to the inference process. Normalization is obtained by dividing each firing strength w_i by the sum of the firing strengths of all rules and can be denoted as:

$$O_{3,i} = \underline{w_i} = \frac{w_i}{\sum_{j=1}^{R} w_j}, \qquad i = 1, 2, \ldots, R \tag{3}$$

The fourth layer, which corresponds to the consequent part of the fuzzy logic structure, multiplies the normalized rule weight $\underline{w_i}$ by the consequent function f_i. This function is linear, and its parameters p_i, q_i, and r_i are adjustable during model training. The linear function is defined as:

$$O_{4,i} = \underline{w_i} f_i = \underline{w_i}(p_i x + q_i y + r_i) \tag{4}$$

where

- p_i is the coefficient of the input x in the consequent function of rule i;
- q_i is the coefficient of the input y in the consequent function of rule i;
- r_i is the independent term or bias of the function resulting from rule i.

The fifth layer corresponds to the output node of the ANFIS model. In this layer, all weighted rule outputs are summed, resulting in the overall system output. The output is a continuous score in which each rule contributes in proportion to its normalized firing strength; no clipping or additional transformation is applied. The output function of the fifth layer is given by:

$$O_{5,1} = \sum_{i=1}^{R} \underline{w_i} f_i = \frac{\sum_{i=1}^{R} w_i f_i}{\sum_{i=1}^{R} w_i} \tag{5}$$

2.5. Hyperparameters and Training Procedure

The model was trained by batch gradient descent on the mean squared error between the continuous output and the binary target, for 1200 epochs. In each epoch, the consequent parameters (initialized at zero) were updated with a learning rate of 0.1, and the premise parameters—the center c and the width a of each membership function—with a learning rate of 0.01; the shape parameter b was kept fixed at 2. The membership functions were initialized with centers evenly spaced between the training minimum and

maximum of each standardized input and with widths equal to half the spacing between centers. The final model configuration, hyperparameters and test-set metrics are summarized in Appendix A (Table A1).

The number of membership functions per input was evaluated using five-fold chronological cross-validation (scikit-learn TimeSeriesSplit) within the 2002–2018 training period. Unlike the antecedent window-length selection procedure, this analysis was conducted on a previously balanced training dataset, obtained by random undersampling the majority non-precipitation class before constructing the folds. Consequently, the validation subsets had approximately balanced class distributions rather than the natural precipitation frequency. Standardization was fitted separately using the training portion of each fold. Configurations with two, three, and four membership functions per input were evaluated, corresponding to 4, 9, and 16 fuzzy rules, respectively. Because the three configurations yielded practically identical mean validation AUC-ROC values (0.774, 0.775 and 0.775; Appendix A, Table A2), the number of membership functions could not be selected on discrimination alone. Three membership functions per input were adopted because they provide a linguistically interpretable partition (low, medium and high) with a moderate number of rules. As a consistency check, the comparison was repeated with validation subsets retaining the natural precipitation frequency, which yielded mean AUC-ROC values of 0.775, 0.777 and 0.775 and led to the same choice (Figure 3). The final ANFIS was subsequently trained on the balanced 2002–2018 training dataset and evaluated on the chronologically separated, non-balanced 2019–2024 test period.

The hybrid least-squares estimator of Jang [14] was not used. Inputs were standardized with the mean and standard deviation of the balanced training set. The table containing hyperparameters and summarized model results is provided in Appendix A.

2.6. Model Evaluation

Model evaluation was performed in two complementary steps: analysis of the continuous output produced by ANFIS and evaluation of the binary classification after applying a decision threshold. The model's continuous output was interpreted as an index of favorability for precipitation occurrence, ranging from absence of precipitation to precipitation occurrence. This output was subsequently converted into binary classes, with 0 corresponding to the absence of precipitation and 1 corresponding to the occurrence of precipitation.

Initially, the performance of the continuous output was evaluated using mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). These metrics were used to monitor error during training and examine the stability of the model-fitting process [27,28]. The metrics are defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (7)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (8)$$

where y_i represents the observed value, $\hat{y}_i$ represents the output estimated by the model, and n corresponds to the total number of samples evaluated.

To evaluate the discriminatory capacity of the model, the receiver operating characteristic (ROC) curve and the area under the curve (AUC) were used. The ROC curve allows analysis of the relationship between the true positive rate and the false positive rate at different decision thresholds. The decision threshold was defined without using the test set, by maximizing Youden's J statistic (sensitivity + specificity − 1) on an internal validation subset: the last 30% of the training period before balancing (1 March 2014 to 31 December 2018; 11,778 windows; 9.1% with precipitation) was predicted by an auxiliary model trained on the first 70% (balanced), and the resulting threshold was applied to the final model trained on the whole training period. The selected threshold was subsequently applied to the final ANFIS model and evaluated on the chronologically separated 2019–2024 test period. The test observations were not used to fit the preprocessing transformations, optimize the classification threshold, or estimate the parameters of the final model. However, because this period had been examined during earlier exploratory model comparisons, it should not be regarded as a completely untouched holdout throughout the entire development process. The reported performance therefore characterizes the final configuration on a later chronological evaluation period, subject to this development-history limitation.

$$\text{Accuracy} = \frac{TP + TN}{TP+TN+FP+FN} \quad (9)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (11)$$

$$\text{F1} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (12)$$

The reported classification and discrimination metrics were calculated on the chronologically separated 2019–2024 evaluation period, subject to the development-history limitation described above. Because precipitation occurred in only approximately 8.0% of the test windows, the area under the precision–recall curve (AUC-PR) was also reported. The 95% confidence intervals were estimated using a block bootstrap with daily blocks and 2000 resamples, preserving the dependence between observation windows belonging to the same day. Differences in AUC-ROC between models were assessed using the same resampling procedure, with days sampled jointly for each pair of models (paired bootstrap).

ANFIS performance was compared with that of a support vector machine with a radial basis function kernel (SVM-RBF) and Random Forest (RF), both using the same two meteorological inputs as ANFIS. The hyperparameters of SVM-RBF (C and γ) and RF (number of trees, maximum depth, minimum samples per leaf, and maximum features) were evaluated by grid search using five-fold time-series cross-validation restricted to the 2002–2018 training period, with AUC-ROC as the selection criterion. In each fold, class balancing was applied exclusively to the training subset, while the subsequent validation subset retained the observed precipitation frequency. Standardization was also fitted exclusively on the corresponding training subset.

