European Water 83/84: 33-40, 2023.

Synergizing machine learning and hydrological models: Enhancing early warning systems

E. Rozos* and P. Dimitriadis

Institute for Environmental Research & Sustainable Development, National Observatory of Athens, 15236 Athens Greece
* *e-mail: erozos@noa.gr*

Abstract: This study aims to enhance the reliability of early warning signals for potentially catastrophic high flows by integrating a machine learning (ML) model with a conceptual hydrological model. The hydrological model underwent an independent calibration process, including standard stages of calibration and verification. Both the inputs (precipitation and evapotranspiration) and the outputs (simulated discharge) of the hydrological model were utilized as inputs for the ML model, resulting in a significant improvement in the detection of extreme events, such as the exceedance of a predetermined discharge threshold, and more accurate estimation of their likelihood. A novel approach employed in this study involved utilizing different thresholds during the training and test periods, which effectively enhanced the training of the ML model. Achieving a desirable balance between false negatives and false positives is crucial in configuring early warning systems, and the proposed ML model provides essential information to accomplish this. This empowers the modeler to achieve the desired balance and enhance the overall performance of the early warning system.

Key words: Early warning systems; hydrological model; machine learning; logistic regression; classification

1. INTRODUCTION

More than 50 years have passed since the introduction of the multilayer perceptron machine learning model (Minsky and Papert, 1969), while the first applications in hydrology started to emerge more than 25 years ago as rainfall-runoff models. Specifically, Minns and Hall (1996) were among the first to apply recurrent neural networks (RNN) in a hydrological application. These early data-driven models were relatively simple, with at most two hidden layers and up to a dozen hidden nodes, and lagged behind in performance compared to traditional hydrological models. However, with advancements in computational power, more complex machine learning (ML) networks, such as long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997), started appearing in various hydrological applications, outperforming traditional approaches in areas like flood prediction, water resources management, stochastic analysis, etc. (Lin et al., 2020; Xu et al. 2020; Rozos 2019; Rozos et al. 2021). LSTM networks provide nodes with dynamic states, acting as memory, to enhance overall performance. Studies have demonstrated the superior generalization capacity of LSTM models over hydrological models in streamflow simulation and catchment analysis (Lees et al. 2021). While LSTM models surpass conventional models in capturing information from extensive hydrological datasets, the effectiveness of LSTM models comes at the expense of increased computational complexity. Alternative approaches involve using ML as a pre-processing tool for data analysis, or as a post-processing tool for the outputs of a hydrological model. ML has also been used to detect anomalies, estimate probability distributions, quantify uncertainty, and improve peak flow estimation and flood forecasting. In this study, it is showcased that the combined use of a machine learning (ML) model and a hydrological model surpasses the performance of using the hydrological model alone, particularly in detecting extreme events within an early warning system.

Let's consider a well-calibrated hydrological model that has reached its optimal overall performance. For instance, attempting to improve the simulation of peak flows results in a deterioration of the base flow simulation. Let's also assume that it is important to predict the exceedance of a specific discharge threshold at a particular cross-section. One could recalibrate the hydrological model by selecting a suitable objective function that places greater emphasis on errors at high flows. However, creating a specific and different objective function for each case study may not be feasible or easy. On the other hand, employing removable layers or easily changeable add-ons within the operational application of a hydrological model provides greater flexibility compared to the extensive reconstruction of the model. This approach allows for efficient adaptation to multiple purposes without the need for modifying the hydrological model setup, which can be challenging, especially in certain commercial packages with limited customization options.

For this reason, this study proposes processing the results of a hydrological model using an ML model to estimate the likelihood of surpassing a specific flow threshold at a designated control point, which represents a critical cross-section of a river segment or canal. It should be noted that a combination of a hydrological model with an ML model has already been used by other researchers. For example, Senent-Aparicio et al. (2019) used ML models to estimate the instantaneous peak flow from the daily average obtained from the SWAT model. Similarly, Noymanee and Theeramunkong (2019) compared several alternative statistical and ML techniques to improve flood forecasting performance based on MIKE 11 simulations. Also, Rozos et al. (2022) and Rozos (2023) have employed ML in order to assess the performance of hydrological models and the potentials for further performance improvement. The present study differs from the aforementioned in that instead of trying to improve the performance of a hydrological model, we directly focus on improving the reliability of the warning signal. This concept was introduced by Rozos and Dimitriadis (2022) and is further elaborated and tested in this study.

