


Article

An Integrative Remote Sensing Application of Stacked Autoencoder for Atmospheric Correction and Cyanobacteria Estimation Using Hyperspectral Imagery

JongCheol Pyo ¹, Hongtao Duan ² , Mayzonee Ligaray ¹, Minjeong Kim ¹, Sangsoo Baek ¹, Yong Sung Kwon ¹, Hyuk Lee ³, Taegu Kang ³, Kyunghyun Kim ⁴, YoonKyung Cha ^{5,*} and Kyung Hwa Cho ¹

¹ School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 689-798, Korea; jcp01@unist.ac.kr (J.P.); mvligaray@unist.ac.kr (M.L.); paekhap0835@kaeri.re.kr (M.K.); kbcqr12@unist.ac.kr (S.B.); wizkys@unist.ac.kr (Y.S.K.); khcho@unist.ac.kr (K.H.C.)

² Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China; htduan@niglas.ac.cn

³ Water Quality Assessment Research Division, National Institute of Environmental Research, Environmental Research Complex, Incheon 22689, Korea; ehyuk72@korea.kr (H.L.); taegu98@korea.kr (T.K.)

⁴ Watershed and Total Load Management Research Division, National Institute of Environmental Research, Incheon 22689, Korea; matthias@korea.kr

⁵ School of Environmental Engineering, University of Seoul, Dongdaemun-gu, Seoul 130-743, Korea

* Correspondence: ykcha@uos.ac.kr; Tel.: +82-2-6490-2872

Received: 9 January 2020; Accepted: 25 March 2020; Published: 27 March 2020



Abstract: Hyperspectral image sensing can be used to effectively detect the distribution of harmful cyanobacteria. To accomplish this, physical- and/or model-based simulations have been conducted to perform an atmospheric correction (AC) and an estimation of pigments, including phycocyanin (PC) and chlorophyll-a (Chl-a), in cyanobacteria. However, such simulations were undesirable in certain cases, due to the difficulty of representing dynamically changing aerosol and water vapor in the atmosphere and the optical complexity of inland water. Thus, this study was focused on the development of a deep neural network model for AC and cyanobacteria estimation, without considering the physical formulation. The stacked autoencoder (SAE) network was adopted for the feature extraction and dimensionality reduction of hyperspectral imagery. The artificial neural network (ANN) and support vector regression (SVR) were sequentially applied to achieve AC and estimate cyanobacteria concentrations (i.e., SAE-ANN and SAE-SVR). Further, the ANN and SVR models without SAE were compared with SAE-ANN and SAE-SVR models for the performance evaluations. In terms of AC performance, both SAE-ANN and SAE-SVR displayed reasonable accuracy with the Nash–Sutcliffe efficiency (NSE) > 0.7. For PC and Chl-a estimation, the SAE-ANN model showed the best performance, by yielding NSE values > 0.79 and > 0.77, respectively. SAE, with fine tuning operators, improved the accuracy of the original ANN and SVR estimations, in terms of both AC and cyanobacteria estimation. This is primarily attributed to the high-level feature extraction of SAE, which can represent the spatial features of cyanobacteria. Therefore, this study demonstrated that the deep neural network has a strong potential to realize an integrative remote sensing application.

Keywords: deep learning; stacked autoencoder; cyanobacteria; hyperspectral image; feature extraction; dimensionality reduction

1. Introduction

Toxic cyanobacterial blooms have been threatening water resource sustainability and water usage, making it a paramount social and economic problem [1,2]. Over-eutrophication and global warming are the main factors that promote cyanobacteria proliferation [3–7], which occurs due to an excess supply of phosphorus and nitrogen fixation ability. A combination of these conditions accelerates cyanobacterial growth [8–10]. In such circumstances, efficient management strategies are required to prevent freshwater resources from harmful blooms. However, the patchy characteristic of blooms renders huge uncertainties in conventional water sampling [11–13]. Thus, a synoptic monitoring program is considered appropriate for precisely identifying the periodic and spatial proliferation of harmful cyanobacterial bloom.

Remote sensing has been applied to determine the spatial characteristics of harmful cyanobacteria, which can be used to generate a quantitative map of cyanobacteria concentration [14]. Multi-spectral and hyperspectral sensors have been employed to detect cyanobacteria bloom, using phycocyanin (PC) and chlorophyll-a (Chl-a) [15]. In particular, despite the relatively high costs of sensor and image processing, hyperspectral sensing offers high spectral and spatial resolution for capturing the optical and distributional features of cyanobacteria in detail. Several studies have used the hyperspectral images for Chl-a monitoring [16–18]. These studies reveal that airborne hyperspectral imagery can provide accurate spectral and spatial information of the harmful cyanobacteria bloom in optically complex fresh waters.

The preprocessing of hyperspectral images is necessary to retrieve useful information from each image. Specifically, the atmospheric correction (AC) is one of the most important image processing methods that can transform digital signals into light intensity, by eliminating the atmospheric interferences. Commonly used commercial software such as MODTRAN, 6SV, LibRadtran, ATCOR, and FLAASH have been used for implementing AC. However, these models are undesired to perform the AC in certain cases, due to the difficulty of representing dynamically changing atmospheric aerosols and water vapor, adjacency effect, and heterogeneous land surface effect [19]. Previous studies have attributed the AC error to the lack of observed gases in the atmosphere. With the default parameter library of the models, poor atmospheric representation, such as for vapor and liquid cloud, leads to the low simulation accuracy of aerosol and water vapor scattering [20–24]. Moreover, after implementing AC, bio-optical algorithms have been applied, with atmospherically corrected reflectance to estimate cyanobacteria concentration [15]. However, the spectral mixture of phytoplankton, debris, and colored matter is relatively intense in fresh water, thereby necessitating a better understanding of the intricacies of a bio-optical model calibration, to achieve an improved estimation performance. [25–28].

In this context, a data-driven model is an alternative to the deterministic or empirical approach for AC and cyanobacteria estimation. This model can effectively estimate remote sensing reflectance and cyanobacteria pigment without any formulation [29,30]. Conventional neural networks are applied to remote sensing data for estimating water surface reflectance. An artificial neural network (ANN) has been used to calculate remote sensing reflectance, using medium resolution imaging spectrometer (MERIS) data [31]. In addition, the quantification of phytoplankton pigments with high accuracy has been implemented. As seen in [32], MERIS reflectance data on the neural network has been utilized for estimating the Chl-a concentration. Moreover, PC concentration was estimated in [33], using an ANN with water surface reflectance in a hyperspectral image. Furthermore, deep learning has been introduced to strengthen the performance of conventional machine learning. An example of a deep learning method is the application of stacked autoencoder (SAE) to feature learning to reduce the dimensionality of the high-dimensional dataset [19]. As seen in [34], SAE with a backpropagation neural network could predict floods using 10 years of flow datasets. On the other hand, it was proven in [35] that SAE with a deep neural network accurately predicted the hourly passenger flow in a transportation hub by utilizing flow data for 11,996,975 passengers. Based on these findings, the atmospheric influence and light-induced information from AC can be considered by the SAE network for estimating water surface reflectance. To estimate the cyanobacteria concentration and

generate the concentration map, the optical feature of the reflectance spectra is taken into account, wherein the optical feature bands can be reduced in the SAE. In other words, SAE is a promising tool that can be implemented for AC and cyanobacteria estimation. However, only a few studies have been performed using autoencoder with hyperspectral images. Moreover, an integrated remote sensing application for AC and cyanobacteria estimation using a deep neural network has not been realized yet. To address these challenges, this study aimed at achieving the following goals: 1) development of an SAE network for AC and cyanobacteria estimation; 2) generation of quantitative cyanobacteria bloom maps; and, 3) comparison of the SAE models with a conventional machine learning model for model performance evaluation.

2. Materials and Methods

2.1. Study Site

The Baekje reservoir is located at the Geum River in South Korea, particularly at the mid- western region ($36^{\circ} 31' 87.75''\text{N}$, $126^{\circ} 93' 90.52''\text{E}$) (Figure 1). Baekje Weir has a length of 23 km, a basin area of $7,976 \text{ km}^2$, and a water storage capacity of $24.2 \text{ million}\cdot\text{m}^{-3}$. Most of the water is consumed for domestic, industrial, and agricultural purposes. Recently, cyanobacterial blooms have been occurring at Baekje Weir during summer, mainly due to the excessive nutrient supply from non-point sources including soil erosion and runoff from livestock farms, as well as rural and domestic wastes [36].

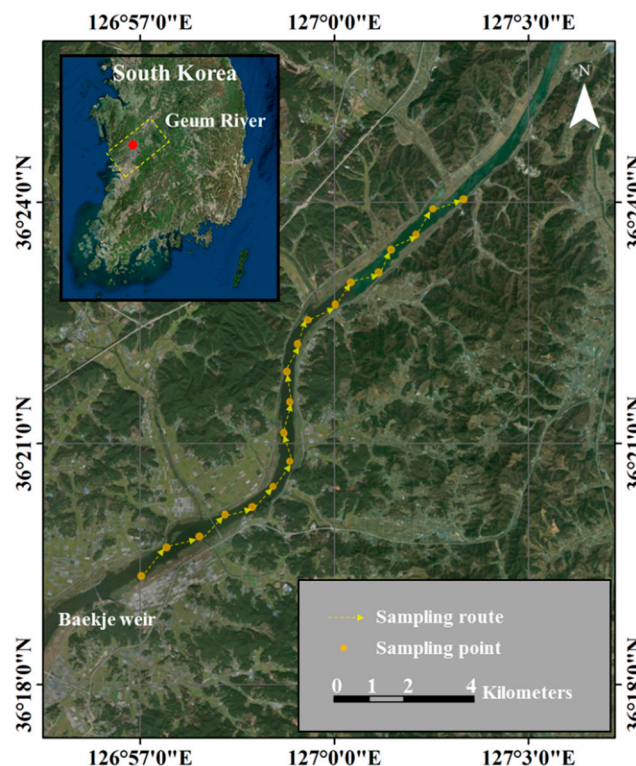


Figure 1. Study area, Baekje Weir in Geum River, with sampling route. The sampling points were mainly assigned in river side. Map showing location of Baekje Weir region (Google Earth, earth.google.com/web/).

2.2. Dataset

2.2.1. Field and Experimental Data

This study implemented field monitoring and experimental analysis four times in 2016 and five times in 2017. Table 1 shows the overall sampling information of algal pigments, water temperature,

and the number of sampling points. The total number of sampling points allowed for the identification of the amount of data that was utilized in the training and validation of the deep learning model. Each monitoring point contained observed data, including that of PC, Chl-a, and surface reflectance spectra. During the field sampling, airborne monitoring was conducted for hyperspectral imagery sensing along the Baekje Weir region. Monitoring was conducted under clear sky conditions. A field spectroradiometer (ASD FieldSpec 4 Hi-Res; ASD Inc., USA) was used to measure the optical parameters, such as downwelling irradiance, downwelling radiance, and water-leaving radiance. The spectroradiometer had a spectral range from 350 nm to 2500 nm, with optical data being recorded at 1 nm interval. The spectral resolution of the device was 3 nm at approximately 700 nm and 8 nm at 1400 nm and 2100 nm, with the spectral bandwidth being 1.4 nm from 350 nm to 1000 nm and 1.1 nm from 1001 nm to 2500 nm. The measured optical parameters were used to calculate the remote sensing reflectance of the water surface, using the following equation:

$$R_{rs} = \frac{L_w - 0.025 \times L_{sky}}{E_d} \quad (1)$$

where R_{rs} is the remote sensing reflectance (sr^{-1}), L_w is the water-leaving radiance ($W \cdot sr^{-1} \cdot m^{-2}$), L_{sky} is radiance from the sky ($W \cdot sr^{-1} \cdot m^{-2}$), and E_d is irradiance from the sky ($W \cdot m^{-2}$). The downwelling irradiance was measured by a cosine detector fore-optic. And, the radiance data were measured with bare fiber fore-optic. The measurement positions of the field spectroradiometer were strictly maintained for the zenith angle less than 42° and azimuth angle less than 135° [37]. This study adopted the skylight correction as a constant value of 0.025, by considering clear sky and gentle breeze condition (i.e., wind speed $< 5 \text{ m s}^{-1}$) [38]. The remote sensing reflectance data was then used to evaluate the AC performance of deep learning approaches.

