Original Research Paper

A Hybrid Model of Bidirectional Long-Short Term Memory and CNN for Multivariate Time Series Classification of Remote Sensing Data

Sawsan Morkos Gharghory

Department of Computers and systems, Electronics Research Institute, Cairo, Egypt

Abstract: Classification of multivariate time series has got massive attention in the last decade. The traditional modeling classifiers are complicated patterns and are incompetent to capture the dependencies of multivariate time series data. To include both of the effective features and the embedded relationships in the multivariate time series, a new hybrid model which incorporates both Convolutional Neural Network (CNN) and Bidirectional long-short term memory (BiLSTM) named Conv-BiLSTM is proposed in this study. The proposed Conv-BiLSTM is carried out for classifying the land cover multivariate time series of Landsat 8 satellite images. The efficacy of the proposed network is verified through its comparison with the-state-of-the-art methods using different cases of training dataset. The suggested network outperforms the classification techniques as Random Forest (RF), BiLSTM and the CNN and it has classification accuracy on average 6.5, 8 and 8.7% over that of those classifiers respectively. Moreover, the classification accuracy of the proposed Conv-BiLSTM network in F-Score metric is larger than that value of the state-of-the-art WEASEL+MUSE technique in average by 1.38%.

Keywords: Multivariate Time Series of Landsat 8 Satellite Images, Convolutional Neural Network, Recurrent Neural Network, Bidirectional Long-Short Term Memory, Land Cover Classification

Introduction

Time series data captures a particular attention in different studies fields such as weather readings and psychological signals (Cui *et al.*, 2015) (Kadous, 2002) (Sharabiani *et al.*, 2017). A data set of time series can be classified into univariate, where a series of measurements is collected from the one variable, or multivariate where a series of measurements is collected from multi variables or multiple sensors (Prieto *et al.*, 2015). Therefore, multivariate time series are more complicated and are not easy to predict comparing to the time series of univariate. Multivariate Time Series (MTS) has received a massive attention in the field of data mining, because of its wide applicability in different domains such as medical diagnosis, motion detection, anomaly detection, financial prediction and remote sensing fields that uses satellite images, etc. (Huai-Shuo Huang *et al.*, 2018) (Wang *et al.*, 2016) (Spiegel *et al.*, 2011). Over the past decade, the classification of multivariate time series has a great attention in many domains such as healthcare (Kang and Choi 2014), sound classification (Kang and Choi 2014),

phoneme classification (Graves and Schmidhuber, 2005), object recognition, human activity recognition and actions recognition (Fu, 2015) (Geurts, 2001) (Yu and Lee, 2015). One of the common approaches of multivariate time series classification is by applying dimensional reduction techniques or concatenating all dimensions of multivariate time series into a univariate time series (Karim, 2019). Symbolic representation approach applied the random forest on Multivariate Time Series named (SMTS) by (Baydogan and Runger, 2015) to partition it into leaf nodes and each leaf is represented by a word to form a codebook. Then, every word was utilized with another random forest for the multivariate time series classification. A similar classifier model that is Learned Pattern Similarity (LPS) by (Baydogan and Runger, 2016) extracted segments from the multivariate time series. The regression trees were trained by these segments to explore the dependencies between them. Each node in the tree was represented by a word and a similarity measure in conjunction with these words were used to classify unknown multivariate time sequence. Ultra-Fast Shapeless (UFS) technique (Wistuba *et al.*, 2015)

extracted random shape let from the multivariate time series and used the linear SVM or the Random Forest as classifier. Thereafter, UFS was improved by calculating the derivatives as features (dUFS). Also, the Auto-Regressive (AR) kernel approach (Cuturi and Doucet, 2011) performed an AR kernel based on distance measure to classify the multivariate time series. On the other hand, Auto-Regressive forests (Tuncel and Baydogan, 2018) used a tree ensemble for modeling multivariate time series, where the trees were trained with different time lags. Most recently, the method combining WEASEL and MUSE techniques (Schäfer and Leser, 2017) build a multivariate feature vector by using a classical approach of bag of patterns on each variable to capture the discrete features such as words and pairs of words. The final classification result is produced by using a logistic classifier on the final feature vector.

In recent years, deep learning has gained incredible popularity and many achievements can be found in literatures Längkvist *et al*., 2014) (Bengio, 2013) (Bengio, 2009) (Schmidhuber, 2015). As one of deep networks, Convolutional Neural Network (CNN) has been successfully used in pattern recognition field. Instead of human-designed features, CNN can automatically mine and generate deep features of input images. Besides, it has a strong robustness against transformation, scaling and rotation. Three important points make CNN discriminative than other traditional feed-forward neural networks which are: Local receptive field, weights sharing and pooling (Swietojanski *et al*., 2014). Inspiring from the CNN structure for image recognition, a deep learning framework for time series classification was proposed by the work (Zhao B. *et al*., 2017), in which a novel CNN framework was presented for this domain. Both of convolution and pooling operations are used as a substitution to create deep features and then features are connected to a Multilayer Perceptron (MLP) to perform classification. Experimental results on both simulated and real data sets prove that CNN achieves best results comparing to the state-of-the-art methods in terms of the classification accuracy and noise tolerance. Among the previous work of CNN for multivariate time series classification was summarized as follows.

Zheng *et al*. (2014) proposed a deep convolution neural network with Multi-Channels Named (MC-DCNN) which separated multivariate time series into univariate ones ones and thereafter latent features are detected from an input of each variable. Finally, the classification is performed through the latent features which fed into MLP laye and it was found from experiments that a good classification performance was achieved by the presented method. However, one major limitation of this method is that it cannot extract the interrelationship between different univariate time series.

The work by (Zaho *et al*., 2017) modified the previous algorithm, instead of feature learning individually, the multivariate time series was jointly trained for feature extraction using a typical CNN architecture for three-variate time series classification with two convolutional layers and two pooling layers as illustrated in his work.

In this study, a new hybrid model that incorporates both Convolutional Neural Network (CNN) and Bidirectional long-short term memory (BiLSTM) named Conv-BiLSTM is suggested for classifying landcover multivariate time series. The proposed Conv-BiLSTM network has the advantages of combining the effective features and the embedded relationships in the long time sequences over other techniques. Comparing to Random Forest (RF), BiLSTM and CNN techniques, the proposed Conv-BiLSTM has classification accuracy on average 6.5, 8 and 8.7% respectively over that of those classifiers.

Related Work

For the last decades, Earth observation has been utilized to study our planet surface and follow its development. The surface changes characterization resulted from deforestation, evolution of agricultural activity or urbanization, are essential to estimate population increasing and climate changes (Running, 2008). Land cover and Land use are among the applications of remote sensing field for Earth surface monitoring. The data and information concerning the land cover such as buildings, rivers, fields, mountains, trees and others are very important for studying the human resources and beings life. As the use of unclean energy sources, human activity, rapid urbanization and climate changes, the earth's surface changes largely on both scales of regional and universal (Schafer *et al*., 2018). Therefore, the accurate analysis in a proper time to the land resources use and any happen surface changes plays an important role in enhancing human societies.

Remote sensing satellites are among the tools used to capture and collect time series data especially for the applications of earth's surface monitoring. They produce up-to-date maps for land use/land cover changes such as urbanization, deforestation and desertification (Schafer *et al*., 2018). Earth's surface data collected by satellites is large sets of hyper-spectral or multi-spectral data with a spatial resolution and a temporal density either in form of images or time series and needs to the classification. So, the balancing between the runtime and accuracy of classification becomes more challenging with existing of multiple readings for pixel via the periodic scans of satellite (Schafer *et al*., 2018). Automatic classification especially of remote sensing data has a very attention because the object label phase is restricted by the following issues: (a) The high cost in both of human

resources and time either through using the field campaigns or the labelling method by experts, (b) by the huge amount of data to be labelled and (c) in addition to the fast changes in landscape which required to the updating of maps and this cannot be done manually (Bailly, 2018). The machine learning methods used in land use and land cover (Rajendran, 2020) classification are as follows.

