Journal of Computer Science and Cybernetics, V.38, N.1 (2022), 85–102
DOI no. 10.15625/1813-9663/38/1/16092

A UNIFIED FRAMEWORK FOR WATER SURFACE EXTRACTION AND CHANGE PREDICTION IN IMAGERY DATA STREAMS

NGUYEN THANH TAM[1,*], NGUYEN THANH TOAN[1], PHAN THANH CONG[1], NGUYEN QUOC VIET HUNG[2]

[1]*Ho Chi Minh City University of Technology (HUTECH), Vietnam*
[2]*Griffith University, Australia*

Abstract. Changes in surface water might result in natural disasters such as floods, water shortages, landslides, waterborne diseases, which lead to loss of lives. Timely extracting for surface water and predicting its movement is essential for planning activities and decision-making processes. Most existing works on extracting water surface using satellite images focus on static spectral images and ignore the temporal evolution of data in streams, leading to less accuracy and lack of prediction power. Although some works realize that modeling temporal information of satellite signals could boost the forecasting capability on environmental changes, most of them only focus on prediction tasks independently and separately from the extraction task. In this paper, we propose **WECP**—a unified framework for **W**ater **E**xtraction and **C**hange **P**rediction—which is built on top of a data stream of satellite images. The framework locates the water surface and predicts its changes over time simultaneously. We evaluate **WECP** using real datasets from different regions, and the results show that our framework is robust in extracting and capturing spatio-temporal changes in the water surface.

Keywords. Deep learning; Satellite imagery mining; Spatio-temporal change prediction; Water surface extraction.

1. INTRODUCTION

Environmental changes and their impacts attract much research in multidisciplinary areas [25, 34]. Water is one of the essential natural resources for social evolution, agricultural production, and human life. Water surface changes are the primary factor for environmental changes, and they result in natural disasters such as floods, water shortages, landslides, waterborne diseases, which lead to loss of human lives [2]. These disasters from the dynamic of water are so severe that authorities had to release regulation for water monitoring frequency as the European Union's Water Framework Directive [19]. Besides the immediate damage, water changes could also cause long-term consequences [30] such as local and global weather impacts, land user/cover (LULC) changes, and coastline changes. Timeliness in monitoring

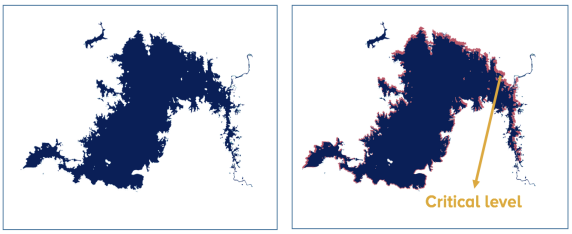
*Corresponding author.
E-mail addresses: nt.tam88@hutech.edu.vn (N.T.Tam), quocviethung.nguyen@griffith.edu.au (N.Q.V.Hung)

and alerting on the dynamics of surface water is, therefore, an essential and sustainable solution for policy makers [10] to mitigate the unexpected and unrecoverable damages.

Despite the significant task, monitoring surface water using *traditional* ground survey techniques is notoriously hard, time-consuming, and even infeasible for large regions such as countries or continents. This is due to the extremely large region of water bodies, the complexity of the coastal line, and the fast-moving of water in floods and storms.

Modern remote sensing technologies enable water monitoring at a global scale using satellite images. Common approaches are to use spectral indices from satellite images to extract water body such as Normalized Difference Water Index (NDWI) [9,24,42]. However, such methods require intensive expert knowledge and the optimal threshold needs to re-define whenever we shift between different regions. Contemporary deep learning techniques allow the perfect generalisation in extracting water surface crossing multiple regions [21,40], they ignore the temporal evolution of data in streams. This leads to less accuracy and lack of the ability of some predictive tasks for water change prediction. Some works realize that surface water change data has seasonal characteristics [41]. Although leveraging such temporal information of satellite signals into the works could boost the performance and even enable the prediction power about the future environmental changes, most of them only focus on prediction task independently and separately from the water extraction task [33].

This study aims to propose a framework to tackle the water extraction and change prediction simultaneously. The system not only accurately locates the water surface using satellite image (Figure 1a), but also can answer such a question: *With the current situation, how long will the water surface reach the critical level?* (Figure 1b). The answer to such a question shall provide useful information for the decision-making process; however, unifying such two different tasks into a framework poses some challenges. First, deep learning building blocks are designed for specific tasks [11]: Convolutional Neural Network (CNN) is designed to capture spatial information in common images, Recurrent Neural Network (RNN) is designed for extracting hidden temporal information. Combining such different tasks into an end-to-end trainable model that learns a universal representation for water extraction and change prediction is a non-trivial task. Second, satellite image often features low spatial resolution, which suffers many adversarial situations such as ice, cloud, solar distribution [13]. Finally, universal features for water extraction may differ from various locations and regions around the world.



(a) Water extraction (b) Change prediction

Figure 1: Motivation example

To overcome these challenges, we propose a unified framework for **W**ater **E**xtraction and **C**hange **P**rediction (**WECP**) in an end-to-end manner. These two tasks are coupled together and mutually enhanced by the same learning process. The universal representation is shared between the two tasks helps to save computational cost when training them separately. Our method is orthogonal to domain-specific approaches employing spectral indices as universal features are extracted directly from input data without any hand-engineering or feature selection. We use Landsat 8 because it is the state-of-the-art satellite system in image capturing [28] and could provide a high standard of resolution. We summarise the

contributions of our work as follows.

- *The unified problem:* Section 3.1 formulates the unified problem for water surface extraction and change prediction.
- *Data streaming and processing pipeline:* Section 3.2 presents a pipeline to incrementally gather data from satellite image sources, clean adversarial conditions and normalize to keep invariant features from different regions around the world.
- *The WECP Framework:* Section 4 establishes a multi-task learning algorithm to optimize the water surface extraction and change prediction in the same process.

The remaining sections of the paper include: Section 2 discusses related work. Section 5 reports the empirical evaluations with real-world datasets, before Section 6 summarises the paper.

2. RELATED WORK

This section briefly presents the related work to the topic of water monitoring using satellite images.