At the same location of measuring the optical data, Water samples were collected from the same locations as the optical data to determine the algal pigment concentrations at Baekje Weir. Water bottles of 2 L capacity were for the sample collection for Chl-a analysis. In addition, plankton net (DAIHAN CHEMLAB Inc. South Korea) having $20 \mu\text{m}$ mesh size was used to concentrate water of 10 L. The 100 mL water bottle contained concentrated samples for PC analyses. All water samples were preserved in an ice box and transported to the laboratory immediately after field sampling for pigment extraction. Chl-a concentration was analyzed as the biomass indicator of algae [39]. The solvent extraction method was used to extract the Chl-a pigment [40].

A freezing and thawing method was implemented to extract PC pigment, which is an indicator of cyanobacteria biomass [41]. The water samples were homogenized using a sonicator (Sonictopia Inc., South Korea), and were then centrifuged at 4000 rpm at 4°C for 15 minutes. 5 mL of phosphate buffer (pH 7.2) was then added to the remaining pellets. These resulting samples were then stored in a dark room for 24 h at -20°C . After the freezing step, the samples were thawed at room temperature. The samples were agitated using a shaking incubator (N-BIOTECK Inc. South Korea), at a speed of 150 rpm. The combination of freezing, thawing, and shaking processes facilitated the release of the PC pigment, without releasing the Chl-a pigment. After which, the samples were centrifuged at 4000 rpm at 4°C for 15 minutes. The absorbance of the supernatant of the samples was measured using a Cary-5000 UV-VIS-NIR spectrophotometer. The following equation was used to determine the PC concentration:

$$PC \text{ (mg m}^{-3}\text{)} = \frac{OD(620) - (q \times OD(652))}{p} \quad (2)$$

where $OD(620)$ is the optical density at 620 nm; $OD(652)$ is the optical density at 652 nm; q is 0.474; and p is 5.34 referred by [41].

Table 1. Phycocyanin (PC) and Chlorophyll-a (Chl-a) results of each sampling event.

	PC (mg m ⁻³)		Chl-a (mg m ⁻³)		Point	AT (°C)	C * (Cell mL ⁻¹)	D ** (Cell mL ⁻¹)	G *** (Cell mL ⁻¹)
	Range	Mean	Range	Mean					
08.12.2016	6.04–146.99	35.46 ± 36.10	14.19–111.40	40.65 ± 23.38	18	31.06	4,224–35,584	192–2,304	384–5,888
08.24.2016	12.48–100.00	39.43 ± 23.40	25.95–61.44	37.39 ± 8.21	19	30.33	2,048–20,544	96–672	512–20,640
09.20.2016	0.83–1.64	1.23 ± 0.27	11.85–60.88	25.51 ± 11.32	17	22.13	0–128	1,888–4,672	512–5,376
10.14.2016	0.19–0.88	0.34 ± 0.17	13.74–46.17	28.21 ± 9.38	20	17.97	0–224	640–3,968	0–512
09.15.2017	7.41–9.66	8.34 ± 0.66	30.24–61.52	47.28 ± 8.94	12	22.30	-	-	-
09.22.2017	7.64–21.69	12.63 ± 3.96	14.08–27.89	17.57 ± 3.98	12	23.60	-	-	-
10.25.2017	2.64–4.56	3.51 ± 0.67	10.56–20.92	13.18 ± 2.99	12	17.25	-	-	-
10.28.2017	1.18–14.77	4.35 ± 4.52	8.45–16.73	10.54 ± 2.39	12	16.55	-	-	-
11.11.2017	0.23–0.71	0.34 ± 0.14	12.76–38.43	22.58 ± 6.95	12	12.93	-	-	-

AT indicates average temperature, * Cyanobacteria is *Microcystis aeruginosa* for August 12, September 20, and October 14 in 2016 and *Oscillatoria sp.* for August 24, 2016, ** diatom is *Aulacoseira granulata*, and *** green algae is *Coelastrum cambricum*.

2.2.2. Hyperspectral Image Sensing

The AISA eagle sensor (SPECIM Inc., Finland) attached to an aircraft captured the hyperspectral images of the Baekje Weir. The airborne monitoring was conducted when the zenith angle of the sun was between 35° and 65°, in order to minimize the sun glint and shading effect. The flying time was less than 3 hours and the flying altitude was 3 km above the ground. The hyperspectral sensor has a full width at half maximum (FWHM), from 4.36 nm to 4.82 nm. A spectral information of the sensor has signal to noise ratio (SNR) as 1,250:1. The field of view (FOV) and instantaneous field of view (IFOV) were 39.7 degrees and 0.039 degrees, respectively. In addition, the swath width of the AISA eagle sensor had 1024 pixels, with a spatial resolution of 2 m. The sensor had a spectral range from 400 nm to 970 nm, with a spectral resolution of 4-5 nm. Image processing was implemented using MODTRAN 6 software. The MODTRAN is a scalar radiative transfer code calculating AC parameters (i.e., path radiance, solar flux, direct transmittance, diffuse transmittance, and spherical albedo) [42]. The default atmospheric condition was assigned, simulating the software. The statistical band model assigned the radiative transfer algorithms for generating atmospheric correction parameters from MODTRAN 6. Specifically, the multiple scattering algorithm was selected for the discrete ordinate radiative transfer algorithm. A mid-latitude summer atmospheric profile was selected and 400 ppmv of CO₂ concentration was set for the atmospheric profile. The aerosol specification was set to rural boundary aerosol. Furthermore, the sampling time and geographic coordinates were used for solar geometry, including solar zenith angle and azimuth angle. More detailed information of MODTRAN 6 implementation is described in [43]. The AC parameters and digital numbers from 400 nm to 800 nm were then used as the input dataset of the data-driven models, to directly estimate water surface reflectance, thereby sequentially estimating cyanobacteria concentrations.

2.3. Data-Driven Models

2.3.1. Autoencoder

Autoencoder is a neural network for unsupervised feature learning [44]. The typical structure of the autoencoder is presented in Figure 2a. The representative layers of the autoencoder are composed of an encoder and a decoder, that are composed of the following nonlinear autoencoder functions:

$$f(x) = e_f(W_e x_i + b_e) \quad (3)$$

$$g(x) = d_f(W_d f(x_i) + b_d) \quad (4)$$

where $f(x)$ and $g(x)$ are the encoder and decoder functions, respectively; W_e and W_d represent the weight matrices, while b_e and b_d are the bias vectors. For the activation function, sigmoid function was utilized by the encoding and decoding layers, as given in Equations (5) and (6):

$$e_f = \frac{1}{1 + e^{-(W_e x_i + b_e)}} \quad (5)$$

$$d_f = \frac{1}{1 + e^{-(W_d f(x_i) + b_d)}} \quad (6)$$

In the encoding layer, the image pixels are fed as input feature. Spectral and spatial information of the input pixels are then compressed and encoded to the middle layer, thereby reducing the number of hidden nodes. In the decoding layer, the terminal nodes are reconstructed to be identical to the original input image.

The encoding layer transforms high-dimensional data into low-dimensional data, while decoding recovers the low-dimensional data and turns it into a high-dimensional data that is identical to the original input structure [19]. Herein, the hidden nodes of the autoencoder layers deal with manifold features from the hypercubes of the input image. In particular, the autoencoder has advantages in

feature extraction and dimensionality reduction of nonlinear data [45]. However, it is only limited to a small number of spectral bands. Handling hundreds of hyperspectral bands would be inadvisable for the autoencoder, since the data complexity causes difficulty in extracting proper abstractions of the input feature. Thus, this study introduced a variant autoencoder network in the form of the SAE. Detailed information and the mathematical formula of the SAE are explained in the following section.

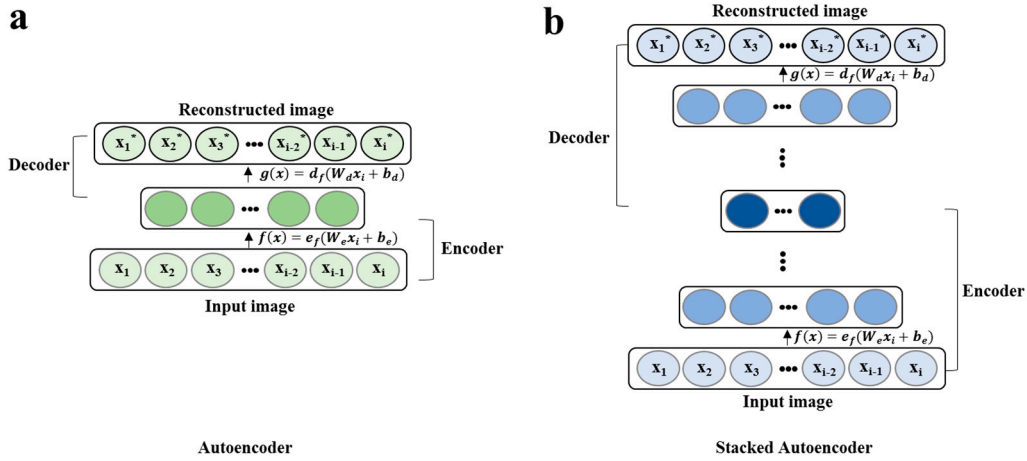


Figure 2. Neural network structure: (a) autoencoder with encoder and decoder and deep neural network structure; (b) stacked autoencoder with multiple encoders and decoders. Rectangular boxes and circles represent the internal hidden layers and nodes, respectively.

2.3.2. Stacked Autoencoder (SAE)

The fundamental principle of the SAE is similar to that of the original autoencoder network. SAE is an alternative to the basic autoencoder network, when dealing with complex feature information of the hyperspectral image cube [46]. Contrary to the autoencoder that has a single hidden layer, SAE consists of multiple encoding and decoding layers (Figure 2b), as represented by the following equations:

$$f_k(x) = e_{k,f}(W_{k,e}x_{k,i} + b_{k,e}) \tag{7}$$

$$g_t(x) = d_{t,f}(W_{t,d}f(x_{t,i}) + b_{t,d}) \tag{8}$$

where $f_k(x)$ and $g_t(x)$ are the encoder and decoder functions in the k -th and t -th layer, respectively, $W_{k,e}$ and $W_{t,d}$ represent the weight matrices in the k -th and t -th layer, while $b_{k,e}$ and $b_{t,d}$ are the bias vectors.

