

# Smart Dam: Upstream Sensing, Hydro-Blockchain, and Flood Feature Extractions for Dam Inflow Prediction

Takato Yasuno<sup>1</sup>, Akira Ishii<sup>1</sup>, Masazumi Amakata<sup>1</sup>, Junichiro Fujii<sup>1</sup>

<sup>1</sup> Research Institute for Infra. Paradigm Shift, 5-20-8 Asakusabashi Taito-ku Tokyo, Japan  
Email: {tk-yasuno, akri-ishii, amakata, jn-fujii}@yachiyo-eng.co.jp

**Abstract.** Heavy rain occurs frequently during extreme weather, and the associated resultant flood damage represents a social problem. The present study aims to redefine the dam watershed as a smart dam and attempts to systematize the technology for flood prediction by integrating upstream sensing, dam inflow prediction, and a hydro-blockchain. In order to detect high water levels, we devise an upstream sensing method to observe the water level at the uppermost stream of a dam watershed, and summarize potential implementation hurdles. We also propose a hydro-blockchain configuration that provides a basis for the fair transaction of water rights. We implement a field study in the Kanto region, Japan, to observe the upstream water level using the devised water level sensor. We analyze the relationship between the measured water levels and the dam inflow and also analyze the response of the time difference. Furthermore, we propose a flood-feature extraction using a 20 year hydro-dataset of rainfall and water levels, and propose dam inflow prediction models using various time series machine learning algorithms. We demonstrate the application of our model results and discuss their usefulness. (182 words).

**Keywords:** River Sensing, Hydrologic Blockchain, Flood Feature Extraction, Time Series Machine Learning.

## 1 Introduction

### 1.1 Background and Related Work

**Floods in a Changing Climate and Hydrological Modeling.** In recent years, extreme flooding has caused massive damage worldwide [1, 2]. In Japan, sudden rainfall and extreme river flooding have occurred in regions such as Joso City, Ibaraki Prefecture [3]. Owing to such extreme and sudden rainfall events, it is difficult to predict the inflow to a dam and to set the operational dam outflow. In flood hydrology, empirical models such as regression models and artificial neural networks (ANNs) have gained popularity [4, 5]. There are several approaches for water resource time series prediction modeling that use statistical and machine learning methods. For example, ANNs have been applied to monthly reservoir inflow time series [6, 7] using data that included 500 monthly streamflow measurement over a period of 40 years as well as climate and land cover data, whereby rainfall-runoff modeling simulates the streamflow of watersheds.

There are currently more than 70 hydrologic models that have varying degrees of data requirements and can be used for numerous applications, for example, estimation of flood runoff and assessment of inundation. An exhaustive review of hydrological models was published in 2002 [8]. The temporal and spatial derivatives of hydrological processes that occur in nature are important in modeling, and models are classified as either distributed (spatial) or lumped (non-spatial). Commonly used temporal scales used in hydrological models are daily or monthly scales [9, 10, 11]. In Japanese watersheds, the distance between upstream and downstream locations is usually relatively short; thus, the time scale should be on the scale of hours or even minutes for flood forecasting in Japan. Hydrological models are physically based and can perform simulations for the estimation of water quantity. However, model accuracy is insufficient for recent flood forecasting due to the extreme nature of phenomena related to climate change.

**Water Smart City and Smart Dam.** Today, smart urbanization is part of thousands of urban projects around the world. A decade ago, the promise of improving and optimizing urban services through the application of information and communication technology (ICT) was largely a techno-utopian fantasy [12]. Many examples of empirical studies of smart cities from 2013 to 2019 are detailed elsewhere [12]. Rob et al. [13] outlined the typology of a smart city for a region of Dublin, Ireland, which included entrepreneurship innovation, green energy, sustainability, resilience, e-government, open data, transparency, accountability, evidence-informed decision making, better service delivery, quality of life, safety, security, risk management, intelligent transport, inclusiveness, and empowerment.

Other studies have focused on water resources, for example, Karwot et al. reported the problems of water distribution and usage as well as problems relating to the management of the technical infrastructure in Poland [14]. The authors provided some ideas and concepts associated with the use of modern tools for supporting water management and the creation of end-user participation for improving the water distribution system [15]. More recent concepts and projects include those in Australia (water-supply city, water-cycle city and water-sensitive city), Korea (water-efficient infrastructure), and in Denmark and the Netherlands (water-smart cities).

However, methods for implementing and realizing water-smart cities remain unclear. Dams have key roles in water-smart cities because they create reservoirs, and enable the optimization of reservoir operation to prevent floods and also generate hydropower. Despite these roles, there is a need for practical methods for the development of a smarter dam for a sustainable basin and the efficient use of water resources (i.e., for agriculture, industry, everyday life, and hydropower).

## 1.2 Study Objectives

The present study aims to redefine the dam watershed as a smart dam and attempts to organize the technology for flood forecasting through the development of key requirements, for example, upstream sensing, dam inflow prediction, and hydro-blockchain. Specifically, in order to detect flood signs, we propose an upstream sensing method to observe the river water level in the uppermost stream of a dam watershed and implement this in a watershed in the Kanto region, Japan. We summarize five main

hurdles for implementation the upstream sensor based on lessons learned in our study installation. In addition, we propose a hydro-blockchain structure that is critical for the fair, conflict-free transaction of water rights among stakeholders. We report on a field study in the selected region whereby the upstream water level was measured by the devised sensor. We aim to analyze the co-relations between the midstream water level and the dam inflow, as well as the response of the time difference between upstream and downstream locations. Furthermore, we propose i) a flood-feature extraction using a 20 year hydro-dataset of rainfall and water levels from the study area, and ii) dam inflow prediction models via time series machine learning and deep learning algorithms with the aim of applying the models to the study area to demonstrate their usefulness.

## **2 Upstream River Sensing and Hydro-Blockchain**

### **2.1 Advantage of Upstream River Sensing**

The flood control of a dam is based on operational rules and is carried out according to the amount of flood inflow into a reservoir. Dams have a role as the most upstream point of flood countermeasure because they allow the lowering of the downstream river water level before shifting to disaster prevention operations during a dam flood. Dams can therefore secure sufficient time for the safe evacuation of residents. However, in recent years, localized and concentrated heavy rains in Japan, which are understood to be related to climate change, have in some cases caused a larger than anticipated inflow into the dam reservoir. Therefore, it is necessary to prepare countermeasures against flooding on the downstream side of the case study dam in the event that the usual flood control operation cannot be handled because of localized and concentrated heavy rains, or in the event that the dam itself is included in flood protection plans.