Water indexing approaches. While the problem of water surface monitoring using satellite images has received considerable attention in recent years, much interest is placed on spectral thresholding methods [14, 15, 43]—or, in other words, on the so-called "water indices", as McFeeters [24] calls it. For example, McFeeters et al. first suggested the Normalized Difference Water Index (NDWI) [24], which used band 2 and band 4 of Landsat image to delineate the water surface. Nevertheless, Xu et al. figured out that NDWI did not efficiently discriminate the built-up surfaces from water surfaces. The authors then devised another index—Modified Normalized Difference Water Index (MNDWI) [42]—which replaced band 4 of NDWI with band 5 of Landsat images and yielded better performance. An ensemble of different methods was also proposed to improve water surface extraction performance [9,17,36]; however, the water distinguishing power of these methods varies with different regions and locations.

Despite existing water extraction approaches in the literature, it is hard to choose one of them for environmental change monitoring on a global scale as they are facing the accuracy and generalisation problems. To be more specific, they heavily rely on domain experts for hand-crafting features and defining the optimal threshold. Environmental monitoring such as water change detection and prediction is less likely to be reliable when such insufficient efficiency methods are employed as a pre-processing step [8, 16].

Machine learning approaches. In order to enhance the performance of water surface extraction, many machine learning approaches have been proposed. A knowledge-based approach [18] leverages both spatial and spectral information to classify the water surface. A support vector machine (SVM) [48] is adopted to derive coastline from sensing images. A clustering technique [44] is applied for extracting spatial information of water body from complex environments. While maintaining satisfactorily high accuracy, these methods still require analysing the spectral bands and selecting the suitable features, thus leading to a low level of automation.

Deep learning, the latest generation of AI technologies, has demonstrated superior performance on various image mining tasks ranging from object detection, image classification

to segmentation [11]. With the advance of remote sensing technology, many studies have been applied to satellite images to extract useful information (e.g., water surface, land use) at a global scale. Lü et al. introduced a design based on the deep belief network (DBN) model [32] to classify satellite images. Chen et al. proposed a model based on stacked autoencoder (SAE) to classify hyperspectral data [7]. A convolutional neural network model (CNN) [45] was devised to extract useful spatial features from Landsat imagery hierarchically, and logistic regression is used as a final layer to classify water and non-water surface.

Most deep learning solutions exploit spatial information, which neglects temporal information. Some works realize that surface water change data has seasonal characteristics [26, 41] and utilising such temporal information of satellite signals into the model could boost the performance and even enable the prediction power about the future environmental changes. However, these studies focus on the prediction task independently and separately from the water extraction task [33]. A recent work aggregates both temporal and spatial information into a machine learning model [29]; however, they address the problem of paddy mapping, and it is orthogonal with our work in this study. Advancing beyond the state-of-the-art, we devise the unified framework to address the water extraction and change prediction task at the same time in an end-to-end process. These two tasks are coupled together and mutually enhanced by the same learning process. The characteristics of our method in comparison with existing techniques are revealed in Table 1.

Table 1: Comparison between water monitoring methods

Method	**Automation Level**	**Water Extraction**	**Change Prediction**
WECP	automatic	✓	✓
Spectral [35]	automatic	✓	✗
w-CNN [45]	automatic	✓	✗
w-SVM [1]	hand-crafted	✓	✗
Threshold [9]	hand-crafted	✓	✗

3. MODEL AND APPROACH

This section briefly reviews some backgrounds on water surface monitoring and then reveals our model together with the problem statement. Finally, we demonstrate our approach for addressing the formulated problem.

3.1. Model

Given a streaming of satellite images $\mathcal{I} = \{I^{(1)}, ..., I^{(t)}, ..., I^{(T)}\}$, where T is the number of timestamps. Each image has the size of $S = S_1 \times S_2$. More precisely, each image is represented as $I^{(t)} = \{i_1^{(t)}, ..., i_{S_1}^{(t)}, ..., i_S^{(t)}\}$, where $i_i^{(t)}$ is a C-dimensional vector. In this study, C is set to the number of spectral bands of Landsat images [22] (i.e. $C = 11$). Let $\mathcal{L} = \{L^{(1)}, ..., L^{(t)}, ..., L^{(T)}\}$ is the label set for $\mathcal{I}$. Each $L^{(t)} = \{l_1^{(t)}, ..., l_{S_1}^{(t)}, ..., l_S^{(t)}\}$ where $l_i^{(t)} = \{1, 0\}$ to indicate the water surface (1) or non-water surface (0), respectively.

Before revealing the problem addressed in this study, we define the two subroutine tasks as follows.

Task 1 (Water extraction)**.** *Given a stream of satellite images $\mathcal{I}$, the task of water surface extraction is to find a mapping function f_e to map from $\mathcal{I}$ to its label set $\mathcal{L}$. The task is formulated as*

$$\mathcal{L} = f_e(\mathcal{I}), \tag{1}$$

where $f_e : \mathbb{R}^{S\times C\times T} \to \mathbb{R}^{N\times T\times 2}$.

In this task 1, we need to define a classification model for f_e and learn an optimal model in such a way that we maximise the performance of the model—the more accurately extracted pixels, the more reliable.

Task 2 (Change prediction)**.** *Given a stream of satellite images $\mathcal{I}$, its label $\mathcal{L}$ to the current timestamp, and a number of timestamps in the future τ to predict, the task of water change prediction is to find a mapping function f_p to map from $\mathcal{I}$ to its predictive label set $\mathcal{L}_\tau$. The task is defined as*

$$\mathcal{L}_\tau = f_p(\mathcal{I}, \mathcal{L}), \tag{2}$$

where $f_p : \mathbb{R}^{S\times C\times T} \to \mathbb{R}^{N\times \tau\times 2}$.

Similar to Task 1, Task 2 requires to define a predictive model for f_p and an aggregation function, and thus we can maximise the predictive performance. We formalise the problem tackled in this paper as follow.

Problem statement. Given a stream of satellite images $\mathcal{I}$ and a number of timestamps in the future τ, the problem of *unifying water surface monitoring* is to find the optimal solution for both *water extraction* and *change prediction* tasks, while requiring only one-pass through the dataset.

3.2. Approach

We propose a framework for unifying water surface monitoring consisted of two stages as in Figure 2.

