Supplement of Nat. Hazards Earth Syst. Sci., 24, 3537–3559, 2024
https://doi.org/10.5194/nhess-24-3537-2024-supplement

Natural Hazards and Earth System Sciences
Open Access

Supplement of

Transferability of machine-learning-based modeling frameworks across flood events for hindcasting maximum river water depths in coastal watersheds

Maryam Pakdehi et al.

Correspondence to: Ebrahim Ahmadisharaf (eahmadisharaf@eng.famu.fsu.edu; eascesharif@gmail.com)

Supplementary Material: Bayesian search for model optimization

We defined a broad search space (pbounds) encompassing the number of layers, setting a range between 1 and 3 layers with the units varied from 10 to 90, regularization rates (0.01 to 0.2), optimizers (Stochastic Gradient Descent (SGD; Bottou 2012) and Adaptive Moment Estimation (Adam; Singarimbun, Nababan, and Sitompul 2019), and activation functions (Exponential Linear Unit (ELU; Trottier, Giguere, and Chaib-draa 2017), and Rectified Linear Unit (ReLU; Agarap 2019) in hidden layers, facilitating a thorough exploration of model architectures. A linear activation function was used for the output layer. The batch size, determining the number of samples processed before the model updates its parameters, varied between 4 and 16, providing a balance between training speed and memory usage. Lastly, the number of epochs, which dictates the number of complete passes through the training dataset, was explored from 100 to 1000, to find the optimal duration for model training. The optimization process, implemented via the Bayesian search framework, systematically evaluated combinations of hyperparameters across the defined space. It began with two initial random evaluations (init_points=2) of hyperparameter sets, followed by three guided evaluations (n_iter=3). Thus, a total of five unique hyperparameter sets were assessed.

Utilizing cross-validation, the dataset was divided into three subsets or 'folds. For each iteration of the optimization process, a different fold was held out as the validation set, while the remaining folds were used for training the model. For each set, we applied 5-fold cross-validation (cv=5), resulting in each set being evaluated five separate times, one for each fold. Consequently, there were 5×5=25 individual model trainings during the optimization process. This approach ensures that each data point contributes to both the training and validation phases, enhancing the reliability of the performance assessment. The Bayesian search process with a cross-validation strategy culminated in identifying an optimal set of hyperparameters that significantly enhanced the model predictive performance. The optimized configuration comprised a specific arrangement of number of layers, units, epochs, batch size, a precise regularization rate, and an optimal combination of optimizer and activation function, tailored to maximize the accuracy of estimations of maximum flood depth.

The optimal hyperparameters identified through the Bayesian optimization method included one hidden layer with 47 units, 636 epochs, a batch size of 8, a regularization rate of approximately 0.07, the SGD optimizer, and the ELU activation function. However, after manually adjusting the number of units to 50 and the regularization rate to 0.104, we achieved the best performance. Additionally, we implemented early stopping, a technique designed to halt the training process when model performance no longer improves on the training and test datasets, further enhancing our ANN-MLP model.

References

1. Bottou, L. Stochastic Gradient Descent Tricks. in *Neural Networks: Tricks of the Trade* (eds. Montavon, G., Orr, G. B. & Müller, K.-R.) vol. 7700 421–436 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012).
2. Singarimbun, R. N., Nababan, E. B. & Sitompul, O. S. Adaptive Moment Estimation To Minimize Square Error In Backpropagation Algorithm. in *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)* 1–7 (IEEE, Medan, Indonesia, 2019). doi:10.1109/ICoSNIKOM48755.2019.9111563.
3. Trottier, L., Giguere, P. & Chaib-draa, B. Parametric Exponential Linear Unit for Deep Convolutional Neural Networks. in *2017 16th IEEE International Conference on Machine*

Learning and Applications (ICMLA) 207–214 (IEEE, Cancun, Mexico, 2017). doi:10.1109/ICMLA.2017.00038.

4. Agarap, A. F. Deep Learning using Rectified Linear Units (ReLU). Preprint at http://arxiv.org/abs/1803.08375 (2019).