IMPROVEMENT OF RIVER FLOW ESTIMATION ACCURACY USING ENSEMBLE LEARNING STACKING

KUMAMOTO UNIVERSITY

DAIJU SAKAGUCHI

TABLE OF CONTENTS

INTRODUCTION

▪It is expected that floods and droughts will become more and more frequent in various places due to the effects of global warming.

▪ The development of highly accurate rainfall-runoff modeling technology is required.

▪In recent years, machine learning, especially a method called deep learning, has been attracting attention in various fields, and many applications have begun to be made in the field of hydrology

▪By introducing ensemble learning in deep learning, estimation accuracy has been improved in various fields.

▪ By using ensemble learning, it is expected that more accurate results can be obtained compared to the estimation accuracy by a single learner.

▪This study will improve the estimation accuracy of rainfall-runoff modeling using stacking, and XGBoost in the second layer.

STUDY METHOD (ENSEMBLE LEARNING)

・Ensemble learning is a method that uses multiple learning devices and integrates their prediction and estimation results by a specific method.

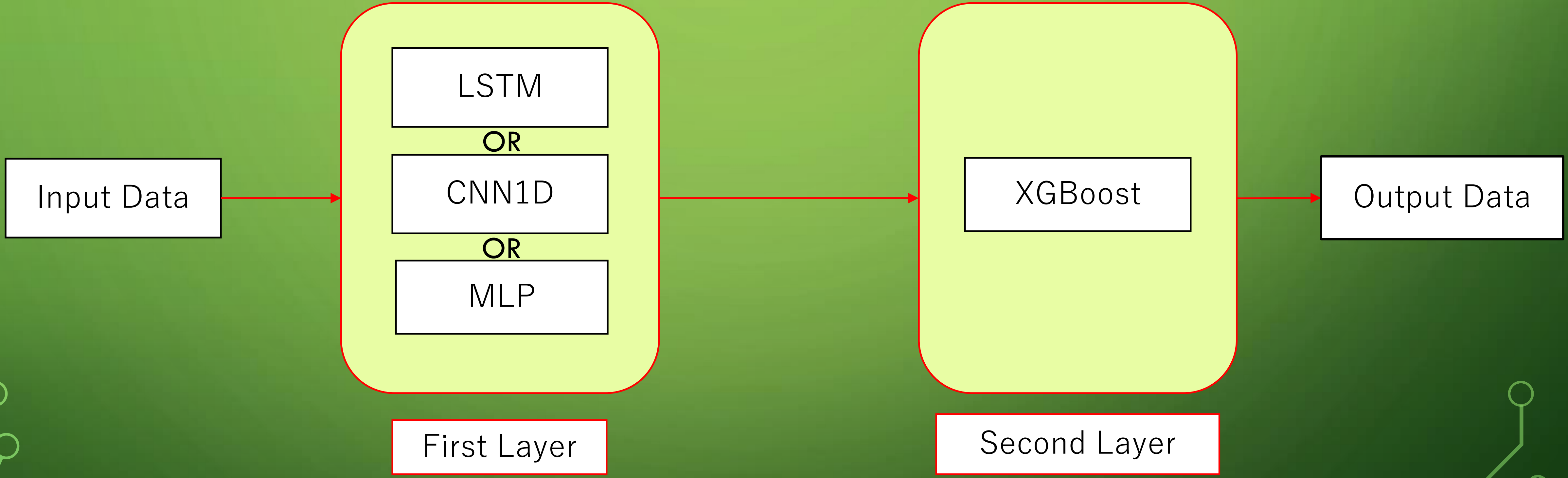

Fig1. Stacking method in this study

STUDY METHOD (LSTM)

- LSTM are mainly used for time series prediction and natural language processing.

