Ensemble Learning via Higher Order Singular Value Decomposition for Integrating Data and Classifier Fusion in Water Quality Monitoring

Zhibin Sun, Ni-Bin Chang, Fellow, IEEE, Chi-Farn Chen, Chandan Mostafiz, and Wei Gao

Abstract—Current multimodal learning in smart feature extraction and classification has reshaped the landscape of remote sensing. The recent developments in smart feature extraction mainly rely on different machine learning and data mining algorithms as powerful classifiers to improve prediction accuracy. This article presents an innovative ensemble learning algorithm for integrated data and classifier fusion via higher order singular value decomposition (IDCF-HOSVD). Based on the fused data, different analytical, semiempirical, and empirical classifiers can be selected and applied to perform information retrieval that can be further synergized via a tensor flow based feature extraction scheme over the classifier space. When preserving core fused image patterns via HOSVD, the final step of IDCF-HOSVD helps rank the contributions from different classifiers via nonlinear correlation information entropy. Practical implementation of the IDCF-HOSVD algorithm was assessed through its application to map the seasonal water quality conditions in Lake Nicaragua, Central America.

Index Terms—Classifier fusion, ensemble learning, higher order singular value decomposition (HOSVD), multimodal learning, water quality monitoring.

I. INTRODUCTION

SATELLITE remote sensing offers various niches to provide timely monitoring of environmental quality and ecosystem state. Various data fusion, data merging, data mining, and ensemble learning techniques have extended our capability to retrieve multifaceted features with better resolution, accuracy, and coverage [1]. For instance, the high spatial resolution Landsat and the high temporal resolution MODIS imageries can be fused to produce improved images with both high spatial and high temporal resolutions in support of better feature extraction [2], [3]. Bai et al. [4] further extended the techniques of data merging and cloudy pixel reconstruction to assimilate heterogeneous satellite images for long-term assessment of air quality with the aid of machine learning skills. Recent regime shifts from traditional feature extraction algorithms to multimodal machine learning algorithms based on either fused or merged images have resulted in significant advancements in thematic information retrieval. Traditional linear regression models (LRMs) have been used on many occasions for feature extraction, and well-known extensions of artificial neural network (ANN) models for feature extraction include morphological shared weight neural networks [5], extreme learning machine (ELM) [6], [7], deep belief nets (DBNs) [8], convolutional neural networks (CNNs) [9], etc. Deep learning, such as DBNs and CNNs, uses multiple-layer architecture to learn representations of input/output data with abstractions from lower to higher levels. The deep learning model was recently used to complete several distinct tasks in different fields, providing better feature extraction capacity [10]–[12].

Feature extraction methods in remote sensing for water quality monitoring can be versatile when using LRMs for mapping Chlorophyll-a or total suspended solid concentrations [13], [14]. The use of genetic programming (GP) further advanced machine learning algorithms to improve the monitoring capability of water quality parameters, such as nitrogen and phosphorus concentrations [15]. In addition, the well-known ANN can be used for inverse modeling [16], [17]. Both ANN and GP models were applied to reconstruct the spatiotemporal distributions of microcystin in Lake Erie based on fused multispectral and hyperspectral remote sensing images [2], [3]. Real world comparative evaluation among single hidden layer feedforward neural network (SLFN), DBN, and ELM was first attempted in Lake Managua by Chang et al. [18] using fused satellite images for classification. Moreover, the CNN for river water quality mapping was implemented by Oga et al. [19] to study landscape complexity in a flood plain.

Although each of the machine learning classifiers (e.g., SLFN, ELM, CNN, and DBN) has its own advantages, intelligent selection and integration of feature extraction algorithms via ensemble learning led to produce more representative content-based mapping. The idea of ensemble learning was proposed by Dasarathy and Sheela [20], who used multiple classifiers to partition the featured space. Later, Hansen and Salamon [21] used a set of similar neural network models to improve network...
performance for classification. Schapire [22] then used an ensemble of classifiers to build a strong model (called “boosting”) with an arbitrarily small error on a problem of binary classification. The ensemble idea has been widely used in different fields. For example, the ensemble Kalman filter [23] is a popular data assimilation technique in atmospheric and oceanic science. Ensemble methods have also been applied in spoken emotion recognition [24], as well as in intrusion detection systems [25].

Although there are many ensemble learning algorithms available at present for processing raw or fused images, they normally fall into one or more of the following three categories.

1) Methods based on the selection of training data for diversity, such as bagging [26], boosting [22], and random subspace methods [27].

2) Methods based on the training approaches for creating ensemble members/classifiers, such as boosting [28], stack generalization [29], and the hierarchical mixture of experts [30].

3) Methods based on the combination of different ensemble members/classifiers, such as major voting [31], algebra combinations such as max/min/sum/product [32]–[34], decision templates [35], and genetic algorithms based on a selective ensemble [36].

When dealing with the image outputs from different ensemble members/classifiers, the common weighting schemes from ensemble learning do not utilize the information from the spectral space of images. Although there are some recently developed ensemble learning methods such as the oblique decision tree ensemble, cotrained kernel ridge regression, and the random vector functional link network ensemble [37], which can outperform many other classification methods, all the ensemble learning methods/techniques mentioned above are designed based on probability and statistical methods that have long and short-term memory dependence issues, thus making their learning capacity unreliable when dealing with high-dimensional issues. Nevertheless, ensemble learning algorithms, which can integrate data and classifier fusion by taking advantage of the merits of different multimodal learning processes, have not yet been thoroughly investigated.

Using tensor flows to perform ensemble learning is an innovative idea. Singular value decomposition (SVD) in matrix theory can not only find the most important information to represent a data matrix but also can reduce redundant/noisy information in the matrix via tensor flows. Many other popular decomposition methods related to SVD mathematically, such as principal component analysis (PCA) and empirical orthogonal functions (EOF), cannot have equivalent power per se. As a specific orthogonal Tucker decomposition [38], high-order singular value decomposition (HOSVD) extends the mathematical domain of SVD, PCA, and EOF from matrix to tensor, which means from two dimensions to N-dimensions. Compared with SVD, HOSVD can extract spectral information in higher dimensional spaces. It has been applied successfully not only in computer vision [39] but also in image denoising of the image’s patches [40]. Additionally, the use of the nonlinear correlation coefficient (NCC) to determine which classifier contributes most to the final mapping after ensemble learning might be a powerful way to screen and prioritize each classifier for comparison [41].

Nevertheless, the impact of high-order tensor flows on either fused or merged data has not yet been investigated and applied for earth observations, let alone the assessment of relative contribution via the use of NCC across the employed classifiers based on the identified tensor flows. This study aims to fill in these technical gaps to advance the capability of multimodal ensemble learning for fused imageries with the aid of integrated HOSVD and NCC to advance earth observations. The goal of this study is a seamless integration of data and classifier fusion, resulting in unified neural network learning with machine intelligence, termed “integrated data and classifier fusion with higher order singular value decomposition” (IDCF-HOSVD). IDCF-HOSVD can magnify a multimodal learning process for images and exerts a powerful tensor flow based feature extraction scheme over the classifier space that can preserve core image patterns with unlimited dimensions. There is no curse of dimensionality issue per se.

In this article, practical implementation of IDCF-HOSVD was ultimately assessed through its application to water quality monitoring in an eutrophied lake, Lake Nicaragua, in Central America. The spatial patterns of Chl-a concentration for both wet and dry seasons in this shallow tropical lake were explored and mapped to answer the following science questions: 1) Can machine learning based empirical classifiers, including SLFN, ELM, and DBN, outperform LRM and TBM in terms of prescribed statistical indices and NCC justification? 2) How does the IDCF-HOSVD harmonize different classifiers and present more informative and representative mapping in earth observations? 3) Which classifier is most influential in ensemble learning in terms of information entropy delivery?

II. METHODS AND DATA

All essential steps are depicted in Fig. 1 and summarized in greater detail in the following. In this study, the high spatial resolution Landsat and the high temporal resolution MODIS imageries can be fused to produce improved images with both high spatial and high temporal resolutions. The five inverse models selected for feature extraction and classification include the analytical classifier (TBM), the semiempirical classifier (LRM), and empirical classifiers (SLFN, ELM, and DBN). These five major classifiers were compared to one another based on a set of prescribed statistical indices for initial diagnosis.