The grid search evaluated 16 SVM-RBF and 64 RF configurations. For SVM-RBF, the optimal configuration remained C = 0.1 and γ = 0.1, with a mean validation AUC-ROC of 0.7755. For RF, the previously adopted configuration of 300 trees, a maximum depth of 5, at least 50 samples per leaf, and $\sqrt{p}$ features per split achieved a mean validation AUC-ROC of 0.7732. Although a configuration with 100 trees achieved a slightly higher mean AUC-ROC of 0.7734, the difference was only 0.0002, and the original 300-tree configuration was retained for consistency with the final model evaluation. The decision thresholds of both models were selected using the same Youden's J criterion adopted for

ANFIS. Training time, inference time per batch and per sample, model size, and number of parameters were measured as medians of repeated runs on the same computer, using a single processing core.

An additional retrospective rolling-origin analysis was conducted to examine the sensitivity of ANFIS discrimination performance to the length of the historical training period. Three strategies were compared: an expanding training window beginning in 2002, a sliding window covering the preceding 10 calendar years, and a sliding window covering the preceding 5 calendar years. For each evaluation year from 2012 to 2024, the ANFIS model was retrained from scratch using observations from the preceding years. Class balancing and standardization were performed separately for each training period, while the corresponding evaluation year retained its natural precipitation frequency. The model architecture and training hyperparameters were held fixed across the strategies. Performance was summarized using the mean annual AUC-ROC over the 13 evaluation years.

This analysis was retrospective and distinct from the five-fold cross-validation procedures used for model configuration. The evaluation included years from the 2002–2018 development period and from the subsequent 2019–2024 test period. Consequently, these results should be interpreted as a sensitivity assessment of historical training-window length rather than as an independent, fully prospective validation of the final model.

Additional diagnostic analyses were conducted on the test period to identify the sources of predictive skill, without refitting the final models. First, two persistence baselines were evaluated: a binary forecast that predicts precipitation at t whenever precipitation was observed at t − 1, and a continuous score defined as the number of hours with precipitation in the antecedent window (0–3). A logistic regression with the same two standardized inputs was trained on the balanced 2002–2018 data, with its decision threshold selected by Youden's J on the same internal validation subset used for ANFIS. Second, the test windows were stratified by the time elapsed since the last observed precipitation hour before t, computed from the complete hourly record, and by the occurrence of precipitation at t − 1 or anywhere in the antecedent window. Third, the relevance of each input was assessed by the AUC of each raw input, by single-input ANFIS models (three membership functions and three rules, trained with the same procedure) and by permutation importance on the test period (100 permutations per input). Fourth, the final models were applied to overlapping windows advanced in 1 h steps over the test period, so that predictions were issued at every hourly timestamp. Differences in AUC were assessed with the paired daily block bootstrap described above. Because these analyses were defined after the test results of the main models were known, they are reported as diagnostic assessments and were not used for model selection.

3. Results

3.1. Discrimination and Classification Performance

Table 1 summarizes the performance of the three models on the test period (1 January 2019–31 December 2024). ANFIS reached an AUC of 0.809 (95% CI 0.794–0.823), statistically indistinguishable from the tuned SVM-RBF (0.807) and RF (0.807): the paired differences in AUC were +0.002 (95% CI −0.001 to +0.005) relative to SVM-RBF and +0.002 (−0.001 to +0.005) relative to RF (Figure 5a). The areas under the precision–recall curve were also similar (0.271, 0.257 and 0.271), about 3.4 times the precipitation frequency of 8.0%, which is the AUC-PR of a non-informative forecast (Figure 5b).

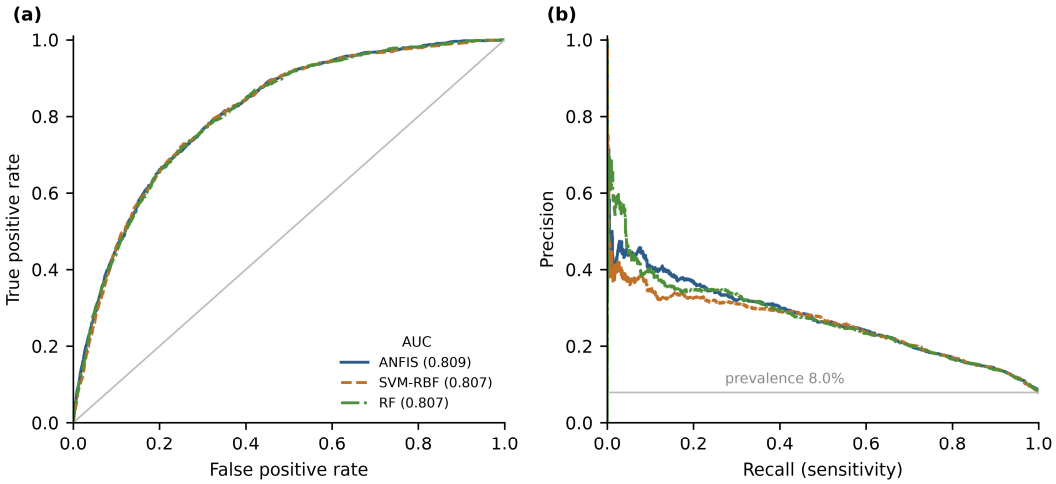


Figure 5. (**a**) ROC curves and (**b**) precision–recall curves of ANFIS, SVM-RBF and RF on the test period (2019–2024). The horizontal line in (**b**) is the precipitation frequency (8.0%), which corresponds to a non-informative forecast.

Table 1. Performance on the test period (2019–2024; 16,326 windows, 8.0% with precipitation). AUC with 95% confidence interval from daily block bootstrap.

Forecast	AUC (95% CI)	AUC-PR	Sens.	Spec.	Precision	F1	Alerts
ANFIS	0.809 (0.794–0.823)	0.271	0.805	0.658	0.169	0.279	6188
SVM-RBF	0.807 (0.792–0.822)	0.257	0.823	0.636	0.164	0.273	6532
RF	0.807 (0.792–0.821)	0.271	0.782	0.681	0.175	0.286	5805

At the decision threshold selected on the internal validation subset (0.489), ANFIS detected 80.5% of the precipitation windows (1045 of 1298) with a specificity of 65.8% (Table 1). At the natural frequency of precipitation, this corresponds to a precision of 0.169: the model issued 6188 alerts in the 2091 days of the test period (about 3.0 per day), of which 1045 were correct. SVM-RBF and RF had similar operating points, with SVM-RBF trading a higher sensitivity (0.823) for a lower specificity (0.636) and RF the opposite (0.782 and 0.681).