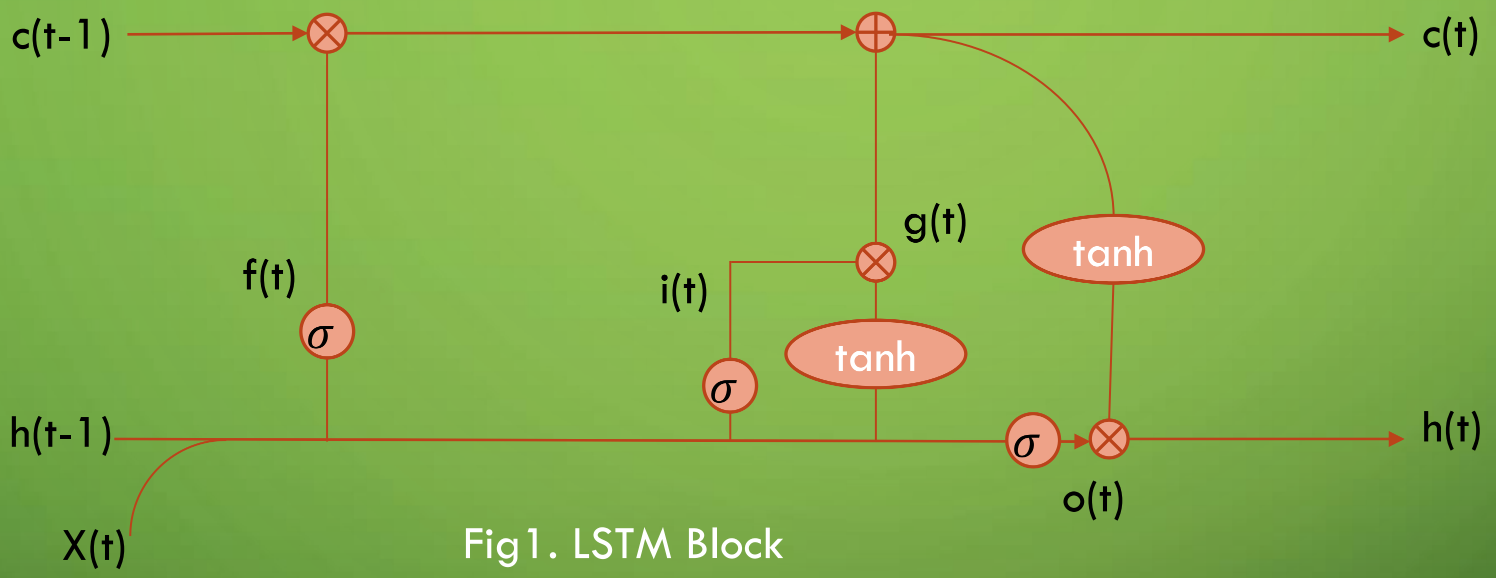

Fig1. LSTM Block

- $\boldsymbol{i}(t) = \sigma(W_{ii}\boldsymbol{x}(t) + \boldsymbol{b}_{ii} + W_{hi}\boldsymbol{h}(t-1) + \boldsymbol{b}_{hi})$
- $\boldsymbol{f}(t) = \sigma\left(W_{if}\boldsymbol{x}(t) + \boldsymbol{b}_{if} + W_{hf}\boldsymbol{h}(t-1) + \boldsymbol{b}_{hf}\right)$
- $\boldsymbol{g}(t) = \tanh\left(W_{ig}\boldsymbol{x}(t) + \boldsymbol{b}_{ig} + W_{hg}\boldsymbol{h}(t-1) + \boldsymbol{b}_{hg}\right)$
- $\boldsymbol{o}(t) = \sigma(W_{io}\boldsymbol{x}(t) + \boldsymbol{b}_{io} + W_{ho}\boldsymbol{h}(t-1) + \boldsymbol{b}_{ho})$
- $\boldsymbol{c}(t) = \boldsymbol{f}(\mathrm{t}) * \boldsymbol{c}(t-1) + \boldsymbol{i}(\mathrm{t}) * \boldsymbol{g}(t)$
- $\boldsymbol{h}(t) = \boldsymbol{o}_t * \tanh\left(\boldsymbol{c}(t)\right)$

- $\boldsymbol{i}(t)$: Input gate
- $\boldsymbol{f}(t)$: Forget gate
- $\boldsymbol{g}(t)$: Cell gate
- $\boldsymbol{o}(t)$: Output gate
- $\boldsymbol{c}(t)$: Cell state
- $\boldsymbol{h}(t)$: Hidden state

STUDY METHOD (CNN)

- CNN is a kind of neural network and is mainly used for image recognition.