As the final step, the selected ensemble learning algorithm (i.e., HOSVD) was applied to synergize the individual contribution from each classifier and perform content-based mapping for Chl-a concentration in Lake Nicaragua. The IDCF approach can infer dynamic water quality conditions using the nonlinear correlation information entropy (NCIE) [41] in concert with HOSVD. Finally, visualization of the water quality concentration maps for Chl-a in both wet and dry seasons was used to confirm the application potential of IDCF-HOSVD with sensitivity analysis of NICE. These steps are described in detail in the following.

A. Data Acquisition at the Study Site

Lake Nicaragua and Lake Managua are two large freshwater lakes in Central America. Both lakes are located at the San Juan watershed, and they are interconnected by a canal. Lake
Managua connects to Lake Nicaragua through the Tipitapa River, creating a pollution plume from Lake Managua to Lake Nicaragua [42]. Since the two lakes are interconnected through a canal, the water quality dataset of the two lakes can be integrated as a single source of a designated water body. Thus, this study grouped all ground truth samples together over these two lakes for model training, testing, and validation toward better feature extraction. Sampling locations and laboratory analysis results are included in Appendixes I and II. Reflectance data for Lake Nicaragua and Lake Managua were collected from the Landsat 7 ETM+ and Terra MODIS (MOD09GA) satellites [43] to generate fused data with higher temporal and spatial resolution.
B. Image Preprocessing [Fig. 1, Step 2]

The acquired MODIS data are at a level-2G basis, wherein the data have been radiometrically calibrated and atmospherically corrected to account for scattering and aerosols [44]. The Landsat data is on a level-1T basis, with radiometric and geometric corrections [45]. As denoted by Step 2 of Fig. 1, ArcGIS mapping and spatial analysis software was used to process the remote sensing images in preparation for the data fusion process, based on the same operation used by Chang et al. in [2].

C. Data Fusion [Fig. 1, Step 3]

Data fusion is the algorithmic fusion of the spectral, temporal, or spatial properties of two or more images into a composite or synthetic image based on the characteristics of the input images. MODIS images with high temporal resolution (i.e., daily) and Landsat images with high spatial resolution (30 m) were fused to create synthetic images with 30 m resolution on a daily basis via the spatial and temporal adaptive reflectance fusion model (STARFM) [46]. The synthetic Landsat image was generated based on the input candidate pixels in the MODIS image taken during the desired prediction date. Data fusion must be verified based on certain indices. This study used Shannon entropy as a representative index for verification. Entropy is a statistical measure of the randomness or richness of information in an image, and hence can be used to characterize the texture of an image [47]. The Shannon entropy of an image is defined as

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i$$  \hspace{1cm} (1)

where $n$ denotes the number of the image’s gray level, and $p_i$ is the probability of the $i$th gray level of the image, which is estimated from the histogram counts of valid image pixels. The higher the value of entropy, the richer the information content.

D. Inverse Modeling for Feature Extraction [Fig. 1, Step 4]

Fused images corresponding to the ground-truthing dates were developed in Section II-C. The ground-truth data were collected from 2011 to 2016. For each ground-truth observation, a MODIS image, as well as a temporally close Landsat image selected for the precondition and postcondition, was acquired. Five different inverse models or classifiers for feature extraction were applied to perform classification with a common ground-truth dataset for content-based Chl-a concentration mapping. These five inverse models or classifiers include analytical, empirical, and semiempirical models, which are suited for determining the relationships between measured water quality parameters at the ground level and fused spectral values from space-borne satellite images [48], [49]. While bands 1, 2, and 3 of the fused images were applied for the development of LRM, only bands 1 and 2 were suited for TBM. For these empirical models (e.g., machine learning models), it is necessary to ensure that the training dataset can be exposed to the widest range of water quality values available to increase the prediction accuracy of the model at extremes. After training and testing, validation can confirm whether the model is suited for calculating water quality concentrations.

The SLFN algorithm is flexible, comprehensive, and effective when large datasets (input/output) are available. The mathematical architecture of SLFN is described as follows: Assume that there is a training set $\{ (x^{(j)}, y^{(j)}) | y^{(j)} \in \mathbb{R}^u, x^{(j)} \in \mathbb{R}^v, j = 1, 2, \ldots, N \}$ with $N$ training samples, which will be used to establish $n$-input and $m$-output. If there are $L$ hidden nodes and the SLFN is the standard one, then the output $y^{(j)} (j = 1, 2, 3, \ldots, N)$ can be modeled as

$$y^{(j)} = \sum_{i=1}^{L} v_i g_i \left( x^{(j)} \right) = \sum_{i=1}^{L} v_i \left( a_i \cdot x^{(j)} + b_i \right)$$  \hspace{1cm} (2)

where $a_i = [a_{i1}, a_{i2}, \ldots, a_{im}]^T$ ($i = 1, 2, \ldots, L$) is the weight vector connecting the $i$th hidden node with all input nodes, and $v_i = [v_{i1}, v_{i2}, \ldots, v_{im}]^T$ ($i = 1, 2, 3, \ldots, L$) is the weight vector connecting the $i$th hidden node with all output nodes. $b_i$ is the bias of the $i$th hidden node, and $g(x)$ is the activation function, which may be sine, cosine, radial basis, or other nonregular functions, as well as the sigmoidal function. The system of these equations shows the forward calculation process of SLFN to generate the predictions $Y$ for $N$ training inputs. Its general expression is

$$HV = Y$$  \hspace{1cm} (3)

where $H$ is the hidden layer output matrix depicted as

$$H = \begin{bmatrix} g \left( a_1 \cdot x^{(1)} + b_1 \right) & \ldots & g \left( a_L \cdot x^{(1)} + b_L \right) \\ \vdots & \ddots & \vdots \\ g \left( a_1 \cdot x^{(N)} + b_1 \right) & \ldots & g \left( a_L \cdot x^{(N)} + b_L \right) \end{bmatrix}_{N \times L}$$  \hspace{1cm} (4)

and $V$ and $Y$ are

$$V = \begin{bmatrix} v_{11}^T \\ \vdots \\ v_{1L}^T \end{bmatrix}_{L \times m}, \quad Y = \begin{bmatrix} (y^{(1)})^T \\ \vdots \\ (y^{(N)})^T \end{bmatrix}_{N \times m}$$  \hspace{1cm} (5)

After completing this process, the BPN algorithm is used to train the SLFN and determine the optimal sets of $\{V^*, a_i^*, b_i^*, \ldots, \} (i = 1, 2, \ldots, L)$ to minimize the difference between predictions and targets

$$\|H(a_1^*, \ldots, a_L^*, b_1^*, \ldots, b_L^*) V^* - Y\| = \min(\|H(a_1, \ldots, a_L, b_1, \ldots, b_L) V - Y\|)_{a_i, b_i, V}.$$  \hspace{1cm} (6)

The minimization process follows the gradient-descent-based algorithms, adjusting the weights and biases through iterations of backward progression. This process is expressed as

$$W_k = W_{k-1} - \eta \frac{\delta E(W)}{\delta W}$$  \hspace{1cm} (7)

where $\eta$ is the learning rate, and $E$ is the error left in each predictive iteration. These gradient-descent-based algorithms work well for a vast majority of problems. However, these algorithms have iterative learning steps, so they are normally slow. They also have problems associated with overfitting and are typically used for local, rather than global, minima.
ELM is a very effective feedforward neural network [7] that inlines the global optimum. It was proposed by Huang et al. [6], [7] to enhance the learning pace and accuracy of SLFN. The input weights and hidden layer biases in ELM can be randomly allocated if the initiation functions are vastly differentiable, with the hidden layer remaining unaltered [6], [7]. Its yield weights \( V \) are the main variable of SLFN to be tuned for the straight framework \( \mathbf{H} \mathbf{V} = \mathbf{Y} \). So, for ELM, (6) now becomes

\[
||\mathbf{H} (a_1, \ldots, a_L, b_1, \ldots, b_L) \mathbf{V}^* - \mathbf{Y}|| = \min(||\mathbf{H} (a_1, \ldots, a_L, b_1, \ldots, b_L) \mathbf{V} - \mathbf{Y}||)_{\mathbf{V}}. 
\] (8)

In fact, it is a linear system, which can be tackled by least square methods, and its output weights are analytically decided by

\[
\mathbf{V}^* = \mathbf{H}^\dagger \tag{9}
\]

where \( \mathbf{H}^\dagger \) is the Moore–Penrose generalized inverse of \( \mathbf{H} \) [6], [7].