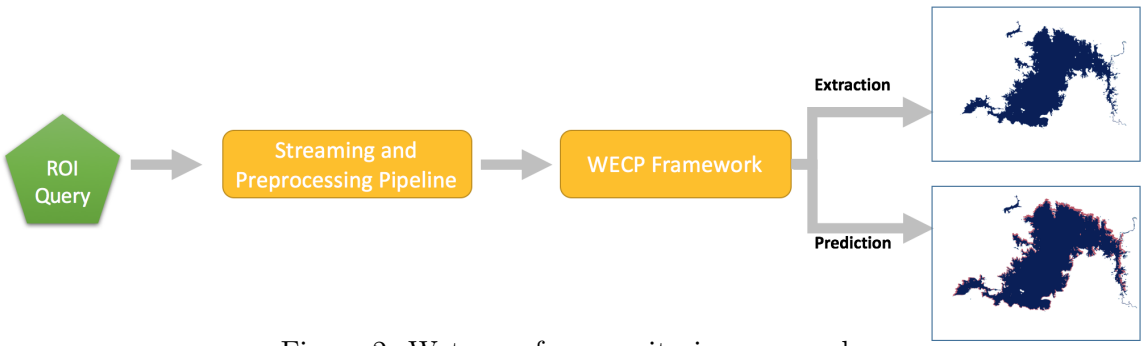


Figure 2: Water surface monitoring approach

The first stage is *Streaming and Processing Pipeline*. This stage helps to load all satellite images within a pre-defined region of interest. The loaded data are further cleaned and corrected according to the requirements about geometric, topographic, radiometric consistency. The second stage is the *Water Extract and Change Prediction (WECP)*, where the processed

data is fed into to produce the extraction and prediction output. Finally, whole data includes images from satellite and extracted and predicted result is combined and visualized in a professional and user-friendly web application, thereby completion of an autonomous and intelligent Water Surface Monitoring. To achieve these goals, the following components require to be realised.

Streaming and Processing Pipeline. Unlike traditional satellite streaming and processing approaches, where all images are acquired in advance before the processing occurs, we develop a stream-based processing pipeline to facilitate real-time monitoring. The processing steps are performed on segments that are assembled in the stream. The processing pipeline consists of three-level. In level I of processing, we apply the Fmask algorithm [49] to increase the quality of images by removing the effect of cloud shadows. Next, in level II of the processing pipeline, the Top of Atmosphere (TOA) technique is adopted to remove solar zenith angle effects and adjust the contrasts in solar irradiance. Finally, in level III, the spatial alignment is conducted to reduce the geo-location inconsistency between satellite images caused by the polar-orbiting.

Water Extract and Change Prediction (WECP). The WECP framework is composed of two key components to perform robust water surface extraction and change prediction. In the first component, the convolutional neural network (CNN) is employed to capture the *spatial-spectral* information. In the second component, bi-directional long short term memory (BiLSTM) is used to apprehend the *spatial-temporal* changes. Together with another supplement building block such as upsampling layers, these components are carefully integrated into a unified framework in such a way that they are closely coupled and mutually enhanced with each other during the optimisation process. We describe this component in details in the next section (Section 4).

4. THE WECP MODEL

4.1. Model structure

Existing water monitoring systems often solve the water extraction (Task 1) and change prediction

(Task 2) separately [15, 33, 43]. Going beyond the state-of-the-art in water monitoring, we study the feasibility of solving the two tasks in an end-to-end model. To this end, we develop a deep neural network model that integrates spatial, temporal, and spectral data into a unified model. The model contains several subnetworks including: *(i) the input module:* augments and feeds the data to subsequent layers; *(ii) the BiLSTM module:* processes temporal dependencies; *(iii) the convolutional module:* captures the spatial and spectral dependencies between pixels of the data; *(iv) the output module:* returns the extraction and prediction results. The overview of our model is presented in Figure 3.

By integrating spatial-temporal information to learn multiple tasks simultaneously—the water extraction and change prediction task—the model tried to optimise multiple loss functions at the same time. By doing so, the two tasks are coupled together, and their performance are mutually enhanced due to several reasons: *i) overcoming the data sparsity.* Although most machine learning techniques exploit the data to solve a single task, these approaches eventually hit a performance ceiling [47]. This is because of the sparsity of the

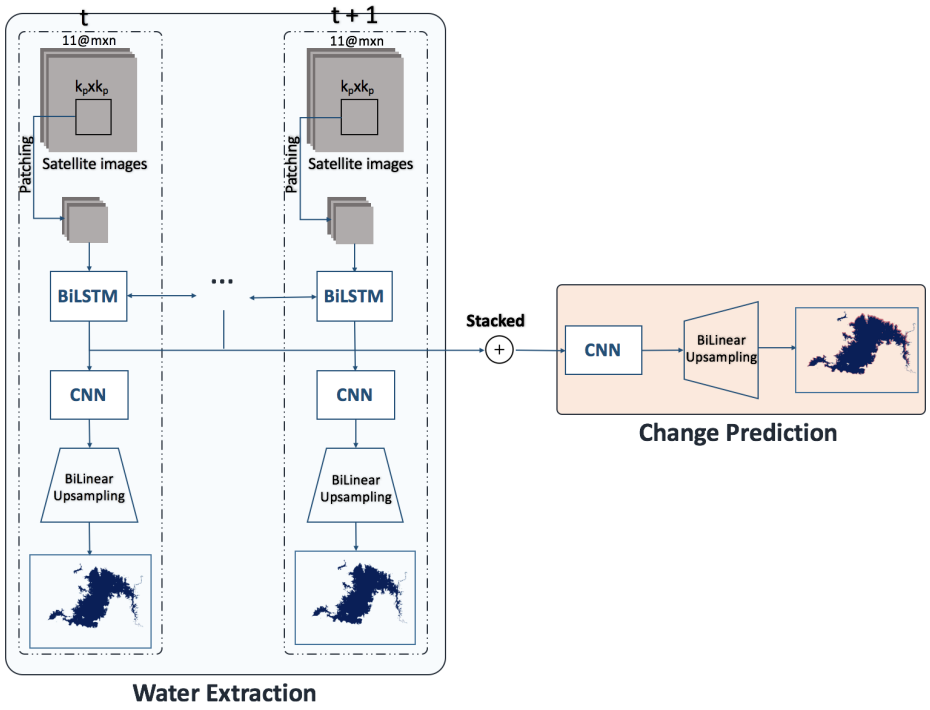


Figure 3: The WECP Framework

dataset or the lack of generalisation of the model in learning meaning representations. In multi-task learning, the model uses all available data and the ground truth across the tasks to learn the generalised representations meaningful for multiple contexts. *ii) performance gain.* A hidden feature computed by one task can be re-used by the other task without redundant computations. This behaviour helps our model gain performance in both training and testing time compared to the total time when the two tasks are undertaken separately. The next section reveals the detailed design of each component of our framework.