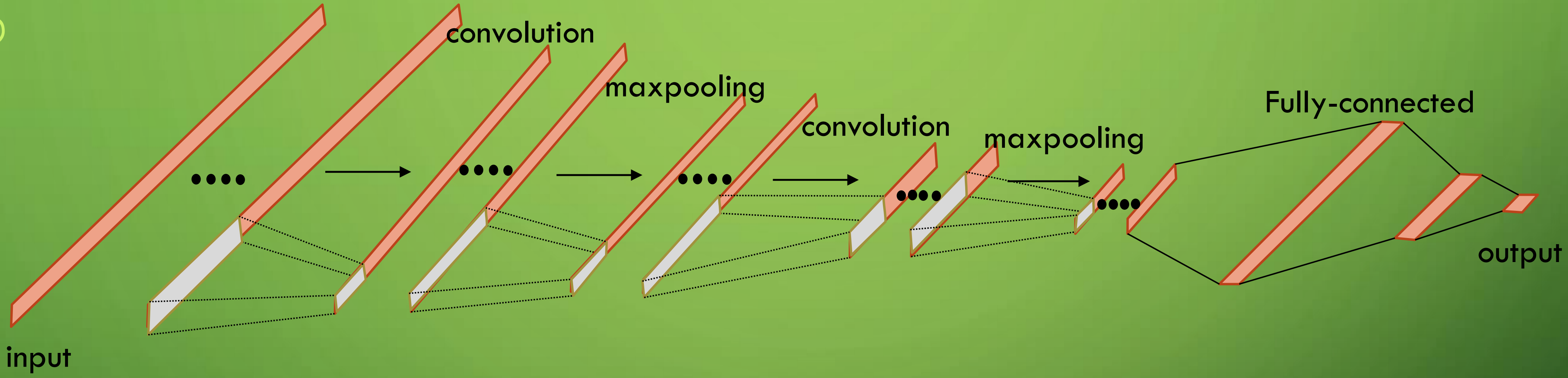

Fig3. Schematic diagram of CNN

- $a_i^{(k)} = \sum_{m=1}^{M} \sum_{n=0}^{N-1} w_n^{(k)} x_{(i+n)} + b^{(k)}$

- $a_i^{(k)}$: Feature map
- $w_s^{(k)}$:Kernel
- x:Input data
- $b^{(k)}$:Bias

STUDY METHOD（MLP）

- MLP is often used to compare deep learning.