Random Forest (RF) is one of the most machine learning algorithms that is widely used (Breiman, 2001) for the purposes of classification and regression Among the RF applications is the Earth science which includes modeling the forest cover (Betts *et al*., 2017), land-use (Araki *et al*., 2018), land-cover (Nitze *et al*., 2015) and object oriented mapping (Kavzoglu, 2017). In the work presented by (Rodriguez-Galiano *et al*., 2012), they proved that RF algorithm outperformed the classification by decision trees algorithm and achieved a high accuracy reached to 92%. The high accuracy of RF algorithm compared to other classification trees was imputed to its sets architecture in which several of them were trained on subsets of training data.

Also, Support Vector Machine (SVM) has a high capability to generalize complicated features, accordingly it outperformed the other classifiers as presented in following works (Shao and Lunetta, 2012) (Mountrakis *et al*., 2011). In a study to the land cover classification which was involved six classes of land-cover of the Landsat-8 data, SVM proved its ability to achieve a relatively high overall accuracy of 88% as mentioned in work (Goodin *et al*., 2015). Recently, a study to analysis the effect of the training size samples on the overall accuracies of both SVM and RF was presented by (Mansaray *et al*., 2020) with application for mapping the paddy rice in China in 2015 and 2016. In this study for mapping the paddy rice in 2015, the overall accuracies of SVM and RF classifiers reached to 90.8 and 89.2%, respectively using 10 observations from Landsat-8 and Sentinel-1A satellites. While in the presented study for mapping the paddy rice in 2016, SVM and RF achieved overall accuracies of 93.4 and 95.2%, respectively by using 14 observations from Landsat-8, Sentinel-1A and Sentinel-2A satellite.

Lately, neural networks have been used for classifying satellite images and were usually using one or two hidden layers which still remained low efficiency due to the expensive of data and the inadequate of computing power (Mas and Flores, 2008). At the beginning of the 21st century and with increasing the earth observations data and the computing resources, the deeper hidden layers and complex network architectures were merited using. Deep Learning (DL) embeds a family of different algorithm architectures which are constructed using neural networks. These architectures include multi-layer perceptrons, deep belief networks, stacked auto-encoders, deep neural networks and restricted Boltzmann machines and others. DL has been widely used in many applications

since 2015 such as mapping land-cover (Li *et al*., 2016) and crops (Kussul *et al*., 2017) (Zhong, 2019), estimating crop yields (Kuwata and Shibasaki, 2015), detecting oil palm trees (Li *et al*., 2017) and plant diseases (Mohanty *et al*., 2016) with accuracies reached to 90%. A review to Different methods of deep learning for classifying land cover and land use of remote sensing data were presented in (Abebaw Alem and Shailender, 2020). An easy systematic review to the application of transfer learning for scene classification using different Dataset of Land cover and land Use and with different models of deep learning were presented in (De Lima and Marfurt, 2020). In this study (Unnikrishnan *et al*., 2019), the Normalized Difference Vegetation Index (NDVI) concept is utilized and consequently only the information of the red and Near Infrared (NIR) bands are used for the classification of the available public SAT-4 and SAT-6 datasets. New deep learning architectures of the three common networks Alex Net, Conv-Net *a*nd VGG were suggested in this study by tuning the hyper-parameters of networks with two the bands data as input.

Data Sets

Monitoring the changes in land use is a significant area of research because the land cover is the main variable that drives balancing in Earth's energy, carbon and hydrological cycle and the supplying of the natural resources (Bengio, 2009). Since land surfaces have various structures and different chemical properties, they absorb and reflect the sunlight differently and dependently on the wavelength and consequently information of land cover can be extracted from these spectral bands. For example, water surface absorbs much of the near infrared radiation; therefore, these wavelengths are benefit for discriminating the boundaries of the land water which are not clear in visible light. Similarly, the green vegetation's absorb much of the arrived radiation in the red spectrum while about 50% from this radiation is reflected in the near-infrared spectrum. Multi-spectral sensors for Earth surface observation receive the sun's energy reflected by a surface with a few distinct spectral wavelengths named bands, e.g. blue, green and red in the visible spectrum bands from (400 to 700 nm), near infrared bands from (700 to 1100 nm) and short-wave infrared with wavelengths from (1100 to 3000 nm). Among the multi-spectral sensors are the American Landsat 8 sensor and the two European Sentinel-2 sensors. Landsat 8 sensor takes images to the Earth with spatial resolution at 30-m every 16 days, while Sentinel-2 sensor acquires images with a spatial resolution of 10−20 m frequently every 5 days.

In this study, a public dataset given in the published research by (Bailly, 2018) for Time Series Land Cover Classification Challenge (TiSeLaC, 2017) is used. This dataset was collected by Landsat 8 satellite over Reunion Island in 2014 during intervals of 16 days where it included time series data of length 23 steps, the Island is shown in Fig. 1.

For each time step, ten spectral features were collected which are seven reflectance bands and three vegetation indices. The vegetation indices are Normalized Difference Vegetation Index (NDVI), the Normalized Difference Water Index (NDWI) and the Brightness Index (BI). Consequently, each time sequence of the collected data consists of 23-time steps with 10 bands for each time step meaning it has length equals 230. The time series data of Landsat 8 was reshaped to be appeared in image format of size 2866 rows and 2633 columns with 10 bands for each of time step as shown in Fig. 2. and 3 shows examples of NDVI spectral band of twenty three time series data for different classes of land cover.

As shown in Fig. 3, NDVI values range from -1.0 to +1.0. Low NDVI values which are 0.1 or less are corresponding to rocks, sand or snow. While, the medium NDVI values are representing the sparse vegetation such as shrubs, grasslands or crops in low season harvesting. Finally, for the high NDVI values which are 0.6 or more are matching the dense vegetation such as forests and crops before harvest.

Each time series of data is labeled with a class name from predefined labels. The reference dataset ((TiSeLaC, 2017) contains 81714 readings categorized to 9 classes and are used for training the classifier model as detailed in Table 1, while the dataset of 17973 readings are used for testing the classifier model as depicted in the same table.

The description of the different classes of dataset by (Bailly, 2018) are as follows. Urban Areas in Reunion Island are not dense like European cities, while Other-built-up surfaces may indicate to greenhouses. Due to the tropical climate on Reunion Island, NDVI values of Forests class have high values but stable. The Sparse Vegetation category combines all samples that do not belong to any of the other classes with intermediate values of NDVI. The Rocks and bare soil category has samples of low NDVI values which close to or below 0. The Grassland

class combines both grazed and mowed grasslands but with different time series signatures due to variation in their re-growth. The time series of Sugarcane crop has the most differentiate profiles because its NDVI values drop after harvests. Sugarcane crops represent 60% of the cultivated area on Reunion Island (Denize, 2015). The Other crops class formed from many crops such as pineapple and bananas crops as well as mangos and this diversity of crops results in high intra-class variability and this faces a difficulty in classification. The Water class has a high intra-class variability which results in NDVI values perturbation.