To achieve such countermeasures, it is important to accurately predict the inflow to the case study dam. However, the upstream area of the dam is mountainous, and the spatial arrangement of the river sensors is comparable to an urban area. Existing sensing systems are insufficient for ensuring the accuracy of predicting the inflow to the dam. Thus, if the water level change in the upstream area of the dam can be detected by a large number of sensors, data would be available for flood forecasting that can provide an indication of the imminence of a flood. Such a system and the generated data would be useful for dam operations and flood countermeasures in the study region and potential also in other regions.

### **2.2 Hurdles for Upstream River Sensing**

Remote sensing using artificial satellites is one method for acquiring information for the upstream area of a dam; however, it is not suitable for predicting inflow to a dam when observation frequency and real-time performance is required. Upstream river sensing is therefore the only effective way to overcome this limitation. In order to install river sensors in the upstream area of a dam, the following five main hurdles may need to be addressed.

**Sensing at an Effective Point for Flood Forecasting.** In steep mountainous areas such as in the present study, the path of the river flow is likely to change due to river bed evolution and sedimentation that occurs during a flood. Since it is impractical and ineffective to monitor all outflows (e.g., from small valleys and swamps), it is preferable to select points with abundant flow rates (i.e., main rivers) that are located downstream of tributaries. However, because it can be difficult to access rivers in mountainous areas, and the water level needs to be measured vertically from directly above, the location for installing water level sensors can be limited (e.g., to bridges) and not necessarily optimal. In some situations, an angle sensor, which can measurement the water level from an oblique angle without restricting the position of the water level sensor, may also be required, and is being development in other research not reported here.

**Construction of a Communication Environment.** A communication environment is necessary for data because flood forecasting requires real-time performance. However, mobile communication radio waves are not generally reliable in mountainous areas. Use of satellite communications, installation of mobile base stations, and networking of low-power wide-area networks (LPWAN) are also unrealistic because of the cost involved. Since digital convenience radios (DCR) depend on geographical features and communication distance their effectiveness can be unknown. In order to reliably transmit water level data, it can therefore be necessary to install a communication line to the location where a water level sensor is installed.

**Securing Power.** The power requirements of a water level sensor and the communication equipment means that solar power generation and batteries are insufficient; thus, it can be necessary to install electrical wires to the sensor location to ensure reliable and continuous measurement throughout the year.

**Installations in Limited Space.** The prop and the box for storing communication equipment must be installed in a limited flat area in steep mountainous areas. Therefore, a quick and simple foundation such as the ground screw method is required.

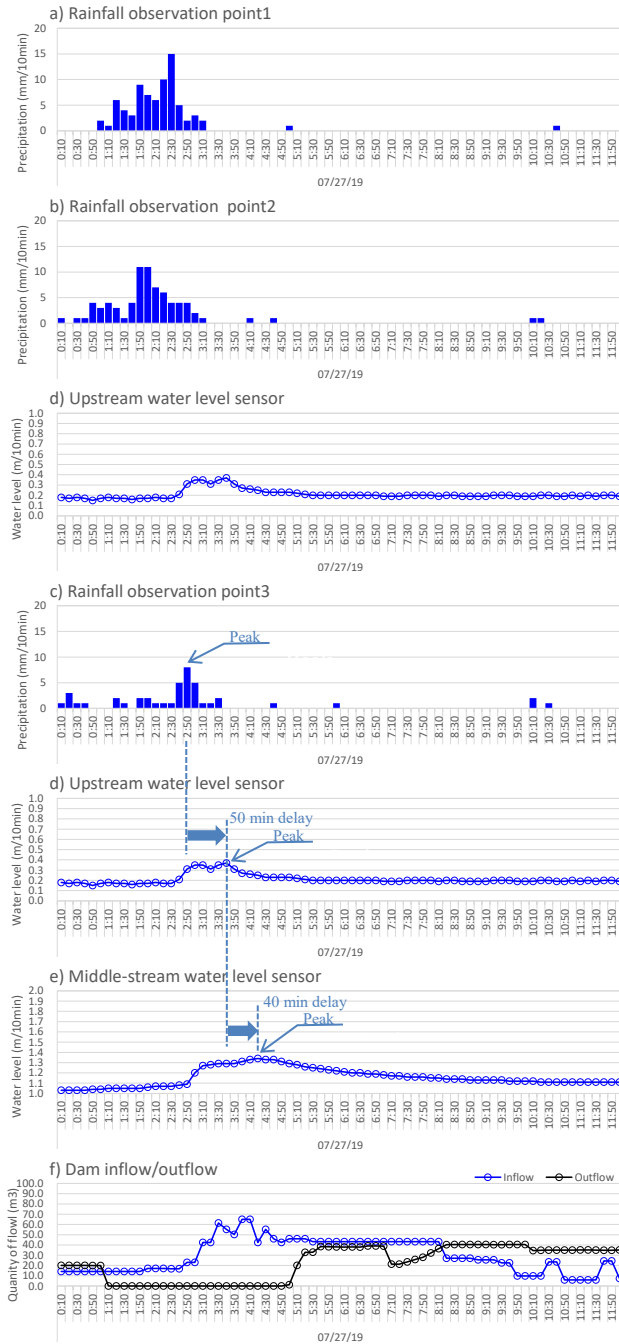
**Application for Installation Permission.** In the present study, the water level sensors were installed on public land, and occupancy permission based on Japanese road and river laws was required for each installation location. If the location of a water level sensor is in a designated national park or conservation area, an application based on relevant laws (i.e., natural park law, forest law) may also be required.

### 2.3 Field Case Study

Figure 1 shows the sensor arrangement in the upstream area of the dam area in the present study. The catchment area of the in Kanto region is  $\sim 100 \text{ km}^2$ . Previous to the present study, there was only one water level sensor in this basin, but on July 10, 2019, an up-stream water level sensor point was installed. On July 27, 2019, there was a flood due to Typhoon No. 6 that was measured by the sensors between 00:00 to 12:00 (Fig.

2). Figure 2 shows that the peak difference between rainfall observation point 3 and the upstream water level sensor point was 50 minutes, and that the peak difference between the upstream water level sensor point and the midstream water level sensor point was 40 minutes. Consequently, we were able to capture a flood signal 40 minutes compared to before by installing the upstream water level sensor, which highlights the effectiveness of the sensor method as a high water countermeasure. With respect to the peak inflow to the dam, we presumed that inflow from the basin where rainfall observation point one is located was dominant since this experienced higher rainfall (Fig. 2).