To optimize the SAE network, the error between input and output data should be minimized. The mean squared error (MSE) of each iteration was determined, while the lowest MSE value was identified using the cost function below:

$$Y = \min \left(\frac{\sum_{i=1}^N (g(x) - I_o)^2}{N} \right) \tag{9}$$

where Y is the cost function, N denotes the number of nodes, $g(x)$ represents the reconstructed input, and I_o is the original input. The input data for AC included the AC parameters and digital number, while remote sensing reflectance with 86 bands between 400 nm and 800 nm was the input for the cyanobacteria estimation. To train the SAE network, the backpropagate error derivatives update the network parameters in the autoencoder layers in the network using the function in Equation (10)

$$\delta = \frac{\partial Y}{\partial A_f} \tag{10}$$

where A_f represents the autoencoder functions.

2.3.3. Stacked Autoencoder with ANN and SVR

This study utilized the feature extraction and dimensionality reduction of the SAE network, to implement AC and cyanobacterial estimation with artificial neural network (ANN) and support vector regression (SVR), as fine-tuning operators of the SAE network. The ANN model is a feedforward neural network capable of the regression task with nonlinear environmental data [47]. The hidden layer of ANN model is composed of trainable weight and biases in the hidden nodes. These nodes capture the input features after which deliver the traits to the consecutive layer, by using the nonlinear activation function. The training of the ANN model optimized the weights and biases, in order to minimize the error between measured and estimated results. The SVR model has been utilized for the regression problem with multivariate datasets. The SVR model projects the training data to the higher dimensional feature space, utilizing nonlinear kernel function [48]. The kernel function makes the nonlinear data into linear in the feature space for solving linear regression. After assigning kernel function, the SVR model is trained to minimize the error between observed and estimated data.

The SAE network with ANN and SVR was able to provide water surface reflectance from AC, and PC and Chl-a pigments from cyanobacteria. The path radiance, solar flux, direct transmittance, diffuse transmittance, and spherical albedo were assigned as atmospheric influence input, and digital numbers were represented to the optical information input for AC. These data were fed into SAE network input for atmospheric and optical feature extraction, after which consecutive ANN and SVR models estimated the surface reflectance spectra. The estimated reflectance data were fed into a sequential SAE model for extracting features of water surface reflectance, thereby estimating algal pigment concentration in the consecutive models. These comprehensive processes and data compositions followed the conventional remote sensing application for water quality estimations.

To run the SAE, ANN, and SVR, the TensorFlow library was adopted. Figure 3 shows the deep neural network structure composed of two SAE networks, which were followed by the fine-tuning layers. The parameters of the data-driven model were adjusted using several empirical experiments [49,50]. The learning rate, number of hidden nodes and layers, activation function, and kernel functions were significant variables for the data-driven model performance. For the convenience of the reader, SAE with ANN and SVR are denoted as SAE-ANN and SAE-SVR, respectively.

2.3.4. Model Comparison

This study evaluated and compared the performances between the conventional machine learning models and deep neural network models. ANN and SVR models without SAE were implemented to estimate water surface reflectance and cyanobacterial concentration. The learning rate and the number of layers and nodes were adjusted iteratively. In addition, the different activation functions of ANN and the different kernel functions of SVR were tested and adopted based on their performances. This study compared the performances between SAE-ANN, SAE-SVR, ANN, and SVR, in estimating the water surface reflectance and cyanobacterial concentration; 70% and 30% of the input data were used as the training and validation dataset, respectively.

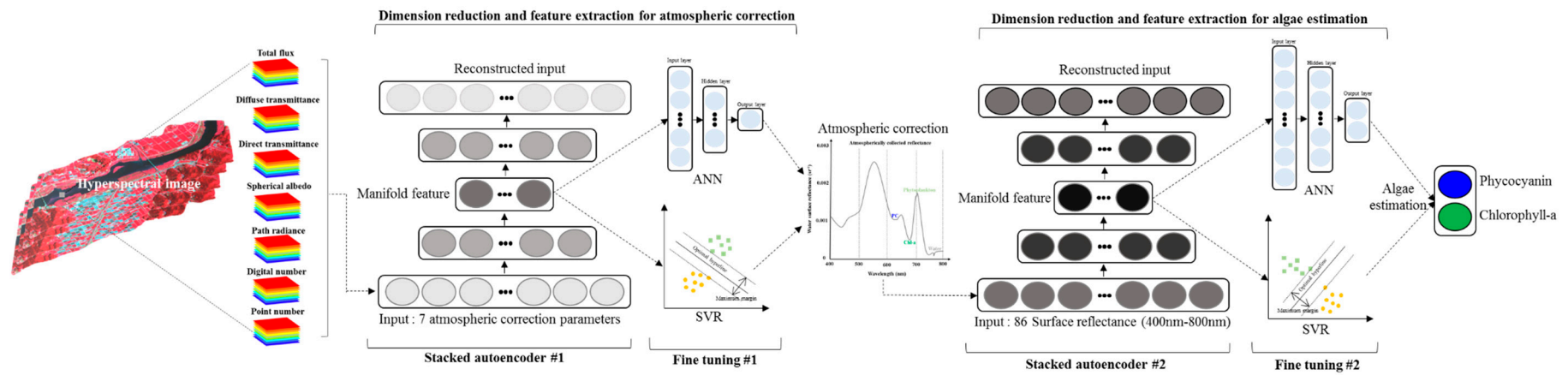


Figure 3. Stacked encoder with fine tunings (i.e., artificial neural network [ANN] and support vector regression [SVR]) for atmospheric correction and cyanobacterial estimation, stacked autoencoder#1 and fine tuning#1, for the water surface reflectance estimation using hyperspectral image data inputs including total flux, diffuse transmittance, direct transmittance, spherical albedo, path radiance, digital number, and point sample number, stacked autoencoder#2 and fine tuning#2 for the PC and Chl-a estimations, using atmospherically-corrected reflectance spectra.

2.4. Accuracy

The performance of the data-driven model was evaluated using the root mean squared error (RMSE), mean absolute error (MAE), and Nash–Sutcliffe efficiency (NSE). The RMSE, MAE, and NSE functions are represented by Equations (11)–(13), respectively:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^n (P_t - O_t)^2}{n}} \quad (11)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{O_t - P_t}{O_t} \right| \quad (12)$$

$$\text{NSE} = 1 - \frac{\sum_{t=1}^n (P_t - O_t)^2}{\sum_{t=1}^n (O_t - O_a)^2} \quad (13)$$

where P_t is the estimated surface reflectance (sr^{-1}), PC (mg m^{-3}), or Chl-a (mg m^{-3}); O_t is the observed surface reflectance, PC, or Chl-a; O_a is the average surface reflectance, PC, or Chl-a; and n is the number of samples.

3. Results

3.1. Variations in Concentrations of the Observed Pigments

Table 1 shows the concentrations of PC and Chl-a. This information was used to identify the temporal variations in the PC concentration as ranging between 0.19 mg m^{-3} to 146.99 mg m^{-3} and Chl-a concentration from 8.45 mg m^{-3} to 111.40 mg m^{-3} during the monitoring periods. Water temperature varied from 12.93°C to 31.06°C during the sampling periods. In particular, the pigments data collected in August 2016 showed considerable variations, with PC ranging between $6.04\text{--}146.99 \text{ mg m}^{-3}$ and Chl-a between $14.19\text{--}111.40 \text{ mg m}^{-3}$. The high PC concentration indicated the outbreak of cyanobacterial blooms. It was found that the dominant cyanobacterial genera were *Microcystis* and *Oscillatoria* (Table 1).

3.2. AC Performance of SAE

This study adopted SAE #1 layer configuration to a 7-6-5-3-5-6-7 hidden node, with encoding and decoding layers (Figure 3). The first layer with seven nodes represents the input layer, consisting of five AC parameters, a digital number, and a sampling event number for each wavelength band. After training the SAE #1, the manifold feature layer (middle layer with three nodes) was used as the input for atmospheric correction of the fine-tuning operators (ANN and SVR). The ANN model had 3-10-5-1 nodes for each layer, wherein the input layer with three nodes corresponded to the results of the manifold feature layer. Meanwhile, the ANN output layer estimated the surface reflectance for each wavelength band, resulting in a total of 86 water surface reflectance values. For the SVR models, radial basis function (RBF) was implemented and optimized as the kernel function. Without SAE, the ANN model had a 7-6-1 node configuration to estimate the water surface reflectance. Furthermore, the SVR model was performed for AC by adopting RBF.

Figure 4 presents the training and validation results of AC, and their respective R^2 and RMSE values. The training and validation results included 134 observation points that had 86 reflectance bands in each point, resulting in a total of 11,524 (i.e., 134×86) comparable points. The SAE-ANN yielded R^2 values of 0.73 and 0.74, and an RMSE of 0.0019 sr^{-1} and 0.0018 sr^{-1} (Figure 4a), while SAE-SVR showed its accuracy by having R^2 values of 0.73 and 0.70, and an RMSE of 0.0019 sr^{-1} and 0.0019 sr^{-1} (Figure 4c) for training and validation, respectively. Figure 5 shows the comparison observed reflectance spectra, with estimated spectra from SAE-ANN. The training and validation results had good agreement with in-situ reflectance spectra. In particular, the estimated spectra from 600 nm to 700 nm was able to describe the PC peaks (i.e., 615 nm - 622 nm) and the Chl-a peaks (i.e., 660 nm - 670 nm).

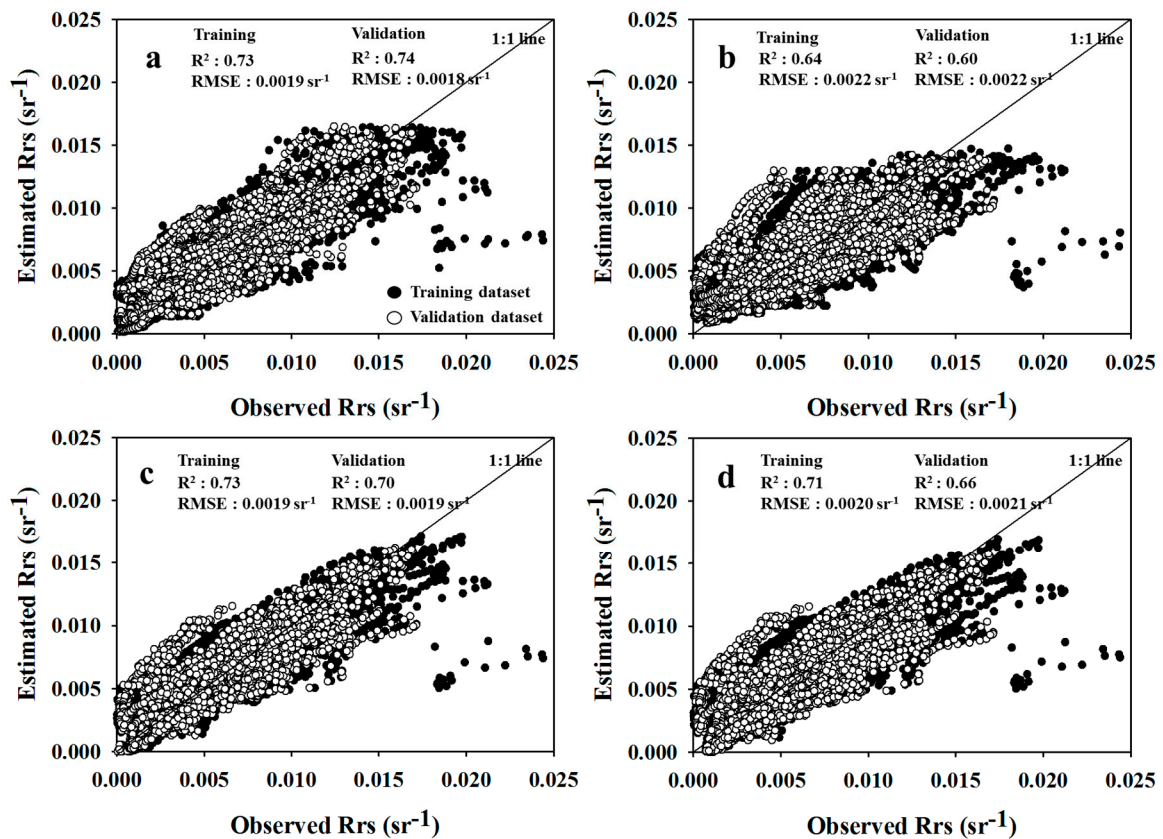


Figure 4. Training and validation results of the atmospheric correction: (a) SAE-ANN, (b) ANN, (c) SAE-SVR, and (d) SVR. A total of 134 observation points is presented. Each point had observed reflectance spectra with 86 bands, indicating 11,524 (i.e., 134×86) points, comparing the estimated reflectance spectra to the observed spectra of total bands. Black and white dots indicate training and validation results, respectively

