HABNet: Machine Learning, Remote Sensing-Based Detection of Harmful Algal Blooms

Paul R. Hill, *Member, IEEE*, Anurag Kumar, Marouane Temimi, and David R. Bull, *Fellow, IEEE*

***Abstract*—This article describes the application of machine learning techniques to develop state-of-the-art detection and prediction system for spatiotemporal events found within remote sensing data; specifically, harmful algal bloom (HAB) events. We propose HAB detection system based on a ground truth historical record of HAB events, a novel spatiotemporal datacube representation of each event (from MODIS and GEBCO bathymetry data), and a variety of machine learning architectures utilizing the state-of-the-art spatial and temporal analysis methods based on convolutional neural networks, long short-term memory components together with random forest, and support vector machine classification methods. This work has focused specifically on the case study of the detection of *Karenia brevis* algae (*K. brevis*) HAB events within the coastal waters of Florida (over 2850 events from 2003 to 2018; an order of magnitude larger than any previous machine learning detection study into HAB events). The development of multimodal spatiotemporal datacube data structures and associated novel machine learning methods give a unique architecture for the automatic detection of environmental events. Specifically, when applied to the detection of HAB events, it gives a maximum detection accuracy of 91% and a Kappa coefficient of 0.81 for the Florida data considered. A HAB forecast system was also developed where a temporal subset of each datacube was used to predict the presence of a HAB in the future. This system was not significantly less accurate than the detection system being able to predict with 86% accuracy up to 8 d in the future.**

***Index Terms*—Convolutional neural networks (CNNs), deep learning, harmful algal blooms (HABs), long short-term memory (LSTMs), random forest (RF), support vector machine (SVM).**

I. Introduction

ALGAL blooms are defined as high concentrations of phytoplankton (algae) (for example, >50000 cells/L [5]). Harmful algal blooms (HABs) are problematic algal blooms causing toxicity and associated environmental impacts. Often termed "Red Tides," HABs have been a significant worldwide research topic over three decades [1]–[7].

They continue to be of major concern, not only due to their considerable environmental and societal impact but also a recent significant increase in frequency reported around the world [2].

Manuscript received April 16, 2020; revised May 16, 2020; accepted June 1, 2020. Date of publication June 10, 2020; date of current version June 23, 2020. This work was supported by The British Council Award under Reference 279334808. *(Corresponding author: Paul Hill.)*

Paul R. Hill and David R. Bull are with the Department of Electrical and Electronic Engineering, The University of Bristol, BS8 1UB Bristol, U.K. (e-mail: paul.hill@bristol.ac.uk; dave.bull@bristol.ac.uk).

Anurag Kumar and Marouane Temimi are with the Khalifa University (KUSTAR), Abu Dhabi 127788, United Arab Emirates (e-mail: anurag.kumar@ku.ac.ae; marouane.temimi@ku.ac.ae).

Digital Object Identifier 10.1109/JSTARS.2020.3001445

HABs can cause severe environmental and human health problems together with associated economic impacts [8]. Environmental impacts include mass fish stock and marine wildlife kills [2]. Human impacts include toxic reactions to affected seafood and, in extreme cases, fatalities [9]. Economic impacts include adverse effects on beach and coastal tourism-based activities together with impacts on coastal-based industries (e.g., fishing) [10]–[12]. Within the United States alone, HABs cause an estimated annual economic loss of at least $82 million [13].

Many factors have been cited as causes of HABs but are generally caused by favorable environmental conditions, including increasing nutrient levels [14], light availability [15], water column stratification, and/or changes in water temperature [16].

Conventionally, the measuring of algae concentrations has relied on direct water sampling for lab-based cell taxonomy. These manual methods of detection and analysis are extremely labor-intensive and are limited spatially and temporally [17]. Conversely, remote sensing-based detection methods have excellent coverage in time and space and offer analysis systems that are not labor intensive. However, remote sensing-based detection methods often rely on estimated remote sensing products such as chlorophyll (Chl-a) that themselves may be unreliable estimates and not a direct measurement (and therefore not precisely accurate) of cell concentrations.

HABs have a spatiotemporal footprint that ranges from weeks to months and from a few square kilometers to thousands of square kilometers [2], [18]. It is implicit that these HABs are spatially and temporally dependent, and for the most effective detection and prediction, a combined spatial and temporal analysis is required.

A. Background and Contributions

HAB monitoring and forecasting using remote sensing data was first proposed by Steidiner and Haddad in 1981 [1], utilizing data from the Coastal Zone Color Scanner sensor onboard Nimbus-7, operational during the 1970s and 1980s.

This work subsequently led to a large number of remote sensing detection, monitoring, and forecasting systems developed for more recent sensors and satellites such as MODIS-Aqua, MODIS-Terra, SeaWiFS, MERIS, and, more recently, Sentinel-3 [2]. The methods used for detection, monitoring, and forecasting of HAB events have included reflectance band-ratio-based detection, reflectance classification (using anomaly detection), satellite product based detection (using thresholds), and spectral band differences.

The most successful and important methods for HAB detection have used spectrally derived products such as Chl-a (chlorophyll concentration estimate) as phytoplankton increases the backscattered light within pigment absorption spectral frequencies. An excellent review of these historical and current methods, sensors, and satellites is given by Blondeau-Patissier *et al.* [2].

There is currently no nationwide or international HAB forecasting system for HABs. However, there are specific areas covered by HAB forecasting systems such as NOAA's HAB forecasting systems [HAB Operational Forecast System (HAB-OFS)] for the Florida region [19]. However, this system only forecasts up to 4 d and focuses mainly on the human impact of HABs (respiratory effects). The HAB Observing System (HABSOS) is a detailed observation system of HABs within the Gulf of Mexico which has also been developed by NOAA [20]. HAB-OFS and HABSOS provide forecasts to stakeholders such as local resource and environmental managers, the seafood industry, and those managing tourism activities.

Within this article, state-of-the-art supervised machine learning systems are proposed for HAB detection and prediction within the region of the Gulf of Mexico and also in an alternative case study within the Arabian Gulf.

We conjecture that large-scale spatial patterns play an extremely significant role in the effective detection and prediction of HABs. We have therefore utilized machine learning tools that not only effectively characterize spatial patterns but combine them with time series analysis machine learning tools such as LSTMs.

B. Article Organization

The rest of this article is organized as follows. Section II gives an overview of the problem and applicable machine learning systems. Section III gives an overview of the proposed HAB detection and prediction methods. Section IV describes the "datacube" data structure and the creation of a large number of datacubes from the ground truth database. Section V then describes the preprocessing of the datacubes in order to make them ready to be ingested into the machine learning system. Then Section VI describes and illustrates the created machine learning structure, and Section VII illustrates how important each modality/feature is for the classification. The classification results are given and discussed within Section VIII. An alternative case study (HABs in the Arabian Gulf) is investigated in Section IX. Finally, a conclusion is given in Section X.

II. Review of HAB Detection Methods

Previous remote sensing-based HAB detection methods have, in the majority of cases, used spatially isolated and single satellite sensor data samples. Many methods have been developed for HAB detection utilizing a wide range of satellite sensors and bands.

Many common methods of HAB detection are currently based on chlorophyll concentration products as Chl-a is, in many cases, a very accurate proxy of local algal activity. Phytoplankton is the primary water constituent [21], [22]; thus, Chl-a can often be accurately estimated using the water-leaving reflectance using relationships (such as remote sensing band ratios) for data from sensors such as SeaWiFS, MERIS, and MODIS [23], [24]. The accuracy of estimating Chl-a by remote sensing sensors has aimed to be within $\pm 35\%$ in deep waters [2]. However, this accuracy has not always been found to have been met by simply using band-ratio algorithms (e.g., [25]).

These simplistic methods in many cases suffer from a large quantity of false positive detections. The most effective updates to these methods further consider measures of carbon dissolved organic matter utilizing backscattering data from SeaWiFS and MODIS [26], [27].

HAB detection using these products often use a chlorophyll anomaly measure that characterizes the difference between today's Chl-a and a background (often monthly or bi-monthly) average concentration [3], [4], [28]. This method is also known as background subtraction [28].