The confirmation of flood propagation (i.e., the flood arrival time) from the uppermost stream to the middle stream suggests that flood forecast signal information can provide an indication of the imminence of a flood. Continued monitoring by the sensors could therefore provide useful data for the future that might be used for studying the relationship between the upstream and midstream water levels in order to set flood thresholds at these locations. To fully understand local flood phenomena, in the future we intend to optimize the balance, position, and number of water level sensors. When forecasting flood events ahead of the propagation time, we note that it is necessary to consider the weather forecast.

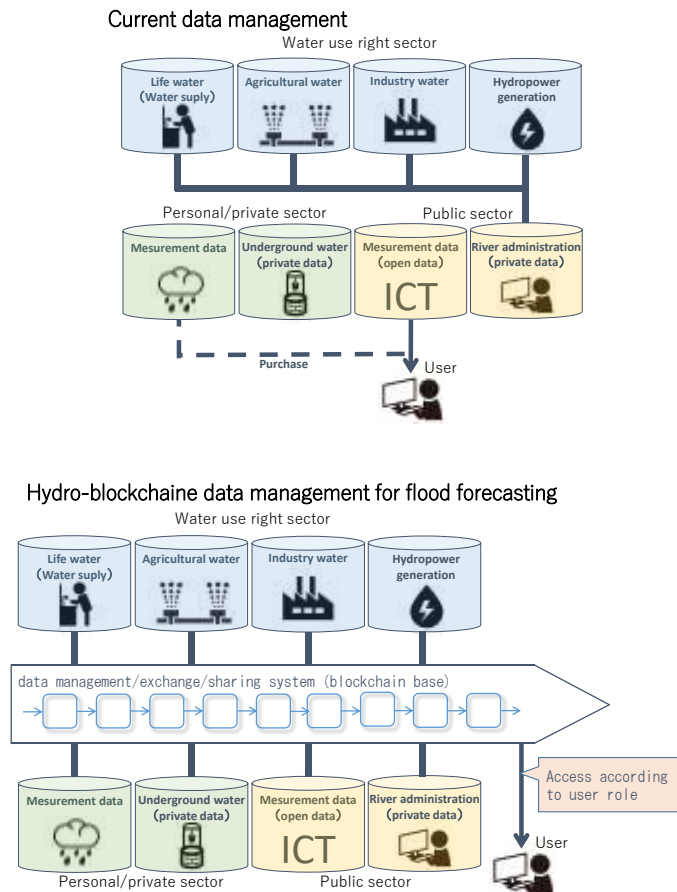


**Fig. 1.** Rainfall at observation points a) one, b) two, and c) three in the study area, Kanto region, Japan. River water levels at the d) upstream, and e) middle stream sensors. f) Dam inflow/outflow data (Typhoon No. 6, July 27, 2019).

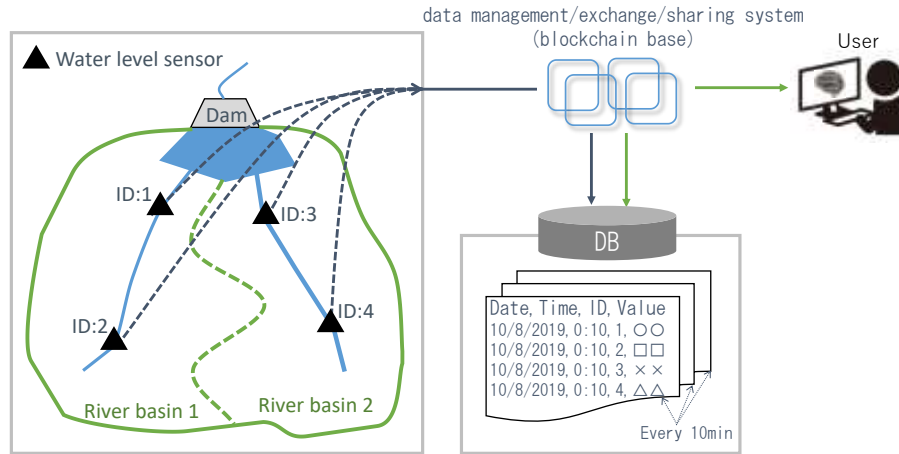
## 2.4 Hydro-Blockchain for Flood Forecasting

In order to improve the accuracy of dam inflow predictions, it is necessary to use various data such as rainfall, water level, artificial water inflow, and reservoir operation. However, various players and stakeholders are involved in a dam, and data is generally distributed and managed by each player's own system. Data is not necessarily shared, and may incur a cost due to restrictions on the use of private and sensitive data. In order to change this situation, it is necessary to build a system that uses data fairly and securely.

Here, we propose a future system based on blockchain technology [16, 17] termed "hydro-blockchain for flood forecasting" (Fig. 3), which can be developed to aggregate the registration and use of large spatiotemporal hydrological datasets. This system enables data management, exchange, and sharing under the robust security of blockchain technology.



**Fig. 2.** Schematic of the hydro-blockchain data management/exchange/sharing system for flood forecasting (Top : Current data management, Bottom : Hydro-blockchain data management).



**Fig. 3.** Schematic of the data registration process from the water level sensors to the database (DB) and data usage via the hydro-blockchain.

Figure 4 shows a schematic of the data registration process from the water level sensors to the database and data usage via the hydro-blockchain. Data is identified by a sensor's ID and recorded in the database via the blockchain every 10 minutes. The data user is given data usage authority according to their role. As data is exchanged through encrypted transactions, it is possible to use data seamlessly and fairly while maintaining privacy and data protection. This system helps to prevent the dispersion and disappearance of data that are useful for learning various flood characteristic, and also improves the accuracy of dam inflow predictions. In addition, by using the smart contract function, information sharing, including dam inflow prediction results, enables i) a quick and smart coordination between conflicting stake holders, and ii) the smooth and timely dissemination of disaster prevention information. Importantly, the system can not only operate in just one basin but also can be integrated to operate in multiple basins.

### 3 Time Series Prediction for Flood Forecasting

#### 3.1 Related Works on Machine Learning Weather Data

**Hydrologic Machine and Deep Learning Modeling.** During the mid-2010s, many machine learning algorithms were applied to water level forecasting problems, for example, Gaussian generalized additive models (GAMs), multivariate additive regression splines (MARSs), random forest (RF), support vector regression (SVR), genetic programming (GEP), and Gaussian process regression (GPR). Several models have been compared by others, including comparisons between Gaussian linear regression, GAMs, MARSs, ANN models, random forests (RFs), and regression tree models [18]. This comparison suggested that GAMs and RF can effectively capture

some non-linear relationships. A study undertaken by Li et al. [19] compared several water level forecasting models (i.e., RF, support vector regression, ANNs, and a linear model) using daily lake water levels from five lake gauges over 50 years. The results suggested that the RF provided a more reliable and accurate prediction of lake water levels [19].