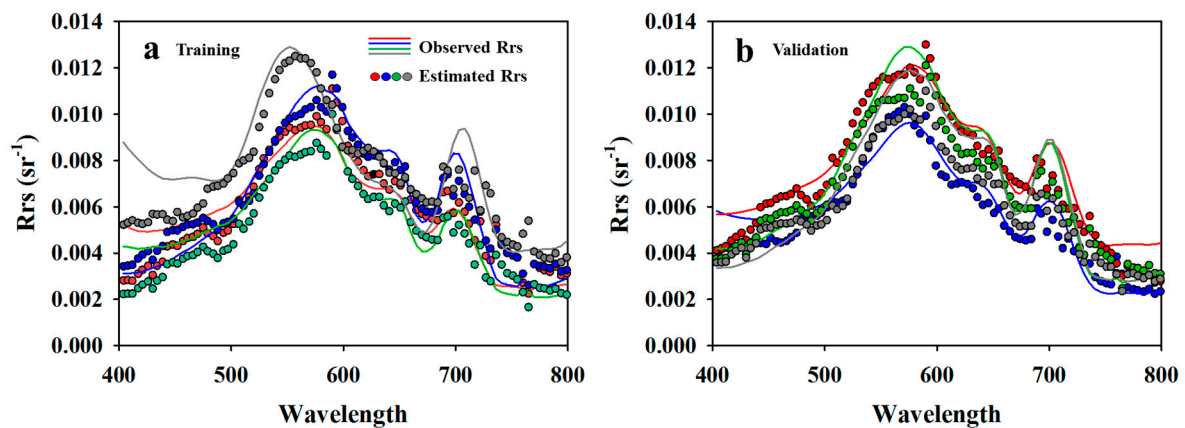


Figure 5. Example of comparison between observed reflectance spectra and estimated spectra: (a) training result of reflectance and (b) validation result. Straight lines indicate the observed reflectance and the marks represent the estimated reflectance from SAE-ANN. For training results, four estimated and observed spectra were selected with respect to 15th point on August 12, 2016, 7th point on September 20, 2016, 7th point on September 11, 2017, and 10th point on October 25th, 2017. For validation results, four estimated and observed spectra were selected with 13th point on September 20, 2016, 20th point on October 14, 2016, 7th point on September 15, 2017, and 1st point on October 25, 2017.

3.3. Cyanobacteria Estimation of SAE

The estimated water reflectance was then used as input of the SAE #2. Seven layers, with 86-60-40-20-40-60-86 node configurations, were adopted. The 86 nodes in the input layer represent the estimated reflectance of the 86 bands. Then, the concentrated feature layer of ANN (middle layer with 20 nodes) was used as input for the second ANN and SVR, that estimated the cyanobacterial concentration. The consecutive ANN model for the cyanobacterial estimation had 20-10-5-2 node configuration, that yielded the PC and Chl-a concentrations. Among the applied activation functions for the ANN model, the sigmoid function was adopted, by showing relatively accurate model performance compared to the other activation functions. Then, the learning rate of 0.0001 was set. For the SVR models, RBF was utilized as the kernel function. The reconstruction of the SAE input showed an RMSE value of $5.4 \times 10^{-7} \text{ sr}^{-1}$. Moreover, the ANN model without SAE was performed by having an 86-2 node configuration for estimates of PC and Chl-a concentrations. The ANN models adopted a sigmoid function as the activation function, with a learning rate of 0.0001. The SVR model without SAE was conducted, to estimate cyanobacteria with RBF.

Figure 6 shows the results of PC estimation. SAE-ANN showed a satisfactory performance with R^2 values of 0.82 and 0.83 and RMSE values of 9.32 mg m^{-3} and 9.76 mg m^{-3} , with respect to training and validation (Figure 6a). For SAE-SVR, the results also showed R^2 values of 0.80 and 0.62 and RMSE values of 15.44 mg m^{-3} and 17.94 mg m^{-3} , as shown in Figure 6c. Figure 7 presents the overall results of SAE with the fine-tuning operators, in terms of the Chl-a estimation. Training and validation results of SAE-ANN had R^2 values of 0.81 and 0.78 and RMSE values of 7.33 mg m^{-3} and 6.34 mg m^{-3} (Figure 7a), while SAE-SVR had 0.79 and 0.78 for R^2 and 9.08 mg m^{-3} and 8.08 mg m^{-3} for RMSE (Figure 7c).

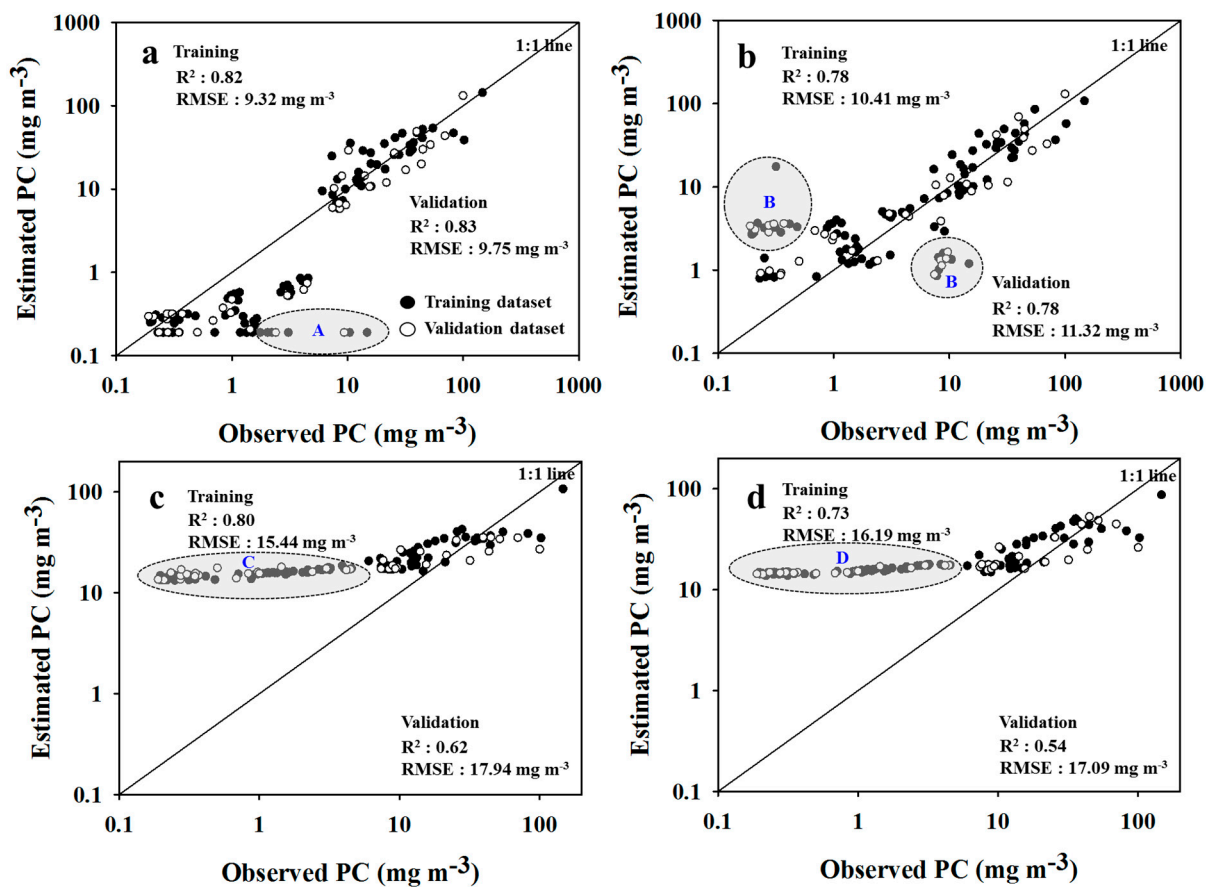


Figure 6. Training and validation results of PC estimation: (a) SAE-ANN, (b) ANN, (c) SAE-SVR, (d) SVR. Black and white dots represent training and validation results, respectively. Highlighted region of A-D indicates discrepancy in the estimated PC and Chl-a, compared to the observed ones.