Another method of reducing the false positives associated with Chl-a HAB detection is the backscattering ratio algorithm [26], [27]. This algorithm utilizes a thresholded ratio formed from Rrs(555) and Chl-a.

Other optical methods have also been used, such as the spectral shape (SS) algorithm [29]. This system was proposed to discriminate *K. brevis* from other blooms creating high Chl-a values. Alternative methods have used both MODIS derived fluorescence line height products and locally tuned algorithms to accommodate common inaccuracies in Chl-a estimation in shallow coastal regions [27].

There are only a limited number of machine learning based HAB detection/prediction systems reported in the literature. Support vector machines (SVMs) have been proposed for this application by Song *et al.* [7] and Li *et al.* [30]. Spatiotemporal analysis using machine learning methods have also been proposed by Gokaraju *et al.* [5], [6]. Other nonmachine learning methods have been proposed for HAB monitoring, detection, and prediction (e.g., [12] and those described within [2]). Machine learning techniques have also been combined with GIS methods to produce interactive predictions [31].

Our work describes the definition of a unique datacube data structure for supervised machine learning of spatiotemporal events; specifically, HAB events together with a novel machine learning architecture to provide optimal HAB event classification and prediction performance.

A. Applicable Machine Learning Methods

Due to improved neural network models and methods combined with improvements in computational power and the availability of extremely large ground truth datasets, image classification performance has recently shown dramatic improvements [32]. Deep convolutional neural networks (CNNs) are now commonly used, simple to understand, and highly effective neural networks for image classification and characterization [33]–[36].

Recurrent neural networks (RNNs) [37] have been used for state-of-the-art classification and characterization of temporally based signals. LSTMs (long short-term memory) are the

dominant RNN form able to characterize and model both long- and short-term dependencies in temporal information [38], [39]. LSTM methods have given excellent results in many temporal characterization problems. However, more recently, alternative methods based on the concept of "Attention" have given better results in many cases [40].

HAB detection requires both spatial and temporal classifications. Previous spatiotemporal characterization methods such as video sequence classification [41] and multiview classification [42] have used and combined CNN and LSTM architectures. HAB event characterization is different from these methods as the input imaging data is multimodal (in our case, it has 12 dimensions). We propose a novel architecture that modifies these previous machine learning models and methods to take into account the multimodal inputs.

Given that the temporal range investigated within the datacubes (see below) is small, we have also investigated flattening the time series sequences and utilized simple high performance non-network classifiers as a last stage: Random forests (RF) [43]; SVMs [44]; and nontemporal, fully connected networks [multilayer perceptrons (MLPs)].

III. Proposed HAB Detection System

The proposed HAB detection system uses a supervised machine learning method. Supervised machine learning requires a detailed ground truth dataset, i.e., labeled positive and negative HAB events defined in time and location together with characterizing remote sensing data.

We have obtained some very large HAB event datasets from Florida Fish and Wildlife Conservation Commission (FWC) [45], the Phytoplankton Monitoring Network [46], and the Harmful Algal Event Database [47].

We have selected the data from the FWC as the dataset is extremely large (of both positive and negative HAB events) together with spanning the dates between 2001 and 2018.

We have chosen a subset of the FWC data from 2003 to June 2018 as it can be effectively characterized by the flight times and data availability of MODIS-Aqua and MODIS-Terra satellites and sensors. More recent and up-to-date satellite sensors such as Sentinel-3 have only recently become active and therefore do not have a large amount of historical sensor data covering the date range within the ground truth.

Only *K. brevis* algae events were extracted from this dataset in order to provide a tractable solution (*K. brevis* is considered to be the most serious cause of HAB events within the Gulf of Mexico region [2], [5], [48]).

In order to further reduce the size of the dataset, a HAB event was considered to have occurred when the event count algae abundance in cells/L is in excess of 50 000. Samples with cells/L between 0 and 50 000 were discounted from the dataset (i.e., neither positive nor negative). This threshold is chosen as it was the threshold used in the previous work by Gokaraju *et al.* [5]. The selection of *K. brevis* events and the 50 000 threshold led to the number of positive events being 1755 (between 2003 and 2018). One thousand one hundred and fourteen negative events were selected from the entire dataset where the algae count in cells/L were 0. It was assumed that the sampling positions and times for the positive and negative events were equivalent (i.e., there was no discrimination possible between the times and places sampled and found to be either positive or negative). Fig. 1 shows the spatial distribution of a selection of these positive events with the circle size reflecting the cells/L count. The positive datapoints have the following statistics (cells/L): Max = 1.62E+8, min = 5.03E+04, median = 2.38E+05, and mean = 1.11E+06.

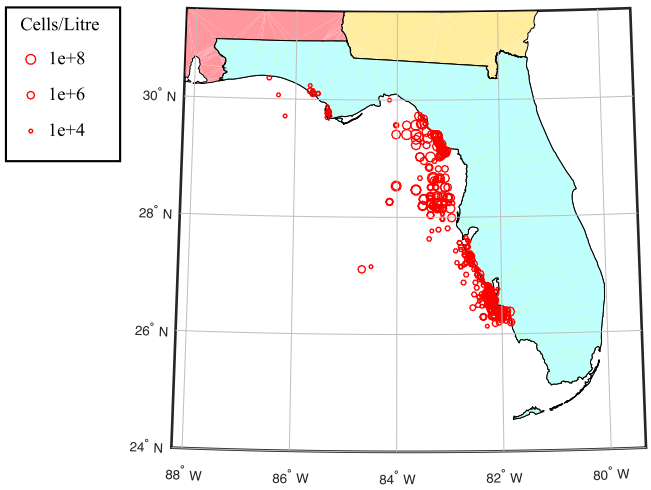

Fig. 1. Selection of HAB events near Florida (2003–2018). Circle radius reflects the log of the algae count (cells/L) of *K. brevis* HAB event as indicated by the legend.

IV. Datacubes for HAB Detection

The most effective characterization of HAB events for HAB event detection needs a "datacube" of remote sensing data that surrounds each HAB event in time and space.

Previous datacube protocols, methods, and code bases have been defined and implemented (e.g., [49]). However, these datacubes are unable to give the required structure and/or access to remote sensing data surrounding spatially and temporally localized events.

We have therefore developed a novel datacube definition as illustrated in Fig. 2. Each datacube associates a range of modalities within a spatial and temporal neighborhood of each data point with the positive and negative HAB ground truth database, i.e., there is a spatiotemporal window defined (in meters and days) surrounding the central ground truth location (in latitude, longitude, and date).

Extraction of remote sensing data is enabled using NASA's Common Metadata Repository (CMR) search facility [50]. OB.DAAC L2 satellite file granule names are obtained, given a target latitude, longitude, and date of a HAB event. These granules, in NetCDF format, are downloaded and the datacube components are extracted within the spatial and temporal neighborhood of each HAB event. Datacube generation is summarized in Algorithm 1.

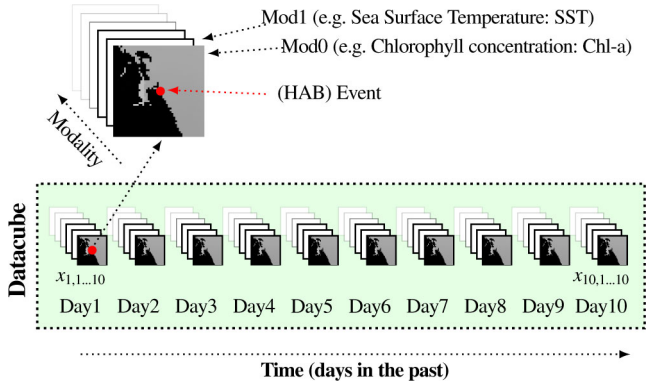

Fig. 2. Structure of a datacube used in this article.