On the other hand, DBN is a probabilistic generative model with multiple layers of stochastic latent variables, and its building blocks are RBMs. There are two layers in each RBM. The lower layer consists of observable data variables (i.e., visible units), and the upper layer consists of latent variables (i.e., hidden units). The energy of the joint configuration of the visible and hidden units in an RBM is expressed by Bishop [50] and Hinton [51]

\[
E(v, h) = -\sum_{i=1}^{V_N} \sum_{j=1}^{H_M} w_{i,j} v_i h_j - \sum_{i=1}^{V_N} a_i v_i - \sum_{j=1}^{H_M} b_j h_j \tag{10}
\]

wherein the model parameters are \( w_{i,j}, a_i, \) and \( b_j; \) \( w_{i,j} \) is the symmetric weight between visible unit \( v_i \) and hidden unit \( h_j, a_i \) and \( b_j \) are the biases for visible and hidden units, respectively. \( V_N \) and \( H_M \) are the numbers of visible and hidden units, respectively. The optimal set of parameters (i.e., \( w_{i,j}, a_i, \) and \( b_j \)) can be achieved by maximizing the probability of visible units, which is described as the following joint probability distribution of the visible-hidden unit pair [50]:

\[
P(v, h) = \frac{e^{-E(v, h)}}{\sum_{v'} \sum_{h'} e^{-E(v', h')}} \tag{11}
\]

whose denominator is the sum of all possible pairs of visible-hidden units. By marginalizing over the space of hidden units, (11) becomes the probability of a visible unit

\[
P(v) = \sum_{h} P(v, h) = \frac{\sum_{h} e^{-E(v, h)}}{\sum_{v'} \sum_{h'} e^{-E(v', h')}}. \tag{12}
\]

The optimal parameter set of \( \{w_{i,j}, a_i, b_j\} \) of an RBM with \( N \) training data samples can be solved via the objective function

\[
\max_{w_{i,j}, a_i, b_j} \left\{ \frac{1}{N} \sum_{k=1}^{N} \log P(v^k) \right\} \tag{13}
\]

the aim of which is to increase the model probability of the training samples [51]. With the training samples and the structure of an RBM, the optimal solution to (13) completes the RBM via the minimization of divergence. In principle, the DBN is stacked by multiples of RBM. Inside DBN’s multiple layer structure, the higher level RBM’s input (via visible units) is taken from the lower level RBM’s output (via hidden units) after the lower level RBM is learned. After layer-by-layer pretraining in the DBN, back-propagation techniques fine-tune all the parameters in the DBN, which include the output weights connecting the network outputs with the top-level RBM outputs [52]. In this study, the DBN is stacked with three RBMs with the size of 180 × 60 × 20.

### E. Concentration Map Generation [Fig. 1, Step 5]

The Lake Nicaragua watershed is very extensive and is situated between Nicaragua and Costa Rica, spanning 13 707 km² in Nicaragua (excluding the lake itself and its islands) and 2577 km² in Costa Rica [42]. Lake Managua has a surface area of 1042 km². The lake is approximately 65 km long and 25 km wide, with an average depth of around 11 m. Managua, the capital city of Nicaragua, lies on the south-western shore of the lake. The wastewater treatment plant has been in operation since 2008. Wastewater discharged into Lake Managua has deteriorated this already polluted lake. Remote sensing can help assess this impact via multitemporal change detection of the water quality in these lakes. There is an acute need to monitor the changing water quality for public health and ecosystem conservation.

Ultimately, an optimal model for feature extraction can translate the surface reflectance values to a map of Chl-a concentrations throughout Lake Nicaragua. Each pixel of the fused reflectance images represents a 30 × 30 m² lake. Within the pixel, there are three bands worth of information characterizing the surface reflectance at different bandwidths. As determined by these models that are functions in terms of surface reflectance values, certain bandwidths may have a stronger explanatory power in the determination of the Chl-a concentration. Finally, it is necessary to use a set of statistical indices to evaluate and rank these five inverse models (i.e., classifiers) after feature extraction. In general, a model is deemed accurate if the predicted values closely match the observed values. Five statistical indices, including root-mean-square error (RMSE), ratio of the standard deviations (CO), mean percentage of error (PE), Pearson product moment correlation coefficient (RSQ or R-value) and coefficient of determination (R²) were used. They are represented by the following, where \( y_i^p \) is the \( i \)th predicted value, \( y_i^o \) is the \( i \)th observed value, \( N \) is the number of samples, \( \bar{y}_p \) is the mean of the predicted value, and \( \bar{y}_o \) is the mean of the observed value.

Model performance is good if RMSE is close to zero, CO is close to one, PE is close to zero, or RSQ (or R²) is close to one.

Before an assessment, these five classifiers were scaled onto \([-1, 1]\) linearly to generate the normalized training dataset for the five statistical indices in different scenarios. For example, the predicted values of ELM are the mean of the training output from 50 experiments. This estimation and evaluation process can even be repeated for each of the pixels of the lake map for different water quality constituents (e.g., Chl-a in this study)

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i^p - y_i^o)^2} \tag{14}
\]
CO = \sqrt{\frac{\sum_{i=1}^{N}(y_{i}^p - \bar{y}_o)^2}{\sum_{i=1}^{N}(y_{i}^o - \bar{y}_o)^2}} \tag{15}

PE = \frac{1}{N} \sum_{i=1}^{N} \frac{y_{i}^p - y_{i}^o}{y_{i}^o} \times 100 \tag{16}

RSQ = \frac{\sum_{i=1}^{N} (y_{i}^p - \bar{y}_o) (y_{i}^o - \bar{y}_o)}{\sqrt{\sum_{i=1}^{N} (y_{i}^o - \bar{y}_o)^2} \sqrt{\sum_{i=1}^{N} (y_{i}^o - \bar{y}_o)^2}} \tag{17}

R^2 = RSQ^2. \tag{18}

F. Ensemble Learning [Fig. 1, Step 6]

Finally, an image-based ensemble learning tool was built to integrate ensemble members (i.e., SFLN, DBN, ELM, LRM, and TBM). The inverse models for classification to be considered included: 1) the same type of inverse model with good performance at different stages (training/testing/validation), or 2) different types of inverse models at the same/different stages. Both cases can be viewed as a combination of different predictions. The following is designed to propose a suitable image combination scheme in spectral space for water quality predictive models by using the HOSVD technique.

Tensor \( A \in \mathbb{R}^{N_1 \times N_2 \times N_3} \) is composed of \( N_1 \) images \((A_1, A_2, \ldots, A_{N_1})\), and the size each image is \( N_2 \times N_3 \). The Tucker decomposition of \( A \) is

\[
A = S \times_1 U^{(1)}(1) \times_2 U^{(2)}(2) \circ U^{(3)}(3)
\]

where \( S \in \mathbb{R}^{N_1 \times N_2 \times N_3} \) is the core tensor, \( U^{(i)} \in \mathbb{R}^{N_i \times N_i} \) is a unitary matrix, \( \times_n \) stands for the \( n \)th mode tensor product, and \( \circ \) is the outer product. The columns of \( U^{(2)} \) and \( U^{(3)} \) are the basis for all images in y-direction and x-direction, respectively. All images (i.e., \( \{A_{i_1}\}_{i_1} \)) share the same base images (i.e., \( \{U^{(2)}_{i_2} \circ U^{(3)}_{i_3}\}_{i_2, i_3} \)) and are the linear combination of the base images. Moreover, (19) can lead to

\[
A = S^{(1)} \times_2 U^{(2)}(2) \times_3 U^{(3)}(3) \tag{20}
\]

tensor \( S^{(1)} = S \times_1 U^{(1)}(1) \in \mathbb{R}^{N_1 \times N_2 \times N_3} \). Therefore, the decomposition of each image is

\[
A_{i_1} = \sum_{i_2, i_3} U^{(2)}_{i_2} \circ U^{(3)}_{i_3} \tag{21}
\]

For a given \( i_1 \), one major part of the leading \( \{S^{(1)}_{i_1}(i_2, i_3)\}_{i_2, i_3} \) (e.g., first \( M \) greatest values, or 95% leading values) can be chosen to approximate image \( A_{i_1} = \sum_{(i_2, i_3) \in I_{i_1}^{2,3}} S^{(1)}_{i_1}(i_2, i_3) U^{(2)}_{i_2} \circ U^{(3)}_{i_3} \), where \( I_{i_1}^{2,3} \) is the chosen set of \((i_2, i_3)\) for the image. Thus, under the basis of \( \{U^{(2)}_{i_2} \circ U^{(3)}_{i_3}\}_{i_2, i_3}, \{S^{(1)}_{i_1}(i_2, i_3)\}_{i_2, i_3} \in I_{i_1}^{2,3} \) represents the approximation of image \( A_{i_1} \).