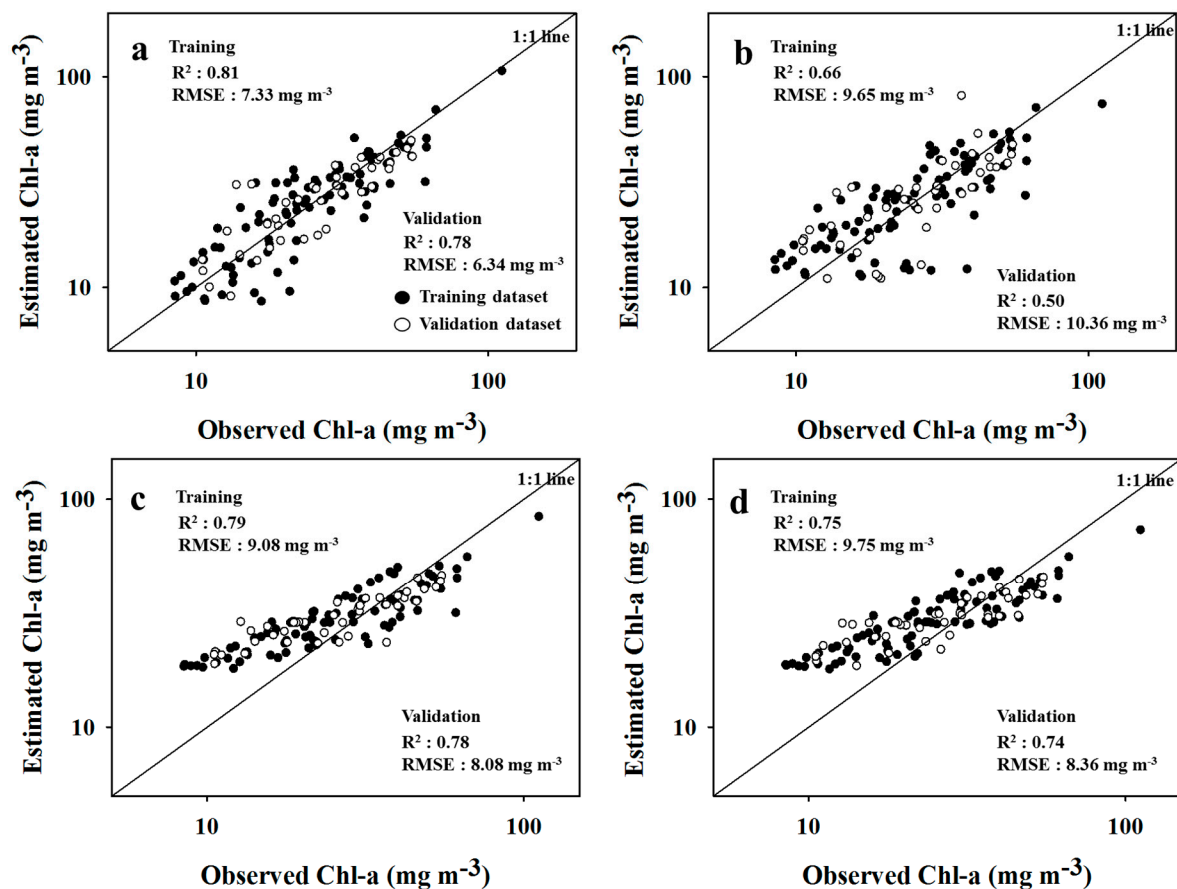


Figure 7. Training and validation results of Chl-a estimation, (a) SAE-ANN, (b) ANN, (c) SAE-SVR, and (d) SVR. Black and white dots represent training and validation results, respectively.

The trained SAE-ANN and -SVR models were applied to generate the PC and Chl-a maps shown in Figure 8. In the figure, the PC and Chl-a concentration levels of SAE-SVR were lower than those of SAE-ANN, due to the tendency of SVR to underestimate PC and Chl-a. Regardless, both models were still able to generate spatial distribution maps, indicating that SAE has the capacity to represent the nonlinear spatial feature of the cyanobacteria by comparing RGB images. (Figure S1).

The SAE-ANN model was able to capture the temporal variation of the cyanobacteria, in terms of the PC and Chl-a concentrations. A relatively low concentration was observed in autumn compared to summer. However, the Chl-a maintained a concentration level $> 10 \text{ mg m}^{-3}$ in autumn (Figure 8f–h). Meanwhile, the spatial dynamic of the cyanobacteria peaked in summer, which eventually lessens in autumn (Figure 8).

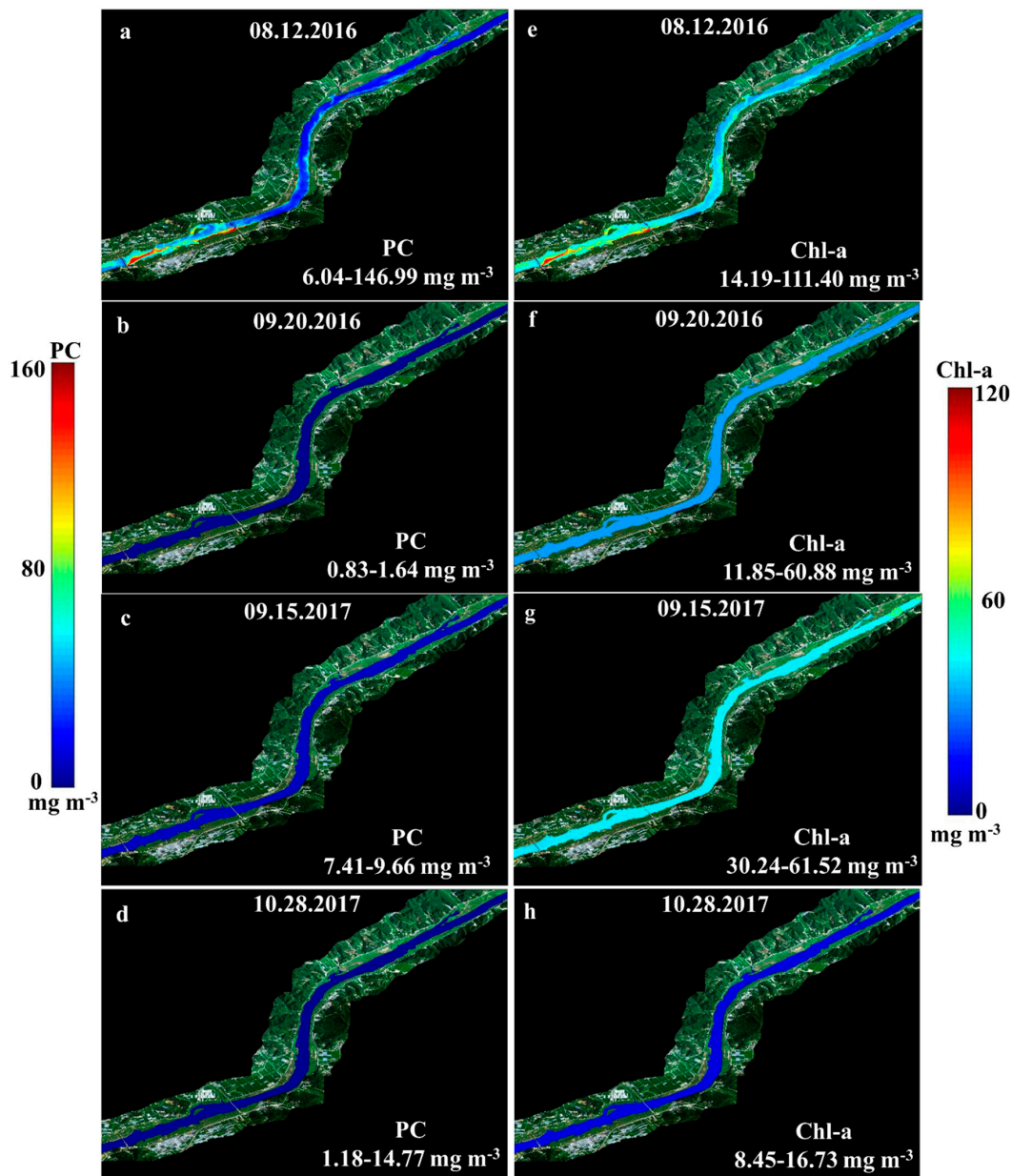


Figure 8. Temporal variations in cyanobacterial maps of SAE-ANN: (a–d) PC variation for August 12 and September 20 in 2016, and September 15 and October 28 in 2017; (e–h) Chl-a variation for August 12 and September 20 in 2016, and September 15 and October 28 in 2017.

3.4. Model Comparison

The ANN and SVR models estimated the water surface reflectance and cyanobacteria concentration, without the feature extraction. The evaluation results of both models for surface reflectance are presented in Figure 9a,b. Compared to SAE-ANN and SAE-SVR, the ANN and SVR models showed higher MAE values, > 0.75 for training and > 0.58 for validation (Table 2). This study also ran the conventional model MODTRAN 6 from [51] for AC, to compare the results from the data-driven models. The accuracy of MODTRAN 6 based-AC showed an R^2 of 0.69 and RMSE of 0.0021 sr^{-1} (Figure 9a,b). Among the models, SAE-ANN showed the best AC performance in terms of training and validation.

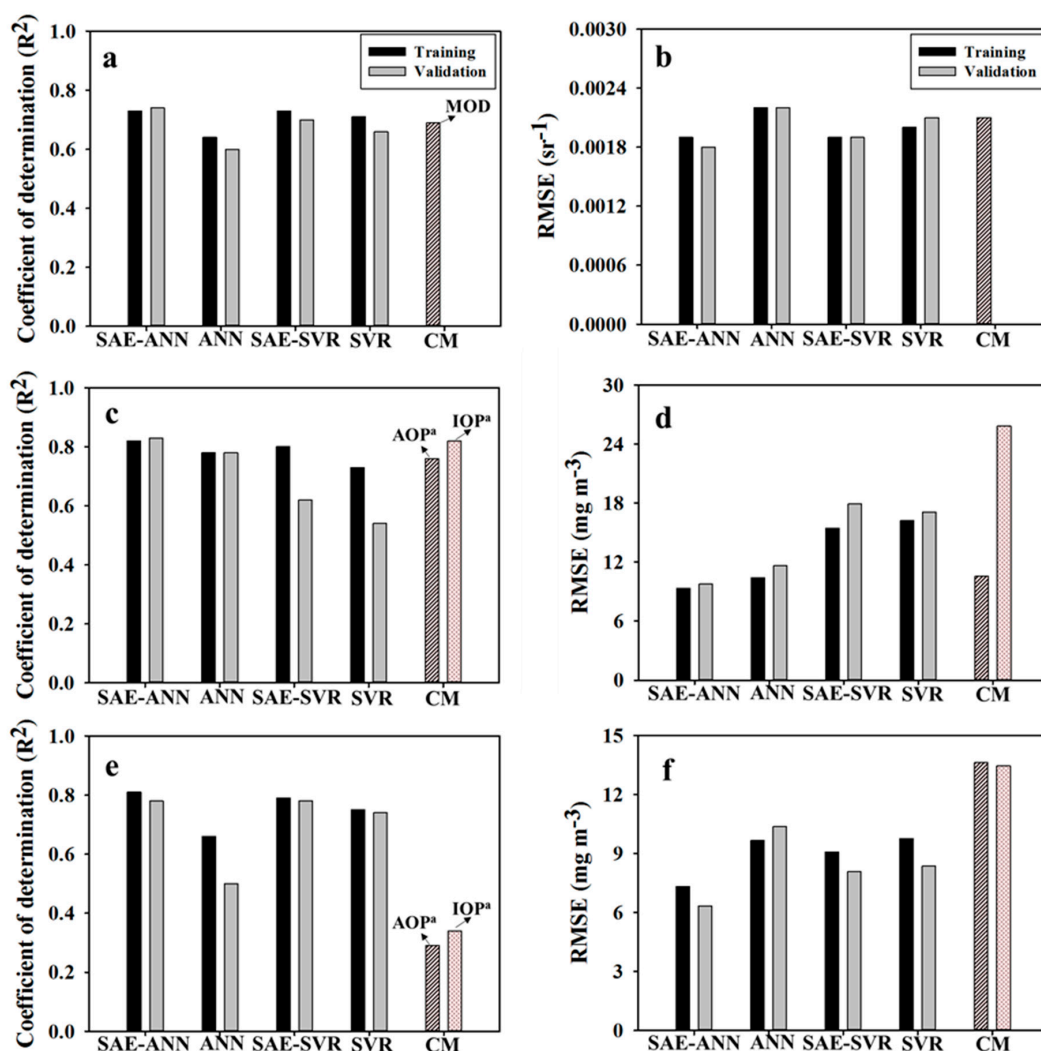


Figure 9. Accuracy of the data-driven model and the conventional model (CM): (a–b) AC, (c–d) PC estimation, and (e–f) Chl-a estimation. Black and gray bars represent training and validation accuracy of the data-driven models (MOD indicates water surface reflectance simulated by MODTRAN 6, AOP^a and IOP^a presents two-band ratio algorithm and the Simis algorithm for PC and Chl-a estimation).