Algorithm 1: Creation of ML Datacube.

```
Input  : Groundtruth File
for ∀ HAB events in Groundtruth File do
    Extract HAB event Lat, Lon, Date Window
    for ∀ List of Modalities do
        Generate list of granules using NASA CMR search
          (within 10 days previously of HAB event date)
        for ∀ NetCDF Granules in Date Range do
            wget NetCDF Granule
            Extract modality data in spatial window
            Place cropped data in output Datacube
        end
    end
    Output  Datacube
    :
end
```

A. Selected Modalities

The datacube architecture and the subsequent machine learning processes should be flexible in supporting a wide range of input modalities. However, to make the system tractable and utilize data that is available across the whole temporal range of the ground truth, only selected sensor data from MODIS-Aqua and MODIS-Terra has been used. Although using both of these satellite sensors provides improved temporal resolution, redundancy of information has led to only Chl-a being used from MODIS-Terra (due to a redundancy of information and the degradation of the Terra sensor over time relative to the Aqua sensor). Higher level (higher than level-2) products were not considered as they lacked the temporal and spatial resolution required for effective HAB detection and prediction. The level-2 products utilized are illustrated in Table I.

Although there is possible redundancy between the chosen bands and products such as Chl-a, the inclusion of these bands was intended to provide fine-grain classification and discrimination for the use of this key available data. All these modalities have been used by previous spatiotemporal HAB detection methods [5], [6]. Bathymetry obviously did not vary over time, but the same modality format was used as shown where the image samples were resampled/interpolated to be co-located as the images of the other modalities. Bathymetry was chosen as it has been noted that estimated chlorophyll concentrations are often inaccurate in shallow water. Including bathymetry should allow the machine learning algorithm to characterize such variations. A list of the utilized modalities is shown in Table I.

TABLE I
LIST OF UTILIZED MODALITIES

Modality	Description
1	**Bathymetry** (GEBCO quantised from 500m grid [51])
2	**MODISA Bimonthly Chl-a** (Estimated Chlorophyll concentration)
3	**MODISA Chl-a** (Estimated Chlorophyll concentration)
4	**MODISA Rrs(412)**
5	**MODISA Rrs(443)**
6	**MODISA Rrs(488)**
7	**MODISA Rrs(531)**
8	**MODISA Rrs(555)**
9	**MODISA PAR** (Daily Mean Photosynthetically Available Radiation)
10	**MODISA SST** (Sea surface Temperature)
11	**MODIST Chl-a** (Estimated Chlorophyll concentration)
12	**MODISA Background Anomaly** (3-2)

V. PREPROCESSING OF DATACUBES

In order to utilize the high-performance characterization performance of pretrained image-based CNNs, the sparse input data of the datacubes were reprojected to a spatially consistent UTM raster image (using the standard WGS84 Ellipsoid).

The spatial representation of the input raster formats within the level-2 MODIS based products are not spatially consistent and therefore not useful for effective machine learning characterization, detection, and prediction. This was due to the capture methods and artifacts such as the "bowtie" effect where horizontal or vertical lines can be repeated in the input 2-D raster arrays. The input datapoints were therefore reprojected to locally spatially consistent UTM reprojection.

The projected datapoints were resampled onto a spatiotemporal grid where each grid element was of extent (1 km $\times$ 1 km $\times$ 1 d) using triangulation-based linear interpolation (the default method of the MATLAB `griddata` function). The default triangulation linear interpolation method uses a convex hull for interpolation. The use of a convex hull generates inaccurate resampled values where it would be better that they were discounted. Holes and disjoint regions are further discounted using the MATLAB `alphashape` function with a threshold of 0.2. Discounted samples were set to zero for input to the neural network (see below).

Given a temporal–spatial span of 100 km and 10 d, the output size of each resampled modality datacube is of dimension ($100 \times 100 \times 10$): (width $\times$ height $\times$ days). This structure is illustrated in Fig. 2.

A. Dealing With Sparse Data

Due to the dependence on sea surface reflectance on the majority of the chosen modalities, a large amount of the data is missing due to cloud cover. This is problematic in effectively characterizing any of the HAB events where there is little data.

In order to most effectively characterize events in the ground truth database, datacubes containing less than a threshold of data are discarded from the training/testing process. As the estimated chlorophyll concentration (Chl-a) is the most important modality within the chosen set, it is used to indicate unacceptable sparsity within dataset events. The threshold chosen was: If more than half of the datapoints within the Chl-a modality (averaged over the entire temporal range) are missing, the datacube is discarded from training. The threshold was varied in various experiments, but classification performance was significant for values near a half (of datapoints).

Furthermore, missing datapoints are usually indicated as NaN values in the original OB.DAAC L2 granules. When resampling (and the use of `alphashape`) the original data, grid points that are not able to be resampled are set to zero. The use of unique flags such as zero representing nondata is common within machine learning systems (e.g., [52]). As these zero values will not be correlated with either class (HAB or non-HAB), they will not affect the characterization and classification of the machine learning system. There have been a few CNN methods proposed specifically for coping with sparse data [53], [54]. However, more recent works have indicated that CNNs are able to learn from sparse representations directly without explicitly changing the network structure and design [52].

VI. Machine Learning Structure

To fully exploit the spatial and temporal discrimination information contained within all the modalities of each datapoint, a novel machine learning structure has been designed and implemented within this work. The application within this article requires both spatial and temporal characterization and classification. CNNs and LSTMs have often been combined to provide such characterization in applications such as video sequence classification [41], [42]. However, our application has the added complexity of multimodal 2-D inputs from each quantized time step within a temporal sequence. We, therefore, propose the novel machine learning structure illustrated within Fig. 3. This figure shows that a single feature vector is extracted from each single image modality at each time step. This is achieved through a form of transfer learning. It has been recognized that utilizing pretrained layer weights of existing CNNs can provide an effective characterization of visual features in new domains where (as is the case in this application) there is a limited amount of training data and computational resources. For each evaluated CNN, the final classification layer is removed and a flattened subset of the penultimate was used as a feature extractor (see Fig. 5). These feature vectors are then concatenated for all modalities. This concatenated feature vector is then the input to a sequenced LSTM across all of the quantized time range (in our case for example, 10 d). A single classification output of the LSTM is used as a binary classifier $\in \{\text{HAB}, \text{No HAB}\}$.

The index of the considered time sequence is denoted as t where $\forall t \in \{1, 2, \ldots, T\}$ where, in this case, $T = 10$. The modality index is denoted as m where $\forall m \in \{1, 2, \ldots, M\}$ where, in this case, $M = 12$. There are therefore 120 input images (12 modalities per each of the 10 time steps) per HAB

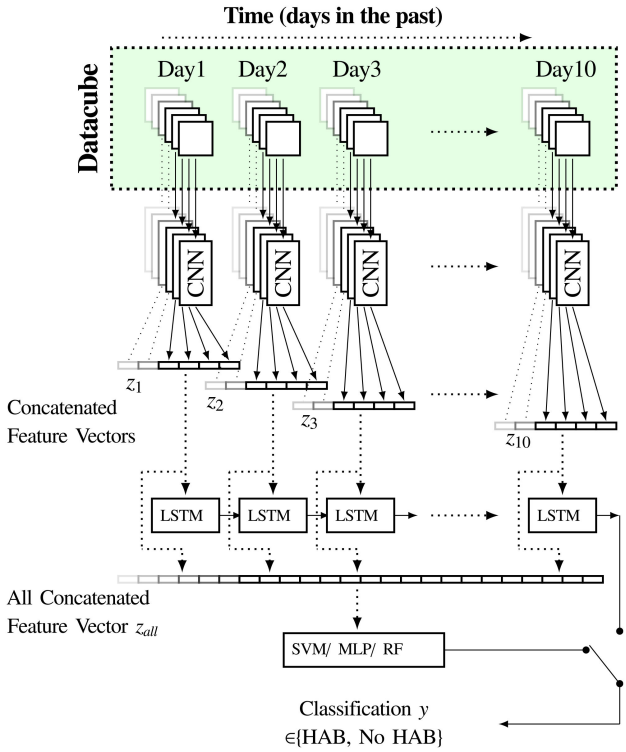

Fig. 3. Structure of HABNet machine learning system for datacube classification: CNN spatial characterization followed by either temporal classification by LSTM or nontemporal classification using multilayer perceptron, support vector machines, or random forest time series classification.