The union of each \( I_{i_1}^{2,3} \) (i.e., \( I_{i_1}^{2,3} = \bigcup_{i_1} I_{i_1}^{2,3} \)) can represent the majority of base indices, and the base set of \( \{U^{(2)}_{i_2} \circ U^{(3)}_{i_3}\}_{(i_2, i_3) \in I_{i_1}^{2,3}} \) can capture the most important features of the entire image set. Therefore, the base set is used to generate a target image, which can incorporate the important features from the entire image set. Assume the target image is

\[
A = \sum_{(i_2, i_3) \in I_{i_1}^{2,3}} a_{i_2, i_3} U^{(2)}_{i_2} \circ U^{(3)}_{i_3} \tag{22}
\]

where \( a_{i_2, i_3} \) is the undetermined coefficient. The optimum set of coefficients can be obtained by minimizing the difference between \( A \) and \( \{A_{i_1}\} \), i.e.,

\[
\{a_{i_2, i_3}^*\} = \arg \min_{\{a_{i_2, i_3}\}} \sum_{i_1} \|A - A_{i_1}\|. \tag{23}
\]

The resulting image

\[
A_* = \sum_{(i_2, i_3) \in I_{i_1}^{2,3}} a_{i_2, i_3}^* U^{(2)}_{i_2} \circ U^{(3)}_{i_3}
\]

is the final result of the IDCF-HOSVD algorithm, which is used to combine information from different fused images collected from different sensors onboard UAVs and satellites in general. The philosophical framework is delineated by Tables I and II.

In Table I, the algorithm is described stepwise based on HOSVD. In the numerical experiments in this study, the input threshold \( T \) was chosen as the 90th percentile (i.e., leading values) in the beginning stage, and the MATLAB codes for Tucker decomposition from Bader et al. [53] were used. The HOSVD method is different, as compared to the logits of a
multistream neural network (e.g., [36]). HOSVD can extract the useful information entropy given different thresholds (i.e., filter rate) to denoise different classifiers via the higher order tensor flows. However, multistream CNNs for group activity recognition might have equivalent learning power via multiple pooling stages [54], which deserves further attention.

After the completion of the ensemble learning via HOSVD, mutual information assessment can be carried out using NCC and NCIE [41]. These two unique indicators can help identify which classifier contributes most to the final Chl-a concentration for our study area with 30-m resolution on a daily basis. The data fusion was carried out for the dates corresponding to the ground-truth data collection dates based on the data fusion basis. The data fusion was performed to meet the goal.

The results of the data fusion were validated by using the Shannon entropy index. The Shannon entropy of the fused image is generally between that of the two source images. Both metrics are defined by relying closely on the entropies of variables, indicating their nonlinear relationship. After a 2-D image is reshaped as a 1-D vector, the NCC between two same-length vectors can be treated like those between two same-size images. Thus, the NCIE among multiple same-size images can also be determined. When the images for computing NCC and NCIE consist of input images and the output/target image from IDCF-HOSVD, the values in the NCC matrix can show the correlation differences between the target image and each input image. Fig. 2 is an illustration of the implementation of IDCF-HOSVD in concert with estimating NCC and NCIE, using three input images as an example.

G. Holistic View of IDCF-HOSVD Algorithm

This comprehensive algorithm is summarized in Table II and depicted by Fig. 1 via the integration of data and classifier fusion for ensemble learning with IDCF-HOSVD.

III. RESULTS AND DISCUSSION

A. Data Fusion

The temporal limitation of the Landsat satellite and the spatial limitation of the MODIS satellite have motivated us to use data fusion as a gap-filling technique to create synthetic high-resolution images that can be used to predict water quality concentrations for our study area with 30-m resolution on a daily basis. The data fusion was carried out for the dates corresponding to the ground-truth data collection dates based on the data fusion steps, as described in Fig. 1. Both Landsat and MODIS images taken before and after the ground-truth data acquisition date are necessary. For each ground-truth observation, a fused image was processed to meet the goal.

The results of the data fusion were validated by using the Shannon entropy index. The Shannon entropy of the fused image is generally between that of the two source images. Both the probability density function (PDF) and entropy properties imply that the data fusion can maintain important properties via STARFM, from which the fused images exhibit information entropy patterns similar to the Landsat images, although with much higher spatial resolution (see Fig. 3). The more the bell-shape information entropy distribution is present, the less noise in the fused or final images via data fusion. The fused image should correctly conform to the ground-truth data of the
Fig. 2. Demonstration uses three input images ($U_1$, $U_2$, and $U_3$) to generate one target image (A) via IDCF-HOSVD. The estimated NCC among these four images and the two connection lines at (3, 1) or (4, 2) in the NCC matrix are shown as examples of two source input images of one NCC value. The resulting symmetric matrix with NCC values is also used for the estimation of NCIE.

Fig. 3. Estimated PDF for the Landsat image (band 1), MODIS image (band 3), and fused image (band 1) at Lake Nicaragua on March 25, 2015. Their estimated Shannon entropies are 4.3007, 6.0496, and 4.5582 for Landsat, MODIS, and fused images, respectively.

original high-resolution Landsat image. The CPU run time is about 25 min for fusing Landsat and MODIS images for four days, including all the corresponding bands for Lake Managua, and about an hour for Lake Nicaragua. The above computational time was obtained on a PC (CPU: Intel Core i7-4790U (four physical cores, 3.60 GHz), memory: 16 GB) using STARFM V1.2.1.

B. Analytical and Semiempirical Feature Extraction Methods

The LRM and TBM can be used as two baselines for comparison against the machine learning classifiers. With the ground-truth dataset, model calibration and validation were conducted for the LRM and TBM:

1) The LRM was calculated as follows:

$$\text{Chl–a} = 19.12076 + 0.00726 \times \text{band1} + 0.00816 \times \text{band2} - 0.00204 \times \text{band3}. \quad (30)$$

2) The TBM was calculated as follows:

$$\text{Chl–a} = 0.0166 \times (\text{band2}/\text{band1}) + 25.122. \quad (31)$$

C. Machine Learning Based Empirical Feature Extraction Methods

1) Comparison of ELM and SLFN: Two scenarios were used for deliberation to gain more insight about each model’s performance with different numbers of hidden nodes. In Scenario 1, 70% of the samples for training (28 samples) were fixed samples
chosen by manual selection. In Scenario 2, 70% of the samples for training (28 samples) were randomly selected by a MATLAB program based on the predicted values of the training output. DBN’s results were not comparable in this case, because DBN’s network architecture was fixed by the nodes over three hidden layers instead of using flexible numbers of hidden nodes. In the numerical experiments, the remaining 15%/15% samples were used for testing/validation.