In Figure 9c,d, the ANN model showed an PC estimation with $R^2 > 0.78$, for both training and validation, while the SVR model showed a lower validation performance. Although the coupling of SAE and SVR improved the accuracy of the SVR model, SAE-SVR still needs further development. Likewise, though the ANN model showed MAE value > 2.54 , it was still higher than that shown by SAE-ANN. The SVR model showed Chl-a results by yielding R^2 values > 0.74 for training and validation (Figure 9e,f); however, SAE-SVR showed a better performance than SVR, by having higher R^2 and NSE values. SAE-ANN also significantly improve the accuracy of ANN results, and showed the best performance, as well as the lowest MAE < 0.22 for Chl-a estimation among the four models (Table 2). In addition, the SAE-ANN showed a relatively better performance for cyanobacterial estimation, compared to the conventional bio-optical algorithms, the two-band ratio and the inherent optical property (IOP) algorithms [14]. The accuracy of the two-band ratio algorithm for the PC estimation showed an R^2 of 0.76 and an RMSE of $10.56\ mg\ m^{-3}$, while the IOP algorithm yielded 0.82 for R^2 and $25.83\ mg\ m^{-3}$ for RMSE (Figure 9c,d). For the Chl-a estimation, the conventional algorithms showed relatively low R^2 and high RMSE values, having 0.29 and $13.62\ mg\ m^{-3}$ for the two-band ratio and 0.34 and $13.45\ mg\ m^{-3}$ for the IOP algorithm (Figure 9e,f), respectively.

Table 2. Deep neural network and conventional machine learning performance.

		SAE-ANN				ANN				SAE-SVR				SVR			
		R ²	NSE (Nash-Sutcliffe Efficiency)	RMSE	MAE (Mean Absolute Error)	R ²	NSE	RMSE	MAE	R ²	NSE	RMSE	MAE	R ²	NSE	RMSE	MAE
R_{rs}	T [*]	0.73	0.73	0.0019	0.68	0.64	0.63	0.0022	0.75	0.73	0.73	0.0019	0.75	0.71	0.70	0.0020	0.78
	V ^{**}	0.74	0.73	0.0018	0.41	0.60	0.60	0.0022	0.59	0.70	0.69	0.0019	0.50	0.66	0.65	0.0021	0.58
PC	T	0.82	0.82	9.32	0.49	0.78	0.78	10.41	2.47	0.80	0.51	15.44	13.37	0.73	0.46	16.19	13.50
	V	0.83	0.79	9.76	0.47	0.78	0.72	11.62	2.54	0.62	0.31	17.94	17.02	0.54	0.37	17.09	16.59
Chl-a	T	0.81	0.81	7.33	0.22	0.66	0.66	9.65	0.28	0.79	0.70	9.08	0.37	0.75	0.66	9.75	0.38
	V	0.78	0.77	6.34	0.21	0.50	0.38	10.36	0.31	0.78	0.63	8.08	0.36	0.74	0.60	8.36	0.36

* indicates training result and ** represents validation result.

4. Discussion

4.1. AC and Cyanobacteria Estimation

The NSE values of SAE-ANN and SAE-SVR were over 0.70, for both training and validation for AC (Table 2), implying that the feature extraction and dimensionality reduction of SAE resulted in accurate performance. Moreover, [52] and [53] mentioned that precise AC was necessary to achieve reliable cyanobacteria estimation. Additionally, [43] suggested that the AC with high accuracy has an influence on the accuracy of the bio-optical algorithm for PC and the reliability of the PC map. However, a few outliers were observed, which resulted from an abnormal reflectance peak beyond 700 nm (Figures 4 and 5). The outlier peaks were caused by high phytoplankton scattering from high algae presence on August 12 in 2016. SAE-ANN and -SVR models underestimated the peaks, because the models may be difficult to learn the specific abnormal features of high phytoplankton scattering.

SAE-ANN has proven to be acceptable for estimating cyanobacteria, compared to previous studies that applied the conventional bio-optical algorithms. The R^2 values of the conventional bio-optical algorithms for cyanobacteria estimations are as follows: 0.76 [54]; 0.71 [55]; 0.77 [14]; 0.55 [56]; and 0.65 [57]. When the PC concentration is greater than 10 mg m^{-3} , most model performances showed a good agreement with the observed PC, while inaccurate PC estimations can be observed for low PC concentrations of less than 10 mg m^{-3} . In particular, the A-D region in Figure 6a–d enclosed in broken circles indicates the region with a discrepancy between the estimated pigments and the observed ones. This could be caused by the relatively weak relationship between the corrected reflectance and low PC concentration. In addition, by comparing the disagreement levels, SAE-SVR and SVR models had higher uncertainties, compared to the SAE-ANN and ANN models (Figure 6c,d). The corrected reflectance error at each band may result in the incorrect feature extraction of low PC concentrations in the models (Figure 5), since the reflectance spectra is affected by the pigment concentration [52,58]. On the other hand, the SAE-SVR showed an underestimation of high PC concentrations greater than 40 mg m^{-3} . This can be attributed to the occurrence of scum during an intense cyanobacterial bloom, leading to the reduced accuracy of cyanobacterial estimation. Overall, the feature extraction with dimensionality reduction of SAE was able to estimate both PC and Chl-a. The encoding layer showed a well-defined temporal variation within the observed range of PC and Chl-a.

A high cyanobacteria concentration was mainly observed near the Baekje Weir region, due to the high flow velocity caused by the hydraulic gate operation (i.e., hydraulic power plant), which gathered the cyanobacteria from the upstream to the back of the Weir [33] (Figure 8a,e). After the gate operation, the cyanobacteria temporarily disappeared in front of the Weir, by the flushing and dilution effect [5]. The gate operation released a substantial amount of water, which generates water turbulence, thereby increasing the turbidity. This occurrence resulted in the decrease of light availability, which eventually hindered cyanobacterial growth [59]. The turbulent flow also physically inhibited cyanobacterial growth by damaging the phytoplankton cells [60]. On the other hand, a high concentration of cyanobacteria can be observed at the river side, which was mainly caused by a longer residence time. The cyanobacteria favor low flow velocity for blooming, since the temperature stratification zone and colonial formation are easily developed without flow suppression. Moreover, [61] suggested that a critical flow velocity $< 0.06 \text{ m s}^{-1}$ would be proper condition for cyanobacterial growth. Likewise, other previous studies found that flow velocity and cyanobacteria concentration have a negative relationship [60]. The decrease in cyanobacteria was driven by the unsuitable growth conditions, primarily due to the decreasing temperature and low light intensity [62,63]. Furthermore, [64] also proved that the main control factors of cyanobacteria growth were temperature and light availability, with 15-year MODIS imagery and the temperature dataset.

4.2. Data-Driven Model Comparison

For AC, SAE-ANN and -SVR models were not comparable. In addition, SAE-ANN showed more accurate AC than the conventional commercial software, even when the training dataset was limited. In this regard, the data-driven model could be used as an alternative to the physical based-model for accurate AC results, when the data of the atmosphere library of the commercial software could not reflect real atmospheric conditions.

SAE-ANN showed higher pigment estimation accuracy than the SAE-SVR model. Similar results were also found in a previous study. The ANN and SVR results without SAE were due to the limitation of the conventional model in reflecting the temporal variability of the optically complexed inland water [28]. Notably, [19] demonstrated that the stacked denoising autoencoder coupled with ANN fine-tuning showed the highest accuracy, compared to conventional contrast models in predicting water quality parameters of the biofilm system. In their study, the encoding layers were able to produce a high-level feature representation of the input imagery, which made the coupling of the models more efficient [65]. The SAE confronted the original input feature into smaller data and reconstructed the reduced data to original data in the training process [45]. The internal parameters of the SAE were updated to retrieve minimum error by comparing them to the output data. After training the SAE, the similarity between original input and reconstructed input implied that the trained parameters ensure the internal features in each layer that can represent the original input features. Accordingly, the input data used for AC and estimation of cyanobacteria was present in the middle of the SAE network. Thus, this confronted layer resulted in reduced data complexity and improved data abstraction, thereby contributing higher regression accuracy than conventional machine learning regression without SAE.

Previous studies showed that ANN has a better regression performance than SVR [66,67]. The performance difference between ANN and SVR models depends on the data. The model performance cannot be generalized, due to inconsistencies in the data behavior [68]. When coupled with the SAE network, the high dimensionality of the input data is compressed to a relatively low dimension with abstracted feature representation. The ANN model might reflect the PC and Chl-a features at low concentrations, with multiple nodes and layers, better than the SVR model. Moreover, [69] discussed the local underestimation of SVR, in which the kernel location was supposed to be the center of the epsilon-tube, but the SVR only allowed a small number of estimated values to fall below the observed values.

4.3. Deep Neural Network for Remote Sensing Application

In several previous studies, AC [21,31,70] and cyanobacterial estimations [33,71,72] have been performed using conventional machine learning models. However, a deep neural network yields a relatively high accuracy compared to the conventional models, owing to the utilization of high-level feature learning from the data [73,74]. Although deep neural networks with large datasets require high-end infrastructure facilities, such as a graphical processing unit (GPU), and a long model training time, the testing time for the trained model can be quite less. This aspect was identified by determining the training time for AC to be 1045.96 s and 2.78 s, respectively, and those for pigment estimation to be 508.54 s and 3.28 s, respectively. In addition, the SAE-ANN model improved the accuracy of surface reflectance estimations by 23% and that of pigment estimations by 26%, compared to conventional ANN models. This is because SAE provides higher level features for the robust representation of the temporal surface reflectance and pigment variations. However, it is difficult to accurately identify the function of neurons and their layers in the network architecture to be modeled [75].

As a deep neural network is suitable for complex image processing, it has been implemented for comprehensive remote sensing application (i.e., AC and cyanobacteria estimation) in this study. When SAE is coupled with ANN, a high estimation accuracy of water surface reflectance and cyanobacteria concentrations is possible. During the training process, the encoding layers learned the abstract features of the input data by reducing their dimension. For AC training, the SAE extracted the optical features (i.e., digital numbers) and atmospheric features (i.e., total flux, diffuse transmittance,

direct transmittance, spherical albedo, path radiance), by reconstructing the original input data. The optical and atmospheric features were utilized to estimate water surface reflectance in the consecutive ANN model. During this process, the digital numbers with atmospheric effect were directly transformed into surface reflectance that rarely possessed the effects. For pigment estimation, the estimated reflectance features were concentrated by the SAE, to estimate Chl-a and PC concentrations. This process also provided an efficient representation of the spatial distribution of the pigments during different periods. In short, the data-driven models provided implicit methods that only considered the relationship between the remote sensing input and target, without any complex formulations and parameterization of AC and bio-optical algorithms.

For study areas that have input data ranges similar to this study, the trained model can provide robust performance, whereas, for study areas that have different data ranges, the model can be used as a pre-trained one that requires additional model tuning without initiating end-to-end model configurations. As many researches have utilized a pre-trained model for their studies [76–78], the application of such a model is the primary benefit for a data-driven model to rapidly achieve reasonable outcomes. In addition, future research using deep learning can be conducted, by referring to the structures and internal parameters of this study for regression tasks using remote sensing data. Thus, we conclusively demonstrated the potential of a deep learning network in providing reliable and comprehensive remote sensing applications.