In classification process, a classifier is a module with a function that is learned via a set of a labeled time series data which is the training data set and then in test phase it takes an un-labeled time series data as input and results a labeled data. The proposed network for land cover classification is detailed in the next section.
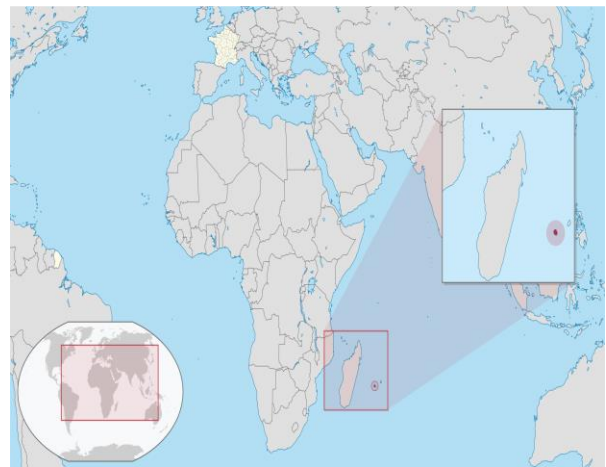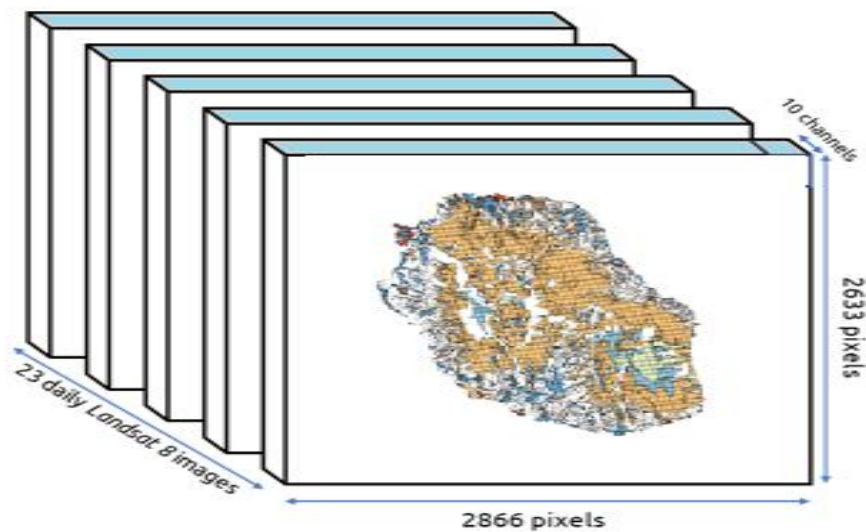

Fig.1: Reunion Island localization

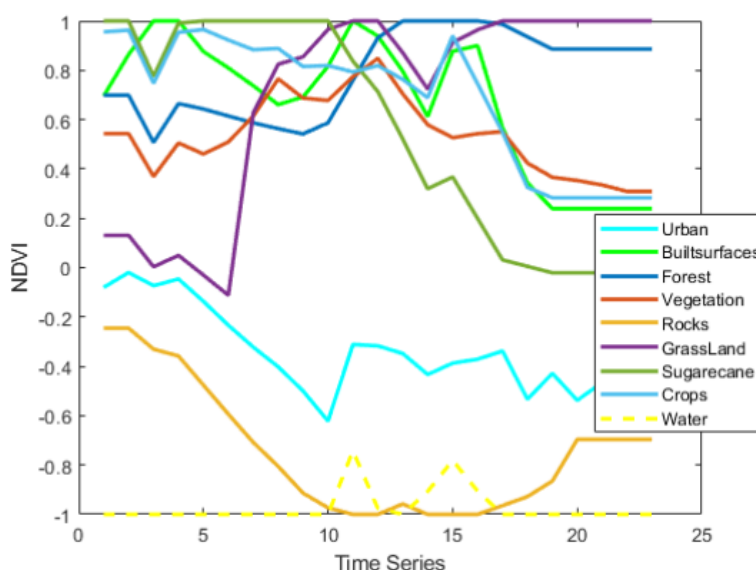**Fig. 2:** Time series Landsat 8 images with 10 Bands for each time step

Fig. 3: NDVI band to different classes of time series data

The Proposed Conv-BiLSTM Network

In this study new structure of a hybrid model combining both of Convolutional Neural Network (CNN) and Bidirectional long Short-Term Memory (BiLSTM) named Conv-BiLSTM is proposed. The proposed Conv-BiLSTM network is trained for classifying the multivariate time series of land cover with its different categories. The proposed Conv-BiLSTM network differentiates itself from other classification techniques by combining the effective features and the embedded relationships in the long time series which results in increasing its classification accuracy. Both of BiLSTM and CNN is detailed in the following sections.

Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of multi-layers neural network and it usually includes two main parts (Pires de Lima *et al*., 2020). One is a feature extractor, which learns features automatically from initial data. And the other part is a trainable fully connected Multi-Layer Perceptron (MLP), which implements the classification process depend on the learned features from the preceding part. Generally, the feature extractor consists of multiple identical stages and each stage is comprised of three cascade layers: Filter layer, activation layer and pooling layer. The output from each layer is called feature map (LeCun and Bengio 1995). CNN is with more detailed in (Krizhevsky 2012).

Recurrent Neural Network

Recurrent Neural Network (RNN) is a category of Adaptive Neural Network (ANN) and differentiates itself from the neural network by it depending on feed forward through the possibility of utilizing its internal state (memory) to handle sequences of inputs. RNN can remember the state of an input from the preceding time steps which helps it to take a decision for the posterior time step. Recently, the Long Short-Term Memory (LSTM) network is a new structure of Recurrent Neural Network (RNN) and it further addresses the problem of vanishing gradients of the previous RNN and holds the long short-term memory. LSTM has two aspects differentiate it from RNN which are as follows. First, the cell state is divided into two parts, the long-term state $c(t)$ and the short-term state h(t). Second, LSTM has three control gates which are the forget gate, the input gate and the output gate, all of them are along the state path to regulate the cell states as introduced in (Gharghory, 2020).

Bidirectional Long Short-Term Memory

Bidirectional Long Short-Term Memory (BiLSTM) learns bidirectional long-term dependencies between time steps of time series data and this is useful to a network needed to be learned at each time step from the complete time series. BiLSTM processes input sequences in both directions with two sub-layers. It consists of two recurrent network layers (Schuster and Paliwal, 1997), in which the first one processes the sequence of inputs in forwards direction, while the second processes the inputs sequence in backwards. Both layers connected to the same output layer and consequently, BiLSTM network reaches to the total information about previous and future sequence of data points. These two sub-layers compute forward and backward hidden sequences $\vec{h}$ and $\overleftarrow{h}$ respectively, which are then combined to compute the output sequence as depicted in the following equations. Also, both the structures of LSTM cell and BiLSTM network are shown in Fig. 4 and 5, respectively:

$$\vec{h}_t = \tan ch\left(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}\right) \tag{1}$$

$$\vec{h}_t = tanch\left(W_{x\vec{h}}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t+1} + b_{\vec{h}}\right) \tag{2}$$

$$y_t = W_{\vec{h}y}\vec{h}_t + W_{\vec{h}y}\vec{h}_t + by \tag{3}$$

Where: $W_{x\vec{h}}$ denotes the connection weight between input layer and forward hidden layer, $W_{x\vec{h}}$ denotes the connection weight between input layer and backward hidden layer, $W_{\vec{h}\,\vec{h}}$ denotes the connection weight between the sequences among forward hidden layer and $W_{\vec{h}y}$ denotes the connection weight between forward hidden layer and output layer.