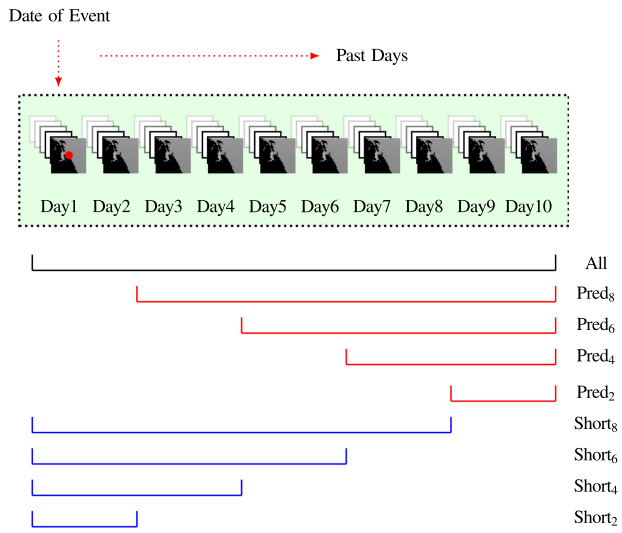

Fig. 4. Variation in temporal prediction structure using datacubes: All use the entire captured datacube. pred$_{\{2,4,6,8\}}$ sequences vary the number of days into the future for the training set, i.e., when testing, the models can predict multiple days into the future. short$_{\{2,4,6,8\}}$ sequences vary the number of days trained but do not predict into the future. This illustrates how the number of days in the training set affects classification.

event (each image denoted as $x_{t,m}$). The concatenated outputs z_t of the CNNs are therefore created as follows:

$$z_t = \{\phi(x_{t,1}), \phi(x_{t,2}), \ldots, \phi(x_{t,M})\} \quad (1)$$

TABLE II
LIST OF CONSIDERED TEMPORAL CLASSIFIERS

Temporal Classifier	Description
RF	**Random Forest:** Standard python (`sklearn`) implementation of RF with grid search of best parameters using validation set
SVM	**Support Vector Machines:** Standard python (`sklearn`) implementation of SVM with grid search of best parameters using validation set
MLP0	**Multi-Layer Perceptron**: Two dense layers each with 256 nodes. Each layer combined with batch normalisation
MLP1	**Multi-Layer Perceptron:** Two dense layers each with 256 nodes. Each layer combined with dropout (0.5)
MLP2	**Multi-Layer Perceptron:** Two dense layers each with 256 nodes. L2 normalisation
LSTM0	**LSTM based network:** One LSTM layer and one dense layer each with 512 nodes each. Each layer combined with batch normalisation and dropout (0.5).
LSTM1	**LSTM based network:** One LSTM layer and one dense layer each with 128 nodes. Each layer combined with dropout (0.5).
LSTM2	**LSTM based network:** One LSTM layer and one dense layer each with 512 nodes both returning sequences. Each layer combined with batch normalisation.
LSTM3	**LSTM based network:** Two LSTM layers and one dense layer each with 256 nodes both returning sequences. This is flattened and then fed to fully connected layer (128 nodes). There is dropout (0.5) and batch normalisation between each layer.
LSTM4	**Attention based model:** An attention layer [40] is combined with an LSTM layer. Dropout and batch normalisation between each layer.

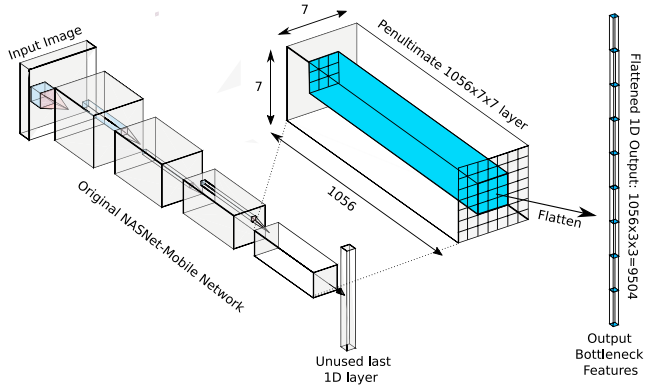

Fig. 5. NASNet-Mobile extraction of translationally invariant bottleneck features. The last NASNet layer has feature length 1056 and has the spatial dimensions of 7×7. The central region selected (to be flattened) is the central 3×3 regions (as illustrated).

where $\phi(\cdot)$ is the operation of the CNN that outputs flattened output as illustrated in Fig. 2.

For nontemporally based classification, the concatenated outputs z_t are themselves concatenated into a single vector z_{all}. The LSTM temporal classification models take as input all of the concatenated outputs z_t to generate the classification y where $y \in \{\text{HAB, No HAB}\}$

$$y = \Psi(\{z_1, z_2, \ldots, z_T\}) \tag{2}$$

where $\Psi(\cdot)$ is the LSTM temporal classification operation that outputs the HAB/No HAB classification.

Conversely, for the nontemporally based classifiers (RF, SVM, and MLPs), the classifiers take the fully concatenated vector z_{all} as input to generate y

$$y = \psi(z_{\text{all}}) \tag{3}$$

where $\psi(\cdot)$ is the nontemporal classification operation (RF, SVM, and MLPs) that outputs the HAB/No HAB classification.

A large number of components were tested within the architecture depicted in Fig. 3. A variety of LSTM structures were also tested alongside simple alternatives including MLPs and RF classifiers. Initial tests showed that NASNet:Mobile produced the best results for the spatial CNN stage. Table III shows the classification results using a NASNet:Mobile CNN and a variety of temporal classification methods.

TABLE III
CLASSIFICATION ACCURACY RESULTS: FLORIDA

Features (Selected from NASNet:Mobile)	Temporal Classifier	Accuracy Mean±SD	F1 Mean±SD	Kappa (K) Mean±SD
{1 to 12}	LSTM0	0.89±0.02	0.86±0.03	0.77±0.05
{1,2,3,9,11}	LSTM0	0.88±0.04	0.84±0.04	0.74±0.08
{1,2,3,9,11}	LSTM1	0.89±0.03	0.86±0.04	0.78±0.06
{1,2,3,9,11}	LSTM2	0.88±0.04	0.84±0.05	0.74±0.08
{1,2,3,9,11}	LSTM3	**0.91**±0.02	**0.88**±0.03	**0.81**±0.05
{1,2,3,9,11}	LSTM4	0.91±0.03	0.88±0.05	0.80±0.05
{1,2,3,9,11}	RF	0.87±0.04	0.82±0.06	0.72±0.08
{1,2,3,9,11}	SVM	0.90±0.03	0.88±0.04	0.80±0.06
{1,2,3,9,11}	MLP0	0.89±0.02	0.86±0.02	0.77±0.04
{1,2,3,9,11}	MLP1	0.73±0.21	0.46±0.75	0.37±0.61
{1,2,3,9,11}	MLP2	0.89±0.03	0.86±0.04	0.77±0.06

Regularization using L_1 and L_2 norm conditions did not improve the results and therefore was not used. Standard ADAM optimization was used with a (decaying) learning rate of 10^{-5}.

A. List of Considered Temporal Classifiers

Table II shows a list of the considered temporal classifiers (i.e., the classifiers to take the bottleneck features generated by the CNNs and generate binary classifications (i.e., $\{\text{HAB}, \text{No HAB}\}$). All of the LSTM-based methods take the concatenated bottleneck outputs as a time series ($z_{1\ldots10}$), whereas the remaining methods take the totally concatenated bottleneck features (z_{all}).