To evaluate the performance among the two classifiers (SLFN, ELM) using the scaled training data, five statistical indices (RMSE, CO, PE, RSQ, and $R^2$) were calculated for different numbers of hidden nodes for Scenario 1 (note: result figures are not shown here). SLFN’s learning capacity is governed by $N \leq N \leq (1 + n/m)N$ [55], so the lower bound of the number of hidden nodes is $N/(1 + n/m) = 28/(1 + 3/1) = 7$ for this experiment. This indicates that SLFN with nine hidden nodes can be used for prediction because its RMSE of training error is small. ELM models require more neurons to reach the same level of prediction accuracy as SLFN, although its computational time is much less. Note that the PE index of ELM approaches zero when the number of hidden nodes is greater than 30, signifying the faster learning capacity. As for Scenario 2, the fluctuated results driven by random groupings for data used in testing and validation imply that the ELM has a smaller mean training error than the originally scaled random weight algorithm of SLFN, although its variation is sometimes a bit greater. In principle, as the number of hidden node increases, the performance of ELM and SLFN improves in both scenarios. In terms of the five statistical indices, the performance of ELM is better than that of SLFN. When the ELM model has 20–25 neurons in the hidden layer, the performance of CO, PE, and RSME can reach a stable level for Scenario 1. Although the fluctuations of indices in Scenario 2 reflect the impact of dynamic choices via the random sampling, unlike the cases in the smooth pattern in Scenario 1, the overall trend associated with each statistical index remains the same. Therefore, a random sampling scheme is more appropriate in a comparative analysis for testing the robustness of a feature extraction model based on the five statistical indices. The comparison in the following section provides further implications.

2) Comparison of Feature Extraction Models: Following the random sampling scheme, a holistic comparison of all feature extraction models (classifiers) is possible. In Table III, five statistical indices are used for comparative evaluation, with bold numbers representing the best result in each category (training/testing/validation). The values marked in bold represent the best value in each category (training/testing/validation). DBN used the samples assigned by the SLFN with the same number of hidden nodes (14).

| Training | Testing | Validation |
|----------|---------|------------|
|          | SLFN    | ELM        | DBN |
|          | RMSE    | 0.30       | 0.17 | 0.38 |
|          | CO      | 0.57       | 0.90 | 0.39 |
|          | PE      | -24.67     | -1.92 | -15.61 |
|          | RSQ     | 0.73       | 0.91 | 0.41 |
|          | $R^2$   | 0.53       | 0.82 | 0.17 |

Note: In total, 14 hidden nodes were chosen in experiments of both SLFN and ELM. The values marked in bold represent the best value in each category (training/testing/validation).

In Scenario 2, 70% of the samples were chosen in experiments of both SLFN and ELM. The values marked in bold represent the best value in each category (training/testing/validation). DBN used the samples assigned by the SLFN with the same number of hidden nodes (14).

| Training | Testing | Validation |
|----------|---------|------------|
|          | SLFN    | ELM        | DBN |
|          | RMSE    | 0.30       | 0.17 | 0.38 |
|          | CO      | 0.57       | 0.90 | 0.39 |
|          | PE      | -24.67     | -1.92 | -15.61 |
|          | RSQ     | 0.73       | 0.91 | 0.41 |
|          | $R^2$   | 0.53       | 0.82 | 0.17 |

Note: In total, 14 hidden nodes were chosen in experiments of both SLFN and ELM. The values marked in bold represent the best value in each category (training/testing/validation). DBN used the samples assigned by the SLFN with the same number of hidden nodes (14).

The columns of SLFN, ELM, and DBN represent the results of hidden nodes for Scenario 1 (note: result figures are not shown here). SLFN’s learning capacity is governed by $N \leq N \leq (1 + n/m)N$ [55], so the lower bound of the number of hidden nodes is $N/(1 + n/m) = 28/(1 + 3/1) = 7$ for this experiment. This indicates that SLFN with nine hidden nodes can be used for prediction because its RMSE of training error is small. ELM models require more neurons to reach the same level of prediction accuracy as SLFN, although its computational time is much less. Note that the PE index of ELM approaches zero when the number of hidden nodes is greater than 30, signifying the faster learning capacity. As for Scenario 2, the fluctuated results driven by random groupings for data used in testing and validation imply that the ELM has a smaller mean training error than the originally scaled random weight algorithm of SLFN, although its variation is sometimes a bit greater. In principle, as the number of hidden node increases, the performance of ELM and SLFN improves in both scenarios. In terms of the five statistical indices, the performance of ELM is better than that of SLFN. When the ELM model has 20–25 neurons in the hidden layer, the performance of CO, PE, and RSME can reach a stable level for Scenario 1. Although the fluctuations of indices in Scenario 2 reflect the impact of dynamic choices via the random sampling, unlike the cases in the smooth pattern in Scenario 1, the overall trend associated with each statistical index remains the same. Therefore, a random sampling scheme is more appropriate in a comparative analysis for testing the robustness of a feature extraction model based on the five statistical indices. The comparison in the following section provides further implications.

2) Comparison of Feature Extraction Models: Following the random sampling scheme, a holistic comparison of all feature extraction models (classifiers) is possible. In Table III, five statistical indices are used for comparative evaluation, with bold numbers representing the best result in each category (training/testing/validation). Based on the results of SLFN and ELM, both models’ RMSE, CO, RSQ, and $R^2$ can reach very good values in training, testing, and validation with 14 hidden nodes when compared with the cases with other numbers of hidden nodes. Thus, the best number of hidden nodes was selected as 14. In addition, the performance curves of SLFN with 14 hidden nodes are shown in Appendix III. It is noticeable that most testing samples fall well within the space spanned by the training samples, which makes the testing error less than the training error on some occasions. The TBM and LRM perform worse than machine learning models in terms of RMSE (see Table IV). However, SLFN has better results in validation based on all 40 samples. While ELM stands out in the training stage, its advantage does not persist over all occasions. Hence, the SLFN algorithm resulted in a relatively better water quality assessment.

The training time of DBN includes the computational time of using 200 iterations in the back-propagation algorithm for fine-tuning in MATLAB requires more computational time than SLFN. The main reason for this is that there are three layers, and many more hidden nodes, in DBN than in ELM or SLFN, which use only one hidden layer and at most 28 hidden nodes.
Content-based mapping may be conducted with respect to DBN, SLFN, ELM, LRM, and TBM individually. However, the bold numbers in Table III are very widely distributed, making it difficult to form a conclusion based on any classifier. The proposed IDCF-HOSVD scheme harmonizes the pros and cons of all five classifiers and produces one Chl-a map for the day when the fused satellite image is available. In our case study, there are only two fused maps available: one for the dry season (March 16, 2015), and one for the wet season (October 5, 2015). The final mapping outcome via the multimodal learning (IDCF-HOSVD) algorithm is described in the following section.

3) Remote Sensing Mapping for Water Quality Monitoring:

The predicted Chl-a concentration maps for Lake Nicaragua in the dry season (March 16, 2015) and the wet season (October 5, 2015) can be systematically presented based on SLFN, ELM, DBN, LRM, and TBM to reflect the role of different learning systems, from analytical (i.e., TBM), to semiempirical (i.e., LRM), to empirical (i.e., DBN, ELM, and SLFN), via the ensemble learning power embedded in the HOSVD-EL model (see Figs. 4 and 5). They look quite different depending on the season, which emphasizes the importance of ensemble learning in this case.

In addition to the 2-D maps of the Chl-a concentration of Lake Nicaragua, the estimated PDF (see Figs. 6 and 7) associated with each water quality map driven by a single classifier can indicate the distribution of predicted values over the lake and the impact of white noise. As shown in Fig. 4, the hot spots of pollution in the dry season occurred at the south tip of Lake Nicaragua, except for the map driven by the TBM classifier [Fig. 4(e)]. This snapshot is deemed reasonable, because the water flowing from Lake Managua to Lake Nicaragua through the Tipitapa River carries a pollution plume that amasses around the entrance of the San Juan River at the south tip of Lake Nicaragua, driven by the local landscape of the two-lake system. While the water quality pattern driven by the LRM classifier shows worse water quality, the water quality patterns driven by the three machine learning models look similar. Therefore, the ensemble learning process conducted by IDCF-HOSVD proved effective in harmonizing the discrepancies across all five classifiers and generating a more neutralized water quality pattern in the dry season.

