

COMPARATIVE STUDY ON DEEP NEURAL NETWORK MODELS FOR CROP CLASSIFICATION USING TIME SERIES POLSAR AND OPTICAL DATA

G. S. Phartiyal*, D. Singh

ECE Department, Indian Institute of Technology, Roorkee, India – (gphartiyal, dharm}@ec.iitr.ac.in

Commission V, SS: Emerging Trends in Remote Sensing

KEY WORDS: Deep neural networks, CNNs, LSTMs, ConvLSTMs, Crop classification, PolSAR, Time series satellite data

ABSTRACT:

Crop classification is an important task in many crop monitoring applications. Satellite remote sensing has provided easy, reliable, and fast approaches to crop classification task. In this study, a comparative analysis is made on the performances of various deep neural network (DNN) models for crop classification task using polarimetric synthetic aperture radar (PolSAR) and optical satellite data. For PolSAR data, Sentinel 1 dual pol SAR data is used. Sentinel 2 multispectral data is used as optical data. Five land cover classes including two crop classes of the season are taken. Time series data over the period of one crop cycle is used. Training and testing samples are measured and collected directly from the ground over the study region. Various convolutional neural network (CNN) and long short-term memory (LSTM) models are implemented, analysed, evaluated, and compared. Models are evaluated on the basis of classification accuracy and generalization performance.

1. INTRODUCTION AND RELATED WORK

Crop classification is an important task in many crop monitoring applications such as generation of crop maps, crop yield estimation, crop rotation records, and soil productivity (Löw et al. 2013). Satellite remote sensing has provided easy, reliable, and fast approaches to crop classification task. With the availability of higher spatial, spectral, and temporal resolutions satellites, more and more data is available to improve crop classification accuracy. SAR and optical satellite image data complement each other in agricultural applications like crop classification (Blaes, Vanhalle, and Defourny 2005). Exploiting the two data modalities always helped the cause. Numerous approaches have been developed over the years to utilize both of these datasets in synergy for agriculture applications (Blaes et al. 2005).

Machine learning (ML) also played an important contribution in synergically using both datasets for agricultural applications. Researchers have developed numerous ML algorithms for crop classification using SAR and optical data (Xie et al. 2018) (Wang et al. 2016). Recently, deep neural networks (DNNs) are making its mark as powerful tool for remote sensing applications.

In this study, convolutional deep neural networks (CDNNs) are explored, critically analyzed, and evaluated as a tool for crop classification using SAR and optical satellite data. Generally, CDNNs are good image classifiers, but their applications in remote sensing applications is relatively new. In this study, 2-dimensional (2D), 3-dimensional (3D), and convolutional-long short term memory (Conv-LSTM) neural networks are used for crop classification using sentinel 1 (SAR) and sentinel 2 (Multispectral) time-series data. Study area includes Roorkee city of northern India and its neighboring region. Five land cover classes namely, wheat, sugarcane, bare soil, forest, and

urban are considered. Preliminary study shows good classification accuracy and generalization by all three models.

This article is divided into five sections. Section 1 provides introduction to crop classification and machine learning. Section 2 provides conceptual background on the technologies to be utilized in this study. Further, section 3 describes the methodology proposed in this study. Section 4 is on results obtained and analysis of results. Finally, section 5 concludes the study.

2. BACKGROUND

2.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a special category of feedforward DNNs which are designed specifically to analyse multidimensional images. Since the inception of CNNs into image processing scientific community, they have been the "state of the art" in many image classification applications especially when the dimension of image increases. This property of CNNs made them suitable for various remote sensing applications (Zhu et al. 2017). CNNs learn features from data instead of "hand-engineering" them. This aspect makes the algorithm faster and less "pre-processing" intensive. This also offers less human interaction during processing which is healthy during "process" automation. The architectural and functional components of CNNs are described in the following sub-sections.

2.1.1. Convolutional layer: It is the core functional block of a CNN which consists of several filters/kernels having a limited spatial receptive field but a full spectral receptive field (image channels). During the forward pass, each filter is convolved across the spatial extent of the input volume, computing the dot product between the entries of the filter and the input and

* Corresponding Author

producing a 2-dimensional activation map of that filter. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input. CNN exploits a local connectivity pattern between neurons using this architecture.

2.1.2 Pooling Layer: Pooling is form of down sampling (non-linear). This task is done using numerous methods such as selecting the maximum, or averaging in a predefined spatial pooling window (Gu et al. n.d.).

2.1.3 Activation function: This layer applies some sort of logic (activation function) to the pooling layer neuron output. It increases the non-linear properties of the network without disturbing the convolutional layer's receptive fields (Yann LeCun, Yoshua Bengio, and Geoffrey Hinton 2015). One such popular activation function is "ReLU". The mathematical formulation of a ReLU is given in equation 1.

$$f(x) = \max(0, x) \quad (1)$$

Where, x is the output from the pooling layer neuron. Other popular activation functions used with CNNs are; the hyperbolic tangent, and the sigmoid activation functions.