3.2. Membership Functions and Fuzzy Rules

Table 2 lists the parameters of the final membership functions in physical units, and Figure 6 shows their shape. For pressure, the centers were 991.5, 1014.1 and 1036.8 hPa, with widths of about 11.3 hPa; for relative humidity, they were 23.3%, 61.6% and 100.0%, with widths of about 19.1 percentage points. The premise parameters changed little from their grid initialization, so the partition remained nearly uniform over the training range. In practice, the antecedent partition therefore behaves close to a fixed grid, and most of the adaptation learned from the data is contained in the consequent parameters (Table 3). Because the low and high pressure centers lie at the extremes of the record (991.2 and 1036.8 hPa), most windows have their largest membership in the medium pressure term, with partial membership in the low or high terms. Similarly, the low and high humidity centers lie at the minimum and maximum of the training windows (21.7% and 100%); only 1.3% of the training windows have a relative humidity below 42%, where the low term dominates, so the rules are mainly determined by the medium and high humidity terms.

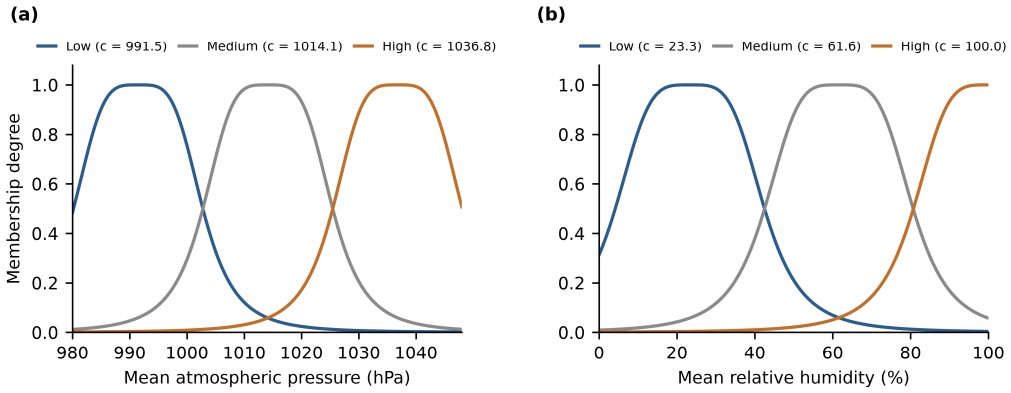


Figure 6. Final bell-shaped membership functions of ANFIS in physical units: (**a**) average atmospheric pressure and (**b**) average relative humidity. The legend gives the center c of each function; all parameters are listed in Table 2.

Table 2. Parameters of the generalized bell-shaped membership functions of the final ANFIS, defined as $\mu(x) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}}$ where (x) is the input variable, (a) controls the width of the membership function, (b) determines the steepness of its slopes, and (c) defines its center. Parameters (a) and (c) are expressed in the physical units of their respective input variables, while (b) is dimensionless and fixed at 2.

Input	Term	c	a	b
Average pressure (hPa)	Low	991.5	11.29	2
	Medium	1014.1	11.34	2
	High	1036.8	11.31	2
Average relative humidity (%)	Low	23.3	19.16	2
	Medium	61.6	19.07	2
	High	100.0	19.22	2

Table 3. Fuzzy rules of the final ANFIS. p, q and r are the consequent parameters ($f = p \cdot x_1 + q \cdot x_2 + r$, with standardized inputs); firing share is the proportion of the total normalized firing strength in the balanced training set; rain ratio is the proportion of precipitation windows weighted by firing strength (0.5 = class balance).

Rule	Pressure (x_1)	Humidity (x_2)	p	q	r	Firing Share (%)	Rain Ratio
R1	Low	Low	−0.010	−0.006	0.006	0.2	0.34
R2	Low	Medium	−0.258	−0.043	0.154	4.0	0.49
R3	Low	High	−0.258	0.067	0.182	6.4	0.72
R4	Medium	Low	0.012	−0.019	0.014	2.1	0.20
R5	Medium	Medium	−0.088	0.120	0.316	33.1	0.34
R6	Medium	High	−0.143	0.233	0.629	44.8	0.62
R7	High	Low	0.004	−0.008	0.004	0.3	0.18
R8	High	Medium	0.135	−0.025	0.081	4.2	0.29
R9	High	High	0.219	0.053	0.122	4.9	0.54

Table 3 lists the nine rules with their consequent parameters, their share of the total normalized firing strength in the balanced training set and the proportion of precipitation windows weighted by firing strength (rain ratio; 0.5 corresponds to the class balance of the training set). Two rules accounted for 78% of the firing strength: medium pressure with high humidity (R6; 44.8%) and medium pressure with medium humidity (R5; 33.1%). The rules followed a consistent physical pattern: all rules with high relative humidity (air

close to saturation) had rain ratios above 0.5 (0.54–0.72), the highest for low pressure with high humidity (R3; 0.72), whereas the rules with low humidity had ratios between 0.18 and 0.34. For each humidity level, the rain ratio decreased with increasing pressure (for example, 0.72, 0.62 and 0.54 for low, medium and high pressure with high humidity), in agreement with the association of lower pressure and moist air with precipitation. These interpretations rest mainly on the rules with substantial support. The three low-humidity rules (R1, R4 and R7) together accounted for only 2.6% of the firing strength, and their consequent parameters remained close to their zero initialization; their low rain ratios therefore reflect the scarcity of such conditions in the record rather than a well-constrained dry regime.

The training error decreased rapidly in the first 100 epochs and remained stable thereafter (Figure 7a), from an RMSE of 0.707 and an MAE of 0.500 at initialization to 0.438 and 0.384 after 1200 epochs. The output of ANFIS is a continuous score that increased monotonically with the observed frequency of precipitation but is not a calibrated probability: because the model was trained on balanced data, the mean score by decile of the test period ranged from 0.04 to 0.81, whereas the observed precipitation frequency in the same deciles ranged from 0.003 to 0.31. Because the output is not clipped, 1.0% of the scores fell slightly outside [0, 1] (from −0.02 to 1.12). The scores of windows followed and not followed by precipitation overlapped substantially (Figure 7b), which explains the moderate precision at the selected threshold.