The Architecture of the Proposed Conv-BiLSTM Network Model

A hybrid model combining CNN and BiLSTM networks named conv-BiLSTM is proposed in this study for the classification of multivariate time series of land cover data. The suggested model is used as a semantic classifier to classify sequences of time series samples of length 23 and multivariable of 10 bands into 9 classes. The proposed network consists of multi-layers included the suggested convolutional layers and BiLSTM layers, thereafter the fully connected layers and the layers of softmax and classification. To perform the convolutional operations on each time step independently, a sequence of folding layer is included before the convolutional layers. To restore the sequence structure and reshape the output of the convolutional layers to sequences of feature vectors, a sequence of unfolding layer and a flatten layer are inserted between the convolutional layers and the BiLSTM layer. The different layers as well as their parameters values of the proposed network structure are depicted as follows:

- A sequence input layer with an input size of dimensions [10 23 1]
- Sequence Folding Layer
- CNN consists of convolutional layer, batch normalization layer and ReLU layer with number of filters equals 30 and with filter size of 5-by-5. Two series of the abovementioned structural layers are used with the proposed network
- Sequence Unfolding Layer restores the sequence structure of the input data after sequence folding
- Flatten Layer
- BiLSTM network with 350 hidden units that outputs the last time step only
- A fully connected layer of size 9 which is the number of classes followed by a softmax layer and a classification layer

The structure of the proposed hybrid Conv-BiLSTM network as well the structure and graph of its layers are shown in Fig. 6 and 7.
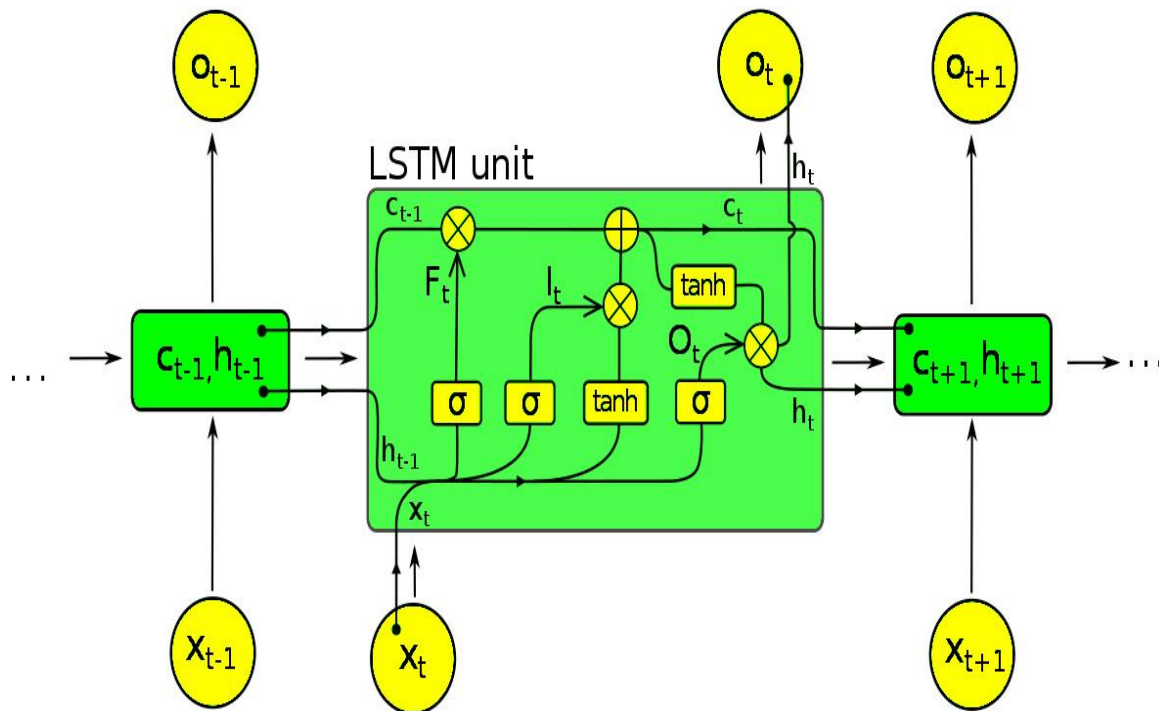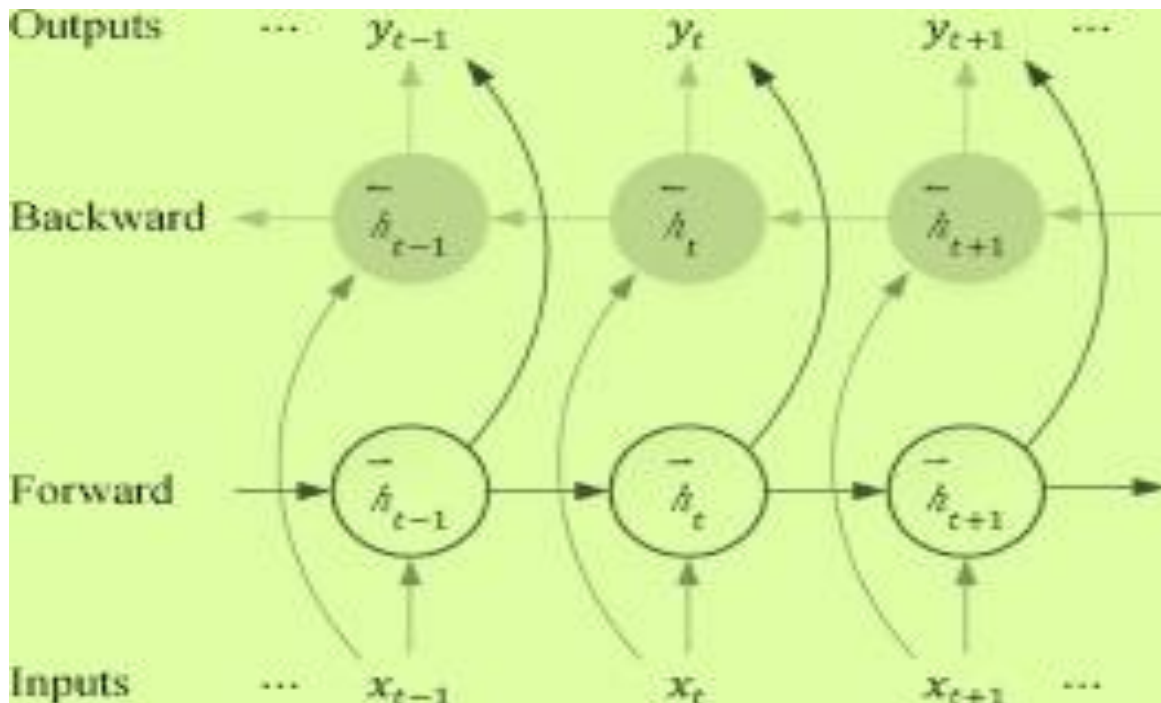