2.1.4 Fully connected layer: After several convolutional and pooling layers, a fully connected conventional perceptron layer is employed. Every neuron in this layer is connected to every activation of the previous layer. The working of this layer is similar to a classical neural network layer. The output of this layer is the desired target.

2.1.5 Hyperparameters and regularization: For efficient working of the CNNs for a particular application, various network parameters are to be set. These parameters such as number of convolutional/pooling layers, number of filters/kernels in each layer, shape of filters (convolutional/pooling) are the hyperparameters of the network. Setting these parameters is known as hyperparameter tuning. This process of 'tuning' is done either manually (hit and error method) or with the help of hyperparameter tuning approaches. Regularization is process, employed during training of the network, to avoid the problem of "overfitting". Dropout is the most preferred regularization method.

In this study, 3D CNNs are used for crop classification purpose. The architecture and functionality of 3D CNNs is similar to as explained above. The only difference is that 3D CNNs incorporate an extra dimension, the temporal dimension, which makes it beneficial for time series analysis of satellite data.

2.2 Long Short-Term Memory

Lang short-term memory (LSTM) are units of a recurrent neural network (RNN). A network of such units is called an LSTM network (Sainath et al. 2015). A common peephole LSTM unit is composed of a cell, an input gate, an output gate and a forget gate as shown in figure 1. In figure 1, each of these gates can be thought as a "standard" neuron in a feed-forward (or multi-layer) neural network: that is, they compute an activation (using an activation function) of a weighted sum. In figure 1, i_t , o_t , and f_t represent the activations of respectively the input, output and forget gates, at time step t . The 3 exit arrows from the memory cell c to the 3 gates i , o , and f represent

the peephole connections. These peephole connections actually denote the contributions of the activation of the memory cell c at time step $t-1$, i.e. the contribution of c_{t-1} (and not c_t , as the picture may suggest). In other words, the gates i , o , and f calculate their activations at time step t (i.e., respectively, i_t , o_t , and f_t) also considering the activation of the memory cell c , at time step $t-1$, i.e. c_{t-1} . The single left-to-right arrow exiting the memory cell is not a peephole connection and denotes c_t . The little circles containing a x symbol represent an element-wise multiplication between its inputs. The big circles containing an S-like curve represent the application of a differentiable function (like the sigmoid function) to a weighted sum.

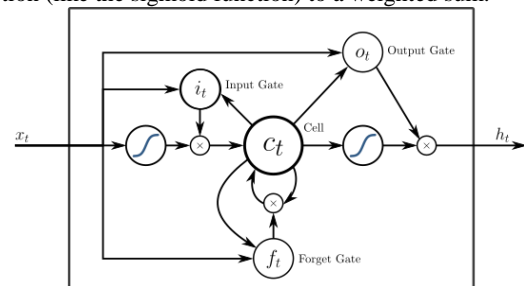


Figure 1 Architecture of an LSTM cell with input (i.e. i), output (i.e. o), and forget (i.e. f) gates.

In brief, the cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. LSTM networks are suitable to classifying, processing and making predictions based on time series data which is the requirement of this study.

In the current study, both CNN-DNN and LSTM-DNNs, separately as well as combined, are evaluated for crop classification task.

3. METHODOLOGY

3.1 Study Area

The study area considered is in the outskirts of Roorkee city, Uttarakhand, India. The area spans about 27 square kilometres. The area includes forests, agricultural lands, built up, and barren lands. Sugarcane, rice and wheat are major crops grown in the area. Google Earth (GE) imagery of the study region is shown in figure 2. The central latitude and longitude of the study area are 29.814692 degrees and 78.054364 degrees respectively. Areas marked in the image (1, 2, 3, and 4) are subsets of study area used for qualitative performance evaluation of the proposed algorithm.

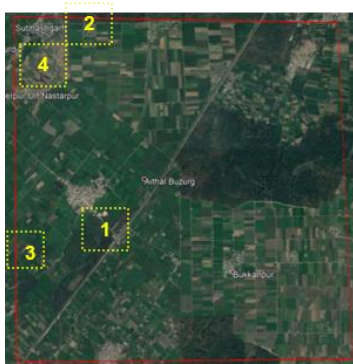


Figure 2 Google Earth image of the study area (GE Image Copyright 2018)

3.2 Data Description and Ancillary Information

Satellite data used for study are both SAR and optical data. SAR data used is Sentinel-1 C-band dual polarimetric SAR data. Sentinel 1 has a temporal revisit of approx. 5 days and data is acquired at a spatial resolution of 10 meters. Optical data used is Sentinel-2 multispectral data. This multispectral data has three spatial resolutions namely 10, 20, and 60 meters. Sentinel-2 also has a temporal revisit time of approx. 5 days at mid latitude regions. Both data are collected over a period of four months (i.e. from October 2017 to January 2018) as this is the season for wheat and sugarcane in the region. A total of 5 sets of data are selected from the observation period. The choice of these data depends on factors such as; low cloud coverage in the optical data or, less difference between SAR and optical data acquisition dates or, enough time difference between two simultaneous data to map growth of crops.