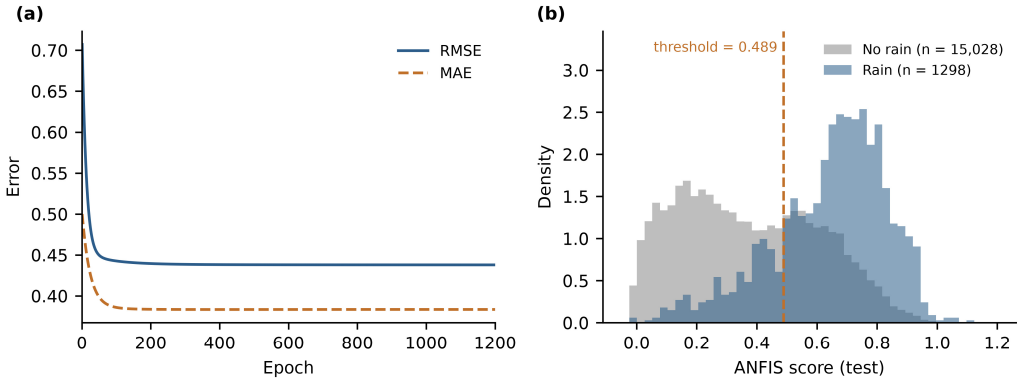


Figure 7. (**a**) RMSE and MAE of the continuous ANFIS output during training (1200 epochs, balanced training set). (**b**) Distribution of the ANFIS score on the test period for windows followed and not followed by precipitation; the dashed line is the decision threshold selected on the internal validation subset (0.489).

3.3. Sensitivity to Temporal Sampling Alignment

The sensitivity analysis showed that the discriminatory performance of ANFIS remained stable across the three starting phases of the non-overlapping 3 h windows (Table 4). The AUC values ranged from 0.808 to 0.810, with a maximum difference of 0.002, indicating limited sensitivity of the model's overall discrimination to the initial sampling hour.

Table 4. Sensitivity of ANFIS performance to the starting phase of non-overlapping 3 h windows.

Phase	Test Windows	Rain	AUC	Threshold	Sensitivity	Specificity	Precision
t − 1	16,326	1298	0.809	0.489	0.805	0.658	0.169
t − 2	16,331	1299	0.808	0.473	0.818	0.642	0.165
t − 3	16,332	1297	0.810	0.455	0.843	0.610	0.157

However, differences were observed in the classification metrics obtained at the decision thresholds selected independently for each configuration. The selected thresholds ranged from 0.455 to 0.489, with sensitivity varying between 0.805 and 0.843, specificity between 0.610 and 0.658, and precision between 0.157 and 0.169.

These findings indicate that, although the overall discriminatory capacity was relatively stable across the evaluated sampling phases, the classification performance at the selected operating point exhibited some variability. Therefore, the choice of the initial sampling phase had a limited effect on AUC but influenced the balance between precipitation detection and false alarms.

The results support the robustness of ANFIS discrimination with respect to the evaluated temporal alignments, while reinforcing the importance of considering threshold-dependent metrics when assessing its potential operational applicability.

3.4. Computational Cost

Table 5 compares the computational cost of the three models. ANFIS was the most compact model (39 parameters and 22 KB, versus 12,430 stored values for the support vectors of SVM-RBF and 14,720 tree nodes for RF) and the fastest at inference: 4.8 ms to score the whole test period (16,326 windows) and 0.15 ms per individual window, compared with 3.2 s and 0.34 ms for SVM-RBF and 185 ms and 13.2 ms for RF. Its training time (5.6 s for 1200 epochs of batch gradient descent) was longer than that of SVM-RBF (3.5 s) and RF (0.7 s). For all models, these costs are negligible compared with the hourly update of the station data, so computational cost is not a limiting factor for operational use; the compactness of ANFIS is mainly relevant for embedded or low-resource applications.

Table 5. Computational cost of the models (medians of repeated runs on the same computer using a single core). Batch: the whole test period (16,326 windows); sample: one window. Parameters: premise and consequent parameters (ANFIS), stored support-vector values and coefficients (SVM-RBF) and tree nodes (RF).

Model	Training (Seconds)	Inference per Batch (Milliseconds)	Inference per Sample (Milliseconds)	Model Size (KB)	Parameters
ANFIS	5.61	4.85	0.15	22.3	39
SVM-RBF	3.46	3178	0.34	147.3	12,430
RF	0.68	184.8	13.20	1233.6	14,720

3.5. Sources of Predictive Skill

Table 6 compares ANFIS with reference forecasts on the test period. A logistic regression with the same two inputs reached an AUC of 0.808 (95% CI 0.793–0.823), indistinguishable from ANFIS (paired difference +0.001; 95% CI −0.003 to +0.005), so the nonlinear rule base did not add discrimination beyond a linear combination of pressure and humidity (standardized coefficients −0.41 and +1.11). Both persistence baselines, which require real-time precipitation observations, performed better. The continuous persistence score reached an AUC of 0.843 (0.829–0.857), higher than that of ANFIS (paired difference −0.034; −0.050 to −0.020), and an AUC-PR of 0.496. Binary persistence issued 1296 alerts with a precision of 0.64 and a sensitivity of 0.63, whereas ANFIS issued 6188 alerts with a precision of 0.17.

Table 6. Comparison of ANFIS with reference forecasts on the test period (2019–2024; 16,326 windows, 8.0% with precipitation). AUC with 95% confidence interval from daily block bootstrap. For binary persistence, which has a single operating point, the AUC equals (sensitivity + specificity)/2.

Forecast	AUC (95% CI)	AUC-PR	Sens.	Spec.	Precision	F1	Alerts
ANFIS	0.809 (0.794–0.823)	0.271	0.81	0.66	0.17	0.28	6188
Logistic Regression	0.808 (0.793–0.823)	0.275	0.80	0.66	0.17	0.28	6153
Binary Persistence	0.801 (0.787–0.816)	-	0.63	0.97	0.64	0.63	1296
Continuous persistence (rainy hours in window)	0.843 (0.829–0.857)	0.496	-	-	-	-	-
Climatology (constant)	0.500	0.080	-	-	-	-	-

The advantage of persistence arises from periods with ongoing precipitation (Table 7). Of the 1298 precipitation windows in the test period, 823 (63%) were preceded by precipitation at t − 1. In these windows, ANFIS discriminated less well than persistence (AUC 0.600 versus 0.652; difference −0.051, −0.089 to −0.013). Conversely, when no precipitation occurred at t − 1 (15,030 windows, 475 with precipitation), ANFIS outperformed persistence (0.739 versus 0.620; +0.119, +0.094 to +0.145), and in fully dry antecedent windows, where persistence carries no information, ANFIS reached an AUC of 0.715 (0.689–0.742). The practical value of this skill is limited, however: in dry windows, precipitation occurred in only 2.4% of the cases, the AUC-PR was 0.062 and the precision at the selected threshold was 0.05. Within each class of time since the last precipitation, the AUC of ANFIS ranged from 0.600 (1 h) to 0.736 (more than 72 h), below the overall value, which therefore reflects to a large extent the separation between rainy and dry periods. SVM-RBF and random forest followed the same pattern, with AUC differences from ANFIS not exceeding 0.011 in any stratum.