4.2. Design of each component

4.2.1. Input module

A region of interest is fed as the input of the framework. The dimension of the studied region is designed as a input block of $m \times n \times 11$ pixels, in which a $m \times n$ matrix represents each spectral band, and we have 11 spectral bands of Landsat images in total. As the number of regions is limited, using a whole region as a training sample makes the model likely to overfit. To overcome this issue and to augment the training set, we employ a *patch normalisation* layer to enhance the generalisation of the framework. This layer consists of two steps: *(i) patching:* the region is divided into patches with a size of $k_p \times k_p$ pixels, and every two patches have 50% overlap between them; *(ii) normalisation:* a normalisation [4] step is then separately performed on each spectral band to ensure that the patches are in the same domain before fetching into the next layer of the training pipeline.

4.2.2. BiLSTM module

A common approach for visual imagery analysis starts by employing a Convolutional Neural Network (CNN) to derive spatial features. But, the extracted feature by CNN is shrunk to specific semantics such as edge detection, curve detection [23], which neglects the temporal information. On the other hand, if using the sequential model alone, the output might crack the 2D structure of imagery domains [6]. Moreover, a simple sequential network such as Recurrent Neural Network (RNN) may suffer the vanishing gradients problem, which hampers its ability to learn temporal dependency in a long-term period. Long short-term memory (LSTM) is introduced to overcome this problem to capture long-term dependency [3]. However, one thing is still missing. That is the future context, essential for water and environmental change data with seasonal characteristics such as floods, tides, and storms.

We overcome these limitations by leveraging Bidirectional LSTM (BiLSTM) [12] to encode temporal dependency of satellite images. The basic idea of BiLSTM to present each data input forwards and backwards to two hidden and separated layers of LSTM. The two hidden layers are then connected to the output layer. After both the past and the future information are remembered, the output is then reconstructed to imaginary space (2D spatial space) before feeding into the next layer of the framework (i.e., the convolutional layer). It has been studied that such a bidirectional model could capture long-range dependencies [37] of the training sequence of satellite images.

Similar to LSTM, the key idea of BiLSTM nets [3] is to continuously update the memory m_t in such a way that it partially forgets the memory(m_{t-1}) and adds the new memory($\tilde{m}_t$) between two consecutive time steps

$$m_t = f(\boldsymbol{x}_t, \boldsymbol{h}_{t-1}) m_{t-1} + a(\boldsymbol{z}_t, \boldsymbol{h}_{t-1}) \tilde{m}_t, \tag{3}$$

where $\boldsymbol{x}_t$ is the time step t-th of the input sequence, and $\boldsymbol{h}_{t-1}$ is the hidden output of the earlier time step. The function $f(.)$ is the forgetting function and the function $a(.)$ is the adding function. Both of these functions are a single layer of neural network that employs a sigmoid function as an activation. As being the sigmoid regression, they always yield the output between 0 and 1. This output controls how much information is allowed to pass each component. The new memory is defined as

$$\tilde{m}_t = \tanh(b_c + \boldsymbol{w}_c^1 \boldsymbol{x}_t + \boldsymbol{w}_c^2 \boldsymbol{h}_{t-1}), \tag{4}$$

where the tanh(.) is exploited to remember both "good" and "bad" memories by smoothing the memorized value into $\{-1, 1\}$. In summary, the output of the current LSTM cell is the combination of three factors, including the previous output, the current memory and the current input

$$h_t = o(\boldsymbol{x}_t, \boldsymbol{h}_{t-1}) \tanh(c_t), \tag{5}$$

where $o(.)$ is the sigmoid regression as well. Unlike the original LSTM building block, the output $\boldsymbol{z}_t$ of BiLSTM is a parameterized function of the bi-directional latent states—the forward h_t^f and the backward hidden state h_t^b

$$\boldsymbol{z}_t = \sigma(w_f h_t^f + w_b h_t^b + b_h), \tag{6}$$

where w_b and w_f are the backward and forward parameters, respectively, and σ is a softmax.

Scalable training. Since the frequency of streaming Landsat 8 data is 16 days, we chunk 22 temporally consecutive images into a sequence. Such sequence represents an observation of a region of interest throughout a year long. This chunking mechanism for the training set of the BiLSTM module allows the model to scale to multiple regions crossing multiple years. We flatten the input before fetching to the BiLSTM module to follow the mentioned mathematical definition. The output is then re-constructed as the equal size to the size of the input. This helps to reshape the output back to the 2D dimension to preserve the spatial indices of the original image.This output is then fed into convolutional module, which is described in the next section.

4.2.3. Convolutional module

To capture spatial dependency, the convolution module is designed to reduce the complexity of fully connected networks [3]. It consists of two levels of spatial resolution: *(i) extraction convolution* and *(ii) prediction convolution.*

Extraction convolution. The produced output of BiLSTM cell is fed to the convolution block for water extraction. In this block, the neutrons of each layer are connected to a few neurons—within a receptive field—of the next layer. [37]. The size of the receptive field is defined by $n_c \times n_c$, which are tunable hyperparameters. The receptive field size depends on how concise or abstract features that layer desires to extract. The deeper the level of convolution goes, the more abstract features are extracted. The receptive field is then walked crossed the input to successfully captures both spectral and spatial dependency of the preceding layer.

Unlike traditional fully-connected networks, a convolutional network is much more computationally efficient due to its shared-weight mechanism during the learning process. More precisely, although multiple convolutional operations happen within a convolutional layer, such operations share the weights and biases. Formally, the output value of each layer is defined as [3]

$$a_{ij} = \sigma\left(b_i + \sum_{m=1}^{C} w_{im} z_{j+m-1}\right) = \sigma(b_i + \boldsymbol{w}_i \boldsymbol{z}_j), \tag{7}$$

where a_{ij} is j-th neuron's output of i-th filter, σ is the activation function, b_i is the i-th filter's shared bias, C is the number of filters, $\boldsymbol{w}_i = [w_{i1}, \ldots, w_{iC}]$ is the shared weight, and $\boldsymbol{z}_j = [z_j, \ldots, z_{j+n_c-1}]$ is the receptive field. By way of explanation, the next layer yields the local spatial feature, the feature map [3], from the previous layer.