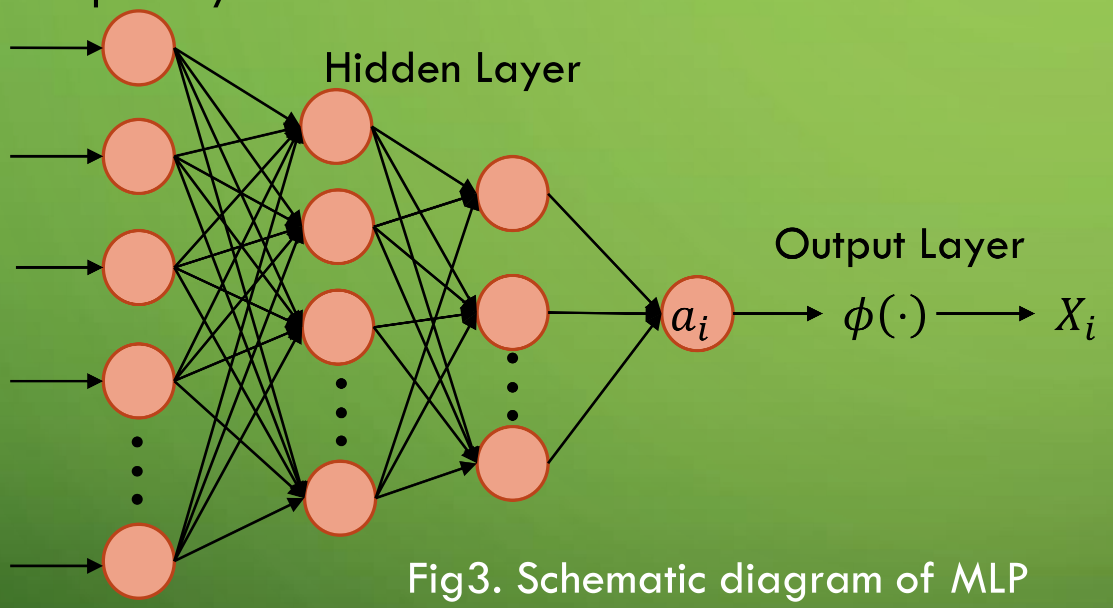

Fig3. Schematic diagram of MLP

- $a_i^{(l)} = \sum_k W_{ik}^{(l)} x_i^{(l)} + b_i^{(l)}$
- $x_k^{(i+1)} = \phi^{(l)}\left(a_i^{(l)}\right) = \phi^{(l)}\left(\sum_k W_{ik}^{(l)} x_k^{(l)} + b_i^{(l)}\right)$

- $a_i^{(l)}$ ：Output
- $W_{ik}^{(l)}$ ：Weight
- $x_i^{(l)}$ ：Output in the front layer
- $b_i^{(l)}$ ：Bias

STUDY AREA

- Target area: Tedori River basin, Ishikawa Prefecture
- Basin area: 809 km²
- Main channel length:72km
- Ishikawa Prefecture's longest river