A variable selection approach can improve the forecasting efficiency of daily reservoir discharge by machine learning methods. Yang et al. [20] used 2854 daily reservoir levels covering a period of eight years to test five methods: ANN, instance-based classifier, k-nearest-neighbor classifier, RF, and the random tree. Key variables that influenced the reservoir water level were selected and improved models were developed. The experimental results indicated that the RF forecasting model, when applied to a variable selection with full variables, performed better than the other models [20]. Vamsi proposed the use of GPR for forecasting weather conditions [21], but used a narrow dataset of just two years. The parameters that were considered for the weather prediction included precipitation, wind speed, and temperature. The main advantage of GPR is that it is easily understood. Adiya used a hybrid approach by combining discriminatively trained predictive model GPR with a deep neural network that modeled the joint statistics of a set of weather-related variables [22]. However, the authors hybrid modeling only used five years of data, which consisted of balloon observations at 60 locations. Unfortunately, the applicability of GPR is unclear for flood forecasting to predict the inflow to a dam based on a long-term dataset.

In the late 2010s, deep learning approaches such as long short-term memory (LSTM) were applied to flood forecast datasets. Zhang et al. [23] compared the simulation performance of a recurrent neural network, LSTM, and a gated recurrent unit. The authors showed that the three models could accurately predict reservoir inflow; however, they only collected data every four or six hours over four years. Xuan-Hien et al. [24] suggested a LSTM model for flood forecasting that used the daily discharge and rainfall as input data. The authors used datasets from 1961 to 1984 (24 years) for the Da River basin, Vietnam, to forecast flow rates over one, two, and three days. However, the historical dataset was daily and many under-prediction error (UPE) points remained in their forecast, which would lead to a delay in reservoir operations in the event of a flood hazard. Unfortunately, the usefulness of LSTM is unclear for recent flood forecasting for the prediction of inflow to a dam. Hence, many researchers have undertaken studies on pre-processing, combined modeling, and evaluation index in order to improve modeling accuracy. However, knowledge on time series machine learning modeling for flood forecasting remains incomplete.

In the following Sections 3.2, 3.3, and 3.4, we propose three methods, respectively: 1) flood feature extractions as pre-processing, 2) time series regression modeling using machine learning algorithms towards a hydro-dataset, and 3) computing the sum of an under-prediction index for evaluating flood hazard. We also demonstrate several applied results for the study area.

### **3.2 Flood Feature Extraction**

**Trend Feature Extractions Using Weighted Least Squares.** Existing time series analysis of hydrological data relies on various filtering procedures for pre-processing,

for example, standardized, auto-regressive, moving average, and weighted moving average. However, these procedures do not focus on the trend and peak of a hydrograph, which could represent key features of a flood's scale and continuous high water. Here, we highlight how to extract features associated to the hydrograph peak and trend change, which can in turn be used for supporting decision making regarding announcing a flood alert.

Firstly, we propose the trend estimates of coefficients from the historical dam inflow dataset for the study area. For example, the past 3 hours before the current time step, which has a batch dataset containing 12 inflow time steps and 15 minute units. This trend coefficients can be estimated every time step using the weighted least squares (abbreviated as WLS) algorithm. The older an observation is, the smaller the WLS becomes, which infers a weak memory. The WLS trend feature can also be used to represent the shift-up change among past time steps, and can predict the kurtosis change around the peak of a flood hydrograph.

**Volume Extractions Using Log-Trapezoids Integration.** Secondly, we propose the trapezoids integration as the hydro-stock volume from the past time steps. If the trend features are extracted in a limited way it may result in the hydro-volume features being mistaken as the flood scale and the continuous high water. Hydro-volume features can be computed at every time step using the trapezoids integration from the historical dam inflow values, for example, for the past three hours with 12-time steps and 15-minute units. These values of trapezoids integration are of a higher numerical order than of those in the hydro-dataset (e.g., rainfall, river level, and historical dam inflow). Thus, we propose the logarithm of the trapezoids integration and transform them into standardized variables whose original values are subtracted by the mean and divided by the standard deviation based on the integration series from the past time steps. In Section 3.3, this hydro-volume feature extraction is termed the log of trapezoids integration (abbreviated as LTI).

### 3.3 Dam Inflow Prediction Method for Flood Forecasting

**Regression Tree and Ensemble Algorithms.** We employed a hydrological time series as a baseline model using a regression tree [25]. Although a regression tree model can provide a fast computation, it offers a limited representation of flood features. We also simulated the hydrological dataset using the ensemble models of bagging and least square-base boosting (abbreviated as LSBoost) [26, 27, 28], which involve optimizing hyper parameters based on the cross-validation function.

**GPR Algorithm.** GPR is a nonparametric Bayesian learning model that can be placed directly over the space or nonlinear function instead of specifying a parametric family of nonlinear functions [29, 30]. The covariance matrix is a kernel function (e.g., linear, exponential, squared exponential, Gaussian, Uhlenbeck, and quadratic functions). An exponential kernel function is more accurate than other kernels for dam inflow time series. The hyper parameter sigma is optimized using the cross-validation function.

**LSTM Regression Algorithm.** In the case of a simple recurrent neural network, long-term dependencies are impossible to learn because the network of vanishing gradient problem has deep layers. The theoretical reasons for this effect were studied by Hochreiter and Schmidhuber in the 1990s [31]. The LSTM is a deep learning model for representing a long-term dependence between inflow time steps, and is designed to solve the vanishing gradient problem [31, 32]. One block of the LSTM framework provides a good performance for a lumped dam region. Before the final regression layer, it is possible to insert a dropout layer at a scale of 0.3 in order to avoid a limited pattern of training data every mini-batch.

### 3.4 Flood Risk Index UPE

**Under-Prediction Index for Evaluating Flood Hazard.** Table 1 shows two types of dam inflow prediction error. Minimization of the UPE (shown in the upper right-hand side of Table 1) is important for minimizing the risk of flood damage [33, 34, 35].