B. Translationally Variant Features

Commonly, bottleneck features are extracted from one of the flat final 1-D layers of a CNN. However, these features are translation invariant due to the max pooling, i.e., they do not discriminate between objects and features found in different spatial positions. This is not what is required in this work as the HAB has been identified as occurring specifically in the spatial center of the input image. In order to make the most use of the abstractions found in the lower layer of a CNN, the central spatial region of the penultimate layer is flattened to form

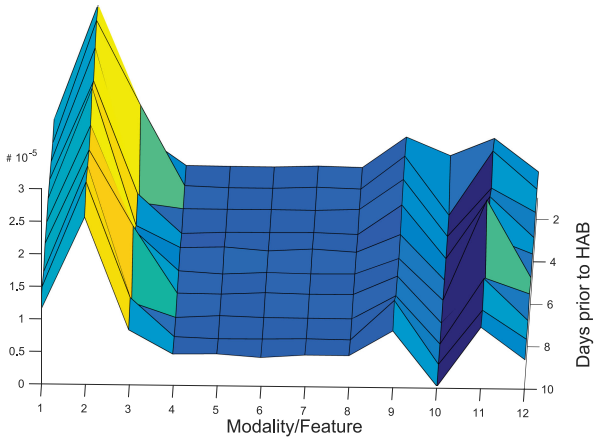

Fig. 6. Feature importances (from random forest analysis [59]): Modality versus time from events. Modality described in Table I. This shows that modalities {1, 2, 3, 9, 11} give the most significant contribution to HAB detection.

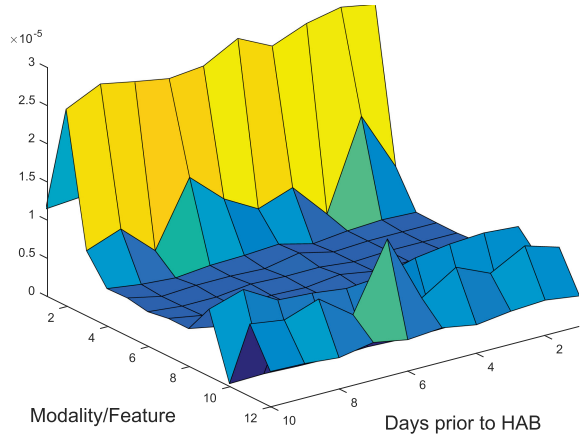

Fig. 7. Feature importances (from random forest analysis [59]): Modality versus time from event. Modality described in Table I. This shows that there is slight decrease in feature importance (for all features) over time.

translationally variant features. By using these flattened features without max pooling the spatial arrangement of CNN outputs is characterized.

C. Choice of CNN and Bottleneck Features

Many different CNN models were tested including a variety of Inception [55], VGG [56], and NASNet [57] architectures. The choice for all the subsequent experiments was NASNet:Mobile as it gave the best results and had a smaller architecture than most other models. Large flattened central regions will lead to excessively large bottleneck features. For the architectures described below, the choice of region size was 3×3. As the last layer of the NASNet-Mobile model has a size feature size of 1056, all of the temporal models input bottleneck features of size $1056 \times 3 \times 3 = 9504$. This is illustrated in Fig. 5.

VII. Feature Classification Importance

To evaluate the classification importance of the features given in Table I, feature vector (input into the last temporal classifier) importance is estimated using an RF classifier [43] and its associated capability at estimating feature importances [58]. Due to the very large feature vector length of the CNN bottleneck features, the most effective way to determine modality importance is to extract bottleneck feature importances for the entire bottleneck feature vector and average the importances for each combination of modality and day. Figs. 6 and 7 show the averaged importances for each modality and day. The modality index shows those modalities labeled in Table I and the day indicates the number of days in the past (from the HAB event). These figures illustrate the slight decrease in feature importance the further into the past the features are. Additionally, it is apparent that the most important features are those indexed {1,2,3,9,11}. This indicates that the individual remote sensing reflectance (Rrs) {4,5,6,7,8}, sea surface temperature (SST) {10}, and background anomaly {12} features are relatively unimportant (in terms of classification).[1]

TABLE IV
Classification Accuracy Results: Florida Results for Varying Temporal Ranges (Ranges Shown in Fig. 4)

Temporal Range	**Accuracy** Mean±SD	**F1** Mean±SD	**Kappa (K)** Mean±SD
All(LSTM3)	**0.91**±0.02	**0.88**±0.03	**0.81**±0.05
$pred_8$	0.90±0.02	0.87±0.02	0.79±0.04
$pred_6$	0.90±0.01	0.87±0.01	0.79±0.02
$pred_4$	0.89±0.03	0.86±0.04	0.76±0.07
$pred_2$	0.88±0.02	0.85±0.03	0.74±0.05
$short_2$	0.87±0.01	0.84±0.03	0.73±0.04
$short_4$	0.88±0.05	0.85±0.06	0.75±0.10
$short_6$	0.89±0.02	0.86±0.04	0.76±0.05
$short_8$	0.90±0.01	0.87±0.03	0.79±0.03

VIII. Results

The most effective way to evaluate the performance of a classifier is using nested cross validation [60]. Nested cross validation utilizes two nested cross validation stages. The outer cross validation iteratively splits the whole dataset (2869 datapoints) into five folds in our case (each fold having separate training and testing subsets). For each of the outer folds, the outer training set is further split into training and validation sets. The inner validation set is used to validate and optimize model and parameter choice. The results shown in Tables III–V show the average and standard deviation across all five outer folds.

The best results were obtained using the LSTM3 model combined with a NASNet:Mobile CNN (mean accuracy 91% and Kappa coefficient 0.81). Although not comparable in terms of dataset location, time, and size, these results are relative improvements to correct classification rates of 83% [4] and 75% [3] for standard chlorophyll anomaly detection methods. These compared methods both used extremely small datasets compared

[1]These "unimportant modalities" are therefore omitted in a subset of subsequent experiments in order to reduce the concatenated feature length and therefore the computational load and memory requirements.

TABLE V
CLASSIFICATION USING CONVENTIONAL THRESHOLDS: SS(488) DEFINED AS 4, Bp_{RATIO} DEFINED AS 5 AND CHLA_ANOM IS DEFINED AS FEATURE 12 IN TABLE I (I.E., FEATURE 3-FEATURE 2 IN THE SAME TABLE)

Method	Accuracy	F1	Kappa (K)
$SS(488) < 0.0$	0.59	0.13	0.00
$Bp_{Ratio} < 1.0$	0.61	0.44	0.15
$Bp_{Ratio} < 2.0$	0.55	0.53	0.13
$Chla\_Anom > 1.0$	0.58	0.29	0.04
$Chla\_Anom > 10.0$	0.62	0.23	0.08
$Chla\_Anom > 100.0$	0.61	0.10	0.03

to our work. Furthermore, the results show that the reduction in the number of features does not significantly decrease the performance of the classifiers. Additionally, these results show that considering the outputs of the bottleneck features as a time series and analyzing them with an appropriate time series analysis tool such as an LSTM or attention-based network does not give significant improvements in performance compared to nontime series classification tools such as SVMs. This is considered because time series of length 10 (temporal data points) are difficult to evaluate using such temporal analysis tools.

A. Conventional Comparative Methods

Comparable methods of HAB detection include SS, thresholded backscattering ratio [26], [27], and thresholded Chl-a anomaly [3], [4], [28]. These three methods are used to compare the performance with our developed methods.

The SS gives a measure of the SS centered on a specific band. It is based on a simple measure combining the Rrs of neighboring reflectance bands

$$\text{SS}(\lambda) = nLw(\lambda) - nLw(\lambda^-) - (nLw(\lambda^+) - nLw(\lambda^-)) \times \left(\frac{\lambda - \lambda^-}{\lambda^+ - \lambda^-}\right) \quad (4)$$

where nLw is water leaving radiance at wavelength in nanometers (nLw is linearly related band reflectances: Rrs). Equation (4) is a second derivative measure centered on wavelength λ. The results in Table V for SS are for MODISA [$\text{SS}(488)$] as used by Stumpf and Werdell [61] and close to $\text{SS}(490)$ used for SeaWifs sources by Tomlinson *et al.* [27] for *K. brevis* in the Florida region.

Another method of reducing the false positives associated with Chl-a HAB detection is the backscattering ratio algorithm [26], [27]. This algorithm utilizes a thresholded ratio formed from Rrs(555) and Chl-a. For SeaWiFS bands, the backscatter ratio is given as

$$bp_{\text{ratio}} = \frac{b_{bp}(555)}{b_{bp}(555)_{\text{Morel}}} \quad (5)$$

where $b_{bp}(555) = -0.00182 + 2.058 \times \text{Rrs}(555)$[62], [63] and $b_{bp}(555)_{\text{Morel}} = 0.3 \times \text{Chl} - a^{0.62} \times (0.002 + 0.02 \times (0.5 - 0.25 \times \log_{10}(\text{Chl} - a)))$ [64].