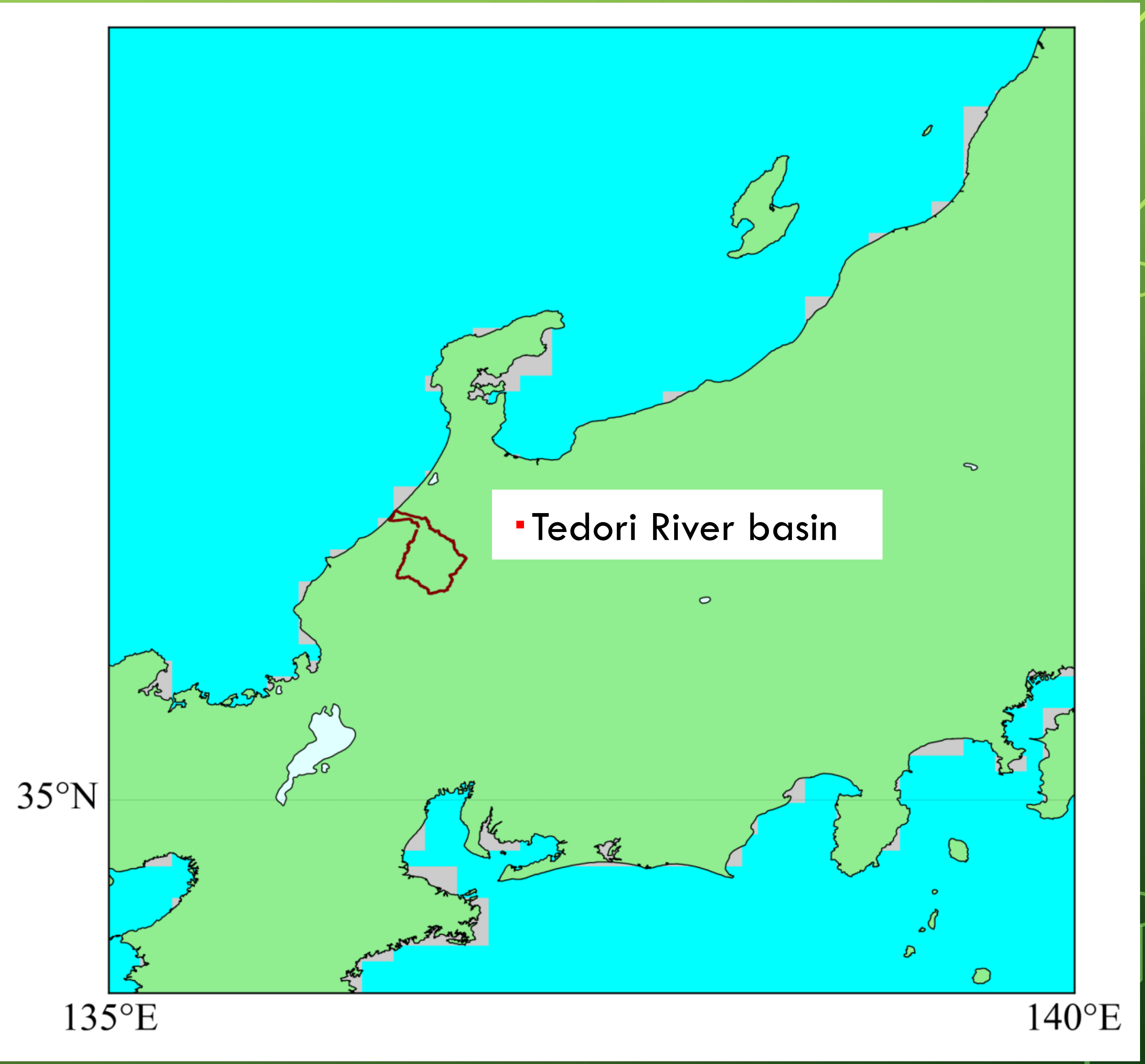

Fig4. Study area

DATASET(USE DATA)

Table1. Usage data and acquisition destination

Input data	Acquisition destination
Daily average precipitation	APHRODITE (Asian Precipitation - Highly Resolved Observational Data Integration Towards Evaluation of Water Resources)
Daily average temperature	ERA5 （European Centre for Medium-Range Weather Forecasts（ECMWF） Reanalysis v5 ）
Target data	**Acquisition destination**
Daily average flow rate(Tsururai point)	Ministry of Land, Infrastructure, Transport and Tourism Hydrological Water Quality Database （WIS）

･APHRODITE is grid data with a horizontal resolution of about 4 km based on observation data, and the data from 1900 to 2015 have been released.

・ ERA5 has a horizontal resolution of 0.25 ° x 0.25 ° and a time resolution of 1 hour is open to the public. The period of ERA5 is from 1979 to the present.

・ WIS has released the flow data of every hour from 2002 to 2019 at Tsurugi observation station. For the flow rate data, the daily average flow rate was calculated from the hourly values. However, if there was a defect, it was excluded from learning and calculation of statistics.

DATASET(DEEP LEARNING)

Table2. Target period

Target period（2002～2015year）	
Training period	2002～2009
Validation period	2010～2012
Test period	2013～2015

Evaluation index

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(Q_{obs}^{i} - Q_{sim}^{i})}$$

Table3. Parameters common to deep learning methods

Hyperparameter	
Hidden Satate Length（HSL）	10,25,50,75,100
Input data length（IDL）	10,50,100180,365
Batch size	256
Early Stopping Patience	30
Loss function	MSE
Optimizer	Adam
Batch size selection method	Shuffle

Table4. Parameters of deep learning method

	Hyperparameter	Condition
LSTM	Hidden Satate Length（HSL）	10,25,50,75,100
CNN	Number of Layers（NLY）	3,4,5
MLP	Hidden Node Size（HNS）	8,16,32

DATASET

Table5. Combination of hyperparameter

	Combination	
	IDL	HSL
LSTM	180	10
		25
		50
		75
		100
	365	10
		25
		50
		75
		100
	IDL	NLY
CNN	180	3
		4
		5
	365	3
		4
		5
	IDL	HNS
MLP	180	8
		16
		32
	365	8
		16
		32