**Sum of UPE for the Flood Risk Index.** The UPE at time  $t$  is indicated by Eq. 1:

$$UPE_t = \hat{y}_t - y_t \quad (1)$$

where  $\hat{y}_t$  is the dam inflow prediction, and  $y_t$  is the actual inflow value. If UPE is negative then  $\hat{y}_t < y_t$ , which appears to be an optimistic forecast at time  $t$ . The sum of under-prediction errors (sUPE) among a term  $t \in \{1, \dots, T\}$  is formulated using Eq. 2:

$$sUPE = \sum_{t=1}^T \min(\hat{y}_t - y_t, 0) \quad (2)$$

A negative value can be computed as the predicted case when there are only inflows that are lower than the actual inflow. If under-predictions occur frequently, i.e. if the absolute value of  $UPE_t$  is due largely to these negative values, then the mitigation policy may be insufficient for the risk of an over-flow scenario. These errors would correspond to the risk of flood damages.

**Table 1.** Two types of dam inflow prediction errors.

		True, actual value	
		Low water	High water
Prediction	Low water	True Low water	<b>Under-prediction Error (Flood Damage Risk)</b>
	High water	<b>Over-prediction Error (Hydropower Opportunity Loss )</b>	True High water

### 3.5 Applied Results

**Hydro-Dataset for Flood Forecasting.** The case study dam is a 32 m high concrete gravity dam (PG) with the purpose of hydropower generation (electrical capacity of 279 MWh/year). The reservoir capacity is 1.5 million m<sup>3</sup>, and the dam catchment area is 112 km<sup>2</sup>. There are three river height sensors are installed. It takes 90 minutes for river water to flow from sensor one to the dam, 60 minutes from sensor two to the dam, and 20 minutes from sensor three to the dam.

Table 2 lists the hydro-dataset names and variable profiles. The model is set to a 3-hour forward forecast. Each of the window sizes is set to 12 (i.e. the historical data of the past three hours), and the unit time intervals are 15 minutes.

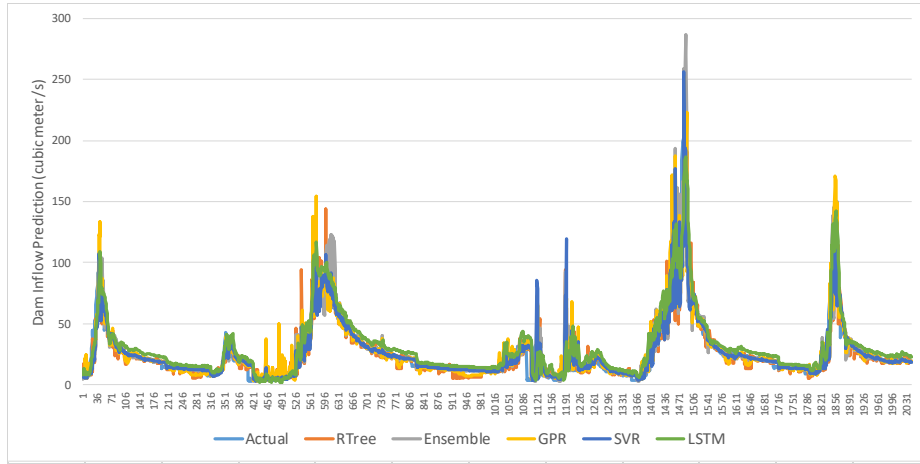
**Table 2.** Hydro-dataset names and variable profiles.

Data name	Data role and feature profile
<i>Dam inflow time series</i>	<i>Target variable. The time series of the dam inflow volume from the upstream region with a measurement interval of 15 minutes.</i>
<i>Rain gauge</i>	<i>The quantity of rainfall is measured at nine points in the upstream region.</i>
<i>River height sensor</i>	<i>Main stream and river features. The water height is sensed at three points and includes an upstream sensor.</i>

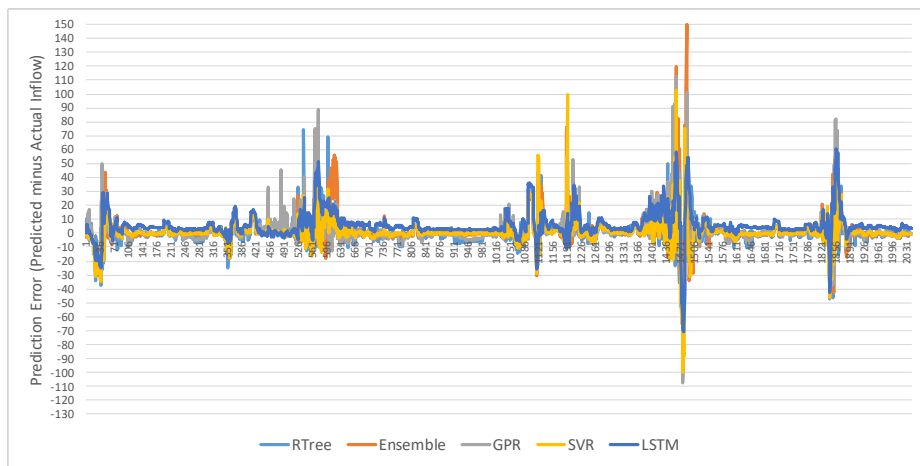
**Applied Result One: Use of the Hydro-Dataset.** Table 3 shows root-mean-squared error (RMSE) and runtime results for five models: regression tree, ensemble, GPR, SVR, and LSTM that were developed using a standard hydro-dataset. This is the benchmark scenario for comparison to scenarios with additions. Figure 5 shows dam inflow predictions results using a hydro-dataset, where these are applied by methods such as regression tree, ensemble Gaussian process regression, support vector regression, and long short-term memory. Figure 6 shows the prediction error plot of predicted values minus the actual dam inflow using a standard dataset. There are a lot of under-predictions less than the actual dam inflow, in other words this means that these errors influences much flood damage risk.

**Table 3.** Test evaluations and Bayes optimization based on the cross-validation function applied to a standard hydro-dataset. RMSE is the root-mean-squared error.

Model	RMSE	Runtime (minutes)
Regression tree	17.49	1
Ensemble model (bagging, LSBoost)	14.01	26
Gaussian process regression (GPR)	12.05	55
Support vector regression (SVR)	12.58	43
Long short-term memory (LSTM)	11.09	70



**Fig. 4.** Predictions (regression tree, RTree; ensemble; Gaussian process regression, GPR; support vector regression, SVR; long short-term memory LSTM) and actual dam inflow series using a hydro-dataset.