Fig. 4: The structure of LSTM cell

Fig. 5: The structure of BiLSTM network
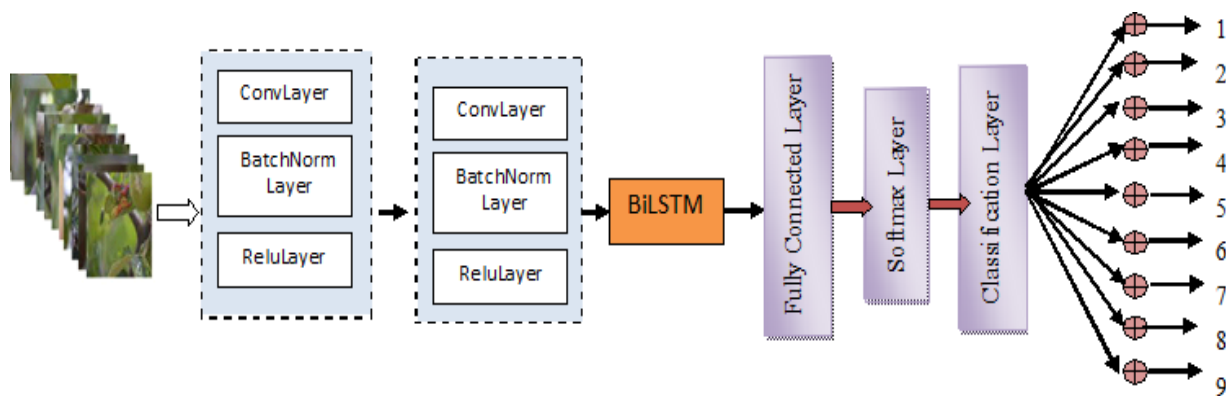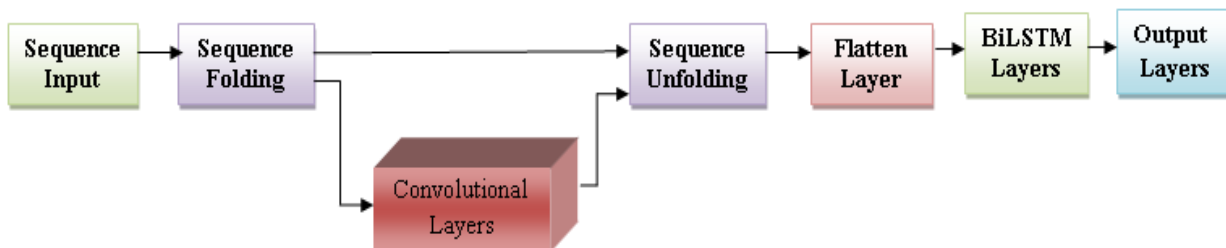

Fig. 6: The structure of convolutional network and BiLSTM network
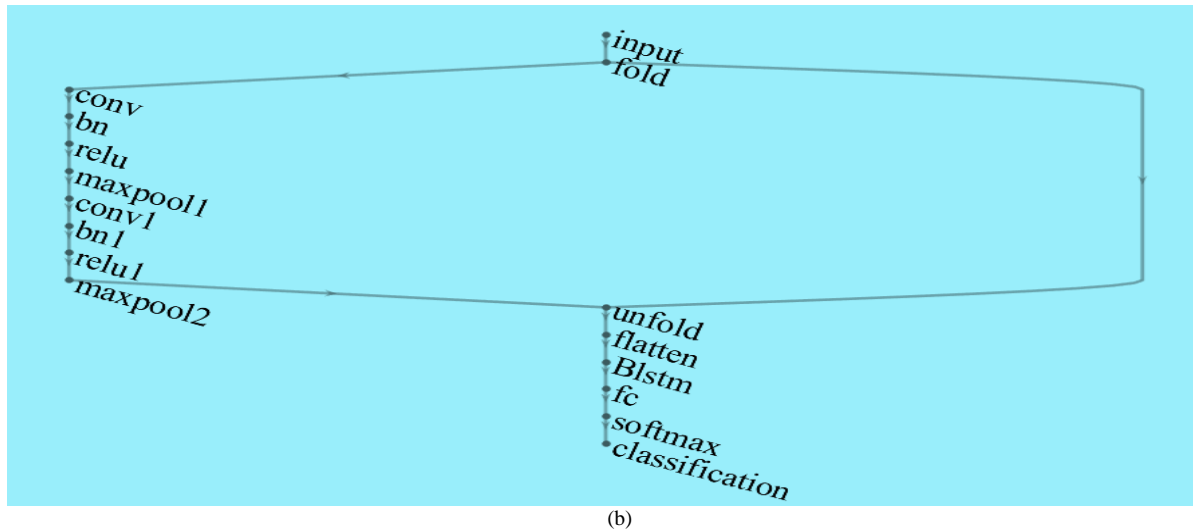

(a)

(b)

Fig. 7: (a) Layers structure of the suggested network; (b) the graph layers of the suggested network

The Exeperimental Results

The suggested Conv-BiLSTM classifier is verified using four different training data sets of land cover which are presented as follows:

- In first case, all training samples 81714 depicted in Table 1 are used for training the suggested classifier, while all test samples descrbed in the same table are used for testing the proposed Conv-BiLSTM network
- In the second case, half the training samples which are 40875 samples are used for training the proposed classifier, while the other half training samples are used for testing the suggested Conv-BiLSTM network . The training dataset used in this case represents 0.5 from the total training dataset demonstrated in Table 1
- In the third case, 1600 samples from each class of the training data s*et a*re used for trianing the model, while 999 samples from each class of the training data set is utilized for testing the suggested network. This means that the training dataset from all class in this case are 14400 samples which represent 0.18 from the total training dataset depicted in Table 1
- In the fourth case, 1000 samples from each class of the training data set are used for trianing the network, while the Conv-BiLSTM network is tested using 600 samples from each class of the training data set. This means that the training dataset from all class in this case are 9000 samples which represent 0.11 from the total training dataset

Also, to evaluates the correctness and efficiency of proposed Conv-BiLSTM network compared to the classifiers of BiLSTM and Random Forest (RF) (Breinman, 2001), Precision, Recall and F-score metrics are used as metric for this purpose as depicted in equations from (4-6).

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Re}call = \frac{TP}{TP + FN} \tag{5}$$

$$F - Score = \frac{1}{|C|} \sum_{CEC} \frac{2xPrecision \times Recall}{Precision + Recall} \tag{6}$$

Where:

TP = The positive samples which are classified truly,
FP = False classified positive samples and FN are the false negative classified samples

The classification results of the suggested Conv-BiLSTM classifier with the above mentioned four training cases compared to BiLSTM classifier and the Random Forest classifier in metric of Precision, Recall and F-score are depicted from Table 2 to 5. Also, both the average of F-Score results and the accuracy of the suggested Conv-BiLSTM network compared to that of BiLSTM, CNN and RF classifiers versus the ratio of the used training dataset in each aforementioned case to the total training data are shown in Fig. 8 and 9 respectively. The classification accuracy of suggested Conv-BiLSTM classifier compared to the the accuracy of the other classifiers is depicted in Table 6.
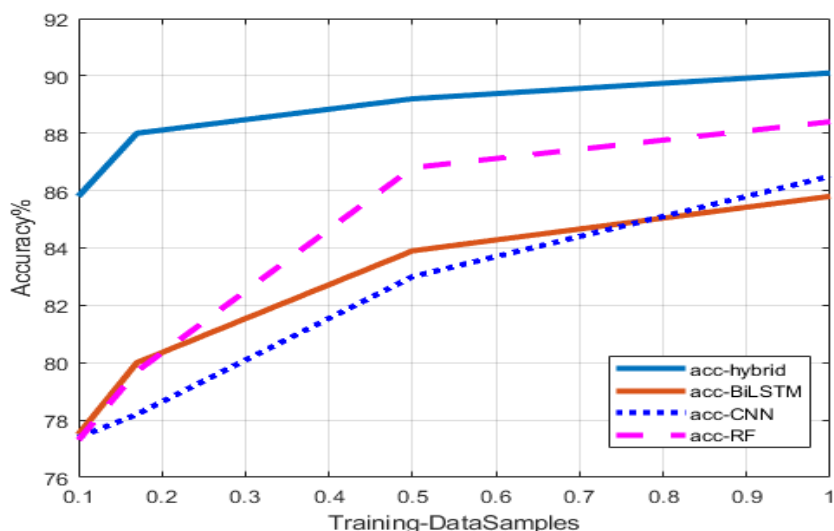
Fig. 8: The classification accuracy of the suggested Conv-BiLSTM network compared to the BiLSTM, CNN networks and the traditional RF technique versus the ratio of training samples used in different training cases
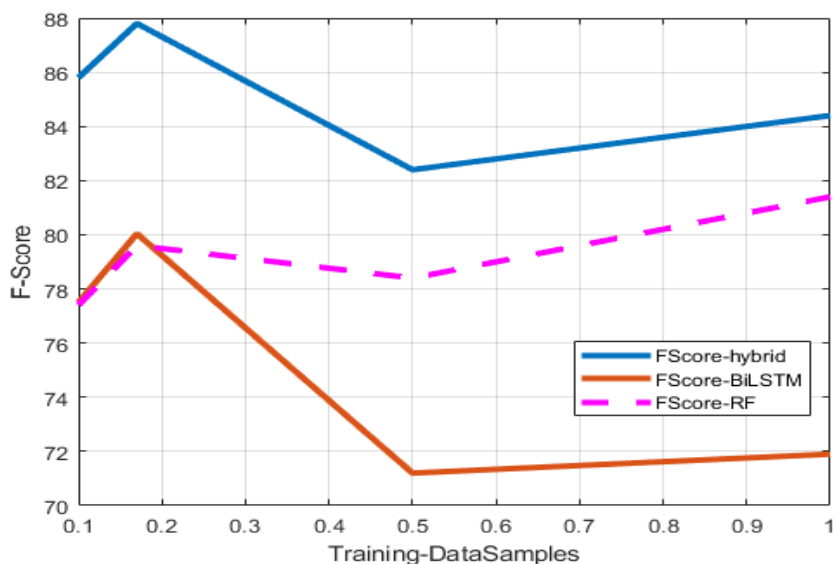
Fig. 9: The classification accuracy of of the suggested Conv-BiLSTM network compared to the BiLSTM nework and the traditional RF technique in metric of F-Score versus the ratio of training samples used in different training cases