2. MATERIALS AND METHODS

2.1 The hydrological model

The hydrological model employed in this study was the LRHM (Rozos, 2020), which integrates two fundamental structural modelling components, namely the direct runoff and soil moisture models, and utilizes linear regression techniques to simulate observed runoff. This approach is based on the concept of genetic programming models (Herath et al., 2021). The schematic representation of LRHM is shown in Figure 1.

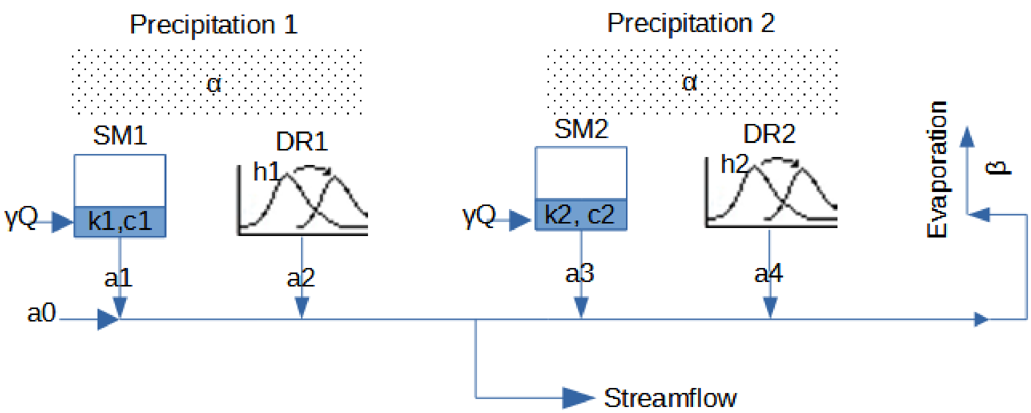

Figure 1. Schematic representation of the hydrological model LRHM.

2.2 The integration with the ML model

Early warning systems send alarm signals that are binary in nature, where a value of 1 indicates an exceedance and 0 represents a non-exceedance event. Therefore, a suitable classification method such as logistic regression (Anderson, 1982) is considered the fundamental choice. Because the inputs of the ML model are time series, i.e., are not independently and identically distributed, a recurrent layer should be employed (Zhang et al., 2021). The topology of the ML network is simple. It employs a hidden layer with a minimal number of nodes (Figure 2) of the LSTM type. The inputs of the ML network include the inputs and outputs of the hydrological model. More specifically, the inputs are the precipitation, the evapotranspiration, and the simulated values by the hydrological model LRHM. The activation function between the layers is the sigmoid function used in logistic regression to estimate the conditional probability of a binary variable taking the values {0, 1} (see Equation 2.5 in Anderson, 1982; Serpa, 2022).

$$S(x) = 1 / (1 + \exp(-x)) \tag{1}$$

For training the ML model, the logistic regression cost function is used (see Equation 4-17 in Géron, 2019), also known as the binary cross-entropy cost function, which is derived using the maximum likelihood estimation method (see Equation 2.6 in Anderson, 1982). If y is the target time series (taking values 0 and 1), p is the time series with the likelihood of y being 1, as estimated by the ML model, and m is the length of the time series, then the cost function is given by:

$$J = -1/m \, \Sigma \, (y \log(p) + (1-y) \log(1-p)) \tag{2}$$

The ML model was implemented using the Cortexsys tool, which runs on MATLAB or GNU Octave. The hydrological model used in this study is LRHM (Rozos, 2020), which takes two inputs: precipitation time series (R) and evapotranspiration (E). The hydrological model produces the simulated discharge, Qh. To prepare the inputs for the ML model, R, E, and Qh were subjected to z-score normalization. These normalized values were then used as inputs for the ML model, as shown in Figure 2.