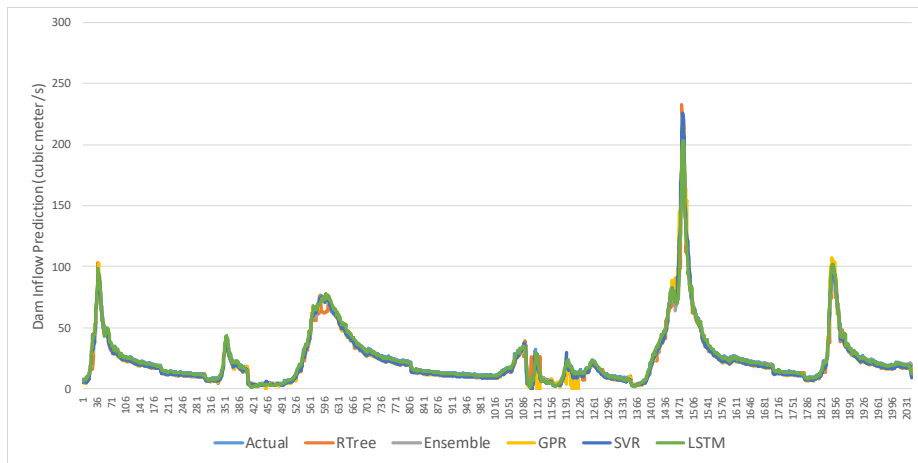
**Fig. 6.** Prediction error plot of predicted values minus the actual dam inflow using a standard dataset. Showing regression tree (RTree), ensemble, Gaussian process regression (GPR), support vector regression (SVR), and long short-term memory (LSTM) models.

**Applied Result Two: Flood-Feature Extraction Added.** Table 4 shows the RMSE and runtime results for the five models that were developed using a hydro-dataset with WLS trend and LTI volume flood-feature extractions added. The RMSE in Table 4 are lower than those obtained using the standard hydro-dataset (Table 3). The flood feature extractions are effective for improving the accuracy of each model, especially the LSTM model. Figure 7 shows dam inflow predictions results using a hydro-dataset also added with hydro-features extractions, where these are applied by the above noted five

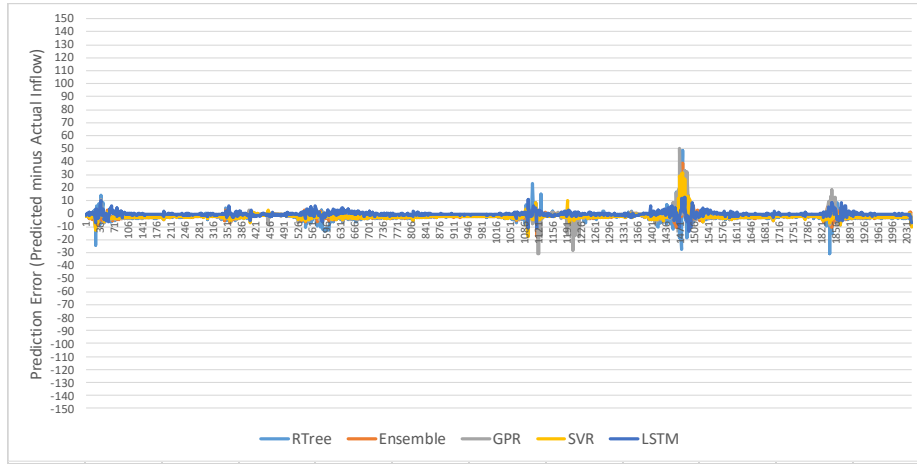
methods. Figure 8 shows the prediction error plot of predicted values minus the actual dam inflow using a standard dataset added with hydro-features extractions. Compared with the previous applied results one, under-predictions less than peak inflow are reduced, so this indicates that the better predictions improve the less flood damage risk, although under-prediction errors are not completely eliminated.

**Table 4.** Test evaluations and Bayes optimization based on the cross-validation function applied to a hydro-dataset with weighted least squares (WLS) trend and log of trapezoids integration (LTI) volume flood-feature extractions added. RMSE is the root-mean-squared error.

Model	RMSE	Runtime (minutes)
Regression Tree	5.39	2
Ensemble model (bagging, LSBoost)	2.72	34
Gaussian Process regression (GPR)	5.54	56
Support vector regression (SVR)	4.99	48
Long short-term memory (LSTM)	1.80	71



**Fig. 7.** Predictions (regression tree, RTree; ensemble; Gaussian process regression, GPR; support vector regression, SVR; long short-term memory LSTM) and actual dam inflow series added with hydro-features extractions.



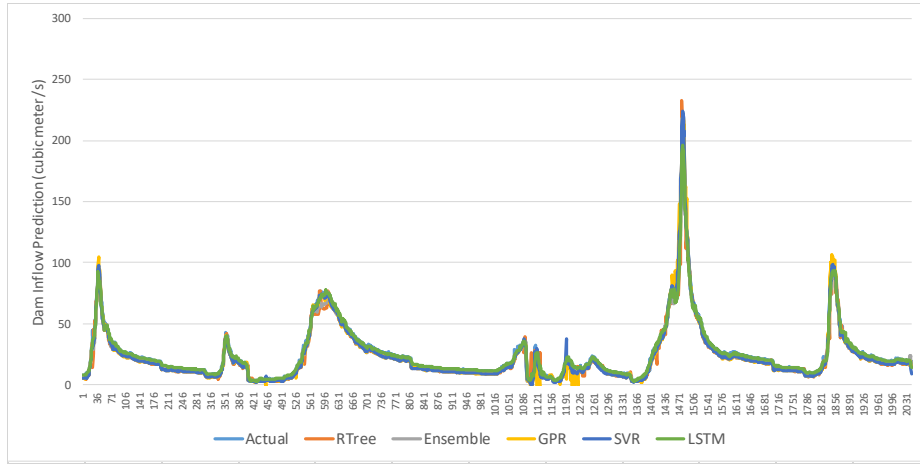
**Fig. 8.** Prediction error plot of predicted values minus actual dam inflow added to hydro-features extractions. Showing regression tree (RTree), ensemble, Gaussian process regression (GPR), support vector regression (SVR), and long short-term memory (LSTM) models.

**Applied Result Three: Upstream Sensor Variable Added.** Table 5 shows the RMSE and runtime results for the five models when developed using the hydro-dataset with WLS trend and LTI volume flood-feature extractions and an upstream sensor variable added. The RMSE in Table 5 show a slight improvement in comparison to those in Table 4. Thus, the upstream sensor variable is somewhat effective for improving the accuracy of each model, with the LSTM model being the most accurate. Figure 9 shows dam inflow predictions results using a hydro-dataset also added with hydro-features extractions and an upstream sensor variable, where these are applied by the previous noted five models. Figure 10 shows the prediction error plot of predicted values minus the actual dam inflow using a standard dataset added with hydro-features extractions and an upstream sensor variable. Therefore, under-predictions less than extreme inflow are strongly decreased, so this suggests that the higher accuracy improves the smaller flood damage risk among three applied results.