We develop the water extraction block with three consecutive convolutional layers. We use many filters with different sizes in each layer. The reason behind being that we have different spatial features' types that need to be identified. More precisely, we use $M_1 = 128$ filters in the first layer, $M_2 = 64$ filters in the second layer and $M_3 = 32$ in the third layer. All the filters are the same size of $n_c = 3$. We choose the same of all kernels because the spatial resolution of Landsat 8 is the same for all pixels (i.e., $30m$ per pixel).

Differently from threshold-based water extraction methods, which leverage only a few spectral bands, we consider the dependency of all spectral bands simultaneously. Practically, these dependencies may vary significantly from different regions or locations throughout the

globe. Thus after each convolutional layer, we use a pooling layer that helps to smooth the output from the convolutional layer by sub-sampling. An *average pooling* is selected for this pooling operator. As the water surface does not include sharp features, we choose *average pooling* rather than the popular *max pooling* [39] to avoid loss of information.

Prediction convolution. The prediction block forecasts how the water changes in the next step based on observations from the previous steps. To this end, the outputs of BiLSTM blocks across multiple time points are stacked as filters in the first convolutional layer. Similar to extraction network, we select the prediction block with three consecutive convolutional layers. We use $M_1 = 128$ filters in the first layer, $M_2 = 64$ filters in the second layer and $M_3 = 32$ in the third layer. All the filters are the same size of $n_c = 3$.

The difference between the extraction and prediction convolution modules is the domain of their feature inputs. While the input feature of the extraction convolution is the hidden feature of a time-step in the sequence, the input feature of the prediction convolution is an aggregation of all time-steps within a year long. However, despite the difference in the structure, two modules are tightly coupled and mutually enhanced as they shared the intermediate features and the optimised metrics.

4.2.4. Output module

Bilinear upsampling. Because the convolutional building block transforms the original input space, we need to restore the input state for pixel-wise water classification. To this end, we leverage a bilinear upsampling layer which is state-of-the-art on 2D image data, to reconstruct the final hidden features to imagery spaces. More precisely, from the final convolutional output, upsampling layer generates $C \times S_1 \times S_2$ pixels where $C = 11$ is the number of proposals [31]. More precisely, each pixel in the proposal is interpolated from its neighbor pixels, and the aggregations are formulated as

$$p^c_{ix} = \frac{v_{(i-1,x)} + v_{ix} + v_{(i+1,x)}}{3} \quad \text{for } j-1 \le x \le j+1, \tag{8}$$

$$p^c_{xj} = \frac{v_{(x,j-1)} + v_{xj} + v_{(x,j+1)}}{3} \quad \text{for } i-1 \le x \le i+1. \tag{9}$$

Eventually, the output of the bilinear upsampling layer is transfered into the final layer to produce the classification score [31]

$$y = wp + b, \tag{10}$$

where w and b are the learnable parameters.

Activation function. The classification score is then presented to a soft-max layer. This layer plays the role of an activation function to non-linearize and normalises the classification scores for differentiating between classes

$$y^c_{ij} = \frac{e^{y^c_{ij}}}{\sum_{c' \in L} e^{(y^{c'}_{ij})}}, \tag{11}$$

where y^c_{ij} is the probability of the pixel (i, j) to be classified as class c. Throughout testing phase, the final class of each pixel is estimated as $y^*_{ij} = \text{argmax}_{c \in L}\, y^c_{ij}$.

Multi-tasking loss. To address the *water surface extraction* and *change prediction* tasks simultaneously, the framework is optimized under two loss functions: The former one is L_e and the latter one is L_p. As definiction in Figure 3, each task of the model will be influenced by each loss function independently. In practice, a ratio of the loss-share β between the prediction and extraction loss necessitates to be employed to find the trade-off loss of two tasks

$$L = \beta L_p + (1 - \beta) L_e. \tag{12}$$

Both functions L_e and L_p exploy the cross-entropy loss function to maximize the estimation of the true class at pixel-level.

5. EMPIRICAL EVALUATION

This section presents the comprehensive empirical evaluation of our proposed framework. We first discuss the experiment setting (Section 5.1), and then assess multiple aspects of our approach:

- The end-to-end performance of the framework (Section 5.2).
- The performance on water extraction task (Section 5.3).
- The performance on change prediction task (Section 5.4).
- The importance of each component of the framework (Section 5.5).

5.1. Experiment setting

Datasets. To assess the robustness of our framework, different regions in Vietnam are collected and used as the real-world datasets for the evaluation as follows:

- Tri An Dam: This is a hydroelectric lake and dam on the Dong Nai River, which was built and became operational in 1988.
- Dau Tieng Reservoir: This is the largest irrigation reservoir in Tay Ninh Province, which has a capacity of 1.6 billion cubic meters.
- Thac Ba Lake: This is an artificial lake in Yen Bai Province, which was built and became operational as a hydroelectric plant in the 1960s.

The datasets were acquired using the Landsat 8 satellite, which showed the changes in water surface in these regions from 2018 to 2020. In more detail, the process of collecting the data and setting up the ground truth follow the pipeline as follows:

1. *Data collection:* Satellite data in this study is collected from the data streams available on the Earth Explore service [38]. Basically, this imagery data is a digital map of radiance values at the top of Earth's atmosphere under the form of wavelengths. Then, these samples are packaged, compressed and transmitted to the ground station, from which they are transformed into geospatial and calibrated pixels.

2. *Data storage:* The collected data is then stored in the format of Georeferenced Tagged Image File (GeoTIFF), which is an international interchangeable format for raster satellite data and widely adopted in NASA's Earth Science systems [27]. Each channel of an image sample is stored in a separate GeoTIFF file to facilitate the subsequent steps of the framework.

3. *Ground truth setup:* The stored data is then forwarded to several pre-processing routines, including spectral normalisation [46], geometric correction [20] solar correction [5]. Finally, we use high spatial resolution images from Google Earth™as the reference data for on-screen digitising the true boundaries of all the test sites. Similar approaches have been used in the literature on water mapping using Landsat imagery [9, 42].

Cross-validation. To ensure a fair evaluation, we apply K-folk validation to split data into a training set and testing set. Particularly, the dataset is shuffled and randomly partition into k parts of the same size, in which $(k-1)$ parts are employed for training and the remaining part is used for testing. We repeat such evaluation k times, and we compute the average results. A common practice is to choose $k = 10$ to trade off the amount of data for training and testing.