However, as shown in Fig. 5, the water quality patterns in the wet season have no such strong hot spots due to the hydrodynamic dilution effect from intensive rainfall and runoff. While the ELM-based map exhibits high-value patches in the middle of the lake [Fig. 5(b)], resulting in an unsymmetrical PDF curve [Fig. 7(b)], the more homogeneous distributions of the SLFN-based map [Fig. 7(a)] and the DBN-based map [Fig. 7(c)] exhibit more symmetric PDF curves. Obviously, the TBM cannot work well, as it has no sensitivity to water quality gradients with scales. Thus, the ensemble learning process conducted by IDCF-HOSVD helps present a relatively homogeneous water quality distribution [Fig. 5(d)] evidenced by a symmetrical PDF curve [Fig. 7(d)]. Results indicate that the IDCF-HOSVD scheme can keep the hydrodynamic pattern driven by the ELM classifier while benefiting from the less heterogeneous patterns driven by the SLFN [Fig. 7(a)] and DBN [Fig. 7(c)] classifiers. Note that TBM has a low learning capacity compared to machine learning classifiers. This observation is evidenced by the fact that the histograms in Figs. 6(e) and 7(e) are almost constant. The values in Figs. 6(b) and 7(b), which have a much broader range, imply that ELM has a relatively weak learning capacity due to its sensitivity to outliers (i.e., noise).

This kind of learning power was proven effective through IDCF-HOSVD for different seasons after assimilating all inputs from the five classifiers, including SLFN, ELM, DBN, LRM, and TBM embedded in their leading eigenspace. This is a unique finding of its kind. The wet-season figures in Fig. 5 differ due to the varying learning capacities embedded in each classifier and disturbed by storm events in the watershed, resulting in much uncertainty in each learning algorithm. If multiple input images of HOSVD have high correlation, then it will result in an output...
Fig. 5. Prediction of the Chl-a value for Lake Nicaragua on October 5, 2015 (wet season) by using (a) SLFN, (b) ELM, (c) DBN, (d) LRM, (e) TBM, and (f) IDCF-HOSVD.

image that approaches a similar PDF. On the other hand, PDF is just one physical property of each output image, and thus two maps with similar PDF could have either a high or low correlation coefficient.

All the results derived from using single classifier can be assimilated and correlated into the final outputs, as shown in Figs. 6(f) and 7(f). As this is a water quality mapping of a big lake with no specific objects for segmentation step wise, there is no need to employ CNN. However, the use of CNN for river water quality mapping could be a powerful solution to landscape complexity [19]. CNN can be added on an as-needed basis at any time, further magnifying the advantage of a multimodal IDCF-HOSVD relative to other ensemble learning algorithms such as bagging, boosting, and random subspace methods.

To further assess the relationship among these inputs from the five different classifiers and the IDCF-HOSVD output, the NCC between any two (pair) was estimated following Wang et al. [41], with 100 bins in the algorithm, and the corresponding NCIE of the resulting NCC symmetric matrix was also estimated for advanced assessment in geophysics. Results of NCC and NCIE for dry and wet seasons are listed in Tables V and VI, respectively. The NCCs between IDCF-HOSVD images and images produced from each classifier suggest that the IDCF-HOSVD output follows the predicted patterns from SLFN, ELM, DBN, and TBM simultaneously during the dry season, and from ELM and TBM during the wet season. This implies that IDCF-HOSVD can synergize extreme information from different classifiers when facing a more dynamic situation (without an obvious pattern) and can harmonize diversified discrepancies when facing a less dynamic situation (with an obvious pattern). In this study, we used MATLAB R2019b to run the HOSVD algorithm with Intel i7-4700 (3.40 GHz, 8 cores) hardware and 24 GB memory. The HOSVD algorithm needs 5 h in computation when using a 65% threshold for Lake Nicaragua images (4303 × 4074 pixels) in our numerical experiments. If using a 90% threshold, it will take more than one day to complete. Thus, the tradeoff between cost and benefit is obvious.

In addition, according to (1), the regional Shannon entropies of Chl-a value at 11 observational locations [as indicated by Appendixes I(b) and II(b)] for different classifiers and IDCF-HOSVD were also estimated for both seasons (see Fig. 8). The regional Shannon entropy was estimated within a circular region with 0.1° radius (in latitude–longitude coordinate), wherein each Chl-a pixel’s probability was adopted from the corresponding PDF of the entire Chl-a map associated with one classifier. According to the results in Fig. 8, DBN with deep learning

| SLFN | ELM | DBN | LRM | TBM | IDCF-HOSVD |
|------|-----|-----|-----|-----|------------|
| 1    |     |     |     |     |            |
| 0.5382 | 0.1178 | 1 |
| 0.0716 | 0.1062 | 0.0549 | 1 |
| 0.2791 | 0.1187 | 0.1825 | 0.0605 | 1 |
| 0.1470 | 0.1625 | 0.1338 | 0.0679 | 0.1368 | 1 |

| SLFN | ELM | DBN | LRM | TBM | IDCF-HOSVD |
|------|-----|-----|-----|-----|------------|
| 1    |     |     |     |     |            |
| 0.0574 | 1 |
| 0.0419 | 0.1869 | 1 |
| 0.0713 | 0.0550 | 0.0589 | 1 |
| 0.0951 | 0.3459 | 0.1612 | 0.0310 | 1 |
| 0.0742 | 0.3249 | 0.1701 | 0.0280 | 0.3815 | 1 |
capacity can acquire more entropy than others, followed by either SLFN or TBM over the two seasons.

Overall, the entropy values of IDCF-HOSVD are relatively smaller in both seasons, implying that IDCF-HOSVD emphasizes pattern recognition rather than information retrieval based on these classifiers’ predictions at 11 locations in an aquatic environment. It is noticeable that the lake water quality has been declining because of surrounding anthropogenic activities, deforestation in the watershed, and the uncontrolled disposal of municipal wastewater into the lake. Big cities like Granada, Rivas, and Juigalpa, along with some other small towns surrounding the lake, divert their sewage and industrial wastes directly toward Lake Nicaragua or toward the channels that discharge into the lake. Moreover, fertile soil conditions near the lake, and the growing agricultural industry in the coastal region, are other sources of contamination for the lake water. Agricultural activities, plantation, and cattle farming near the Chontales, Boaco, and Rivas areas input an enormous amount of fertilizer.

Fig. 6. Estimated PDF of the Chl-a value within [0, 100] for Lake Nicaragua on March 16, 2015 by using (a) SLFN, (b) ELM, (c) DBN, (d) LRM, (e) TBM, and (f) IDCF-HOSVD.

Fig. 7. Estimated PDF of the Chl-a value within [0, 100] for Lake Nicaragua on October 5, 2015 by using (a) SLFN, (b) ELM, (c) DBN, (d) LRM, (e) TBM, and (f) IDCF-HOSVD.
directly into the lake, significantly increasing the eutrophication potential in wet seasons. These observations are consistent with the concentration maps of Chl-a.

4) Sensitivity Analysis in Multimodal Learning via Tensor Structure: In the IDCF-HOSVD algorithm, the threshold \( T \) can be chosen as a variable for sensitivity analysis. Theoretically, a greater percentile of threshold \( T \) results in better optimization output of image \( A^* \) in the algorithm (see Table I) in terms of the cost function of (23). This is because more base-pixels may be introduced to optimize \( A^* \) at the expense of more computations. Users must always balance the cost (computational load) and benefit (prediction accuracy) by choosing a certain threshold. A sensitivity analysis may further help identify or verify the optimal threshold.

Six different percentiles (i.e., 65%, 70%, 75%, 80%, 85%, and 90%) were used for demonstration in this sensitivity analysis. Their RMSEs corresponding to (23) were calculated and compared to show the changes of RSME due to the selection of different percentiles (threshold) relative to the baseline (90%). These relative RMSE values (ratio) and their performance are depicted in Fig. 9 for both dry and wet seasons, in which a 90% threshold in the IDCF-HOSVD experiments is the comparative basis. Results indicate that the RMSE decreases with the increasing percentile (more stringent threshold or larger filter rate), implying that higher prediction accuracy does occur along with a greater threshold (higher percentile).

However, the magnitude of RMSE changes is very small over the range when it approaches 100%. This is exactly the case of the marginal rate of return, implying that a smaller percentile (i.e., 65%) can achieve almost the same results as a higher one (i.e., 90%), because the base images associated with a smaller percentile can capture enough features in the context of content-based mapping via IDCF-HOSVD. Thus, a smaller percentile can be adopted in practice to reduce computational load.