Learn 100 times with each combinationA total of
2200 learning sessions

1000case

600case

600case

Input Data

Estimate daily average flow rate with XGBoost

RESULT（DEEP LEARNING）

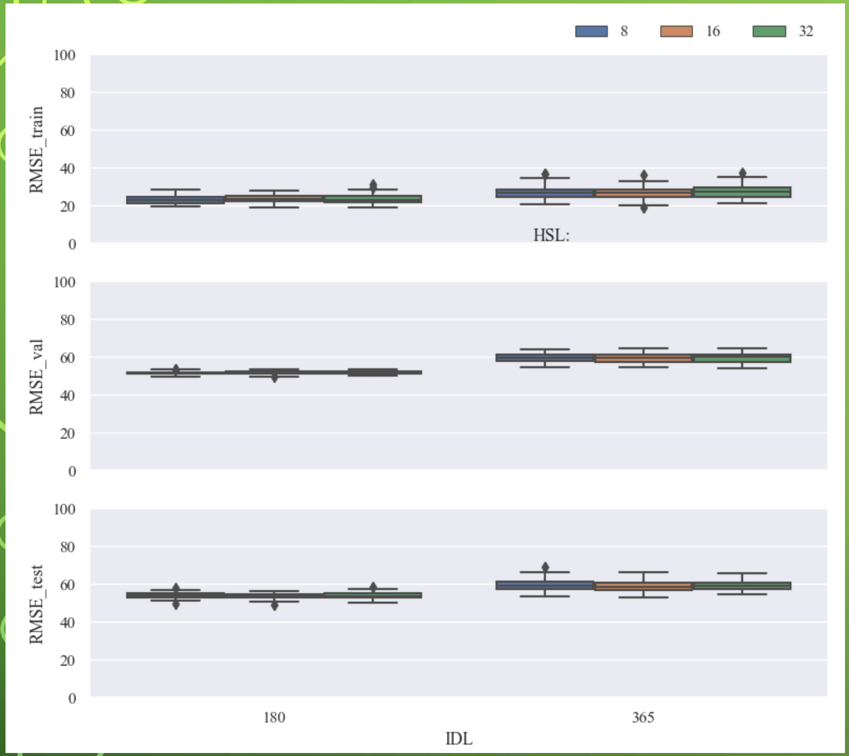

Fig5. MLP_boxplot

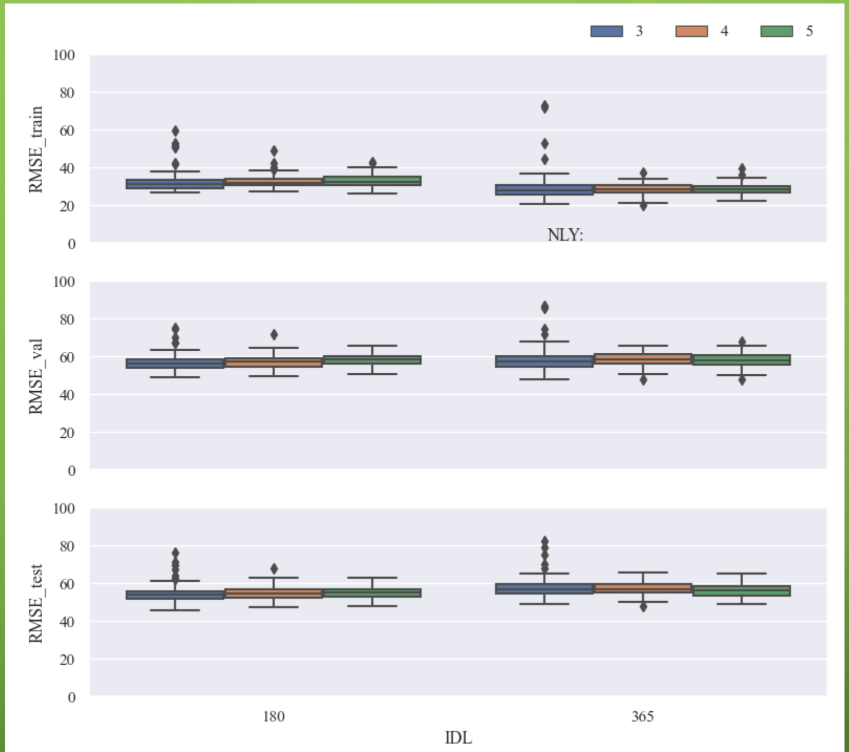

Fig6. CNN_boxplot

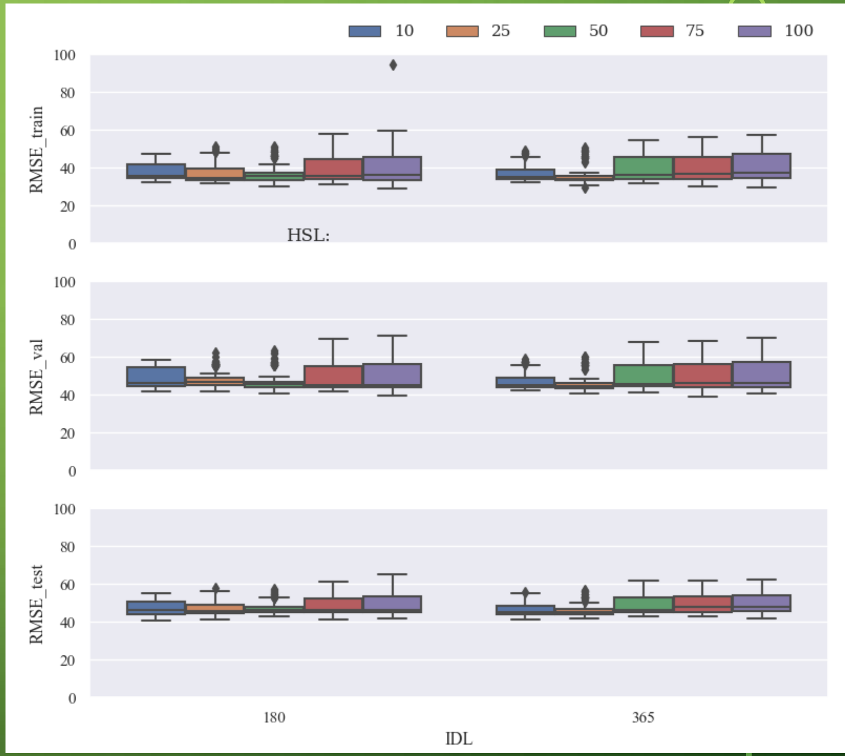

Fig7. LSTM_boxplot

RESULT（DEEP LEARNING）

Table6. LSTM result

IDL	HSL	Training period Min	Training period Median	Validation period Min	Validation period Median	Test period Min	Test period Median
180	10	32.16	35.74	41.41	46.22	40.44	45.85
	25	31.53	34.56	41.63	46.43	41.25	45.62
	50	30.10	35.30	40.34	45.43	42.66	46.35
	75	30.92	35.62	41.49	44.89	41.24	45.87
	100	28.84	36.23	39.51	45.17	41.77	46.28
365	10	32.40	34.78	42.47	45.21	41.00	44.92
	25	29.35	33.74	40.31	44.53	41.55	44.90
	50	31.51	35.85	41.27	45.51	42.54	46.27
	75	30.13	36.93	38.66	46.32	42.52	47.66
	100	29.67	37.21	40.79	45.90	41.52	47.55

Table7. CNN result

IDL	NLY	Training period Min	Training period Median	Validation period Max	Validation period Median	Test period Min	Test period Median
180	3	26.55	30.93	56.27	48.83	53.79	45.83
	4	27.29	31.85	49.20	56.98	47.47	57.69
	5	26.01	32.39	50.52	58.39	47.88	55.04
365	3	20.81	27.87	47.51	58.39	49.09	56.91
