Drinking Water Engineering and Science Discussions

Drink. Water Eng. Sci. Discuss., author comment AC1
https://doi.org/10.5194/dwes-2021-7-AC1, 2021

Reply on RC1

Ahmad Ravanbakhsh et al.

Author comment on "Implementing and evaluating various machine learning models for pipe burst prediction" by Ahmad Ravanbakhsh et al., Drink. Water Eng. Sci. Discuss., https://doi.org/10.5194/dwes-2021-7-AC1, 2021

1-There is very little novelty in this paper, which is essentially a poorly done machine learning exercise that would should be rather featured in a blog on Kaggle, rather than in a scientific journal; known techniques, different dataset, no advancing of the state-of-the-art.

Response:

The focus of this paper is to evaluate the performance of RCNN-SVR method on water distribution networks. During the studies conducted by the authors, the use of RCNN-SVR was observed in some engineering sciences, but the ability of this method in estimating the failure rate of water network pipes (in real case study) has not been measured in any scientific article. Therefore, the present study shows the high accuracy of this innovative method. Researchers can use this method as an efficient method in their researches. On the other hand, there are several articles in high-credited scientific journals that have compared outdated conventional machine learning methods. In this paper the new method that has much higher accuracy and efficiency is applied.

2-1 Most of the paper is about the description of existing techniques, which have been described several times in the literature. The authors employ the RCNN-SVR technique but they fail to cite the correct paper [1] both in the references (Line 375, a paper from 2016) and in the text (Line 148, a paper from 2018).

Response:

Description of existing techniques makes some of the newest and most important machine learning methods available for readers. Comparing RCNN-SVR with these known methods has also been performed to demonstrate the strength of the RCNN-SVR method. However, if in the opinion of the honorable referee there is no need to explain the methods, these methods can be removed from the paper and only shown in result tables in future reviews. The referencing will be corrected in future review too.

2-2 They lack to provide the required details to understand the model, e.g., are those 1D convolutional layers? How many trainable parameters total?

Response:

Due to the limited number of words in the article, the full description of the RCNN-SVR method was omitted, which accords to the referee viewpoint, the description of other methods can be removed and the RCNN-SVR method will be explained in full detail.

3-1 The latter information is particularly important because Deep Learning models usually have hundreds of thousand of parameters, and the proposed dataset is made of barely 200 data points. Deep Learning is employed when BIG DATA is available, this is clearly not the case. This raises the issue of whether the models have been trained appropriately or not.

Response:

An outstanding feature of the RCNN-SVR method is the accurate learning with limited data because other machine learning methods require a big dataset for training. Due to spending 8 years to collect failure data of water pipes in the case study network, little information is available, so the RCNN-SVR method has been selected which provides a more accurate answer with low data. Obviously, the RCNN-SVR method is more efficient if there is a lot of input data.

3-2 In line 52, the authors claim that 85% of data have been selected for training and the rest of them have been used to test the models. Was the training data further split to create a separate validation dataset? Or is the test set used for "validation" and model selection? Is this a fair comparison?

Response:

The training data is not split to create a separate validation dataset. The test set is used for validation. 85% of data have been selected for training and the rest of them have been used to test the models. All the machine learning methods that described in this paper have been compared in the same way and a fair comparison has been made. Also the same has been done in other scientific articles [1], [2].

3-3 Given the data is tabular, not an image or a time-series, I am confident tree-based algorithms such as Random Forest, Extra Trees or Gradient Boosting trained using a single line of Python would provide at least the same results.

Response:

Although it is possible to use a single line of Python for tree-based algorithms but the results are not as accurate as RCNN-SVR. In this paper (line 86), a tree-based algorithms (m5') is used which has a more inaccurate answer than the RCNN-SVR method. The provided codes can also be cited.

3-4 Why did the authors use outdated benchmarks for comparison?

Response:

Criteria used to compare machine learning methods are commonly used in many new scientific papers and can be cited. However, the use of other criteria can be used according to the suggestion of the honorable referee.

3-5 Most importantly, the authors failed to provide the details needed to understand how they developed and compared the models.

Response:

RCNN-SVR model development has been done in other engineering sciences but its use in water networks is innovative. The purpose of this article is to implement and compare some machine learning methods and developing of models was not considered. In this paper, an innovative new method is used to provide accurate results with low dataset which has not been used to predict the failure rate of pipes in any researches.

4- The paper lacks organization and structure. Experimental setup (e.g., train/test division) is presented in the Methodology, so are the final trained models. The dataset is cited in the Methodology but introduced in a subsequent section. The results and discussion section is minimal, and there is no discussion at all.

Response:

The structure will be correct in next revision.

5- The literature review is outdated and incomplete, see for instance the papers [2-4] reported at the end of this review.

Response:

Suggested articles will use in next revision. Thanks for introducing useful articles.

6- The paper is difficult to read even for an expert in the field.

Response:

Due to the difficulty of the content, an attempt has been made to express the content in the most appropriate way. However, it will be expressed more simply in the next revision.

[1] Shirzad, Akbar, and Mir Jafar Sadegh Safari. "Pipe failure rate prediction in water distribution networks using multivariate adaptive regression splines and random forest techniques." *Urban Water Journal* 16.9 (2019): 653-661.

[2] Shirzad, Akbar, Massoud Tabesh, and Raziyeh Farmani. "A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks." *KSCE Journal of Civil Engineering* 18.4 (2014): 941-948.