IV. CONCLUSION

This study demonstrated a new ensemble learning approach for IDCF, in which multimodal learning with machine intelligence was emphasized via IDCF-HOSVD. In the beginning stage, the high spatial resolution Landsat imageries and the high temporal resolution MODIS imageries were fused to produce images with both high spatial and high temporal resolutions. Then, in the feature extraction stage, five inverse models (classifiers) were applied for processing fused images toward feature extraction and water quality mapping. The performance of the five classifiers was evaluated in terms of a set of statistical indices, computational time, water quality patterns, and corresponding PDFs. Finally, the IDCF-HOSVD algorithm utilizes the eigenspace information to synthesize the core information entropy, harmonize the prediction discrepancies across the five classifiers, and produce the final water quality maps. Practical implementation was assessed using the final spatial patterns of Chl-a concentration maps in both the wet and dry seasons of Lake Nicaragua, Central America. Paired water quality maps derived by the ensemble learning algorithm and any of the classifiers were further investigated with the aid of NNC and NCIE to assure the individual contribution of each classifier to the final pattern recognition via IDCF-HOSVD.

From the RMSE comparison for multimodal learning in Table IV, the machine learning based empirical classifiers (i.e., SLFN, ELM, and DBN) outperformed the semiempirical and analytical classifiers (i.e., LRM and TBM). DBN can play a critical role in extracting inherent patterns with a higher level of information entropy under a more dynamic situation. TBM might be hampered by its insensitivity from low-level to
high-level features, which deeply affects prediction accuracy. This indicates that the ensemble learning model can harmonize some irregular patches in water quality maps and present a synthesized pattern via an intrinsic assortment scheme in an eigenspace with the aid of the eigenvector via HOSVD. According to the NCC results in Tables V and VI, all classifiers except LRM can play a significant role in different scenarios. DBN and TBM can contribute to more information entropy than their counterparts. Thus, to answer question 1 of this study, empirical classifiers can outperform LRM, but not TBM. To answer question 2, HOSVD can pick up all leading base images from the predicted images associated with different classifiers in the spectral space, preserving the most important information related to inherent patterns from those images/classifiers. This is true for both the dry season and the wet season. To answer question 3, this study suggested that DBN most benefits IDCF-HOSVD in terms of information entropy. Future efforts will be directed regarding the information flows of multibands from visible bands to other (e.g., microwave) bands within the IDCF-HOSVD framework and the comparison with multistream CNNs for group activity recognition.

APPENDIX I
GROUND-Truth DATA

29 SAMPLES FROM LAKE MANAGUA FOR TRAINING, TESTING, AND VALIDATION

| No. | Date      | Band 1 | Band 2 | Band 3 | Chl. n | Train | Test | Valid |
|-----|-----------|--------|--------|--------|--------|-------|------|-------|
| 2   | 08/30/2014| 541    | 616    | 540    | 33.68  |       |      | x     |
| 2   | 01/15/2015| 825    | 819    | 819    | 26.64  |       |      | x     |
| 3   | 08/30/2014| 943    | 999    | 884    | 32.56  |       |      | x     |
| 5   | 08/30/2014| 617    | 727    | 727    | 55.54  |       |      | x     |
| 4   | 01/15/2015| 490    | 721    | 596    | 28.88  |       |      | x     |
| 4   | 04/15/2015| 577    | 540    | 649    | 18.50  |       |      | x     |
| 5   | 04/09/2015| 451    | 806    | 868    | 25.16  |       |      | x     |
| 6   | 07/15/2015| 712    | 906    | 797    | 28.88  |       |      | x     |
| 7   | 04/09/2015| 497    | 459    | 742    | 14.64  |       |      | x     |
| 7   | 07/15/2015| 543    | 797    | 794    | 20.42  |       |      | x     |
| 7   | 04/09/2015| 1954   | 9590   | 1820   | 19.98  |       |      | x     |
| 8   | 07/15/2015| 704    | 981    | 841    | 25.16  |       |      | x     |
| 7   | 04/09/2015| 4581   | 4552   | 5103   | 14.64  |       |      | x     |
| 9   | 07/30/2015| 404    | 355    | 328    | 78.44  |       |      | x     |
| 9   | 04/08/2015| 1247   | 1335   | 1340   | 25.16  |       |      | x     |
| 9   | 07/15/2015| 1278   | 1131   | 941    | 28.88  |       |      | x     |
| 10  | 04/09/2015| 4056   | 3851   | 4743   | 8.88   |       |      | x     |
| 10  | 05/13/2015| 774    | 810    | 476    | 22.20  |       |      | x     |
| 11  | 04/08/2015| 1378   | 1396   | 1550   | 12.59  |       |      | x     |
| 12  | 07/10/2015| 514    | 819    | 1081   | 14.64  |       |      | x     |
| 12  | 04/09/2015| 1230   | 1335   | 1488   | 33.50  |       |      | x     |
| 13  | 07/10/2015| 952    | 3612   | 1826   | 22.94  |       |      | x     |
| 13  | 04/09/2015| 1152   | 1399   | 1283   | 19.98  |       |      | x     |
| 14  | 04/08/2015| 1278   | 1401   | 1263   | 27.58  |       |      | x     |
| 15  | 04/08/2015| 1174   | 1418   | 1253   | 20.72  |       |      | x     |
| 16  | 05/13/2015| 798    | 799    | 570    | 22.20  |       |      | x     |
| 17  | 07/29/2015| 1118   | 1041   | 933    | 29.96  |       |      | x     |

Note: The checked marks in the training/testing/validation column indicate the assignment scheme for fixed samples.
ACKNOWLEDGMENT

The authors would like to acknowledge suggestions from Dr. W. Zhang who is funded in part by the UCF Pre-eminent Postdoctoral Program.

REFERENCES

[1] K. X. Bai, N. B. Chang, and C. F. Chen, “Spectral information adaptation and synthesis scheme for merging cross-mission consistent ocean color reflectance observations from MODIS and VIIRS,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 1, pp. 311–329, Jan. 2016.

[2] N. B. Chang, B. Vannah, Y. J. Yang, and M. Elovitz, “Integrated data fusion and mining techniques for monitoring total organic carbon concentrations in a lake,” Int. J. Remote Sens., vol. 35, no. 3, pp. 1064–1093, 2014.

[3] N. B. Chang, B. Vannah, and Y. J. Yang, “Comparative sensor fusion between hyperspectral and multispectral remote sensing data for monitoring microcystin distribution in Lake Erie,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2426–2442, Jun. 2014.

[4] K. Bai, N. B. Chang, H. Yu, and W. Gao, “Statistical bias corrections for creating coherent total ozone records with OMI and OMPS observations,” Remote Sens. Environ., vol. 182, pp. 150–168, 2016.

[5] Y. Won and P. Gader, “Morphological shared-weight neural network for pattern classification and automatic target detection,” in Proc. ICNN’95, Int. Conf. Neural Networks, Univ. Missouri-Columbia, Columbia, MO, USA, vol. 4, Nov. 1995, pp. 2134–2138.

[6] G. B. Huang, Q. Y. Zhu, and C. K. Siew, “Real-time learning capability of neural networks,” IEEE Trans. Neural Netw., vol. 17, no. 4, pp. 863–878, Aug. 2006.

[7] G. B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: Theory and applications,” Neurocomputing, vol. 70, no. 1, pp. 489–501, 2006.

[8] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” J. Neural Comput., vol. 18, no. 7, pp. 1527–1554, 2006.

[9] K. Nogueira, W. O. Miranda, and J. A. Dos Santos, “Improving spatial feature representation from aerial images by using convolutional networks,” in Proc. 28th IEEE SIBGRAPI Conf. Graph., Patterns Images, 2015, pp. 289–229.

[10] Y. Bengio, “Learning deep architectures for AI,” Found. Trends Mach. Learn., vol. 2, no. 1, pp. 127–129, 2009.

[11] O. Penatti, K. Nogueira, and J. Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2015, pp. 44–51.