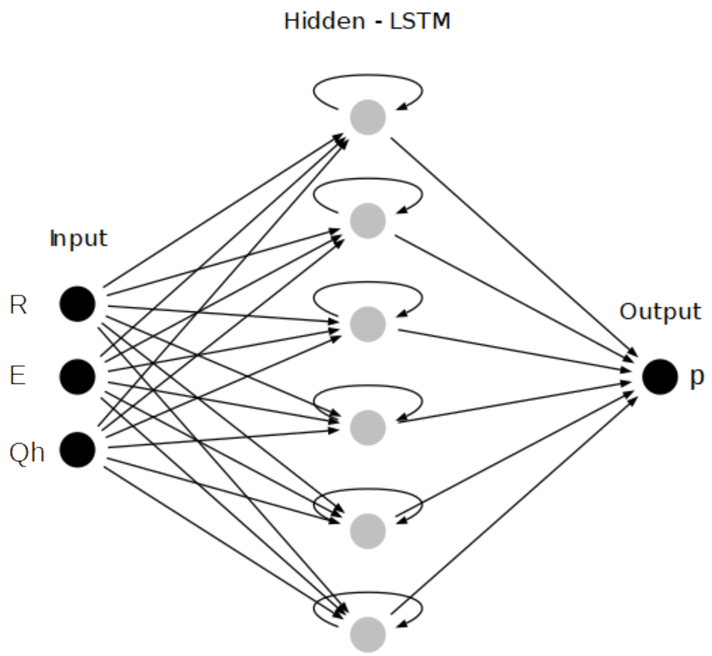

Figure 2. ML network topology.

The output of the ML model, i.e. the p in Equation (2), is a time series with values between 0 and 1, where 1 corresponds to a 100% probability of exceeding the threshold. An empirical rule to be considered is the Rule of 10 (a.k.a one in ten rule), according to which there should be at least 10 events per explanatory variable in the observed time series (Harrel et al., 1984). For the three inputs of the ML network used in this study, this rule implies a minimum of 30 events in the training set.

To determine whether an alarm signal should be issued based on the output *p* of the ML model, the output is compared to a predetermined cut-off value. This value balances the trade-off between true positives (desired) and false negatives (undesired) values. Increasing one of these values often leads to an increase in the other, creating a delicate balance. Selecting the ideal cut-off value becomes a policy decision, as it involves finding the optimal trade-off point. While it is crucial for the model to detect all relevant events, an excessive number of false alarms can lead to desensitization of the response from social mechanisms, rendering the system ineffective in practice.

2.3 Case study – Sieve River

In line with the objective to enhance the reliability of early warning signals for potentially catastrophic high flows, Rozos and Dimitriadis (2022) tested the integration of a conceptual hydrological model with a machine learning (ML) model in Karveliotis River, Greece. In this study, this integration is tested in Sieve River at Fornacina, Tuscany, Italy. The catchment area of Sieve River is 846 km^2. The observed data include the mean areal hourly rainfall, evapotranspiration and discharge at the basin exit. The period of the available data starts on 3 June 1992 and ends on 2 January 1997 (36,554 time steps, with a gap in the data from 1 January 1995 to 2 June 1995). The flow regime of Sieve River is intermittent. The annual rainfall is 1190 mm/year. More information regarding this location (map, description of geomorphological characteristics, source of data) can be found in Koutsoyiannis and Montanari (2022). The observed and simulated discharge of Sieve River is displayed in Figure 3. From this figure, it becomes evident that the hydrological model fails to detect the high flows taking place before January 1997.

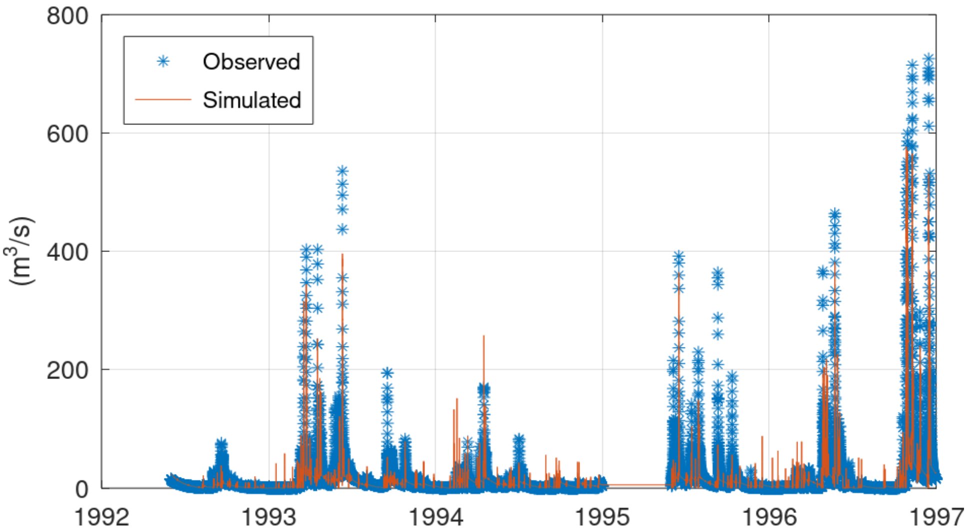

Figure 3. Simulated, with the LRHM hydrological model, and observed discharge of Sieve River.