	4	20.22	28.45	47.80	58.24	48.02	56.80
	5	22.06	28.30	47.99	57.86	48.88	56.14

Table8. MLP result

IDL	HNS	Training period Min	Training period Median	Validation period Min	Validation period Median	Test period Min	Test period Median
180	8	19.72	23.04	49.65	51.36	49.62	53.70
	16	19.02	23.22	49.30	51.64	48.93	53.70
	32	18.63	22.88	49.77	51.62	50.08	53.48
365	8	20.52	26.87	54.54	59.56	53.17	59.15
	16	19.03	26.40	54.25	59.52	52.52	58.39
	32	21.34	27.18	54.00	59.88	54.71	58.89

RESULT（ENSEMBLE LEARNING）

Table9. ensemble learning result

	RMSE Training	RMSE Validation	RMSE Test
LSTM	24.346	39.011	42.8
CNN	3.462	48.52	50.144
MLP	7.737	50.804	52.577
ALL	3.653	47.217	49.579

Table10. The result of each Deep learning in first layer

	Training RMSE_MIN	Validation RMSE_MIN	Test RMSE_MIN
LSTM	28.84~32.40	38.66~42.47	40.44~42.66
CNN	20.22~27.29	47.51~56.27	47.47~53.79
MLP	18.63~21.34	49.30~54.54	48.93~54.71

CONCLUSION

- There was no major improvement, but there was a possibility that it could be improved.
- The lower the RMSE in the first layer, the lower the NSE during the test period

- This result indicates that the selection of weak learners is important.
- Deep learning methods with higher accuracy are more suitable as weak learners in this study