5. Conclusions

This study utilized the deep neural network in implementing AC and cyanobacteria estimation using hyperspectral images. To accomplish this, field and airborne monitoring, water sample collection, and optical measurement of the water were implemented. After which, phytoplankton pigments were analyzed (i.e., PC and Chl-a). To perform AC and estimate cyanobacteria, we developed the SAE-ANN and SAE-SVR models. The input data for AC consists of AC parameters driven by MODTRAN 6, digital numbers from hyperspectral imagery, and the number of sampling events. The input parameters were fed into the first SAE-ANN and -SVR models to produce the estimated surface reflectance, which was consequently assigned as input for the second SAE-ANN and -SVR, to estimate PC and Chl-a concentrations. The ANN, SVR, SAE-ANN, and SAE-SVR models were evaluated by R^2 , RMSE, NSE, and MAE. The major findings of this study are the following:

1. SAE-ANN and -SVR models for AC showed good agreement with the observed reflectance spectra (i.e., $NSE > 0.7$); the SAE-ANN model estimated the cyanobacteria concentrations with the highest accuracy.
2. The encoding layers of the SAE-ANN and -SVR models were able to contribute to the generation of cyanobacterial distribution maps, that represented actual cyanobacterial distribution, by reflecting the varied spatial and spectral features of the input data.
3. The SAE-ANN and -SVR models showed an improved accuracy of 23% and 6% for surface reflectance, and 26% and 9% for cyanobacteria estimation, respectively, due to the high-level feature extraction of SAE, compared to the single model performances of ANN and SVR.

This study demonstrated an integrative implementation of AC and cyanobacteria estimation with high accuracy, by developing deep neural networks. Thus, we hope that this study will provide the preceding information to a comprehensive remote sensing application for cyanobacteria management to future researches.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-4292/12/7/1073/s1>, Figure S1: Cyanobacteria map on August 12, 2016, RGB image.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “conceptualization, J.P., Y.C., and K.H.C.; methodology, J.P., M.K., S.B., and Y.S.K.; software, J.P.; validation, J.P., H.D., Y.C., and K.H.C.; formal analysis, J.P.; investigation, J.P., H.L., T.K., and K.K.; resources, J.P.; data curation, J.P. and H.D.; writing—original

draft preparation, J.P.; writing—review and editing, H.D., M.L., Y.C., and K.H.C.; visualization, J.P.; supervision, Y.C., and K.H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research was supported by the National Institute of Environmental Research funded by Ministry of Environment [NIER-2018-03-01-005] and was also supported by the 2016 Research Fund of University of Seoul for YoonKyung Cha.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hudnell, H.K. The state of U.S. freshwater harmful algal blooms assessments, policy, and legislations. *Toxicon* **2008**, *55*, 1024–1034. [[CrossRef](#)] [[PubMed](#)]
2. Lee, T.A.; Rollwagen-Bollens, G.; Bollens, S.M.; Faber-Hammond, J.J. Environmental influence on cyanobacteria abundance and microcystin toxin production in a shallow temperate lake. *Ecotoxicol. Environ. Saf.* **2015**, *114*, 318–325. [[CrossRef](#)] [[PubMed](#)]
3. Cho, K.H.; Kang, J.H.; Ki, S.J.; Park, Y.; Cha, S.M.; Kim, J.H. Determination of the optimal parameters in regression models for the prediction of chlorophyll-a: A case study of the Yeongsan Reservoir, Korea. *Sci. Total Environ.* **2009**, *407*, 2536–2545. [[CrossRef](#)]
4. Heisler, J.; Gilbert, P.M.; Burkholder, J.M.; Anderson, D.M.; Cochlan, W.; Dennison, W.C.; Dortch, Q.; Gobler, C.J.; Heil, C.A.; Humphries, E.; et al. Eutrophication and harmful algal blooms: A scientific consensus. *Harmful Algae* **2008**, *8*, 3–13. [[CrossRef](#)]
5. Paerl, H.W.; Hall, N.S.; Calandrino, E.S. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Sci. Total Environ.* **2011**, *409*, 1739–1745. [[CrossRef](#)]
6. O'neil, J.M.; Davis, T.W.; Burford, M.A.; Gobler, C.J. The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. *Harmful Algae* **2012**, *14*, 313–334. [[CrossRef](#)]
7. Rigosi, A.; Carey, C.C.; Ibelings, B.W.; Brookes, J.D. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa. *Limnol. Oceanogr.* **2014**, *59*, 99–114. [[CrossRef](#)]
8. Lehtimäki, J.; Moisaner, P.; Sivonen, K.; Kononen, K. Growth, nitrogen fixation, and nodularin production by two baltic sea cyanobacteria. *Appl. Environ. Microbiol.* **1997**, *63*, 1647–1656. [[CrossRef](#)]
9. Stewart, W.D.P.; Alexander, G. Phosphorus availability and nitrogenase activity in aquatic blue-green algae. *Freshw. Biol.* **1971**, *1*, 389–404. [[CrossRef](#)]
10. Wasmund, N. Occurrence of cyanobacterial blooms in the Baltic Sea in relation to environmental conditions. *Int. Rev. Ges. Hydrobiol. Hydrogr.* **1997**, *82*, 169–184. [[CrossRef](#)]
11. Kutser, T.; Metsamaa, L.; Strombeck, N.; Vahtmae, E. Monitoring cyanobacterial blooms by satellite remote sensing. *Estuarine Coast. Shelf Sci.* **2006**, *67*, 303–312. [[CrossRef](#)]
12. Randolph, K.; Wilson, J.; Tedesco, L.; Li, L.; Pascual, D.L.; Soyeux, E. Hyperspectral remote sensing of cyanobacteria in turbid productive water using optically active pigments, chlorophyll a and phycocyanin. *Remote Sens. Environ.* **2008**, *112*, 4009–4019. [[CrossRef](#)]
13. Agha, R.; Cirés, S.; Wörmer, L.; Domínguez, J.A.; Quesada, A. Multi-scale strategies for the monitoring of freshwater cyanobacteria: Reducing the sources of uncertainty. *Water Res.* **2012**, *46*, 3043–3053. [[CrossRef](#)]
14. Simis, S.G.H.; Peter, S.W.M.; Gons, H.J. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnol. Oceanogr.* **2005**, *50*, 237–245. [[CrossRef](#)]
15. Kudela, R.M.; Palacios, S.L.; Austerberry, D.C.; Accorsi, E.K.; Guild, L.S.; Torres-Perez, J. Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters. *Remote Sens. Environ.* **2015**, *167*, 196–205. [[CrossRef](#)]
16. Jupp, D.L.; Kirk, J.T.; Harris, G.P. Detection, identification and mapping of cyanobacteria—Using remote sensing to measure the optical quality of turbid inland waters. *Mar. Freshw. Res.* **1994**, *45*, 801–828. [[CrossRef](#)]
17. Kutser, T. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnol. Oceanogr.* **2004**, *49*, 2179–2189. [[CrossRef](#)]
18. Kallio, K.; Kutser, T.; Hannonen, T.; Koponen, S.; Pulliainen, J.; Vepsäläinen, J.; Pyhälähti, T. Retrieval of water quality from airborne imaging spectrometry of various lake types in different seasons. *Sci. Total Environ.* **2001**, *268*, 59–77. [[CrossRef](#)]

19. Shi, S.; Xu, G. Novel performance prediction model of a biofilm system treating domestic wastewater based on stacked denoising auto-encoder deep learning network. *Chem. Eng. J.* **2018**, *347*, 280–290. [[CrossRef](#)]
20. Adler-Golden, S.M.; Matthew, M.W.; Bernstein, L.S.; Levine, R.Y.; Berk, A.; Richtsmeier, S.C.; Acharya, P.K.; Anderson, G.P.; Felde, J.W.; Gardner, J. Atmospheric Correction for Shortwave Spectral Imagery Based on Modtran4. In *Imaging Spectrometry V International Society for Optics and Photonics*; SPIE: Bellingham, WA, USA, 1999; pp. 61–70.
21. Bernstein, L.S.; Adler-Golden, S.M.; Sundberg, R.L.; Levine, R.Y.; Perkins, T.C.; Berk, A.; Ratkowski, A.J.; Felde, G.; Hoke, M.L. Validation of the Quick Atmospheric Correction Algorithm for Vnir-Swir Multi-and Hyperspectral Imagery. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI International Society for Optics and Photonics*; SPIE: Bellingham, WA, USA, 2005; pp. 668–679.
22. Gao, B.C.; Montes, M.J.; Davis, C.O.; Goetz, A.F. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sens. Environ.* **2009**, *113*, S17–S24. [[CrossRef](#)]
23. Hunter, P.D.; Tyler, A.N.; Carvalho, L.; Codd, G.A.; Maberly, S.C. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. *Remote Sens. Environ.* **2010**, *114*, 2705–2718. [[CrossRef](#)]
24. Van Laake, P.E.; Sanchez-Azofeifa, G.A. Simplified atmospheric radiative transfer modelling for estimating incident PAR using MODIS atmosphere products. *Remote Sens. Environ.* **2004**, *91*, 98–113. [[CrossRef](#)]
25. Allali, K.; Bricaud, A.; Claustre, H. Spatial variations in the chlorophyll-specific absorption coefficient of phytoplankton and photosynthetically active pigments in the equatorial Pacific. *J. Geophys. Res.* **1997**, *102*, 12413–12423. [[CrossRef](#)]
26. Kimes, D.S.; Knyazikhin, Y.; Privette, J.L.; Abuelgasim, A.A.; Gao, F. Inversion methods for physically-based models. *Remote Sens. Rev.* **2000**, *18*, 381–439. [[CrossRef](#)]
27. Le, C.F.; Li, Y.M.; Zha, Y.; Sun, D.; Yin, B. Validation of a Quasi-Analytical Algorithm for highly turbid eutrophic water of Meiliang Bay in Taihu Lake, China. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2492–2500.
28. Odermatt, D.; Gitelson, A.; Brando, V.E.; Schaepman, M. Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote Sens. Environ.* **2012**, *118*, 116–126. [[CrossRef](#)]
29. Camps-Valls, G.; Bruzzone, L.; Rojo-Alvarez, J.L.; Melgani, F. Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 339–343. [[CrossRef](#)]
30. Kwon, Y.S.; Eunna, J.; Im, J.; Baek, S.H.; Park, Y.E.; Cho, K.H. Developing data-driven models for quantifying *Cochlodinium polykrikoides* in coastal water. *Int. J. Remote Sens.* **2018**, *39*, 68–83. [[CrossRef](#)]
31. Schroeder, T.; Behnert, I.; Schaale, M.; Fischer, J.; Doerffer, R. Atmospheric correction algorithm for MERIS above case-2 waters. *Int. J. Remote Sens.* **2007**, *28*, 1469–1486. [[CrossRef](#)]
32. Doerffer, R.; Schiller, H. The MERIS Case 2 water algorithm. *Int. J. Remote Sens.* **2007**, *28*, 517–535. [[CrossRef](#)]
33. Park, Y.; Pyo, J.; Kwon, Y.S.; Cha, Y.; Lee, H.; Kang, T.; Cho, K.H. Evaluating physico-chemical influences on cyanobacterial blooms using hyperspectral images in inland water, Korea. *Water Res.* **2017**, *126*, 319–328. [[CrossRef](#)] [[PubMed](#)]
34. Liu, F.; Xu, F.; Yang, S. A Flood Forecasting Model Based on Deep Learning Algorithms Via Integrating Stacked Autoencoders with BP Neural Network. In Proceedings of the 2017 IEEE Third International Conference on Multimedia Big Data (BigMM), Laguna Hills, CA, USA, 19–21 April 2017; pp. 58–61.
35. Liu, L.; Chen, R.C. A novel passenger flow prediction model using deep learning methods. *Transp. Res. Part C Emerg. Technol.* **2017**, *84*, 74–91. [[CrossRef](#)]
36. Kim, D.M.; Park, H.S.; Chung, S.W. Relationship of the thermal stratification and critical flow velocity near the Baekje Weir in Geum River. *J. Korean Soc. Water Environ.* **2017**, *33*, 449–459.
37. Mobley, C.D. Estimation of the remote-sensing reflectance from above-surface measurements. *Appl. Opt.* **1999**, *38*, 7442–7455. [[CrossRef](#)] [[PubMed](#)]
38. Zhang, Y.; Liu, M.; Qin, B.; Van Der Woerd, H.J.; Li, J.; Li, Y. Modeling remote-sensing reflectance and retrieving chlorophyll-a concentration in extremely turbid case-2 waters (Lake Taihu, China). *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1937–1948. [[CrossRef](#)]
39. Boyer, J.N.; Kelble, C.R.; Ortner, P.B.; & Rudnick, D.T. Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA. *Ecol. Indic.* **2009**, *9*, S56–S67. [[CrossRef](#)]
40. APHA (American Public Health Association). *Standard Methods for the Examination of Water and Waste Water*, 21st ed.; APHA-AWWA-WPCF: Washington, DC, USA, 2001.