In the training period (1 June 1992 to 31 December 1995) the ML model was trained to detect exceedances of a discharge threshold equal to 250 m^3/s. This value was selected to ensure an adequate number of events in the training data, following the Rule of 10. Then, the ML model was applied in the test period to detect exceedances of a discharge threshold equal to 580 m^3/s, which is a value dictated by the hydraulic conditions, i.e., the capacity of a critical cross-section.

3. RESULTS AND DISCUSSION

Figure 4 displays the output of the ML model for the events of the training period. The vertical axis corresponds to the likelihood of exceedance, as it is estimated by the ML model. If the estimated likelihood exceeds a cut-off value (0.23 in this case) then a warning signal is issued (red circles on the upper horizontal axis). The yellow cross marks on the upper horizontal axis indicate that the simulated discharge by the hydrological model exceeds the threshold (250 m^3/s for the training period). The blue “×” marks indicate that the observed discharge exceeded the threshold. According to this figure, for the training period, there is one event (close to time step 25000) of which the discharge is underestimated by the hydrological model. The ML model also fails to detect this event. Close to time step 16000 the hydrological model overestimates the discharge. The ML model accurately avoids false positive for this case.

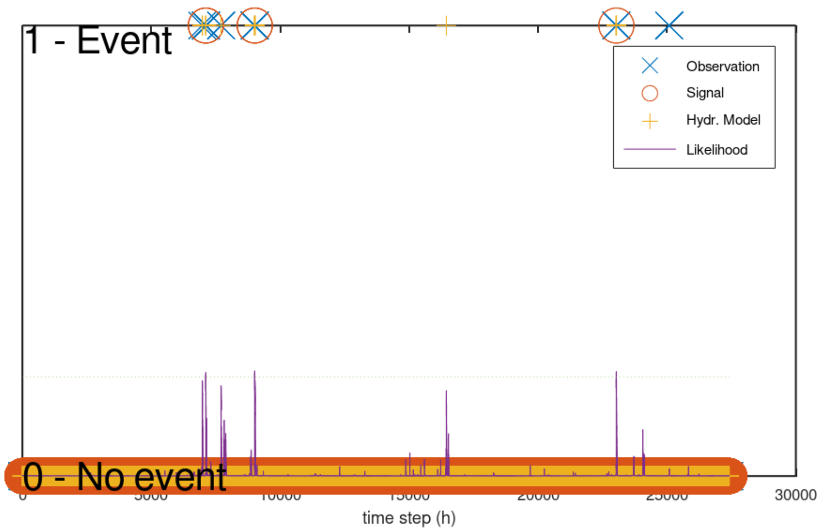

Figure 4. Exceedance of 250 m^3/s during the training period.

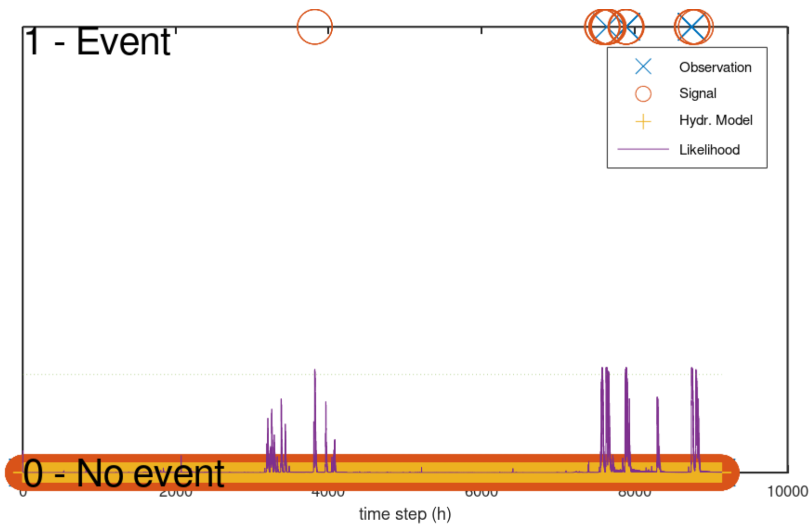

Figure 5. Exceedance of 580 m^3/s during the test period.