Table 7. Discrimination of continuous persistence and ANFIS on the test period, stratified by antecedent precipitation. AUC with 95% confidence interval; Δ is the paired difference ANFIS—persistence (daily block bootstrap). [a] Constant score.

Stratum	Windows	Rain	Persistence AUC	ANFIS AUC	Δ (95% CI)
No precipitation at t − 1	15,030	475	0.620 (0.598 0.641)	0.739 (0.717 0.760)	+0.119 (+0.094 to +0.145)
Dry antecedent window (t − 3 to t − 1)	14,180	339	0.500 [a]	0.715 (0.689 0.742)	+0.215 (+0.189 to +0.242)
Precipitation only at t − 3 or t − 2	850	136	0.564 (0.520–0.607)	0.596 (0.551 0.643)	+0.033 (−0.026 to +0.094)
Precipitation at t − 1	1296	823	0.652 (0.623–0.682)	0.600 (0.566 0.634)	−0.051 (−0.089 to −0.013)
All test windows	16,326	1298	0.843 (0.829–0.857)	0.809 (0.794 0.823)	−0.034 (−0.050 to −0.020)

Relative humidity was the dominant input. Used alone, it reached an AUC of 0.783 (0.768–0.798), compared with 0.647 (0.622–0.671) for an ANFIS using pressure alone; relative to the two-input model, removing pressure reduced the AUC by 0.026 (0.018 0.035) and removing humidity by 0.162 (0.142–0.183). Permuting the inputs reduced the AUC of ANFIS by 0.254 for humidity and 0.049 for pressure, with similar values for SVM RBF and random forest. Discrimination was similar across seasons (AUC 0.781–0.821),

with overlapping confidence intervals. When the final models were applied to overlapping windows advanced in 1 h steps (48,989 test windows, 8.0% with precipitation), 97.3% of the precipitation hours of 2019 2024 became prediction targets, compared with 32.4% under the non-overlapping sampling. The discrimination of ANFIS was unchanged (AUC 0.809, 0.796–0.822; AUC-PR 0.276), with a sensitivity of 0.80, a specificity of 0.66 and a precision of 0.17 at the selected threshold, corresponding to 8.9 alerts per day.

4. Discussion

4.1. ANFIS Compared with SVM-RBF and Random Forest

During the 2019–2024 test period, ANFIS, SVM-RBF, and Random Forest achieved similar AUC-ROC values of 0.809, 0.807, and 0.807, respectively. The paired bootstrap confidence intervals for the differences between ANFIS and the benchmark models included zero, indicating that no statistically significant differences in discrimination were detected under the evaluated conditions. However, these findings do not establish statistical equivalence or the general superiority of any model. The main advantages of ANFIS in this application are associated with its compact architecture, comprising 39 adjustable parameters, its computational efficiency during inference, and the interpretability of its nine fuzzy rules. These rules provide a transparent representation of the statistical relationships between antecedent meteorological conditions and precipitation occurrence, although they do not establish the physical mechanisms responsible for individual precipitation events.

The comparison with reference forecasts qualifies these advantages. ANFIS did not discriminate better than a logistic regression using the same inputs, and it was outperformed by a persistence score based on the number of rainy hours in the antecedent window. Most of the overall skill therefore reflects the identification of ongoing precipitation, during which near-surface air is close to saturation, rather than the anticipation of new events. The contribution of pressure and humidity is clearest where persistence is uninformative: when no precipitation occurred in the preceding hour, ANFIS exceeded the AUC of persistence by 0.12. Even in this regime, however, precision remained low (0.06), so the two surface variables provide a ranking of onset risk rather than reliable onset alerts. In an exploratory analysis, adding the number of rainy hours in the window as a third ANFIS input (18 rules) increased the AUC to 0.897 (0.886–0.907), above persistence (+0.053; +0.045 to +0.061). Because this configuration was defined after the test results were known and its membership functions were not selected by cross-validation, it is reported only as an indication for future work and not as a validated model.

4.2. Comparison with Previous Studies

Table 8 summarizes studies that applied ANFIS or other machine learning methods to precipitation forecasting. A direct comparison of performance values is limited, because the studies differ in forecast horizon (hourly to monthly), target (occurrence or amount), predictors, size and period of the data, definition of precipitation, data splitting and evaluation metrics. In particular, random splits of autocorrelated series (e.g., the random 70/30 splits of daily data in [10,11]) and balanced test sets tend to produce higher scores than those obtained with a chronological split at the natural frequency of precipitation, as in the present study. The contribution of this study is therefore not a higher score, but an evaluation on a continuous, chronologically independent period at the natural frequency of precipitation.

Table 8. Studies on precipitation forecasting with ANFIS and other machine learning methods.

Study	Location	Horizon	Target	Predictors	Data and Split	Methods	Metric
Shamshirband (2015) [7]	Serbia, 29 synoptic stations	Monthly	Precipitation amount	Month of the year (single input)	Monthly records 1946–2012;chronological 50/50 split (training 1946–1979, testing 1980–2012)	ANFIS vs. SVR (linear, polynomial, RBF)	RMSE, r, R^2; ANFIS more accurate than SVR
Hashim (2016) [5]	Patna, India	Monthly	Rainfall amount	Wet-day frequency, vapor pressure, Tmax, Tmin, cloud cover	IMD, 1901–2000; split not specified	ANFIS (with input selection) vs. ANN, SVM, ELM, GP	R^2 = 0.976 (ANFIS) vs. 0.953–0.957
Manandhar (2018) [15]	São Luís, Brazil; Singapore	5 min	Onset of rain	GPS precipitable water vapor and its second derivative	5 min data; Singapore 2010–2015, São Luís 2016–2017; independent years and stations	Threshold rule	Detection 84.7%, false alarm 37% (São Luís); 87.7%, 38.6% (Singapore)
Manandhar (2019) [16]	Singapore	5 min	Occurrence	Precipitable water vapor, solar radiation, day of year, hour of day	2012–2015; repeated 30/70 partitions; balanced classes	SVM	Detection 80.4%, false alarm 20.3%
Benevides (2019) [17]	Lisbon, Portugal	1 h	Intense rain (≥5 mm h^{-1})	Precipitable water vapor, surface pressure, temperature, relative humidity	Hourly, 2011–2015; random split in 2011–2014, 2015 held out	NARX neural network	63–72% of intense events correctly classified; false positives 21–36% of the intense class
Pham (2020) [10]	Hoa Binh, Vietnam	Daily	Rainfall amount	Tmax, Tmin, wind speed, RH, solar radiation	Daily, 2004–2013 (3653 days); random 70/30 split	PSO-ANFIS,ANN, SVM	R, MAE, SS, POD, CSI, FAR; SVM best
Pham (2024) [11]	Cao Phong, Hoa Binh, Vietnam	Daily	Rainfall amount	Tmax, Tmin, wind speed, RH, solar radiation (Savitzky–Golay filtered)	Daily, 2004–2013; random 70/30 split	ANFIS optimized by ABC, GA and SA	R, MAE, RMSE, POD, CSI; ANFIS ABC best (R = 0.82 raw, 0.90 filtered)
This study	Rio Grande, Brazil (INMET A802)	1 h	Occurrence (>0 mm)	Average pressure and relative humidity of a 3 h window	Hourly, 2002–2024; chronological split, natural frequency	ANFIS (9 rules), SVM-RBF, RF	AUC 0.809 (test)