Table 2: The study areas

Study Area	Period	#Images	#Pixels	Class Distribution [1]	Type of Water surface
Tri An	2018-2020	66	7711 × 7541	3,134,190,869: 728,881,597	Dam
Dau Tieng	2018-2020	66	7711 × 7541	3,326,534,624: 536,537,842	Reservoir
Thac Ba	2018-2020	66	7711 × 7541	3,425,370,542: 437,701,924	Lake

[1] The proportion between # water and # non-water pixels

Baselines. We compare our method against several baselines as follows:

- *Threshold:* This is an ensemble of different threshold methods were also proposed to improve water surface extraction performance [9].
- *w-SVM:* This is a machine learning approach [1], which is built on top of spetral-based feature such as (NDWI) [24] and (MNDWI) [42]. This SVM-based is specialised for water classification, which use the linear kernel and L2 regulariser to maintain the regularisation parameter.
- *w-CNN:* This is the ubiquitous deep learning architecture to extract spatial features in image processing, which is specially designed for the water body extraction task using landsat images [45].
- *Spectral*: This is the state-of-the-art deep learning approach for spectral images processing [35].

Metrics. Both water extraction and change prediction tasks are assessed at pixel level using different metrics as follows:

- *Precision:* The amount of positive samples which is estimated accurately as 'water' divided by the total amount of available samples that are estimated as a positive class.
- *Recall:* The amount of the positive samples, which is estimated correctly as 'water' divided the number of actual water samples in the ground truth.
- *Accuracy:* The proportion of samples which are correctly estimated over the number of samples.
- *F1-score:* The harmonic mean of *Recall* and *Precision*, which is computed for each class.

Reproducibility environment. The framework was implemented in Python v3.6 and Keras API. All results were obtained on GeForce GTX 1080 Ti GPU and 32GB of main memory. We ran every experiment 10 times and reported the average results. We include the description of subnetwork architecture for our proposed framework in the table below (Table 3).

Table 3: Description of subnetwork architecture

Subnetwork	**Layer**	**Input**	**Output**
Water Extraction	*Input*	$22 \times 64 \times 64 \times 11$	$4096 \times 22 \times 11$
	BiLSTM	$4096 \times 22 \times 11$	$22 \times 64 \times 64 \times 128$
	CNN	$22 \times 64 \times 64 \times 128$	$22 \times 64 \times 64 \times 2$
	Output	$22 \times 64 \times 64 \times 2$	$22 \times 64 \times 64 \times 2$
Change prediction	*CNN*	$22 \times 64 \times 64 \times 128$	$1 \times 64 \times 64 \times 2$
	Output	$1 \times 64 \times 64 \times 2$	$1 \times 64 \times 64 \times 2$

5.2. End-to-end evaluation

In this section, we present the end-to-end comparison of our framework against the competing baselines over the three real datasets. The comparison results are shown in Table 4.

In general, our model achieves much higher performance in terms of both water extraction and change prediction *F1-score* over the competing baselines. This improvement arises from the mutual enhancement when training the two tasks in an end-to-end fashion, i.e., the prediction task acts as validated information for the result of the extraction task for each step.

Table 4: End-to-end comparison

	F1 extraction	F1 prediction	Training(s)	Testing(s)
WECP	**94.8%**	**92.9%**	1239	9
Spectral	91.5%	89.7%	1325	15
w-CNN	89%	88.3%	970	9
w-SVM	87.5%	86.2%	330	5
Threshold	81.2%	79.6%	50	50

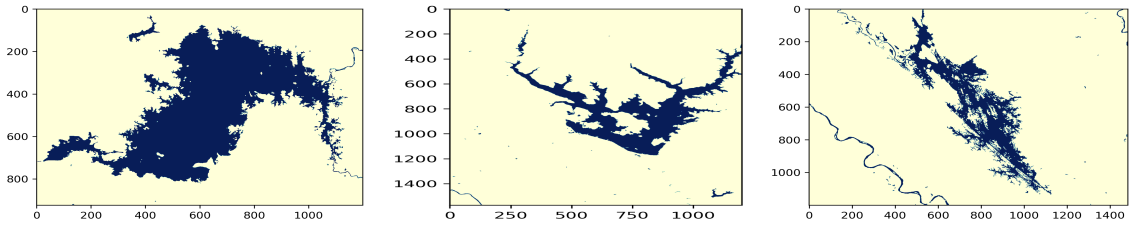
Other baselines are trained separately for each task. Spectral is the best method among the baselines. This is because spectral is a state-of-the-art that, similar to us, adopts an exact order of BiLSTM layers, CNN layers, and upsampling layers in its architecture. Machine learning-based methods (i.e., w-SVM, w-CNN) give moderate performances as they are designed for standard image processing rather than spectral images. The threshold method yields a low performance in overall due to its simplicity in the model design and the lack of generalisation through multiple study regions.

5.3. Evaluation on water extraction task

In this experiment, we further investigate the performance of the water extraction task at a fine-grained level. More precisely, we compute the extraction rates per class (i.e., true positive - TP, false-positive - FP, true negative - TN, false-negative - FN). Table 5 shows the results for Dau Tieng Reservoir, and results for other datasets are omitted as they reveal the same trends. We also visually demonstrate the correctness of the extraction task in Figure 4. We see that our extracted results are almost matched with the ground truth data.

Table 5: Normalized confusion matrices

Baselines	Class	Extracted as Water	Extracted as Nonwater
WECP	Water	97.33% (TP)	2.67% (FN)
	Nonwater	7.93% (FP)	92.07% (TN)
Spectral	Water	94.19%	5.81%
	Nonwater	11.53%	88.47%
w-CNN	Water	90.19%	9.81%
	Nonwater	12.48%	87.52%
w-SVM	Water	88.24%	11.76%
	Nonwater	13.37%	86.63%
Threshold	Water	82.24%	17.76%
	Nonwater	20.37%	79.63%



(a) Tri An (b) Dau Tieng (c) Thac Ba

Figure 4: Qualitative showcases on water extraction task

5.4. Evaluation on change prediction task

In this experiment, we evaluate the performance of our model on the water change prediction task. The baselines are trained separately on each dataset, and the results are presented in Figure 5.

