Open Access Article

Prediction of Chlorophyll-*a* Concentrations in the Nakdong River Using Machine Learning Methods

by Yuna Shin [1,†], Taekgeun Kim [2,†], Seoksu Hong [2,*], Seulbi Lee [2], EunJi Lee [2], SeungWoo Hong [3], ChangSik Lee [3], TaeYeon Kim [3], Man Sik Park [4], Jungsu Park [5] and Tae-Young Heo [2,*]

[1] Department of Water Environment Research, National Institute of Environmental Research, Incheon 22689, Korea

[2] Department of Information & Statistics, Chungbuk National University, Chungbuk 28644, Korea

[3] Intelligent Network Research Section, Electronics and Telecommunications Research Institute, 218 Gajeong-ro, Yuseong-gu, Daejeon 34129, Korea

[4] Department of Statistics, Sungshin Women's University, Seoul 02844, Korea

[5] Department of Civil and Environmental Engineering, Hanbat National University, Daejeon 34158, Korea

[*] Authors to whom correspondence should be addressed.

[†] Both authors contributed equally to this manuscript.

Water **2020**, *12*(6), 1822; https://doi.org/10.3390/w12061822

Received: 18 May 2020 / Revised: 17 June 2020 / Accepted: 23 June 2020 / Published: 25 June 2020

Models: SVR, Bagging, XGBoost, Random Forest, RNN, LSTM

Target: Chl-a Concentration at Dasa weir along Nakdong river, Korea using observations from 6 Weirs upstream.

Data: Collected from 7 weirs.

Daily data from June 2015 to November 2017 (871 days) for training
Daily data from December 2017 for validation (31 days).

1% of measurements replaced by KNN

Weather data: avg temp, sunshine, rainfall, inflow, outflow
Water quality data: water temp., pH, EC, DO, TOC

Matrices:

Table 3. Prediction performance of the studied machine learning models.

Method		1–Step Ahead Recursive Prediction					
		Cumulative Learning			Rolling Window Learning		
		MAPE (%)	RMSE	NSE	MAPE (%)	RMSE	NSE
All variables	SVR	11.02	3.864	0.5308	11.12	3.8950	0.5233
	Bagging	28.58	10.41	−2.4049	33.88	11.6356	−3.2545
	RF	16.53	5.760	−0.0426	16.23	5.5398	0.0356
	XGBoost	8.69	3.5854	0.5960	09.16	3.6702	0.5767
	RNN	11.28	4.2872	0.1765	15.88	5.4196	−0.1164
	LSTM	16.38	5.8634	0.2136	14.15	5.2502	−0.1653
Selected variables based on forward selection	SVR	9.85	3.1717	0.6838	9.85	3.1750	0.6832
	Bagging	16.33	6.3604	−0.2712	13.37	4.9668	0.2247
	RF	8.50	3.1213	0.6939	8.98	3.1920	0.6798
	XGBoost	10.21	3.9305	0.5145	9.53	3.7712	0.5531
	RNN	7.54	2.6843	0.7601	7.27	2.6453	0.7516
	LSTM	14.40	4.6984	0.3783	17.25	5.7119	0.1077

Major conclusions:

RNN better than ML models