Studies that forecast precipitation occurrence at a single station from moisture-related predictors provide a closer, although still partial, reference. Using GPS-derived precipitable water vapor, Manandhar et al. [15] detected 84.7% of the precipitation onsets at a station in São Luís, Brazil, with false alarms in 37% of the non-rainy periods, for a lead time of 5 min and independent validation years. Manandhar et al. [16], with a support vector machine that combined precipitable water vapor, solar radiation and time of day, reported a detection rate of 80.4% and a false alarm rate of 20.3% for a lead time of 5 min in Singapore, evaluated on repeated partitions of class-balanced data. At its selected

threshold, the ANFIS of the present study detected 81% of the precipitation windows, with false alarms in 34% of the dry windows; at false alarm rates of 37% and 20%, it would detect 82% and 66% of the precipitation windows, respectively. Its detection is therefore of the same order as that reported for São Luís, which refers only to onsets and to a lead time of 5 min, and lower than that reported for Singapore at a similar false alarm rate. Both studies, however, forecast 5 min ahead, whereas the present model forecasts the following hour. Part of these differences may be related to the use of vertically integrated water vapor instead of surface variables, but the studies also differ in the definition of events, the composition of the evaluation sets and the threshold selection, so a quantitative comparison is not possible. Benevides et al. [17] forecast hourly intense precipitation (at least 5 mm h^{-1}) in Lisbon with a neural network that combined precipitable water vapor with surface pressure, temperature and relative humidity, and correctly classified 63–72% of the intense events; because the target and the definition of false positives differ from those used here, these values are not directly comparable.

4.3. Limitations

At the natural frequency of precipitation, the precision of the alerts is low (0.17 at the selected threshold), so the threshold must be chosen considering the costs of false alarms and missed events for each application. Raising the threshold increases precision but reduces the detection of precipitation: at a threshold of 0.80, precision rises to 0.36 while sensitivity falls to 0.20. Even when only the 2% of windows with the highest scores are flagged, precision remains approximately 0.41. These findings suggest that adjusting the decision threshold alone does not resolve the substantial false-alarm burden under the evaluated conditions. Finally, the ANFIS output is a favorability score rather than a calibrated probability, and an additional calibration step would be required if probabilistic forecasts were needed.

An additional limitation concerns the temporal sampling strategy and the representation of antecedent atmospheric conditions. Although the prediction target corresponds to precipitation occurrence in the following hour, the use of consecutive, non-overlapping 3 h observation windows means that forecasts are generated at 3 h intervals rather than at every hourly timestamp. Consequently, only 4407 of the 14,244 observed precipitation hours (30.9%) were represented as positive prediction targets in the complete observational record. However, applying the final models to overlapping windows advanced in 1 h steps, which made 97.3% of the precipitation hours of the test period prediction targets, yielded the same discrimination and operating point (Section 3.5), indicating that the non-overlapping sampling did not bias the reported metrics. Furthermore, the model inputs consist of average atmospheric pressure and average relative humidity calculated over each antecedent window. This aggregation does not preserve short-term fluctuations or temporal tendencies within the observation period. Accordingly, the implemented ANFIS operates as a static nonlinear mapping between aggregated meteorological conditions and subsequent precipitation occurrence, rather than explicitly modeling atmospheric temporal evolution or identifying the physical mechanisms responsible for individual precipitation events.

An additional limitation concerns the spatial representativeness of the observational dataset. The present study was conducted exclusively with meteorological observations from INMET station A802, located in Rio Grande, southern Brazil. Although the chronological evaluation provides evidence of predictive performance across different years, it does not establish the spatial generalizability of the proposed model. Precipitation exhibits substantial spatial variability, and atmospheric relationships identified at a single observational site may differ across locations with distinct geographical characteristics and meteorological conditions. Consequently, the reported performance metrics and

fuzzy-rule interpretations should be restricted to the investigated station and should not be extrapolated to the municipality as a whole or to other geographical regions. Future investigations using multiple meteorological stations are necessary to evaluate the spatial transferability, robustness, and potential station-specific calibration requirements of the proposed ANFIS framework.

5. Conclusions

This study evaluated an ANFIS model with two physically interpretable inputs, average atmospheric pressure and average relative humidity, to forecast precipitation occurrence in the following hour using antecedent 3 h observation windows at INMET station A802, located in Rio Grande, southern Brazil. The evaluation followed a chronological experimental design, with training data from 2002–2018 and test data from 2019–2024, preserving the observed precipitation frequency within the sampled test windows.

ANFIS achieved an AUC-ROC of 0.809 on the 2019–2024 test period, with no statistically significant difference in discrimination detected relative to the tuned SVM-RBF and Random Forest models. Its nine fuzzy rules revealed interpretable associations between precipitation occurrence, higher relative humidity, and lower atmospheric pressure. Furthermore, the model required only 39 adjustable parameters and achieved an inference time below one millisecond per individual observation window. However, it did not outperform a logistic regression with the same inputs and was outperformed by a persistence baseline (AUC 0.843); its added value relative to persistence was restricted to periods without precipitation in the preceding hour (AUC 0.739 versus 0.620), where precision remained low.