Detection of HABs utilizes this ratio together with a conventional threshold of 1.0 [62] and a threshold of 2.0 for non *K. brevis* based HAB detection [27]. Table V shows the results on the entire dataset of 2869 datapoints. These values are very low compared to our developed method.

Although difficult to compare directly, Gokaraju *et al.* [5], [6] have developed limited spatiotemporal methods of HAB detection using SVMs and neural networks. These pieces of work both use ground truth in Florida but only a very small dataset (less than 30 datapoints for MODIS-based estimation). The best kappa classification for the MODIS-A data was 0.65.

B. Results Metrics

We define true positives, false positives, true negatives, and false negatives as TP, FP, TN, and FN, respectively. We also define precision $= \text{TP}/(\text{TP} + \text{FP})$ and recall $= \text{TP}/(\text{TP} + \text{FN})$. We use the (global, not user) Accuracy and F1 metrics as the performance metrics in the following tables, defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (6)$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (7)$$

C. Results: Discussion

All of the results for the developed methods for HAB classification shown in Table III gave significantly better results than the conventional methods shown in Table V. It is assumed that adopting a two-stage machine learning-based approach is able to much more effectively characterize spatial and temporal discriminating information for HAB classification relative to these conventional methods. Further observations from the results include the following.

1) Larger spatial areas included in the output of the CNN bottleneck features improve results (i.e., going from 1×1 to 3×3 improves results.)
2) The selection of a smaller number of the most important features does not decrease results (i.e., going from all features {1 …12} to {1, 2, 3, 9, 11} does not decrease classification performance results).
3) Temporally based networks such as LSTMs do not provide significantly better results than "flat" solutions such as MLPs or SVMs. This was assumed because time series with only ten datapoints are not long enough for LSTMs to effectively characterize temporal variations over such a short amount of time.

IX. ALTERNATIVE CASE STUDY: HABS IN THE ARABIAN GULF

As an alternative case study region, the Arabian Gulf was chosen at it has significantly different environmental factors and the availability of ground truth data. HABs occur regularly within the Arabian Gulf with serious outbreaks having happened most years over the last few decades [65]–[68].

Over 38 types of algal taxa have been identified in the Arabian Gulf [69]. A very serious outbreak in 2008 affected over 1200 km of coastline while destroying thousands of tons of fish and marine life. Such serious HAB events can do considerable damage to local aquaculture and can potentially shutdown vital local desalination plants (a major source of local potable water [70]).

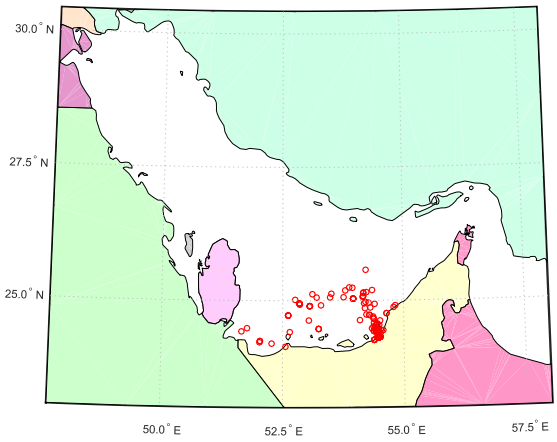

Fig. 8. Selection of positive HAB events recorded in the Arabian Gulf off the coast of the UAE between 2003 and 2018.

A number of local research projects into HAB monitoring and prediction have been undergone in the last ten years [69], [71]. These methods have focused on remote sensing data such as MODIS-A and MERIS. However, these works have not generated quantitative results in the detection and/or prediction of a database of HAB events within the Gulf. They instead have used HAB indicators focused on optical measures such as the modified fluorescence line height and enhanced RedGreenBlue measures together with flow models such as HYCOM [72].

A. Ground Truth

A considerably smaller set of ground truth HAB events (compared to that obtained for the Florida area by the FWC) have been obtained from the Environment Agency-Abu Dhabi between 2002 and 2018 (covered by the flight times of MODIS). This ground truth dataset contains 249 positive events from multiple species (from generic labels such as Cyanobacteria, specific HAB detection species such as Cochlodinium and multispecies detections). Alongside the 249 positive events, 374 negative events were generated that were distinct (in time and space from the positive events) within the same Arabian gulf region (off the coast of Abu Dhabi; see Fig. 8). These detections were not accompanied by concentrations (in cells/L). This means that it was implicitly assumed that they were "significant events." However, this does render it difficult to make this case study analysis comparable to the Florida case above (as within the Florida case, there was an explicit threshold of concentrations that constituted a HAB "event").

B. Results

As per the mechanisms described in section IV, datacubes were extracted for all the spatial locations given by the positive and negative events. The datacubes were composed of the same list of modalities listed in Table I. However, (due to the same analysis given in Section VII), the modalities used for classification were actually a subset of those included in the datacubes (as shown in Table VI).

The results are shown in Table VI giving classification accuracy, F1, and Kappa metrics for the given temporal classifiers.

TABLE VI
CLASSIFICATION ACCURACY RESULTS: ARABIAN GULF

Features Selected (from NASNet:Mobile	**Temporal Classifier**	**Accuracy** Mean±SD	**F1** Mean±SD	**Kappa** Mean±SD
{1,2,3,9,11}	LSTM0	0.91±0.05	0.88±0.06	0.82±0.10
{1,2,3,9,11}	LSTM1	0.88±0.08	0.81±0.14	0.73±0.19
{1,2,3,9,11}	LSTM2	0.86±0.06	0.82±0.08	0.70±0.12
{1,2,3,9,11}	LSTM3	0.91±0.09	0.88±0.11	0.80±0.18
{1,2,3,9,11}	LSTM4	0.91±0.06	0.87±0.13	0.80±0.16
{1,2,3,9,11}	RF	0.91±0.06	0.88±0.08	0.81±0.12
{1,2,3,9,11}	SVM	**0.93**±0.08	**0.91**±0.09	**0.85**±0.16
{1,2,3,9,11}	MLP0	0.91±0.07	0.88±0.11	0.81±0.16
{1,2,3,9,11}	MLP1	0.71±0.23	0.31±0.77	0.27±0.66
{1,2,3,9,11}	MLP2	0.90±0.09	0.85±0.16	0.78±0.21

As with the previous work in Florida, the temporal CNN analysis (to produce bottleneck features) was the NASNet:Mobile CNN producing flattened 3×3 translationally variant features. The temporal classification stages (LSTM0, LSTM1, etc.) are defined as shown in Table II.

This table shows that SVMs in this case give the best classification results (in terms of classifier accuracy, F1, and the Kappa coefficient).

X. CONCLUSION

HABNet is the first machine learning architecture to use remote sensing-based datacubes defined specifically for the classification (using machine learning methods) of temporally and spatially isolated events such as HAB events. A very large database of positive and negative HAB events was utilized over the last two decades off the coast of Florida. SeaDas tools combined with NASA's CMR web-based enquiry method were used to populate a ground truth database of datacubes (one per data point in the ground truth database). Twelve modalities were chosen including estimated sea surface temperature, chlorophyll concentrations, reflectance bands (from MODIS sensors), and bathymetry. A combined CNN/LSTM spatiotemporal classification system was implemented to classify and discriminate between HAB and non-HAB events. Using a small NASNet-Mobile CNN with an LSTM temporal stage, a classification accuracy and Kappa coefficient of 91% and 0.81 were achieved, respectively. This is a significant improvement compared to results generated from historical methods (e.g., Chl-a anomaly: Maximum Kappa $= 0.08$) and other reported state-of-the-art classification methods (spatiotemporal classification method using for a very small dataset: Maximum Kappa $= 0.65$). Our results represent a significant correct classification rate (and Kappa coefficient), given that the number of datapoints is an order of magnitude greater than any previous study. In the future, targeted integration of supplementary modalities and optimization of machine learning methods and structures are anticipated to lead improved classification rates.