Table1: The training samples and test samples with the corresponding class label for each of them.

Class ID	Class name	Number of training samples	Number of test samples
1	Urban area	16000	4000
2	Other-built-up surfaces	3236	647
3	Forests	16000	4000
4	Sparse vegetation	16000	3398
5	Rocks and bare soil	12942	2588
6	Grassland	5681	1136
7	Sugarcane crops	7656	1531
8	Other crops	1600	154
9	Water	2599	519

Table 2: Classification accuracy of the suggested Conv-BiLSTM network compared to BiLSTM network and RF classifier in terms of Precision, Recall and F-Score using the first training case

| Class name | 1- All 81714 samples of training data sets are used for training the classifiers | | | | | | | | |
| | BiLSTM and CNN | | | BiLSTM | | | Random forest | | |
	Precision	Recall	F-Score	Precision	Recall	F-Score	Precision	Recall	F-Score
Urban Areas	93.62	90.4	91.99	92.5	82.7	87.3	90.2	85.7	87.9
Other-built-up surfaces	64.91	80.1	71.70	33.3	65.6	44.2	53.2	72.0	61.2
Forests	89.97	92.04	90.99	87.0	88.5	87.8	91.9	85.6	88.6
Sparse Vegetation	92.70	93.7	93.2	89.8	87.7	88.8	91.8	91.8	91.8
Rocks & bare soil	94.66	95.1	94.9	90.6	94.2	92.4	94.9	95.3	95.1
Grassland	91.02	86.8	88.8	80.3	66.5	72.7	82.6	79.6	81.1
Sugarcane	93.27	96.2	94.7	86.2	96.1	90.9	92.8	93.3	93.08
Other crops	65.58	38.4	48.4	4.5	33.3	8	34.9	91.9	50.6
Water	85.74	84.6	85.1	73.7	76.4	75.1	77.7	89.3	83.1
Average%	85.7	84.4	84.4	70.9	76.8	71.9	78.9	87.2	81.4
Classification Accuracy	90.1%			85.8%			88.4%		

Table 3: Classification acuracy of the suggested Conv-BiLSTM network compared to BiLSTM network and RF classifier in terms of Precision, Recall and F-Score using the second training case

| Class name | 2- 40 857samples of training data sets is used for training the classifiers and 40857 samples for testing the classifiers | | | | | | | | |
| | BiLSTM and CNN | | | BiLSTM | | | Random forest | | |
	Prec.	Recall	F-score	Prec.	Recall	F-score	Prec.	Recall	F-score
Urban Areas	93.38	84.7	88.8	93.46	79.36	85.84	89.2	84.5	86.8
Other-built-up surfaces	59.73	77.3	67.4	30.09	66.35	41.40	49.4	69.3	57.6
Forests	85.59	90.3	87.9	88.72	81.94	85.20	91.3	83.6	87.3
Sparse Vegetation	92.59	91.5	92.06	88.69	85.71	87.17	90.4	90.4	90.4
Rocks and bare soil	95.22	94.3	94.7	87.95	94.89	91.29	93.8	94.6	94.2
Grassland	87.13	82.2	84.6	78.43	71.83	74.98	80.05	76.7	78.6
Sugarcane	91.34	94.9	92.7	84.77	96.01	90.12	91.6	92.6	92.1
Other crops	47.70	52.6	50.06	5.657	81.13	10.57	24.1	84.4	37.5
Water	78.52	89.1	83.5	72.2	76.5	74.3	75.4	87.2	80.9
Average %	81.2	84.1	82.4	70%	81.5%	71.2%	76.1	84.8	78.4
Classification accuracy	89.2%			83.9%			86.8%		

Table 4: Classification accuracy of the suggested Conv-BiLSTM network compared to BiLSTM network and RF classifier in terms of precision, recall and F-score using the third training case

| Class name | 3-1600samples from each class of training data sets are used for training the classifiers and 999 samples for testing the classifiers | | | | | | | | |
| | BiLSTM and CNN | | | BiLSTM | | | Random forest | | |
	Prec.	Recall	F-score	Prec.	Recall	F-score	Prec.	Recall	F-score
Urban Areas	82.58	81.6	82.1	81.6	77.2	79.4	71.77	75.7	73.6
Other-built-up surfaces	79.87	85.7	82.6	65.4	78.6	71.4	70.7	72.6	71.76
Forests	79.27	83.8	81.4	69.5	74.7	72.2	79.6	72.6	75.9
Sparse Vegetation	93.09	93.9	93.5	91.89	92.2	92.07	89.5	92.5	91.04
Rocks & bare soil	89.48	87.0	88.2	82.68	71.4	76.6	81.1	75.0	77.9
Grassland	90.79	87.4	89.1	80.98	85.6	83.2	82.5	86.2	84.3
Sugarcane	87.58	97.4	92.2	86.6	93.6	90.0	88.4	90.6	89.5
Other crops	97.89	83.7	90.2	74.8	65.2	69.7	69.1	67.9	68.4
Water	90.79	92.1	91.4	86.2	85.4	85.7	83.4	85.1	84.2
Average%	87.9	88.1	87.8	80%	80.4%	80.04%	79.3	79.8	79.6
Classification accuracy	88%			79.99%			79.7%		

Table 5: Classification accuracy of the suggested Conv-BiLSTM network compared to BiLSTM network and RF classifier in terms of Precision, Recall and F-Score using the fourth training case

	4-1000 samples from each class of training data sets are used for training the classifiers and 600 samples from each class of training data sets are used for testing the classifiers								
	BiLSTM and CNN			BiLSTM			Random forest		
Class name	Prec.	Recall	F-score	Prec.	Recall	F-score	Prec.	Recall	F-score
Urban areas	80.5	78.1	79.3	80.5	68.1	73.7	73.3	75.9	74.6
Other-built-up surfaces	74.1	83.4	78.5	54.5	73.1	62.4	68.5	67.2	67.8
Forests	81.6	83.4	82.5	67.6	74.3	70.8	76.5	72.6	74.5
Sparse Vegetation	90.6	92.9	91.8	86.6	93.3	89.8	85.1	89.3	87.2
Rocks & bare soil	83.6	86.7	85.1	81.3	71.5	76.1	79.1	71.1	74.9
Grassland	84.6	85.1	84.8	84.1	77.2	80.5	78.3	82.4	80.3
Sugarcane	90.8	93.1	91.9	89	90.9	89.9	87.6	88.7	88.1
Other crops	96.8	82.1	88.9	71.6	66.4	68.9	64	66.6	65.3
Water	89.8	88.5	89.1	82	86.1	84.0	83.1	83.3	83.2
Average %	85.8	85.9	85.8%	77.5	77.9	77.4%	77.3	77.5%	77.4
Classification accuracy	85.8%			77.5			77.3%		