Figure 5 displays the events of the test period. In this period, the observed discharge values are not available to the system, only the precipitation and evapotranspiration, which in operational conditions will be provided by a forecast service. Based on this figure, it is evident that a cluster of events occurred towards the end of the test period, all of which exhibit an underestimation of discharge by the hydrological model. However, the ML (which processes the outputs of the hydrological model along with precipitation and evapotranspiration) successfully issues a signal for these events. The ML model gives one false positive just before time step 4000.

The conventional approach in ML training involves employing a single exceedance threshold for both the training and test periods. This threshold is typically determined based on the specifications of the real-world hydraulic system under study. However, this approach may benefit from further exploration and refinement. Consideration of alternative threshold strategies tailored to the training and test periods could lead to improved performance. Indeed, the preceding figures hint at the viability of a more flexible definition of this threshold during the training period. In this study, a lower threshold (250 m³/s) was employed during the training period, which resulted in a sufficient number of exceedances to effectively train the ML model. Subsequently, this trained model successfully projected and transferred the acquired knowledge to the test period, where higher thresholds of 580 m³/s were applied, aligning with the capacity of the hydraulic system. It is important to note that the threshold chosen for the training period should not only satisfy the Rule of 10 but also be as close as possible to the hydraulic system's actual capacity.

As mentioned earlier, the choice of the cut-off value plays a crucial role in balancing the trade-off between false positive and true positive rates. It is important to note that the training process itself is independent of this value (refer to Equation 2). The receiver operating characteristic (ROC) curve, introduced by Spackman (1989), is the typical approach for a comprehensive analysis of the model's performance across different cut-off values, aiding in the decision-making process. This curve is displayed in Figure 6. The horizontal axis is the false positive rate (FPR), which is the ratio of the false positives to the number of real negative cases in the data, and the vertical axis is the true positive rate (TPR), which is the ratio of the true positives to the number of real positive cases in the data.

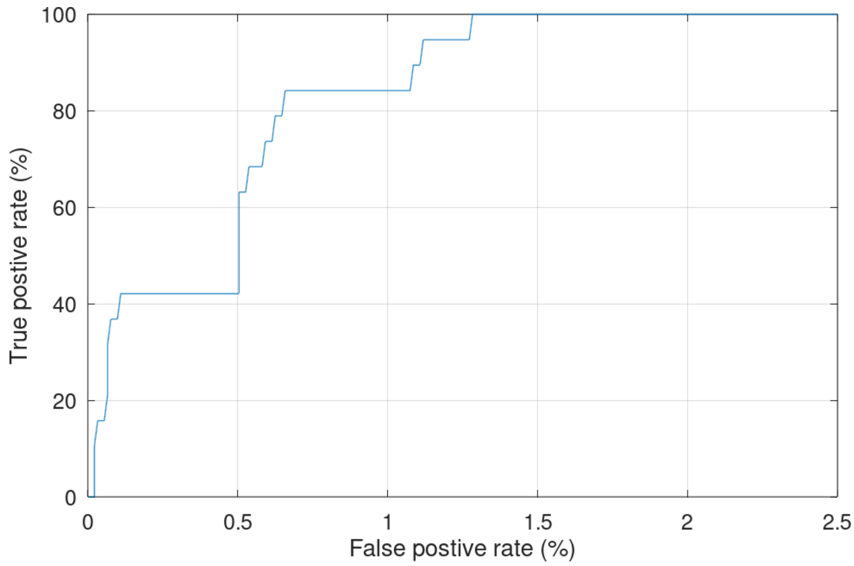

Figure 6. The ROC curve.

According to Figure 6, for this specific case study, the TPR can reach 100% for a 1.2% FPR. These rates can be achieved with a cut-off value equal to 0.15 for the test period (Figure 7). Whereas the maximum F1 score (an evaluation score for classification algorithms that balances between the true positives, false positives and false negatives) is achieved with a cut-off value equal to 0.23, which was used in Figure 4 and Figure 5. This cut-off value results in TPR and FPR values equal to 42% and 0.35% respectively.