[12] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016. [Online]. Available: http://goodfeli.github.io/dlbook

[13] Z. Wang, K. Kawamura, Y. Sakono, X. Fan, Z. Gong, and J. Lim, “Retrieval of Chlorophyll-a and total suspended solids using iterative stepwise elimination partial least squares (ISE-PLS) regression based on field hyperspectral measurements in irrigation ponds in Higashihirashima, Japan,” Remote Sens., vol. 9, p. 264, 2017, doi: 10.3390/rs9030264.

[14] M. A. Matus-Hernández, N. Y. Hernández-Saavedra, and R. O. Martínez-Rincón, “Predictive performance of regression models to estimate Chlorophyll-a concentration based on Landsat imagery,” Plos One, vol. 13, 2018, Art. no. e0205682. [Online]. Available: https://doi.org/10.1371/journal.pone.0205682

[15] N. B. Chang, K. X. Bai, and C. F. Chen, “Smart information reconstruction via time-space-spectrum continuum for cloud removal in satellite images,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 5, pp. 1898–1912, May 2015.

[16] D. Rummelhart, G. Hinton, and R. Williams, “Learning internal representations by error propagation,” in Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations, vol. 1. Cambridge, MA, USA: MIT Press, 1986.

[17] P. J. Werbos, “Back propagation through time: What it does and how to do it,” Proc. IEEE, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[18] N. B. Chang, C. Mostafiz, Z. Sun, W. Gao, and C. F. Chen, “Developing a prototype satellite-based cyber-physical system for smart wastewater treatment,” in Proc. IEEE 14th Int. Conf. Netw. Sens. Control, 2017, pp. 339–344.

[19] T. Oga, Y. Umeki, M. Iwashashi, and Y. Matsuda, “River water quality estimation based on convolutional neural network,” presented at the Asia-Pacific Signal Information Processing Association Annu. Summit Conf., Honolulu, HI, USA, Nov. 2018, doi: 10.23919/APSIPA.2018.8695954.

[20] B. V. Sadasathy and B. V. Sheela, “A composite classifier system design: Concepts and methodology,” Proc. IEEE, vol. 67, no. 5, pp. 708–713, May 1979.

[21] L. K. Hansen and P. Salamon, “Neural network ensembles,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 10, pp. 993–1001, Oct. 1990.

[22] R. E. Schapire, “The strength of weak learnability,” Mach. Learn., vol. 5, no. 2, pp. 197–227, 1990.

[23] G. Evensen, “The ensemble Kalman filter: Theoretical formulation and practical implementation,” Ocean Dyn., vol. 53, no. 4, pp. 343–367, 2003.

[24] D. Morrison, R. Wang, and L. C. De Silva, “Ensemble methods for spoken emotion recognition in call-centres,” Speech Commun., vol. 49, no. 2, pp. 98–112, 2007.

[25] S. Chebrolu, A. Abraham, and J. P. Thomas, “Feature deduction and ensemble design of intrusion detection systems,” Comput. Secur., vol. 24, no. 4, pp. 295–307, 2005.

[26] L. Breiman, “Bagging predictors,” Mach. Learn., vol. 24, no. 2, pp. 123–140, 1996.

[27] T. K. Ho, “The random subspace method for constructing decision forests,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 8, pp. 832–844, Aug. 1998.

[28] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, 1997.

[29] D. H. Wolpert, “Stacked generalization,” Neural Netw., vol. 5, no. 2, pp. 241–259, 1992.

[30] M. I. Jordan and R. A. Jacobs, “Hierarchical mixtures of experts and the EM algorithm,” Neural Comput., vol. 6, no. 2, pp. 181–214, 1994.

[31] L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms. Hoboken, NJ, USA: Wiley, 2004.

[32] T. K. Ho, J. J. Hull, and S. N. Srihari, “Decision combination in multiple classifier systems,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 1, pp. 66–75, Jan. 1994.

[33] J. Kittler, M. Hatef, P. R. Duin, and J. Matas, “On combining classifiers,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 3, pp. 226–239, Mar. 1998.

[34] L. I. Kuncheva, “A theoretical study on six classifier fusion strategies,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 2, pp. 281–286, Feb. 2002.

[35] L. I. Kuncheva, J. C. Bezdek, and R. P. Duin, “Decision templates for multiple classifier fusion: An experimental comparison,” Pattern Recognit., vol. 34, no. 2, pp. 299–314, 2001.

[36] Z. H. Zhou, J. Wu, and W. Tang, “Ensemble learning neural networks: Many could be better than all,” Artif. Intell., vol. 137, no. 1/2, pp. 239–263, 2002.

[37] L. Wang and P. N. Suganthan, “Benchmarking ensemble classifiers with novel co-trained kernel ridge regression and random vector functional link ensembles [research frontier],” IEEE Comput. Intell. Mag., vol. 12, no. 4, pp. 61–72, Nov. 2017.

[38] L. R. Tucker, “Some mathematical notes on three-mode factor analysis,” Psychometrika, vol. 31, no. 3, pp. 279–311, 1966.

[39] M. A. O. Vasilescu and D. Terzopoulos, “Multilinear analysis of image ensembles: Tensorfaces,” in Proc. Eur. Conf. Comput. Vis., 2002, pp. 447–460.
A. Rajwade, A. Rangarajan, and A. Banerjee, “Image denoising using the higher order singular value decomposition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 849–862, Apr. 2013.

Q. Wang, Y. Shen, and J. Q. Zhang, “A nonlinear correlation measure for multivariable data set,” *Phys. D, Nonlinear Phenom.*, vol. 200, no. 3/4, pp. 287–295, 2005.

World Bank, “Policy and investment priorities to reduce environmental degradation of the Lake Nicaragua watershed (Cocibolca),” Latin America Caribbean Region Environ. Water Resources Occasional Paper Ser., World Bank, Washington, DC, USA, Rep. 76886, 2013.

J. G. Masek et al., “A Landsat surface reflectance data set for North America, 1990-2000,” *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 68–72, Jan. 2006.

E. F. Vermote, S. Y. Kotchenova, and J. P. Ray, MODIS Surface Reflectance User’s Guide. MODIS Land Surface Reflectance Science Computing Facility, Ver. 1, 2008.

United States Geological Survey, Landsat Processing Details, Accessed: Jul. 27, 2016. [Online]. Available: http://landsat.usgs.gov/Landsat_Processing_Details.php

F. Gao, J. Masek, M. Schwaller, and F. Hall, “On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.

W. Shi, C. Zhu, Y. Tian, and J. Nichol, “Wavelet-based image fusion and quality assessment,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 6, no. 3, pp. 241–251, 2005.

A. G. Dekker, Ž. Zamurović-Nenad, H. J. Hoogenboom, and S. W. M. Peters, “Remote sensing, ecological water quality modelling and in situ measurements: A case study in shallow lakes,” *Hydrol. Sci. J.*, vol. 41, no. 4, pp. 531–547, 1996.

T. Yu, “Utility of remote sensing data in retrieval of water quality constituents concentrations in coastal water of New Jersey,” Ph.D. dissertation, Civil Environ. Eng. Dept., New Jersey Inst. Technol., Newark, NJ, USA, 2005. [Online]. Available: http://archives.njit.edu/etd/2005/niit-etd2005-081/niit-etd2005-081.pdf. Accessed on: Aug. 2016.

C. M. Bishop, *Pattern Recognition and Machine Learning*. Cambridge, U.K.: Springer, 2006, p. 128.

G. E. Hinton, “A practical guide to training restricted Boltzmann machines,” in *Neural Networks: Tricks of the Trade*, Berlin, Heidelberg: Springer, pp. 599–619, 2012.

M. A. Keyvanrad and M. M. Homayounpour, “A brief survey on deep belief networks and introducing a new object-oriented toolbox (DeeBNet),” 2014, arXiv:1408.3264.

B. W. Bader et al., “MATLAB tensor toolbox,” version 3.1, 2019. [Online]. Available: https://www.tensortoolbox.org

Z. Tu et al., “Multi-stream CNN: Learning representations based on human-related regions for action recognition,” *Pattern Recognit.*, vol. 79, pp. 32–43, 2018.

P. L. Narasimha, M. T. Manry, and F. Maldonado, “Upper bound on pattern storage in feedforward networks,” *Neurocomputing*, vol. 71, no. 16, pp. 3612–3616, 2008.