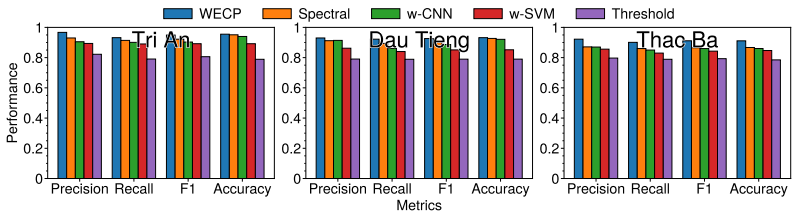


Figure 5: Change prediction performance

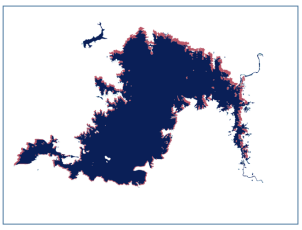


Figure 6: Tri An Dam

In general, our framework performs better compared to the baselines—with over 0.92 F1-score. We further qualitatively demonstrate the correctness of our framework in Figure 6. Although only the showcase for Tri An Dam is presented due to space limitation, other regions reveal the same trends of results.

5.5. Ablation testing

In this section, we assess the quality of each component in our framework, and we, therefore, compare the performance of our framework with several variants. The details of these variants are as follows.

Table 6: Quality of each model component

	F1 extraction	F1 prediction	Training(s)	Testing(s)
WECP	**94.8%**	**92.9%**	1239	9
WECP-1	92.8%	85.4%	605	9
WECP-2	93.8%	88.5%	718	10
WECP-3	91.3%	87.1%	1319	17
Ratio $\beta = 1.0$	95.2%	51.8%	1173	9
Ratio $\beta = 0.9$	94.6%	91.1%	1239	9
Ratio $\beta = 0.5$	92.5%	91.3%	1078	9
Ratio $\beta = 0.1$	92.8%	92.1%	1123	9
Ratio $\beta = 0.0$	64.2%	93.7%	1218	9

- *WECP-1:* We replace the BiLSTM module by a LSTM building block to assess the ability of capturing the past and future observations.
- *WECP-2:* We replace the CNN module by a multi-layer perception network to assess the ability of capturing the spatial information and spectral information.
- *WECP-3:* We replace the upsamping module by a deconvolutional neural network (DNN) to verify the upsampling effect for both extraction and prediction tasks.

Besides, we vary β, which is the trade-off loss shared between the extraction and prediction task. Table 6 summarises the results in terms of F1 prediction score, F1 extraction score, training time, and testing time. We can see that our proposed framework outperforms other variants in both water extraction and change prediction tasks. Another interesting finding is that there is a trade-off when we vary the ratio of the shared loss between the two tasks. When the ratio β is high, meaning that we put more focus on the prediction task, the prediction accuracy is high; and vise-versa. However, with a suitable trade-off (i.e., $\beta = 0.1$), the tasks are coupled together and mutually enhanced, which yields the best performance for both tasks.

6. CONCLUSION

In this work, we developed a specific framework (**WECP**), which can extract the water surface and predict its change over time simultaneously. The framework runs on top of data streams of satellite images to form an end-to-end application for water surface monitoring and forecasting. Intensive evaluations have been conducted on Landsat 8 data, and the results show the advantages of our framework in several different aspects.

The proposed framework has profound implications for authorities that seek sustainable water and environmental monitoring. In the future, we intend to extend our framework in several directions: First, the finding in this study could enhance the accuracy and generalisation of water surface extraction and prediction using optical images. Second, we will extend our works on extracting and predicting more subtle water surfaces, such as the different levels of turbidity and depths. Finally, we aim to build a comprehensive framework for various water mapping applications that supports spatio-temporal analysis.

ACKNOWLEDGMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2019.323.

REFERENCES

[1] T. D. Acharya, A. Subedi, and D. H. Lee, "Evaluation of machine learning algorithms for surface water extraction in a landsat 8 scene of nepal," *Sensors*, vol. 19, no. 12, p. 2769, 2019.

[2] J. C. Aerts, W. J. Botzen, K. C. Clarke, S. L. Cutter, J. W. Hall, B. Merz, E. Michel-Kerjan, J. Mysiak, S. Surminski, and H. Kunreuther, "Integrating human behaviour dynamics into flood disaster risk assessment," *Nature Climate Change*, vol. 8, no. 3, pp. 193–199, 2018.

[3] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Systems with Applications*, vol. 115, pp. 465–473, 2019.

[4] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PloS One*, vol. 12, no. 6, p. e0177544, 2017.

[5] P. Bicheron and M. Leroy, "A method of biophysical parameter retrieval at global scale by inversion of a vegetation reflectance model," *Remote Sensing of Environment*, vol. 67, no. 3, pp. 251–266, 1999.

[6] W. Byeon, M. Liwicki, and T. M. Breuel, "Texture classification using 2d LSTM networks," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 1144–1149.

[7] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[8] R. G. Congalton and K. Green, *Assessing The Accuracy Of Remotely Sensed Data: Principles And Practices*. CRC press, 2019.

[9] G. L. Feyisa, H. Meilby, R. Fensholt, and S. R. Proud, "Automated water extraction index: A new technique for surface water mapping using landsat imagery," *Remote Sensing of Environment*, vol. 140, pp. 23–35, 2014.

[10] C. Giardino, M. Bresciani, P. Villa, and A. Martinelli, "Application of remote sensing in water resource management: The case study of Lake Trasimeno, Italy," *Water Resources Management*, vol. 24, no. 14, pp. 3885–3899, 2010.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT press, 2016.

[12] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[13] R. Gupta, S. J. Nanda, and U. P. Shukla, "Cloud detection in satellite images using multi-objective social spider optimization," *Applied Soft Computing*, vol. 79, pp. 203–226, 2019.

[14] S. K. Jain, A. K. Saraf, A. Goswami, and T. Ahmad, "Flood inundation mapping using noaa avhrr data," *Water Resources Management*, vol. 20, no. 6, pp. 949–959, 2006.

[15] S. K. Jain, R. Singh, M. Jain, and A. Lohani, "Delineation of flood-prone areas using remote sensing techniques," *Water Resources Management*, vol. 19, no. 4, pp. 333–347, 2005.

[16] L. Ji, L. Zhang, and B. Wylie, "Analysis of dynamic thresholds for the normalized difference water index," *Photogrammetric Engineering & Remote Sensing*, vol. 75, no. 11, 2009.