Table 6: Classification Accuracy of the Suggested Conv-BiLSTM network using the different training samples cases and compared to the BiLSTM, CNN and RF classifiers

Classes training samples	Accuracy CNN and BiLSTM	Accuracy BiLSTM	Accuracy CNN	Accuracy RF
81714 samples	90.1%	85.8%	86.5%	88.4%
40587 samples	89.2%	83.9%	83%	86.8%
14400 samples	88%	80%	78.2%	79.7%
9000 samples	85.8%	77.5%	77.4%	77.3%

Besides, the efficiency of the proposed Conv-BiLSTM is verified through its comparison to the state-of-the-art method; weasel + muse (Word ExtrAction for time SEries cLassification + MUltivariate Symbols and dErivatives) (Schäfer and Leser, 2017) in metric of F-Score. The average F-Score of the proposed Conv-BiLSTM network reached to 87.8% when training by 1600 samples from each class of training data sets, while the WEASEL + MUSE technique reached to 86.6% when training with half training samples. The proposed Conv-BiLSTM network has larger F-Score value comparing to the state-of- the- art WEASEL + MUSE by 1.38%.

Results and Dissucision

From the results given in Table 2 to 5, the suggested network for multivariate time series classification of nine land cover categories outperforms the other compared classifiers BiLSTM, CNN and Random Forest. The classification accuracy of the suggested Conv-BiLSTM network reaches to 90.1%. Also, it is demonstrated that the suggested network has best F-Score value when it is trained using equal samples from each class of the training data*set a*nd its F-Score value ranges from 87.8 to 85.8%. While the suggested network has smaller F-Score value when it is trained with non-equal samples from each class of the training dataset either using the total training samples or half samples and its F-Score value ranges from 84.4 to 82.2% respectively. Concerning the other compared classifiers, the accuracy of classification of

both BiLSTM and RF in terms of F-Score have same values when training the two classifiers with 1000 samples from each class of land cover dataset. The accuracy of classification with BiLSTM has larger value in average by 0.55% compared to RF when using the last two cases of training data. While the accuracy of RF classifier in terms F-Score is larger than that of BiLSTM classifier in average 11.65% with the two first cases of training data respectively. On other side, the accuracy in classifying the other crops class with all classifiers using the first cases of training dataset in metric of Precision and F-Score was very poor due to the diversity of crops inside the class which result in high intra-class variability and this faced a difficulty in classification. Generally, the classification accuracy of the suggested Conv-BiLSTM network with all training cases has larger value in average by 6.5, 8 and 8.7% compared to RF, BiLSTM and CNN classifiers respectively. Besides, the accuracy of classification of the proposed Conv-BiLSTM network in metric of F-score is in average larger by 1.38% than that value of the state-of-the-art WEASEL+MUSE technique.

Conclusion

In this study, a hybrid model for multivariate time series classification of land cover is suggested. The suggested model combines both of the convolutional network and BiLSTM network which enable the suggested model to extract the spatial and temporal features of the small land cover data set used. From the

simulation results of training the suggested network in metrics of F-Score, Precision and Recall which are given in previous tables prove that the suggested model outperforms the other comparison techniques. The classification accuracy of the suggested network using the aforementioned different cases of training data set ranges from 90.1 to 85.8% respectively. Moreover, the accuracy of classification of the proposed Conv-BiLSTM network in metric of F-Score is in average 1.38% better than that value of the state-of-the-art WEASEL+MUSE technique in literature.

Acknowledgement

The author would like to thank all my colleagues and the researchers in the department who give me useful comments about the work. Also, I am grateful to the encourage and support of our ERI institute with the required tools and software.

Ethics

The author confirms that the research paper is original, the experimental results are not published before and no ethical issues are involved.

References

Abebaw Alem and Shailender Kumar, (2020). Deep Learning Methods for Land Cover and Land Use Classification in Remote Sensing: A Review. 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)Amity University, Noida, India. June 4-5, 2020, 978-1-7281-7016-9/20/$31.00 ©2020 IEEE.

Araki, S., Shima, M., & Yamamoto, K. (2018). Spatiotemporal land use random forest model for estimating metropolitan NO2 exposure in Japan. Science of The Total Environment, 634, 1269-1277. doi.org/10.1016/j.scitotenv.2018.03.324

Bailly, A. (2018). Time Series Classification Algorithms with Applications in Remote Sensing (Doctoral dissertation, Université Rennes 2). https://tel.archives-ouvertes.fr/tel-02139897/

Baydogan, M. G., & Runger, G. (2015). Learning a symbolic representation for multivariate time series classification. Data Mining and Knowledge Discovery, 29(2), 400-422. doi.org/10.1007/s10618-014-0349-y

Baydogan, M. G., & Runger, G. (2016). Time series representation and similarity based on local autopatterns. Data Mining and Knowledge Discovery, 30(2), 476-509. https://link.springer.com/article/10.1007%2Fs10618-015-0425-y

Betts, M. G., W. J. Christopher Wolf, B. P. Ripple, K. A. Millers, S. H. Adam Duarte, M. Butchart and T. Levi. (2017). Global Forest Loss Disproportionately Erodes Biodiversity in Intact Landscapes. Nature 547: 441. doi.org/10.1038/nature23285

Bengio, Y. (2009). Learning deep architectures for AI. Now Publishers Inc.

Bengio, Y. (2013, July). Deep learning of representations: Looking forward. In International conference on statistical language and speech processing (pp. 1-37). Springer, Berlin, Heidelberg. https://link.springer.com/chapter/10.1007/978-3-642-39593-2_1

Breinman, L. (2001). Random forests. Mach. Learn., 45, 5-32. https://ci.nii.ac.jp/naid/10030586413/

Cui, Y., Shi, J., & Wang, Z. (2015). Complex rotation quantum dynamic neural networks (CRQDNN) using complex quantum neuron (CQN): Applications to time series prediction. Neural Networks, 71, 11-26. doi.org/10.1016/j.neunet.2015.07.013

Cuturi, M., & Doucet, A. (2011). Autoregressive kernels for time series. arXiv preprint arXiv:1101.0673. https://arxiv.org/abs/1101.0673

Denize, J. (2015). Développement d'un modèle de classification d'images SAR pour la cartographie des cultures tropicales: Exemple de l'île de la Réunion (Doctoral dissertation, Université Rennes 2). https://agritrop.cirad.fr/589973/

Geurts, P. (2001, September). Pattern extraction for time series classification. In European conference on principles of data mining and knowledge discovery (pp. 115-127). Springer, Berlin, Heidelberg. doi.org/10.1007/3-540-44794-6_10

Gharghory, S. M. (2020). Deep Network based on Long Short-Term Memory for Time Series Prediction of Microclimate Data inside the Greenhouse. International Journal of Computational Intelligence and Applications, 19(02), 2050013. https://www.worldscientific.com/doi/abs/10.1142/S1469026820500133

Goodin, D. G., K. L. Anibas and M. Bezymennyi. (2015). Mapping Land Cover and Land Use from Object-based Classification: An Example from a Complex Agricultural Landscape. International Journal of Remote Sensing 36 (18): 4702–4723. doi:10.1080/01431161.2015.1088674.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural networks, 18(5-6), 602-610. doi.org/10.1016/j.neunet.2005.06.042.