Ground truth is collected directly from the study site. The ground truth data is used for algorithm training, validation, and testing purposes. Ground truth about three land cover classes (built up, forest, and barren land) and two crop types (sugarcane and wheat) are collected directly from the site. Approximately, 400 samples are collected per class. Of the 400 samples, 300 samples are used for training and validation of the algorithms and 100 samples are used for testing purposes.

3.3 Preprocessing

Both SAR and optical data needs preprocessing before to be used together by DNN models. Sentinel-1 C-band PolSAR data downloaded from ESA data portal is a ground range detected (GRD) product. Hence, data calibration is performed first on the GRD data. The calibrated SAR data is the terrain corrected using SRTM 1 arc second DEM provided in the ESA's SNAP (Python API) (Anon n.d.). Sentinel 2 data collected from ESA data portal is a level 1 product. Hence, first an atmospheric correction is performed on the 'L1C' product using "sen2cor" SNAP plugin. The atmospherically corrected "L2A" bands are resampled to 10 meter using the nearest neighbour interpolation method. Finally, both data are co-registered, stacked and subset is taken according to study area. The final stacked subset is used as input to the various DNN models for crop classification. It is to be noted here that Python is used as working platform during this study and Keras "Deep Learning" library is used for DNN model development which provides support in python (Anon n.d.).

A generic flowchart of the study is shown in figure 3. This flowchart is well suited for all DNN based SAR and optical data processing approaches as it includes only the mandatory steps

and not depicting the internal configuration of DNN models which may vary from application to application. The internal configuration of developed DNNs is discussed separately in the next few sections.

3.4 Three Dimensional Convolutional Neural Networks

The 3D CNN model architecture used in this study is as follows. In the first layer, a bank of 10, $5 \times 1 \times 1$ filters is employed, where 10 is the number of filters, 5 is the number of time stamps and, 1×1 is the spatial window size. Here, 1×1 means per pixel convolution. A window of 1×1 is set in order to capture the smallest spatial details possible. Further, ReLU activation function is used in the activation layer. A pooling layer of size $1 \times 2 \times 2$ is employed. No pooling is done in the time dimension and 2×2 is the spatial extent of pooling.

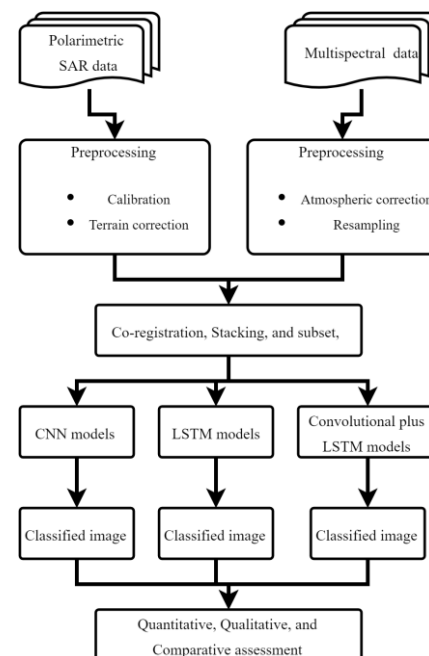


Figure 3 Generic flowchart for DNN based processing of SAR and optical data.

A dropout layer with a dropout value of 0.1 is used for regularization purpose. A flatten layer is used next to transform the 3 dimensional input to one dimensional input vector. Next, a fully connected (FC) layer of 20 neurons with "ReLU" activation function and a dropout layer with value of 0.25 is used. Finally, one more FC layer of 5 neurons (targets) with "softmax" activation function is employed. The model configuration is briefly displayed in figure 4. "Adam" optimizer is used as optimizing technique during model compilation. Optimization is done on the basis of "categorical cross entropy" loss function. It is the most preferred loss function in CNN based classification applications. In the end, it is to be noted that all the hyperparameters are set based on the "trial and error" method.