[17] Z. Jiang, J. Qi, S. Su, Z. Zhang, and J. Wu, "Water body delineation using index composition and his transformation," *International Journal of Remote Sensing*, vol. 33, no. 11, 2012.

[18] C. Jing-bo, L. Shun-xi, W. Cheng-yi, Y. Shu-cheng, and W. Zhong-wu, "Research on urban water body extraction using knowledge-based decision tree," *Remote Sensing Information*, vol. 1, 2013.

[19] G. Kallis and D. Butler, "The EU water framework directive: measures and implications," *Water policy*, vol. 3, no. 2, pp. 125–142, 2001.

[20] R. Lan, Z. Li, Z. Liu, T. Gu, and X. Luo, "Hyperspectral image classification using k-sparse denoising autoencoder and spectral–restricted spatial characteristics," *Applied Soft Computing*, vol. 74, pp. 693–708, 2019.

[21] L. Li, Z. Yan, Q. Shen, G. Cheng, L. Gao, and B. Zhang, "Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks," *Remote Sensing*, vol. 11, no. 10, p. 1162, 2019.

[22] P. Liu, H. Zhang, and K. B. Eom, "Active deep learning for classification of hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 2, pp. 712–724, 2017.

[23] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645–657, 2016.

[24] S. K. McFeeters, "The use of the normalized difference water index (ndwi) in the delineation of open water features," *International Journal of Remote Sensing*, vol. 17, no. 7, 1996.

[25] A. Mitchell, G. H. Romano, B. Groisman, A. Yona, E. Dekel, M. Kupiec, O. Dahan, and Y. Pilpel, "Adaptive prediction of environmental changes by microorganisms," *Nature*, vol. 460, 2009.

[26] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 924–935, 2018.

[27] NASA. (2019) Geotiff. [Online]. Available: https://earthdata.nasa.gov/esdis/eso/standards-and-references/geotiff

[28] ——. (2019) Landsat 8. [Online]. Available: https://landsat.gsfc.nasa.gov/landsat-8/

[29] T. T. Nguyen, T. D. Hoang, M. T. Pham, T. T. Vu, T. H. Nguyen, Q.-T. Huynh, and J. Jo, "Monitoring agriculture areas with satellite images and deep learning," *Applied Soft Computing*, vol. 95, p. 106565, 2020.

[30] S. C. Palmer, T. Kutser, and P. D. Hunter, "Remote sensing of inland waters: Challenges, progress and future directions," 2015.

[31] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I. Kweon, "High quality depth map upsampling for 3d-tof cameras," in *ICCV*, 2011, pp. 1623–1630.

[32] L. Qi, D. Yong, N. Xin, X. Jiaqing, and X. Fei, "Remote sensing image classification based on dbn model," *Journal of Computer Research and Development*, vol. 51, no. 9, p. 1911, 2014.

[33] G. Sarp and M. Ozcelik, "Water body extraction and change detection using time series: A case study of lake burdur, turkey," *Journal of Taibah University for Science*, vol. 11, 2017.

[34] P. F. Scheelbeek, F. A. Bird, H. L. Tuomisto, R. Green, F. B. Harris, E. J. Joy, Z. Chalabi, E. Allen, A. Haines, and A. D. Dangour, "Effect of environmental changes on vegetable and legume yields and nutritional quality," *Proceedings of the National Academy of Sciences*, vol. 115, no. 26, pp. 6804–6809, 2018.

[35] V. Slavkovikj, S. Verstockt, W. De Neve, S. Van Hoecke, and R. Van de Walle, "Hyperspectral image classification with convolutional neural networks," in *MM*, 2015, pp. 1159–1162.

[36] F. Sun, W. Sun, J. Chen, and P. Gong, "Comparison and improvement of methods for identifying waterbodies in remotely sensed imagery," *International Journal of Remote Sensing*, vol. 33, no. 21, pp. 6854–6875, 2012.

[37] S. Thirumuruganathan, N. Tang, and M. Ouzzani, "Data curation with deep learning [vision]: Towards self driving data curation," *arXiv preprint arXiv:1803.01384*, 2018.

[38] USGS. (2019) Earth explore. [Online]. Available: https://earthexplorer.usgs.gov/

[39] S. Van Tran, W. B. Boyd, P. Slavich, and T. M. Van, "Agriculture and climate change: perceptions of provincial officials in vietnam," *Journal of Basic and Applied Sciences*, vol. 11, 2015.

[40] Y. Wang, Z. Li, C. Zeng, G.-S. Xia, and H. Shen, "An urban water extraction method combining deep learning and google earth engine," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 768–781, 2020.

[41] G. Xu, P. Li, K. Lu, Z. Tantai, J. Zhang, Z. Ren, X. Wang, K. Yu, P. Shi, and Y. Cheng, "Seasonal changes in water quality and its main influencing factors in the dan river basin," *Catena*, vol. 173, pp. 131–140, 2019.

[42] H. Xu, "Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery," *International Journal of Remote Sensing*, vol. 27, 2006.

[43] S. Yang, C. Xue, T. Liu, and Y. Li, "A method of small water information automatic extraction from tm remote sensing images," *Acta Geodaetica et Cartographica Sinica*, vol. 39, 2010.

[44] Y. Yang, Y. Liu, M. Zhou, S. Zhang, W. Zhan, C. Sun, and Y. Duan, "Landsat 8 oli image based terrestrial water extraction from heterogeneous backgrounds using a reflectance homogenization approach," *Remote Sensing of Environment*, vol. 171, pp. 14–32, 2015.

[45] L. Yu, Z. Wang, S. Tian, F. Ye, J. Ding, and J. Kong, "Convolutional neural networks for water body extraction from landsat imagery," *International Journal of Computational Intelligence and Applications*, vol. 16, no. 01, p. 1750001, 2017.

[46] M. Zhang, M. Gong, and Y. Chan, "Hyperspectral band selection based on multi-objective optimization with high information and low redundancy," *Applied Soft Computing*, vol. 70, pp. 604–621, 2018.

[47] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[48] C. Zhu, X. Zhang, J. Luo, W. Li, and J. Yang, "Automatic extraction of coastline by remote sensing technology based on svm and auto-selection of training samples," *Remote Sensing for Land and Resources*, vol. 25, no. 2, pp. 69–74, 2013.

[49] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images," *Remote Sensing of Environment*, vol. 159, pp. 269–277, 2015.

Received on May 26, 2021
Accepted on January 13, 2022