The TPR of 42% indicates significant room for improvement in terms of performance. Similarly, the FPR, although relatively low at 0.35%, corresponds to approximately 31 false positive signals per year. However, when evaluating operational reliability, it is crucial to consider the false alarm rate. In Figure 5, which depicts the exceedances during the test period (slightly longer than 1 year), there are 32 false positives (note that multiple circles may appear as a single one due to proximity). Nevertheless, the primary objective is to reliably predict dangerous events rather than precisely forecasting their exact timing and duration. Upon closer examination of Figure 5, the false alarm rate is actually 25%. Therefore, when assessing the performance of an early warning system, it is essential to consider not only the FPR and TPR values but also evaluate them in conjunction with a schematic representation of the exceedances (e.g., Figures 4 and 5).

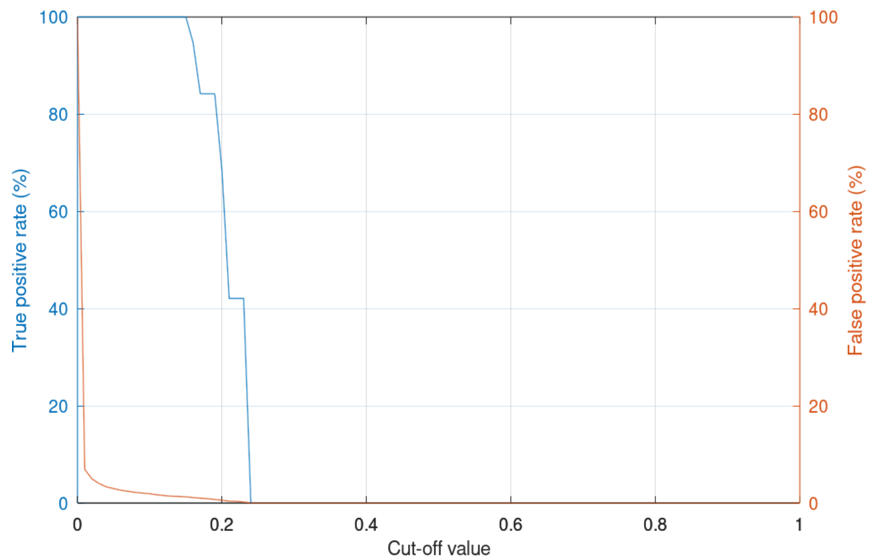

Figure 7. True positive and false positive rates for various cut-off values for the test period.

4. CONCLUSIONS

In this study, a machine learning model was used to process the results of a hydrological model with the aim of extracting alarm signals that can be used in an early warning system. The application of the model in a case study demonstrated that the machine learning model improves the reliability of predictions compared to those obtained directly and solely from the hydrological model. In this particular application, and based on a previous study, this improvement is likely not due to the reduced performance of the hydrological model but rather due to the objective function used in the calibration of the hydrological model that rewards the model's accuracy across all ranges of flows. Therefore, the proposed approach allows for the use of ready-made hydrological models, as configured and calibrated for another application (e.g., water resources management), and their direct implementation in an early warning system.

The discharge threshold that distinguishes between dangerous and trivial events is typically determined based on the specific characteristics of the hydraulic system being studied, such as the capacity of a critical cross-section. However, dangerous events may be very rare, making it challenging to effectively train the machine learning (ML) model. In this study, a lower discharge threshold was employed during the ML training period to ensure an adequate number of events. Subsequently, the ML model was applied in the test period using the discharge threshold that corresponds to the actual dangerous levels. This approach proved successful in training the ML model and enabled accurate risk projections for higher discharge thresholds.

As for future research, the proposed approach should be compared with other methods of early warning systems (e.g., pure machine learning systems or data-driven statistical systems) to identify further weaknesses and comparative advantages. Speaking of performance, during the test period, the ML model exhibited a false positive alarm rate of 25%. Additionally, there was a false negative observed during the training period. It is crucial to recognize that early warning systems alone cannot provide 100% reliability in predicting natural disasters, as there will always be factors that can hinder their effectiveness. Therefore, the primary line of defence should always be the properly designed and implemented civil engineering works, which serve as the cornerstone of disaster prevention and mitigation efforts.

ACKNOWLEDGMENT

An initial, shorter, version of the paper has been presented in the 15th Conference of the HHA, Thessaloniki, Greece, 2–3 June 2022. This research was funded by the Internal Grant/Award of the National Observatory of Athens: Low computational burden flood modelling in small to medium-sized water basins in Greece" – 5080.