Despite its discriminatory capability, the precision of precipitation alerts remained relatively low under the observed class imbalance (0.17 at the selected decision threshold and approximately 0.41 when considering only the 2% of windows with the highest scores). These findings indicate that the two selected surface meteorological variables provide limited information for accurately identifying positive precipitation cases. An evaluation at every hourly timestamp yielded the same discrimination, indicating that the non-overlapping 3 h sampling did not bias these results. The aggregation of antecedent observations also prevents the explicit representation of short-term atmospheric tendencies and sequential dependencies. Therefore, the reported performance is restricted to the observational conditions at station A802.

Overall, the results indicate that a compact ANFIS model can provide an interpretable and computationally efficient approach to short-term precipitation occurrence forecasting at the station scale, although its current predictive limitations restrict immediate operational applicability. Future investigations should consider additional meteorological predictors, including antecedent precipitation and atmospheric tendencies, explicit temporal information, calibration of model outputs, validation across multiple meteorological stations, and a prospective evaluation on observations collected after 2024, which were not available during model development, to assess the potential improvement, temporal robustness and spatial transferability of the proposed approach.

Author Contributions: Conceptualization, J.S.; methodology, A.D.; software, A.D.; validation, J.S., F.B. and J.M.; formal analysis, A.D.; investigation, A.D.; resources, National Institute of Meteorology (INMET), Brazil; data curation, A.D.; writing—original draft preparation, A.D.; writing—review and editing, A.A.L., J.S., F.B. and J.M.; visualization, A.D.; supervision, A.A.L. and J.S.; project administration, J.S.; funding acquisition, Laboratório de Interação Atmosfera–Oceano (Atmosphere-Ocean Interaction Laboratory). All authors have read and agreed to the published version of the manuscript.

Funding: Brazilian National Council for Scientific and Technological Development (CNPq - Process: 308846/2026-5 - DT level C and Project No. 406763/2022-4).

Data Availability Statement: The data used in this study are publicly available from National Institute of Meteorology (INMET), Brazil. The processed dataset generated during the study is available from the corresponding author upon reasonable request.

Acknowledgments: The authors acknowledge the support of the Laboratório de Interação Atmosfera-Oceano (Atmosphere–Ocean Interaction Laboratory) for providing infrastructure and technical support for the development of this study. The authors also thank the Laboratório de Estudos e Pesquisas em Metodologias de Sistemas de Apoio à Decisão (Laboratory for Studies and Research for Decision Support System Methodologies, Labsadi) for guidance and methodological support throughout the development of this research, as well as the National Institute of Meteorology (INMET) for providing access to the meteorological data used in this study. During the preparation of this manuscript, the authors used ChatGPT (GPT-6, OpenAI) and Claude (Claude Opus 5.5, Anthropic) to assist in language refinement, text structuring and drafting of the revised text. The authors reviewed and edited the output and take full responsibility for the content of this publication.:

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANFIS	Adaptive Neuro-fuzzy Inference System
AUC	Area Under the Curve
AUC-PR	Area Under the Precision-Recall Curve
CI	Confidence Interval
FP	False Positive
FN	False Negative
FURG	Federal University of Rio Grande
INMET	National Institute of Meteorology, Brazil
MAE	Mean Absolute Error
MF	Membership Function
MSE	Mean Squared Error
RBF	Radial Basis Function
RF	Random Forest
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
RS	Rio Grande do Sul
SVM	Support Vector Machine
SVM-RBF	Support Vector Machine with Radial Basis Function kernel
TN	True Negative
TP	True Positive
WMO	World Meteorological Organization

Appendix A

Appendix A.1

Table A1. ANFIS model configuration, hyperparameters and performance metrics on the test set.

Category	Description
Model type	ANFIS (Adaptive Neuro-Fuzzy Inference System)
Objective	Binary classification of precipitation occurrence

Input variables	Average atmospheric pressure (x_1) and average relative humidity (x_2), both averages of the 3 h window
Number of inputs	2
Membership function type	Generalized bell, $\mu(x) = \frac{1}{1+\left\|\frac{x-c}{a}\right\|^{2b}}$, with b = 2 fixed
Number of membership functions	3 per input (low, medium, high)
Total number of rules	9
Learning rate (LR)	0.1 (consequent parameters); 0.01 (premise parameters c and a)
Number of epochs	1200
Data split	Chronological: training 1 January 2002–31 December 2018 (balanced, 6218 windows); test 1 January 2019–31 December 2024 (16,326 windows)
Model output type	Continuous favorability score for precipitation (not a calibrated probability)
Decision criterion	Maximum Youden's J on the internal validation subset (last 30% of the training period)
Adopted threshold	0.489
Main evaluation metric	AUC-ROC (with AUC-PR)
AUC-ROC (test)	0.809 (95% CI 0.794–0.823)
Accuracy (test)	66.9%
Precision (test)	0.169
Recall (test)	0.805 (specificity 0.658)
F1-score (test)	0.279
Confusion matrix (test)	TP = 1045/FP = 5143/FN = 253/TN = 9885

Appendix A.2

Table A2. Sensitivity analyses of model configuration and training-window length.

Analysis	Configuration	Cross-Validation AUC	Test/Note
Window length (ANFIS)	3 h (adopted)	0.775 ± 0.031	–
	6 h	0.752 ± 0.034	3 h better in 5 of 5 folds
	12 h	0.711 ± 0.021	3 h better in 5 of 5 folds
	24 h	0.732 ± 0.023	3 h better in 4 of 5 folds
Membership functions per input	2 (4 rules)	0.774 ± 0.018	–
	3 (9 rules, adopted)	0.775 ± 0.016	–
	4 (16 rules)	0.775 ± 0.017	–
Tuned benchmarks	SVM-RBF (C = 0.1, γ = 0.1)	0.775 ± 0.017	0.807
	RF (300 trees, depth 5, 50 samples per leaf)	0.773 ± 0.016	0.807
Training window (ANFIS)	Expanding (adopted)	–	0.792
	Sliding, 10 years	–	0.792 (Δ −0.001)
	Sliding, 5 years	–	0.790 (Δ −0.002)

Note: Antecedent window length and SVM-RBF/RF hyperparameters were evaluated using five chronological folds with natural precipitation frequency in validation. ANFIS membership functions were selected using five-fold cross-validation on previously balanced data. Training-

window strategies were compared separately through retrospective rolling-origin evaluation (2012–2024), reporting mean annual AUC-ROC.