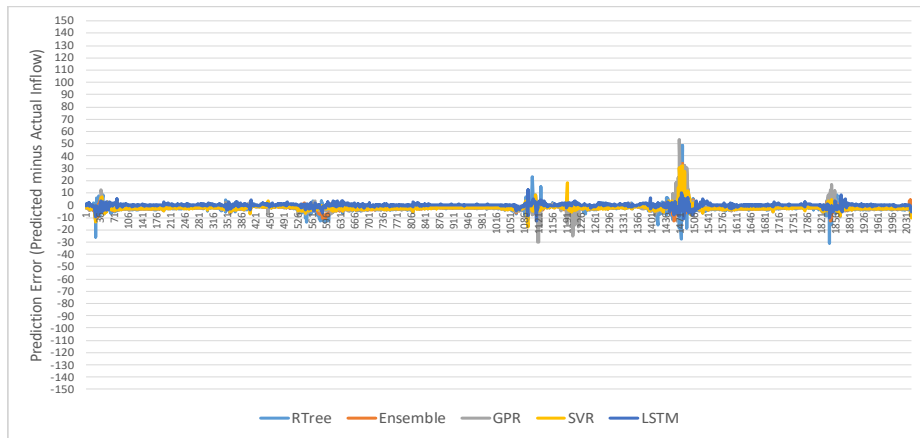
**Table 5.** Test evaluations and Bayes optimization based on the cross-validation function applied to a hydro-dataset with weighted least squares (WLS) trend and log of trapezoids integration (LTI) volume flood-feature extractions and an upstream sensor variable added.

RMSE is the root-mean-squared error.

Model	RMSE	Runtime (minutes)
Regression Tree	5.30	2
Ensemble model (bagging, LSBoost)	2.56	35
Gaussian process regression (GPR)	5.47	57
Support vector regression (SVR)	4.55	49
Long short-term memory (LSTM)	1.61	71



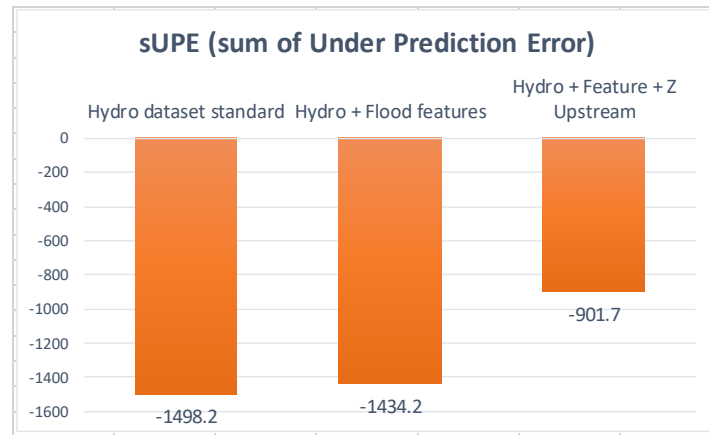
**Fig. 9.** Predictions (regression tree, RTree; ensemble; Gaussian process regression, GPR; support vector regression, SVR; long short-term memory LSTM) and actual dam inflow series added to flood-feature extractions such as weighted least squares (WLS) trend, log of trapezoids integration (LTI) volume, and an upstream sensor variable.



**Fig. 10.** Prediction error plot of predicted values minus actual dam inflow added to flood-feature extractions such as weighted least squares (WLS) trend, log of trapezoids integration (LTI) volume, and an upstream sensor variable. Showing regression tree (RTree), ensemble, Gaussian process regression (GPR), support vector regression (SVR), and long short-term memory (LSTM) models.

**Evaluation sUPE on the Best Model.** The LSTM model was the most accurate of the five machine and deep learning algorithms assessed here. Comparison of the LSTM sUPE for each of the three modeling scenarios is shown in Fig. 11. Comparison of the hydro-dataset standard with the hydro-dataset standard with hydro-feature extraction added, shows that our proposed flood feature extractions improved the sUPE and hence

allows for a reduced risk of flood damage. Furthermore, the addition of the upstream sensing variable is effective for minimizing the sUPE to avoid potential delays that may result in the event that an inflow prediction is lower than the actual dam inflow.



**Fig. 11.** Comparison of sum of under-prediction errors (sUPE) for the long short-term memory (LSTM) model: a) hydro-dataset standard, b) hydro-dataset standard with hydro-feature extraction added, and c) hydro-dataset standard with hydro-feature extraction and an upstream sensor added.

## 4 Conclusion

### 4.1 Concluding Remarks

We applied a smart dam technique for flood forecasting, which involves upstream sensing, flood feature extracted time series regression modeling, and a hydro-blockchain. In addition to the existing 20 year hydro-dataset for our case study, we found that it was possible to add upstream features by installing a water level sensor at the previously unmonitored uppermost stream. We summarized hurdles that may need to be overcome for smart dams, and detail the installation requirements for a water level sensor. Using the water level data measured by a sensor installed in the most upper stream of the dam in the in Kanto study region, we analyze the relationship between the upstream, middle, and downstream water levels during a typhoon in July 2019.

Our study highlights the problems of predicting dam inflow over a forecast period of 3 hours using data from the past 3 hours. We used WLS and LTI flood feature extractions methods to define flood trend and flood volume features, respectively. These methods vastly improved the prediction accuracy of the 3 hour forecast in comparison to that using the standard hydro-dataset only. We also found that the addition of a sensor resulted in a further, slight improvement in the RMSE. Hence, we conclude that upstream sensing and flood feature extractions are useful for improving the accuracy of flood prediction.

## 4.2 Future Works and Opportunities

In the future, we aim to address low water forecasting during periods outside of the flood season. In order to forecast more than 12 hours ahead, a deeper stream network will be constructed by incorporating the neighboring watershed to the study area.

In a situation of multiple dams, water level changes and the volume of reservoir water can become more complex, and can make it difficult for dam managers to make decisions about the discharge operation. Although upstream sensing for dam watersheds is an undeveloped field, it can increase data availability and offers the possibility of improving the accuracy of flood peak and volume predictions that are helpful for flood forecasting. Increasing the density of observations would provide data that could increase understanding of flood features and improve prediction accuracy.

Rainfall, river water levels, and other natural phenomena are complex and watershed-specific, and case studies on floods are site specific. Unsupervised generative learning may allow for the discovery of flood scenarios previously unexperienced. Our study highlights the opportunity that exists for improving the accuracy of flood forecasting by setting up synthetic flood scenarios that use observational data.