REFERENCES

Anderson, J.A. 1982. Classification pattern recognition and reduction of dimensionality. In Handbook of Statistics (Vol. 2); Krishnaiah, P.R. and Kanal, L.N., Ed.; North-Holland Publishing Company: Amsterdam, The Netherlands, 169 – 191.

Géron, A. 2019. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media.

Harrell Jr, F.E., Lee, K.L., Califf, R.M., Pryor, D.B. and Rosati R.A. 1984. Regression modelling strategies for improved prognostic prediction. Statistics in medicine, 3(2), pp.143-152.

Herath, H.M.V.V., Chadalawada, J. and Babovic V. 2021. Genetic programming for hydrological applications: to model or forecast that is the question. Journal of Hydroinformatics, 23, 740–763.

Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.

Koutsoyiannis, D., Montanari, A. 2022. Bluecat: A Local Uncertainty Estimator for Deterministic Simulations and Predictions. Water Resour. Res. 58, e2021WR031215. https://doi.org/10.1029/2021WR031215.

Lin, Q., Leandro, J., Wu, W., Bhola, P. and Disse M. 2020. Prediction of Maximum Flood Inundation Extents With Resilient Backpropagation Neural Network: Case Study of Kulmbach. Frontiers in Earth Science, 8, p.332.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G. and Dadson S.J. 2021. Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. Hydrology and Earth System Sciences, 25(10), pp.5517-5534.

Minns, A.W. and Hall M.J. 1996. Artificial neural networks as rainfall-runoff models. Hydrological sciences journal, 41(3), pp.399-417.

Minsky, M., Papert, S. 1969. Perceptrons: an Introduction to Computational Geometry, Cambridge, Mass.

Noymanee, J., Theeramunkong T. 2019. Flood forecasting with machine learning technique on hydrological modeling. Procedia Computer Science, 156, pp. 377-386.

Rozos, E., Dimitriadis, P. 2022. Early Warning Employing Hydrological And Machine Learning Models, Proceedings of the 15th Conference of the Hellenic Hydrotechnical Association (EYE), pp. 208 – 215, Thessaloniki, June 2022, Greece.

Rozos, E. 2023. Assessing Hydrological Simulations with Machine Learning and Statistical Models. Hydrology, 10(2), p.49.

Rozos, E., Dimitriadis, P. and Bellos V. 2022. Machine Learning in Assessing the Performance of Hydrological Models. Hydrology, 9(1), p.5.

Rozos, E., Dimitriadis, P., Mazi, K. and Koussis A.D. 2021. A Multilayer Perceptron Model for Stochastic Synthesis. Hydrology, 8(2), p.67.

Rozos, E. 2020. A methodology for simple and fast streamflow modelling. Hydrological Sciences Journal, 65(7), pp.1084-1095.

Rozos, E. 2019. Machine learning, urban water resources management and operating policy. Resources, 8(4), p.173.

Senent-Aparicio, J., Jimeno-Sáez, P., Bueno-Crespo, A., Pérez-Sánchez, J. and Pulido-Velázquez D. 2019. Coupling machine-learning techniques with SWAT model for instantaneous peak flow prediction. Biosystems engineering, 177, pp.67-77.

Serpa, Y. 2022. A Comprehensive Guide on Activation Functions, https://towardsdatascience.com/a-comprehensive-guide-on-activation-functions-b45ed37a4fa5 (Πρόσβαση στις 11/1/2022).

Spackman, K.A. 1989. Signal Detection Theory: Valuable Tools For Evaluating Inductive Learning, Editor(s): Alberto Maria Segre, Proceedings of the Sixth International Workshop on Machine Learning, Morgan Kaufmann, pp.160-163,

Xu, W., Jiang, Y., Zhang, X., Li, Y., Zhang, R. and Fu G. 2020. Using long short-term memory networks for river flow prediction. Hydrology Research, 51(6), pp.1358-1376.Yapo, P. O., Gupta, H. V., Sorooshian, S., 1998. Multi – objective global optimisation for hydrologic models. Jurnal of Hydrol.; 204: 83-97.

Zhang, A., Lipton, Z. C., Li, M., Smola, A. J. 2021. Dive into deep learning. arXiv preprint arXiv:2106.11342.