References

1. Doswell, C.A., III; Brooks, H.E.; Maddox, R.A. Flash flood forecasting: An ingredients-based methodology. *Weather Forecast.* **1996**, *11*, 560–581. https://doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2.
2. Parmar, A.; Mistree, K.; Sompura, M. Machine learning techniques for rainfall prediction: A review. In *Proceedings of the International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 17–18 March 2017*; IEEE: New York, NY, USA.
3. de Sousa Araújo, A.; Silva, A.R.; Zárate, L.E. Extreme precipitation prediction based on neural network model: A case study for southeastern Brazil. *J. Hydrol.* **2022**, *606*, 127454. https://doi.org/10.1016/j.jhydrol.2022.127454.
4. da Costa, A.S.; Medeiros, E.S. de; Fragal, V.H.; Silva, A.Q. de; Bicalho, C.C. Algoritmo de machine learning para mapeamento das chuvas extremas em uma região do nordeste brasileiro. *Rev. Bras. Climatol.* **2024**, *34*, 288–305. https://doi.org/10.55761/abclima.v34i20.17378.
5. Hashim, R.; Roy, C.; Motamedi, S.; Shamshirband, S.; Petković, D.; Gocić, M.; Lee, S.C. Selection of meteorological parameters affecting rainfall estimation using neuro-fuzzy computing methodology. *Atmos. Res.* **2016**, *171*, 21–30. https://doi.org/10.1016/j.atmosres.2015.12.002.
6. Dubey, A.D. Comparative analysis of ANFIS and SVR model performance for rainfall prediction. In *Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO 2015)*; Ravi, V., Panigrahi, B.K., Das, S., Suganthan, P.N., Eds.; Advances in Intelligent Systems and Computing; Springer: Cham, Switzerland, 2015; Volume 415, pp. 63–75. https://doi.org/10.1007/978-3-319-27212-2_6.
7. Shamshirband, S.; Gocić, M.; Petković, D.; Saboohi, H.; Herawan, T.; Mat Kiah, M.L.; Akib, S. Soft-computing methodologies for precipitation estimation: A case study. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1353–1358. https://doi.org/10.1109/JSTARS.2014.2364075.
8. Pradip, K.; Kumar, P.; Manojkumar, S. Rainfall forecasting using artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) models. *Int. J. Agric. Sci.* **2018**, *10*, 6153–6159.
9. Hussein, E.; Ghaziasgar, M.; Thron, C.; Vaccari, M.; Jafta, Y. Rainfall Prediction Using Machine Learning Models: Literature Survey. In *Advances in Data Science and Artificial Intelligence*; Springer: Cham, Switzerland, 2022; pp. 75–108. https://doi.org/10.1007/978-3-030-92245-0_4.
10. Pham, B.T.; Le, L.M.; Le, T.-T.; Bui, K.-T.T.; Le, V.M.; Ly, H.-B.; Prakash, I. Development of advanced artificial intelligence models for daily rainfall prediction. *Atmos. Res.* **2020**, *237*, 104845. https://doi.org/10.1016/j.atmosres.2020.104845.
11. Pham, B.T.; Bui, K.-T.T.; Prakash, I.; Ly, H.-B. Hybrid artificial intelligence models based on adaptive neuro fuzzy inference system and metaheuristic optimization algorithms for prediction of daily rainfall. *Phys. Chem. Earth* **2024**, *134*, 103563. https://doi.org/10.1016/j.pce.2024.103563.
12. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X.
13. Xexéo, G. Lógica Fuzzy: Uma Abordagem Prática; LTC: Rio de Janeiro, Brasil, 2003; pp. 1–240.
14. Jang, J.-S.R. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. https://doi.org/10.1109/21.256541.
15. Manandhar, S.; Lee, Y.H.; Meng, Y.S.; Yuan, F.; Ong, J.T. GPS-derived PWV for rainfall nowcasting in tropical region. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4835–4844. https://doi.org/10.1109/TGRS.2018.2839899.
16. Manandhar, S.; Dev, S.; Lee, Y.H.; Meng, Y.S.; Winkler, S. A data-driven approach for accurate rainfall prediction. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9323–9331. https://doi.org/10.1109/TGRS.2019.2926110.
17. Benevides, P.; Catalão, J.; Nico, G. Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors. *Remote Sens.* **2019**, *11*, 966. https://doi.org/10.3390/rs11080966.
18. Silva, T.d.R.; Reis, I.; Klering, E.; Bayer Maier, E. Precipitação em Rio Grande—RS (1913–2016): Análise descritiva e da variabilidade. *Rev. Bras. Geogr. Fís.* **2021**, *14*, 537–553. https://doi.org/10.26848/rbgf.v14.2.p537-553.
19. Britto, F.; Saraiva, J.M. Estudo da precipitação associada aos sistemas frontais na cidade do Rio Grande na estação do inverno. *Geogr. em Atos* **2001**, *1*, 111–123. https://doi.org/10.35416/geoatos.v1i3.7607.
20. Britto, F.; Krusche, N. Frequência e intensidade das frentes frias. In *Proceedings of the Anais do 9º Congresso Brasileiro de Meteorologia, Campos do Jordão, Brasil, 1996*; Sociedade Brasileira de Meteorologia: Rio de Janeiro, Brasil, 1996.

21. Pegoraro, T.; Saraiva, J.M.B. Climatologia dos sistemas precipitantes para o período de fevereiro a dezembro de 2003, no Rio Grande do Sul. In *Proceedings of the Anais do XIII Congresso Brasileiro de Meteorologia, Fortaleza, CE, Brazil, August 2004*; Sociedade Brasileira de Meteorologia, Rio de Janeiro, Brazil, 2004.
22. Krusche, N.; Saraiva, J.M.B.; Reboita, M.S. Normais climatológicas provisórias de 1991 a 2000 para Rio Grande, RS; Departamento de Geociências, Fundação Universidade Federal do Rio Grande: Rio Grande, Brazil, 2002; pp. 8–14.
23. Campello, F.D.; Saraiva, J.M.B.; Krusche, N. Periodicity of atmospheric phenomena occurring in the extreme south of Brazil. *Atmos. Sci. Lett.* **2004**, *5*, 65–76. https://doi.org/10.1016/j.atmoscilet.2003.12.003.
24. Nobre, C.A.; Cavalcanti, I.F.A.; Gan, M.A.; Kousky, V.E. *Climanálise: Boletim de Monitoramento e Análise Climática; Número Especial*; INPE: São José dos Campos, Brasil, 1986.
25. Instituto Nacional de Meteorologia (INMET). Sobre o INMET. Available online: https://portal.inmet.gov.br/sobre (accessed on 2 December 2025).
26. Jang, J.-S.R.; Sun, C.-T.; Mizutani, E. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*; Prentice Hall: Upper Saddle River, NJ, USA, 1997; pp. 1–614.
27. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.
28. Hodson, T.O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.