41. Bennett, A.; Bogorad, L. Complementary chromatic adaptation in a filamentous blue-green alga. *J. Cell Biol.* **1973**, *58*, 419–435. [[CrossRef](#)] [[PubMed](#)]
42. Berk, A.; Conforti, P.; Kennett, R.; Perkins, T.; Hawes, F.; van den Bosch, J. Modtran@6: A Major Upgrade of the Modtran® Radiative Transfer Code. In Proceedings of the Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 6th Workshop on, Lausanne, Switzerland, 24–27 June 2014; pp. 1–4.
43. Pyo, J.; Ligaray, M.; Kwon, Y.; Ahn, M.H.; Kim, K.; Lee, H.; Kang, T.; Cho, S.B.; Park, Y.; Cho, K. High-spatial resolution monitoring of phycocyanin and chlorophyll-a using airborne hyperspectral imagery. *Remote Sens.* **2018**, *10*, 1180. [[CrossRef](#)]
44. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [[CrossRef](#)] [[PubMed](#)]
45. Zabalza, J.; Ren, J.; Zheng, J.; Zhao, H.; Qing, C.; Yang, Z.; Du, P.; Marshall, S. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. *Neurocomputing* **2016**, *185*, 1–10. [[CrossRef](#)]
46. Chan, P.P.K.; Lin, Z.; Hu, X.; Eric, C.C.; Yeung, D.S. Sensitivity based robust learning for stacked autoencoder against evasion attack. *Neurocomputing* **2017**, *267*, 572–580. [[CrossRef](#)]
47. Cho, K.H.; Sthiannopkao, S.; Pachepsky, Y.A.; Kim, K.W.; Kim, J.H. Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. *Water Res.* **2011**, *45*, 5535–5544. [[CrossRef](#)] [[PubMed](#)]
48. Park, Y.; Ligaray, M.; Kim, Y.M.; Kim, J.H.; Cho, K.H.; Sthiannopkao, S. Development of enhanced groundwater arsenic prediction model using machine learning approaches in Southeast Asian countries. *Desalination Water Treat.* **2016**, *57*, 12227–12236. [[CrossRef](#)]
49. Du, C.; Wang, Q.; Li, Y.; Lyu, H.; Zhu, L.; Zheng, Z.; Wen, S.; Liu, G.; Guo, Y. Estimation of total phosphorus concentration using a water classification method in inland water. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *71*, 29–42. [[CrossRef](#)]
50. Gonzalez, P.A.; Zamarreno, J.M. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy Build.* **2005**, *37*, 595–601. [[CrossRef](#)]
51. Duan, S.B.; Li, Z.L.; Tang, B.H.; Wu, H.; Ma, L.; Zhao, E.; Li, C. Land surface reflectance retrieval from hyperspectral data collected by an unmanned aerial vehicle over the Baotou test site. *PLoS ONE* **2013**, *8*, e66972. [[CrossRef](#)]
52. Ogashawara, I.; Mishra, D.; Mishra, S.; Curtarelli, M.; Stech, J. A performance review of reflectance based algorithms for predicting phycocyanin concentrations in inland waters. *Remote Sens.* **2013**, *5*, 4774–4798. [[CrossRef](#)]
53. Matthews, M.W.; Bernard, S.; Robertson, L. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* **2012**, *124*, 637–652. [[CrossRef](#)]
54. Dingtian, Y.; Delu, P. Hyperspectral retrieval model of phycocyanin in case II waters. *Sci. Bull.* **2006**, *51*, 149–153.
55. Li, L.; Li, L.; Song, K. Remote sensing of freshwater cyanobacteria: An extended IOP inversion model of inland waters (IIMIW) for partitioning absorption coefficient and estimating phycocyanin. *Remote Sens. Environ.* **2015**, *157*, 9–23. [[CrossRef](#)]
56. Lyu, H.; Wang, Q.; Wu, C.; Zhu, L.; Yin, B.; Li, Y.; Huang, J. Retrieval of phycocyanin concentration from remote-sensing reflectance using a semi-analytic model in eutrophic lakes. *Ecol. Inf.* **2013**, *18*, 178–187. [[CrossRef](#)]
57. Ali, K.; Witter, D.; Ortiz, J. Application of empirical and semi-analytical algorithms to MERIS data for estimating chlorophyll a in Case 2 waters of Lake Erie. *Environ. Earth Sci.* **2014**, *71*, 4209–4220. [[CrossRef](#)]
58. Mishra, S.; Mishra, D.R.; Schluchter, W.M. A novel algorithms for predicting phycocyanin concentrations in cyanobacteria: A proximal hyperspectral remote sensing approach. *Remote Sens.* **2009**, *1*, 758–775. [[CrossRef](#)]
59. Mitrovic, S.M.; Hardwick, L.; Dorani, F. Use of flow management to mitigate cyanobacterial blooms in the Lower Darling River, Australia. *J. Plankton Res.* **2010**, *33*, 229–241. [[CrossRef](#)]
60. Li, F.; Zhang, H.; Zhu, Y.; Xiao, Y.; Chen, L. Effect of flow velocity on phytoplankton biomass and composition in a freshwater lake. *Sci. Total Environ.* **2013**, *447*, 64–71. [[CrossRef](#)] [[PubMed](#)]

61. Zhang, H.; Chen, R.; Li, F.; Chen, L. Effect of flow rate on environmental variables and phytoplankton dynamics: Results from field enclosures. *Chin. J. Oceanol. Limnol.* **2015**, *33*, 430–438. [[CrossRef](#)]
62. Post, A.F.; Wit, R.D.; Mur, L.R. Interactions between temperature and light intensity on growth and photosynthesis of the cyanobacterium *Oscillatoria agardhii*. *J. Plankton Res.* **1985**, *7*, 487–495. [[CrossRef](#)]
63. Robarts, R.D.; Zohary, T. Temperature effects on photosynthetic capacity, respiration, and growth rates of bloom-forming cyanobacteria. *N. Z. J. Mar. Freshw. Res.* **1987**, *21*, 391–399. [[CrossRef](#)]
64. Duan, H.; Tao, M.; Loiselle, S.A.; Zhao, W.; Cao, Z.; Ma, R.; Tang, X. MODIS observations of cyanobacterial risk in a eutrophic lake: Implications for long-term safety evaluation in drinking-water source. *Water Res.* **2017**, *122*, 455–470. [[CrossRef](#)]
65. Hinton, G.E.; Osindero, S.; The, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
66. Asilturk, I.; Kahramanli, H.; Mounayri, H.E. Prediction of cutting forces and surface roughness using artificial neural network (ANN) and support vector regression (SVR) in turning 4140 steel. *Mater. Sci. Technol.* **2012**, *28*, 980–986. [[CrossRef](#)]
67. Nasir, M.T.; Mysorewala, M.; Cheded, L.; Siddiqui, B.; Sabih, M. Measurement Error Sensitivity Analysis for Detecting and Locating Leak in Pipeline Using ANN and SVM. In Proceedings of the 2014 IEEE 11th International Multi-Conference on Systems, Signals, & Devices, Barcelona, Spain, 11–14 February 2014.
68. Shirzad, A.; Tableesh, M.; Farmani, R. A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks. *KSCE J. Civ. Eng.* **2014**, *18*, 941–948. [[CrossRef](#)]
69. Stockman, M.; Awad, M.; Khanna, R. Asymmetrical and Lower Bounded Support Vector Regression for Power Estimation. In Proceedings of the 2011 International Conference on Energy Aware Computing, Istanbul, Turkey, 30 November–2 December 2011; pp. 1–6.
70. Goyens, C.; Jamet, C.; Schroeder, T. Evaluation of four atmospheric correction algorithms for MODIS-Aqua images over contrasted coastal waters. *Remote Sens. Environ.* **2013**, *131*, 63–75. [[CrossRef](#)]
71. Nieto, P.G.; García-Gonzalo, E.; Fernández, J.A.; Muñiz, C.D. A hybrid wavelet kernel SVM-based method using artificial bee colony algorithm for predicting the cyanotoxin content from experimental cyanobacteria concentrations in the Trasona reservoir (Northern Spain). *J. Comput. Appl. Math.* **2017**, *309*, 587–602. [[CrossRef](#)]
72. Vilán, J.V.; Fernández, J.A.; Nieto, P.G.; Lasheras, F.S.; de Cos Juez, F.J.; Muñiz, C.D. Support vector machines and multilayer perceptron networks used to evaluate the cyanotoxins presence from experimental cyanobacteria concentrations in the Trasona reservoir (Northern Spain). *Water Resour. Manag.* **2013**, *27*, 3457–3476. [[CrossRef](#)]
73. Li, W.; Wu, G.; Zhang, F.; Du, Q. Hyperspectral image classification using deep pixel-pair features. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 844–853. [[CrossRef](#)]
74. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]
75. Barrett, D.G.; Morcos, A.S.; Macke, J.H. Analyzing biological and artificial neural networks: Challenges with opportunities for synergy? *Curr. Opin. Neurobiol.* **2019**, *55*, 55–64. [[CrossRef](#)]
76. Han, X.; Zhong, Y.; Cao, L.; Zhang, L. Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sens.* **2017**, *9*, 848. [[CrossRef](#)]
77. Sherrah, J. Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. *arXiv* **2016**, arXiv:1606.02585.
78. Wang, J.; Luo, C.; Huang, H.; Zhao, H.; Wang, S. Transferring pre-trained deep CNNs for remote scene classification with general features learned from linear PCA network. *Remote Sens.* **2017**, *9*, 225. [[CrossRef](#)]