Furthermore, we have opportunities to create cyber-physical architectures [36], i.e. digital twin control for dam outflow towards various flood scenarios. We continue to be more robust the above proposed techniques to expand upstream sensing, to extract more densely flood features on hydro-blockchain, and to optimize the dam outflow task installing a hybrid learning technique between a relaxed physical generator and a faster and more accurate inflow predictor.

**Acknowledgements:** We thank Mr. Shinichi Kuramoto and Mr. Takuji Fukumoto (MathWorks Japan) for providing stable resources for machine and deep learning.

## References

1. UN News Centre: EM-DAT International Disaster Database, <https://reliefweb.int/disasters>, last accessed 2019/8/10.
2. Insurance Information Institute: World Natural Catastrophes, <https://www.iii.org/fact-statistic/facts-statistics-global-catastrophes>, last accessed 2019/8/10.
3. 2016 September 10, Japan floods: City of Joso hit by 'unprecedented' rain. BBC Homepage, <http://www.bbc.com/news/world-asia-34205879>, last accessed 2019/8/10.
4. ASCE: Task Committee on Application of Artificial Neural Networks in Hydrology, II: Hydrological applications, *Journal of Hydrological Engineering*, 5, 124-137, (2000).
5. Mujumdar P.P. et al.: *Floods in a Changing Climate Hydrologic Modeling*, International Hydrologic Series, Cambridge Univ. Press (2012).
6. Othman, F. et al.: Reservoir Inflow Forecasting using Artificial Neural Network. *International Journal of the Physical Sciences* 6(3), 434-440 (2011).
7. Attygalle D. et al.: Ensemble Forecast for Monthly Reservoir Inflow; A Dynamic Neural Network Approach. doi:10.5176/2251-1938\_ors16.22 (2016).

8. Singh, V. et al.: Mathematical Modeling of Watershed Hydrology, *Journal of Hydrological Engineering*, 7(4), 270-292 (2002).
9. Kouwen, N. et al. : Grouped Response Units for Distributed Hydrologic Modeling, *Journal of Water Resource Planning Management*, 119(3), 289-305 (1993).
10. Ewen, J. et al. : SHETRAN: Distributed River Basin Flow and Transport Modeling System, *Journal of Hydrological Engineering*, 5(3), 250-258 (2000).
11. Fortin, J.P. et al. : A Distributed Watershed Model Compatible with Remote Sensing and GIS Data, I: Description of Model, *Journal of Hydrological Engineering*, 6(2), 91-99 (2001).
12. Andrew K. et al. : *Inside Smart Cities : Places, Politics and Urban Innovation*, Routledge (2019).
13. Rob, K., Claudio, C. et al.: *Actually Existing Smart Dublin : Exploring smart city development in history and context*, edited by Andrew, K.: *Inside Smart Cities* (2019).
14. Karwot, J., Kazmierczak, J. et al.: *Smart Water in Smart City*, *Hydrology and Water Resources*, 16<sup>th</sup> International Multidisciplinary Scientific GeoConference SGEM (2016).
15. *Smart Cities of Tomorrow*, Water, 2014 : <http://www.smartcitiesoftomorrow.com/water/>, last accessed 2019/8/10.
16. Daniel Drescher: *Blockchain Basics:A Non-Technical Introduction in 25 steps*. 1st edn. Apress (2017).
17. Information System Authority: Data Exchange Layer X-tee, <https://www.ria.ee/en/state-information-system/x-tee.html>, last accessed 2019/8/10.
18. Shortridge, J.E. et al.: Machine Learning Methods for Empirical Streamflow Simulation: A Comparison of Model Accuracy, Interpretability and Uncertainty in Seasonal Watersheds. *Hydrology and Earth System Sciences* 20, 2611–2628 (2016).
19. Li, B. et al.: Comparison of Random Forests and Other Statistical Methods for the Prediction of Lake Water Level: A Case Study of the Poyang Lake in China. *Hydrology Research* 47.S1 (2016).
20. Yang, J.-H. et al.: A Time-series Water Level Forecasting Model Based on Imputation and Variable Selection Method. *Computational Intelligence and Neuroscience*, 1–11 (2017).
21. Vamsi, K.G. : Modeling of Weather Data using Gaussian Linear Regression, *International Journal of Computer Science and Information Technologies*, Vol.6(4), 3693-3696 (2015).
22. Adiya, G. et al. : A Deep Hybrid Model for Weather Forecasting, *ACM Conference on Knowledge Discovery in Databases, KDD'15, Sydney Australia* (2015).
23. Zhan, D. et al. : Simulating Reservoir Operation using Recurrent Neural Network Algorithm, *Water*, 11, 865, MDPI (2019).
24. Xuan-Hien, L. et al.: Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting, *Water*, 11, 1387, MDPI (2019).
25. Brieman, L., Friedman, J. et al.: *Classification and Regression Trees*. Wadsworth (1984).
26. Friedman, J., Hastie, T., Tibshirani, R.: Additive Logistic Regression: A Statistical View of Boosting (with discussion). *Annals of Statistics* 28, 307–337 (2000).
27. Friedman, J.: Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* 29(5), 1189–1232 (2001).
28. Scholkopf, B., Freund, Y.: *Boosting: Foundations and Algorithms*, MIT Press (2012).
29. Rasmussen, C.E. and Williams, C.K.I. : *Gaussian Processes for Machine Learning*, MIT Press (2006).
30. Sergios, T. : *Machine Learning : A Bayesian and Optimization Perspective*, Elsevier (2015).
31. Hochreiter, S. and Schmidhuber, J. : Long Short-Term Memory, *Neural Computation*, 9(8), 1735-1780 (1997).
32. Francois, C. : *Deep Learning with Python*, Manning (2018).

33. Domingos, P.: MetaCost: A General Framework for Making Classifiers Cost-sensitive. ACM Conference on Knowledge Discovery in Databases, 165–174 (1999).
34. Ling, C.X. et al.: Cost-Sensitive Learning and the Class Imbalance Problem. In : Sammut, C. (eds.) Encyclopedia of Machine Learning, Springer (2008).
35. Yasuno, T: Dam Inflow Time Series Regression Models for Minimizing Loss of Hydro Power Opportunities, Workshop of Data Mining for Energy Modeling and Optimization, PAKDD2018, Melbourne (2018).
36. Zandi, K. , Ransom, E.H., Topac, T., Chen, R., Beniwal, S., Blomfors, M., Shu, J. and Chang, F-K. : A Framework for Digital Twin of Civil Infrastructure – Challenges & Opportunities, 12<sup>th</sup> International Workshop on Structural Health Monitoring, 1659-1665, DEStech (2019).