Furthermore, our study shows that the datacube method is able to effectively predict HABs up to 8 d in the future without significant degradation of classification accuracy.

An alternative case study was investigated for multispecies HAB ground truth events within the Arabian Gulf. Good results

were also obtained from this study, given a maximum classification rate of 93% and a Kappa coefficient of 0.85 (using a NASNet-Mobile CNN and an SVM temporal stage).

The LSTM3 approach gave the best performance for Gulf of Mexico data and (only marginally) the second-best for Arabian Gulf data. This method is therefore apparently robust to the geographical location of the HAB.

A transfer learning method to use the present work for characterization transferred to the use of more up-to-date sensors such as Sentinel-3 would be an essential follow-up project. Furthermore, a study of the effect of spatial and temporal resolution together with temporal air (wind) and sea (wave) motion data on classification performance would also be key for subsequent studies.

ACKNOWLEDGMENT

The authors would like to thank the FWC for the ground truth data in Florida and EAD (UAE) for the ground truth data in the Arabian Gulf.

REFERENCES

[1] K. A. Steidinger and K. D. Haddad, "Biologic and hydrographic aspects of red tides," *BioScience*, vol. 31, no. 11, pp. 814–819, 1981.

[2] D. Blondeau-Patissier, J. Gower, A. Dekker, S. R. Phinn, and V. E. Brando, "A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans," *Prog. Oceanogr.*, vol. 123, pp. 123–144, 2014.

[3] R. P. Stumpf *et al.*, "Monitoring *Karenia brevis* blooms in the Gulf of Mexico using satellite ocean color imagery and other data," *Harmful Algae*, vol. 2, no. 2, pp. 147–160, 2003.

[4] M. C. Tomlinson *et al.*, "Evaluation of the use of SeaWifs imagery for detecting *Karenia brevis* harmful algal blooms in the eastern Gulf of Mexico," *Remote Sens. Environ.*, vol. 91, no. 3-4, pp. 293–303, 2004.

[5] B. Gokaraju, S. S. Durbha, R. L. King, and N. H. Younan, "A machine learning based spatio-temporal data mining approach for detection of harmful algal blooms in the gulf of mexico," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 710–720, Sep. 2011.

[6] B. Gokaraju, S. S. Durbha, R. L. King, and N. H. Younan, "Ensemble methodology using multistage learning for improved detection of harmful algal blooms," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 827–831, Sep. 2012.

[7] W. Song, J. M. Dola, D. Cline, and G. Xiong, "Learning-based algal bloom event recognition for oceanographic decision support system using remote sensing data," *Remote Sens.*, vol. 7, no. 10, pp. 13564–13585, 2015.

[8] L. E. Fleming *et al.*, "Oceans and human health: A rising tide of challenges and opportunities for europe," *Mar. Environ. Res.*, vol. 99, pp. 16–19, 2014.

[9] S. K. Moore *et al.*, "Impacts of climate variability and future climate change on harmful algal blooms and human health," in *Environmental Health*, vol. 7, no. S2. Berlin, Germany: Springer, 2008, p. S4.

[10] K. Dyson and D. D. Huppert, "Regional economic impacts of razor clam beach closures due to harmful algal blooms (HABs) on the pacific coast of Washington," *Harmful Algae*, vol. 9, no. 3, pp. 264–271, 2010.

[11] P. Hoagland, D. Anderson, Y. Kaoru, and A. White, "The economic effects of harmful algal blooms in the united states: Estimates, assessment issues, and information needs," *Estuaries*, vol. 25, no. 4, pp. 819–837, 2002.

[12] G. A. Carvalho, P. J. Minnett, L. E. Fleming, V. F. Banzon, and W. Baringer, "Satellite remote sensing of harmful algal blooms: A new multi-algorithm method for detecting the florida red tide (*Karenia brevis*)," *Harmful Algae*, vol. 9, no. 5, pp. 440–448, 2010.

[13] P. Hoagland and S. Scatasta, "The economic effects of harmful algal blooms," in *Ecology of Harmful Algae* (Ecology Studies Series), E. Graneli and J. Turner, Eds. Dordrecht, The Netherlands: Springer-Verlag, 2006, ch. 29.

[14] R. Santoleri *et al.*, "Year-to-year variability of the phytoplankton bloom in the southern adriatic sea (1998–2000): Sea-viewing wide field-of-view sensor observations and modeling study," *J. Geophys. Res.*, vol. 108, 2003, Art. no. 8122.

[15] F. Gohin *et al.*, "Satellite and in situ observations of a late winter phytoplankton bloom, in the Northern Bay of Biscay," *Continental Shelf Res.*, vol. 23, pp. 1117–1141, 2003.

[16] A. C. Thomas *et al.*, "Satellite-measured phytoplankton variability in the Gulf of Maine," *Continental Shelf Res.*, vol. 23, pp. 971–989, 2003.

[17] S. E. Craig *et al.*, "Use of hyperspectral remote sensing reflectance for detection and assessment of the harmful alga, *karenia brevis*," *Appl. Opt.*, vol. 45, pp. 5414–5425, 2006.

[18] C. R. McClain, "A decade of satellite ocean color observations," *Annu. Rev. Mar. Sci.*, vol. 1, pp. 19–42, 2009.

[19] "Gulf of Mexico harmful algal bloom forecast." [Online]. Available: https://tidesandcurrents.noaa.gov/hab/gomx.html

[20] "Harmful algal blooms observing system." [Online]. Available: https://habsos.noaa.gov/

[21] A. Morel and L. Prieur, "Analysis of variations in ocean color," *Limnol. Oceanography*, vol. 22, no. 4, pp. 709–722, 1977.

[22] A. Morel, "In-water and remote measurements of ocean color," *Boundary-Layer Meteorol.*, vol. 18, no. 2, pp. 177–201, 1980.

[23] M. W. Matthews, "A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters," *Int. J. Remote Sens.*, vol. 32, no. 21, pp. 6855–6899, 2011.

[24] H. M. Dierssen, "Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate," *Proc. Nat. Acad. Sci.*, vol. 107, no. 40, pp. 17073–17078, 2010.

[25] T. Moore, J. W. Campbell, and M. D. Dowell, "A class-based approach to characterizing and mapping the uncertainty of the modis ocean chlorophyll product," *Remote Sens. Environ.*, vol. 113, no. 11, pp. 2424–2430, 2009.

[26] J. P. Cannizzaro, K. L. Carder, F. Chen, C. A. Heil, and G. A. Vargo, "A novel technique for detection of the toxic dinoflagellate, *Karenia brevis*, in the Gulf of Mexico from remotely sensed ocean color data," *Continental Shelf Res.*, vol. 28, no. 1, pp. 137–158, 2008.

[27] M. C. Tomlinson, T. T. Wynne, and R. P. Stumpf, "An evaluation of remote sensing techniques for enhanced detection of the toxic dinoflagellate, *Karenia brevis*," *Remote Sens. Environ.*, vol. 113, no. 3, pp. 598–609, 2009.

[28] P. I. Miller J. D. Shutler, G. F. Moore, and S. B. Groom, "Seawifs discrimination of harmful algal bloom evolution," *Int. J. Remote Sens.*, vol. 27, no. 11, pp. 2287–2301, 2006.

[29] T. T. Wynne *et al.*, "Relating spectral shape to cyanobacterial blooms in the Laurentian great lakes," *Int. J. Remote Sens.*, vol. 29, no. 12, pp. 3665–3672, 2008.

[30] X. Li, J. Yu, Z. Jia, and J. Song, "Harmful algal blooms prediction with machine learning models in Tolo Harbour," in *Proc. Int. Conf. Smart Comput.*, Nov. 2014, pp. 245–250.

[31] Y. Tian and M. Huang, "An integrated web-based system for the monitoring and forecasting of coastal harmful algae blooms: Application to Shenzhen city, China," *J. Mar. Sci. Eng.*, vol. 7, Sep. 2019, Art. no. 314.

[32] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[33] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.

[34] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2014, pp. 580–587.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Advances Neural Inform. Process. Syst.*, 2012, pp. 1097–1105.