Huai-Shuo Huang, Chien-Liang Liu, Vincent S. Tseng. (2018). ,Multivariate Time Series Early Classification using Multi-Domain Deep Neural Network. IEEE 5th International Conference on Data Science and Advanced Analytics (pp. 90-98), doi.org/10.1109/DSAA.2018.00019.

Kadous, M. W. (2002). Temporal classification: Extending the classification paradigm to multivariate time series. Kensington: University of New South Wales. https://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.67.8007&rep=rep1&type=pdf

Kang, H., & Choi, S. (2014). Bayesian common spatial patterns for multi-subject EEG classification. Neural Networks, 57, 39-50. doi.org/10.1016/j.neunet.2014.05.012

Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2019). Multivariate LSTM-FCNs for time series classification. Neural Networks, 116, 237-245. doi.org/10.1016/j.neunet.2019.04.014

Kavzoglu, T. (2017). Object-oriented random forest for high resolution land cover mapping using quickbird-2 imagery. In Handbook of neural computation (pp. 607-619). Academic Press. doi.org/10.1016/B978-0-12-811318-9.00033-8

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105. https://proceedings.neurips.cc/paper/2012/file/c3998 62d3b9d6b76c8436e924a68c45b-Paper.pdf

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. IEEE Geoscience and Remote Sensing Letters, 14(5), 778-782. https://ieeexplore.ieee.org/abstract/document/7891032

Kuwata, K., & Shibasaki, R. (2015, July). Estimating crop yields with deep learning and remotely sensed data. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 858-861). IEEE. https://ieeexplore.ieee.org/abstract/document/7325900

Längkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognition Letters, 42, 11-24. doi.org/10.1016/j.patrec.2014.01.008

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech and time series. The handbook of brain theory and neural networks, 3361(10), 1995. http://www.iro.umontreal.ca/~lisa/pointeurs/handbo ok-convo.pdf

Li, W., Fu, H., Yu, L., & Cracknell, A. (2017). Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sensing, 9(1), 22. https://www.mdpi.com/2072-4292/9/1/22

Li, W., Fu, H., Yu, L., Gong, P., Feng, D., Li, C., & Clinton, N. (2016). Stacked Autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping. International journal of remote sensing, 37(23), 5632-5646. https://www.tandfonline.com/doi/abs/10.1080/0143 1161.2016.1246775

Mansaray, L. R., Wang, F., Huang, J., Yang, L., & Kanu, A. S. (2020). Accuracies of support vector machine and random forest in rice mapping with Sentinel-1A, Landsat-8 and Sentinel-2A datasets. Geocarto International, 35(10), 1088-1108. https://www.tandfonline.com/doi/abs/10.1080/101 06049.2019.1568586

Mas, J. F., & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. International Journal of Remote Sensing, 29(3), 617-663. https://www.tandfonline.com/doi/full/10.1080/0143 1160701352154

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 1419. https://www.frontiersin.org/articles/10.3389/fpls.201 6.01419/full

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), 247-259. doi.org/10.1016/j.isprsjprs.2010.11.001

Nitze, I., Barrett, B., & Cawkwell, F. (2015). Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. International Journal of Applied Earth Observation and Geoinformation, 34, 136-146. doi.org/10.1016/j.jag.2014.08.001

Pires de Lima, R., & Marfurt, K. (2020). Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. Remote Sensing, 12(1), 86. https://www.mdpi.com/2072-4292/12/1/86

Prieto, O. J., Alonso-Gonzلlez, C. J., & Rodrيguez, J. J. (2015). Stacking for multivariate time series classification. Pattern Analysis and Applications, 18(2), 297–312.

Rajendran, G. B., Kumarasamy, U. M., Zarro, C., Divakarachari, P. B., & Ullo, S. L. (2020). Land-use and land-cover classification using a human group-based particle swarm optimization algorithm with an LSTM Classifier on hybrid pre-processing remote-sensing images. Remote Sensing, 12(24), 4135. https://www.mdpi.com/2072-4292/12/24/4135

Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo and J. P. Rigol-Sanchez. (2012). An Assessment of the Effectiveness of a Random Forest Classifier for Land-cover Classification. ISPRS Journal of Photogrammetry and Remote Sensing 67: 93–104, doi:10.1016/j.isprsjprs.2011.11.002.

Running, S. W. (2008). Ecosystem disturbance, carbon and climate. Science, 321(5889), 652-653. https://science.sciencemag.org/content/321/5889/ 652.summary

Schäfer, P., & Leser, U. (2017). Multivariate time series classification with WEASEL+ MUSE. arXiv preprint arXiv:1711.11343. https://arxiv.org/abs/1711.11343

Schafer, P., Pflugmacher, D., Hostert, P., & Leser, U. (2018, March). Classifying Land Cover from Satellite Images Using Time Series Analytics. In EDBT/ICDT Workshops (pp. 10-15).

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117. doi.org/10.1016/j.neunet.2014.09.003

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE transactions on Signal Processing, 45(11), 2673-2681. https://ieeexplore.ieee.org/abstract/document/650093/

Shao, Y., & Lunetta, R. S. (2012). Comparison of support vector machine, neural network and CART algorithms for the land-cover classification using limited training data points. ISPRS Journal of Photogrammetry and Remote Sensing, 70, 78-87. doi.org/10.1016/j.isprsjprs.2012.04.001

Sharabiani, A., Darabi, H., Rezaei, A., Harford, S., Johnson, H., & Karim, F. (2017). Efficient classification of long time series by 3-d dynamic time warping. IEEE Transactions on Systems, Man and Cybernetics: Systems, 47(10), 2688-2703. doi.org/10.1109/TSMC.2017.2699333.

Spiegel, S., Gaebler, J., Lommatzsch, A., De Luca, E., & Albayrak, S. (2011). Pattern recognition and classification for multivariate time series. In Proceedings of the fifth international workshop on knowledge discovery from sensor data (pp. 34–42). ACM.

Swietojanski, P., Ghoshal, A., & Renals, S. (2014). Convolutional neural networks for distant speech recognition. IEEE Signal Processing Letters, 21(9), 1120-1124. doi.org/10.1109/LSP.2014.2325781.

TiSeLaC, (2017): Time Series Land Cover Classification challenge. https://sites.google.com/site/dinoienco/tiselc

Tuncel, K. S., & Baydogan, M. G. (2018). Autoregressive forests for multivariate time series modeling. Pattern recognition, 73, 202-215. doi.org/10.1016/j.patcog.2017.08.016

Wang, L., Wang, Z., & Liu, S. (2016). An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm. Expert Systems with Applications, 43, 237-249. doi.org/10.1016/j.eswa.2015.08.055

Wistuba, M., Grabocka, J., & Schmidt-Thieme, L. (2015). Ultra-fast shapelets for time series classification. arXiv preprint arXiv:1503.05018. https://arxiv.org/abs/1503.05018

Yu, Z., & Lee, M. (2015). Real-time human action classification using a dynamic neural model. Neural Networks, 69, 29-43. doi.org/10.1016/j.neunet.2015.04.013

Zhao, B., Lu, H., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification. Journal of Systems Engineering and Electronics, 28(1), 162-169. doi.org/10.21629/JSEE.2017.01.18

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014, June). Time series classification using multi-channels deep convolutional neural networks. In International Conference on web-age information management (pp. 298-310). Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-319-08010-9_33

Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. Remote sensing of environment, 221, 430-443. doi.org/10.1016/j.rse.2018.11.032