[36] J. Biggs, N. Anantrasirichai, F. Albino, P. Hill, and V. Ponce Lopez, "Machine learning techniques for detecting slow, sustained deformation in InSAR time series," *AGUFM*, vol. 2019, 2019, Art. no. S54A-03.

[37] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[39] P. Hill, *Audio and Speech Processing With MATLAB*. Boca Raton, FL, USA: CRC Press, 2018.

[40] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.

[41] J. Donahue *et al.*, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 2625–2634.

[42] W. Andrew, C. Greatwood, and T. Burghardt, "Visual localisation and individual identification of Holstein Friesian cattle via deep learning," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 27–29.

[43] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[44] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[45] "Florida fish and wildlife conservation commission monitoring database." [Online]. Available: http://myfwc.com/research/redtide/monitoring/database/
[46] "Phytoplankton monitoring network (PWN)." [Online]. Available: https://coastalscience.noaa.gov/research/stressor-impacts-mitigation/pmn/dat a/
[47] "Harmful algal event database (HAED)." [Online]. Available: http://haedat.iode.org/
[48] L. H. Pettersson and D. Pozdnyakov, *Monitoring of Harmful Algal Blooms*. Berlin, Germany: Springer, 2012.
[49] M. Mahecha *et al.*, "The emerging earth system data cube: Idea, implementation, and first scientific case studies," in *Proc. EGU General Assembly Conf. Abstracts*, vol. 19, 2017, Art. no. 11813.
[50] "NASA's common metadata repository (CMR) Search." [Online]. Available: https://cmr.earthdata.nasa.gov/search/
[51] "General Bathymetric Chart of the Oceans (GEBCO)." [Online]. Available: https://www.gebco.net
[52] M. Jaritz, R. De Charette, E. Wirbel, X. Perrotton, and F. Nashashibi, "Sparse and dense data with CNNs: Depth completion and semantic segmentation," in *Proc. Int. Conf. 3D Vision*, 2018, pp. 52–60.
[53] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger, "Sparsity invariant CNNs," in *Proc. Int. Conf. 3D Vision*, 2017, pp. 11–20.
[54] M. Ren, A. Pokrovsky, B. Yang, and R. Urtasun, "SBNet: Sparse blocks network for fast inference," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 8711–8720.
[55] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 1–9.
[56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[57] B. Zoph, V. Vasudevan, J. Shlens, and V. Le Quoc, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 8697–8710.
[58] L. Breiman, *Classification and Regression Trees*. Evanston, IL, USA: Routledge, 2017.
[59] C. Strobl and A. Zeileis, "Danger: High power–exploring the statistical properties of a test for random forest variable importance," in *Proc. COMPSTAT*, vol. 2, 2008, pp. 59–66.
[60] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[61] R. Stumpf and P. J. Werdell, "Adjustment of ocean color sensor calibration through multi-band statistics," *Opt. Express*, vol. 18, no. 2, pp. 401–412, 2010.
[62] J. P. Cannizzaro, K. L. Carder, F. R. Chen, C. A. Heil, and G. A. Vargo, "A novel technique for detection of the toxic dinoflagellate, *Karenia brevis*, in the Gulf of Mexico from remotely sensed ocean color data," *Continental Shelf Res.*, vol. 28, no. 1, pp. 137–158, 2008.
[63] C. S. Roesler S. M. Etheridge, and G. C. Pitcher, "Application of an ocean color algal taxa detection model to red tides in the Southern Benguela," in *Proc. 10th Int. Conf. Harmful Algae*, 2002, pp. 303–305.
[64] A. Morel, "Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters)," *J. Geophys. Res.: Oceans*, vol. 93, no. 9, pp. 10749–10768, 1988.
[65] S. Rao and F. Al-Yamani, "Phytoplankton ecology in the waters between Shatt Al-Arab and straits of Hormuz, Arabian gulf: A review," *Plankton Biol. Ecology*, vol. 45, no. 2, pp. 101–116, 1998.
[66] C. A. Heil, P. M. Glibert, M. A. Al-Sarawi, M. Faraj, M. Behbehani, and M. Husain, "First record of a fish-killing *Gymnodinium* sp. bloom in Kuwait Aay, Arabian sea: Chronology and potential causes," *Mar. Ecology Prog. Ser.*, vol. 214, pp. 15–23, 2001.
[67] P. M. Glibert *et al.*, "A fish kill of massive proportion in Kuwait Bay, Arabian Gulf, 2001: The roles of bacterial disease, harmful algae, and eutrophication," *Harmful Algae*, vol. 1, no. 2, pp. 215–231, 2002.
[68] M. Moradi and K. Kabiri, "Red tide detection in the strait of Hormuz (east of the Persian Gulf) using modis fluorescence data," *Int. J. Remote Sens.*, vol. 33, no. 4, pp. 1015–1028, 2012.
[69] J. Zhao and H. Ghedira, "Monitoring red tide with satellite imagery and numerical models: A case study in the Arabian Gulf," *Mar. Pollut. Bull.*, vol. 79, no. 1-2, pp. 305–313, 2014.
[70] A. Berktay, "Environmental approach and influence of red tide to desalination process in the middle east region," *Int. J. Chem. Environ. Eng.*, vol. 2, no. 3, 2011.
[71] J. Zhao, M. Temimi, S. A. Kitbi, and N. Mezhoud, "Monitoring HABs in the shallow Arabian Gulf using a qualitative satellite-based index," *Int. J. Remote Sens.*, vol. 37, no. 8, pp. 1937–1954, 2016.
[72] J. Zhao, M. Temimi, H. Ghedira, and C. Hu, "Exploring the potential of optical remote sensing for oil spill detection in shallow coastal waters-a case study in the Arabian Gulf," *Opt. Express*, vol. 22, no. 11, pp. 13755–13772, 2014.

Paul Hill (Member, IEEE) received the B.Sc. degree in computer science from the Open University, in 1996, and the M.Sc. and Ph.D. degrees in image segmentation and fusion from the University of Bristol, Bristol, U.K., in 1998 and 2002, respectively.

He is currently a Research Fellow with the Department of Electrical and Electronic Engineering, University of Bristol, and also lecturing in audio technology. His research interests include image and video analysis, remote sensing, and machine learning.

Anurag Kumar received the B.Sc. and M.Sc. degrees in physics and mathematics from Deen Dayal Updahayay Gorakhpur University, Gorakhpur, India, in 2006 and 2009, respectively.

Since July 2019, he has been working as a Postdoctoral Fellow with Khalifa University, Abu Dhabhi, United Arab Emirates. His research interests are numerical simulation, ocean dynamics, air–sea interaction, and machine learning.

Marouane Temimi received the M.Sc. degree in hydrology and the Ph.D. degree in water resources and remote sensing from the University of Quebec, Montreal, Canada, in 2002 and 2006, respectively.

He worked with the NOAA-CREST Institute, the City University of New York, New York, USA, as Research Associate Professor. He was overseeing research activities with the Remote Sensing of Land and Hydrology Research Group. He taught several courses in water resources, remote sensing, and GIS. He mentored more than 30 graduate and undergraduate students. He has more than 70 peer-reviewed publications and several contributions to books and proceeding. He expanded his experience internationally when he worked with Masdar Institute of Science and Technology, part of Khalifa University of Science and Technology, Abu Dhabi. He established the Coastal and Environmental Sensing and Modeling (CESAM) Lab, which included more than ten scientists active in the field of remote sensing and numerical modeling.

David R. Bull (Fellow, IEEE) received the B.Sc. degree from the University of Exeter, Exeter, England, in 1980, the M.Sc. degree from the University of Manchester, Manchester, England, in 1983, and the Ph.D. degree from the University of Cardiff, Cardiff, Wales, in 1988, all in electrical and electronic engineering.

He was the Head of the Electrical and Electronic Engineering Department, University of Bristol, Bristol, England, from 2001 to 2006. He is currently the Head of the Visual Information Laboratory and also the Director of the Bristol Vision Institute. He has authored over 450 academic papers. His current activities are focused on the problems of image and video communications and analysis for wireless, Internet, military, and broadcast applications.