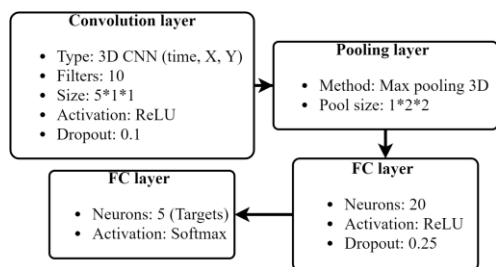


Figure 4 Internal configuration of 3D CNN

3.5 Long Short-Term Memory DNN

LSTM based model developed in this study for crop classification is as follows. In the first layer, 30 LSTM units, each with a size of $32*5*12$ where 32 is the batch size, 5 is the number of time stamps and, 12 the number of bands in one input data. Further, hyperbolic tangent i.e. “Tanh” activation function is used in this layer. A dropout layer with a dropout value of 0.1 is used for regularization purpose. A flatten layer is used next to transform the three dimensional input to one dimensional input vector. Next, a fully connected (FC) layer of 20 neurons with ReLU activation function and a dropout layer with value of 0.25 is used. Finally, one more FC layer of 5 neurons (targets) with “softmax” activation function is employed. The model configuration is briefly displayed in figure 5. “Adam” optimizer is used as optimizing technique during model compilation. Optimization is done on the basis of “categorical cross entropy” error. In brief, apart from the first layer, the model is similar to CNN model.

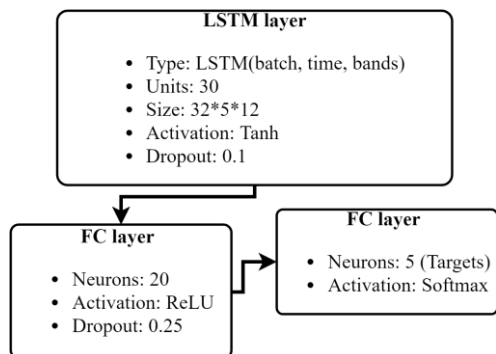


Figure 5 Internal configuration of LSTM model

3.6 Convolutional LSTM (ConvLSTM) models

These models are a hybrid of convolutional models and LSTM models explained in the previous sections. In ConvLSTM models, unlike LSTM, the advantage of convolving with the neighbourhood of a pixel is present. In practice, 2D convolution is used with LSTM. For example, a five dimensional ConvLSTM tensor consists of a time dimension and two spatial dimensions. The other two dimensions are batch size and number of bands. These models are used for object tracking in time series dat. In the current study, ConvLSTMs are to be explored for crop classification task using high dimensional time series satellite image data. In this study, ConvLSTM model is developed for the same purpose. The internal configuration is as follows. The first layer of the ConvLSTM model is a convolutional LSTM layer which consists of 20 ConvLSTM units, each with shape $32*5*1*1*12$ where, 32 is batch size, 5 is the number of time stamps, $1*1$ is the spatial extent of convolutional filter and, 12 is the number of bands. Activation function used is *tanh*. The rest of model configuration is similar

to the LSTM model configuration. The model configuration is summarized in figure 6.

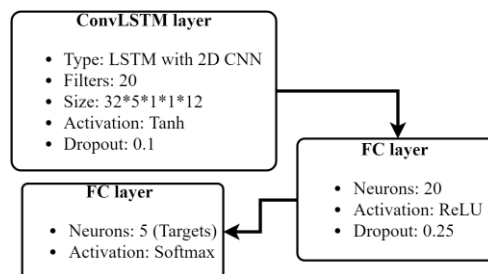


Figure 6 Internal configuration of ConvLSTM model

4. RESULTS AND DISCUSSION

All the models developed in section 3.4, 3.5 and, 3.6 are applied on the data prepared in section3.3. In this section, classification results are analysed and compared on the basis of classification accuracy and generalization capability.

First, the CNN model is applied on the data (stacked SAR and optical) for crop classification. The classification image obtained is shown in figure 7. Overall classification accuracy achieved is 95.02%.

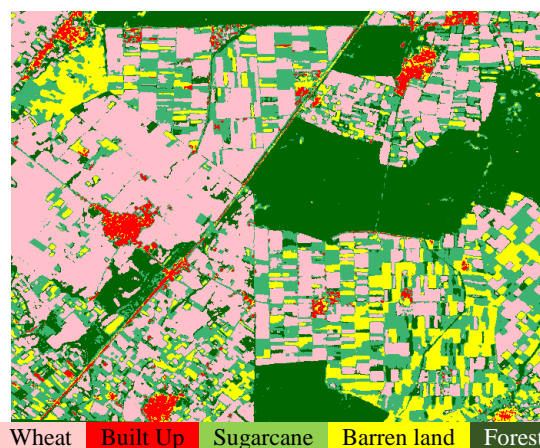


Figure 7 Crop classification image using 3D CNN model.

Next, the LSTM model is used for crop classification. The classification image is shown in figure 8. Overall classification accuracy achieved in this case is 96.8%.

Classification image obtained after applying the ConvLSTM model is shown in figure 9. Overall classification accuracy achieved is 93.6%.

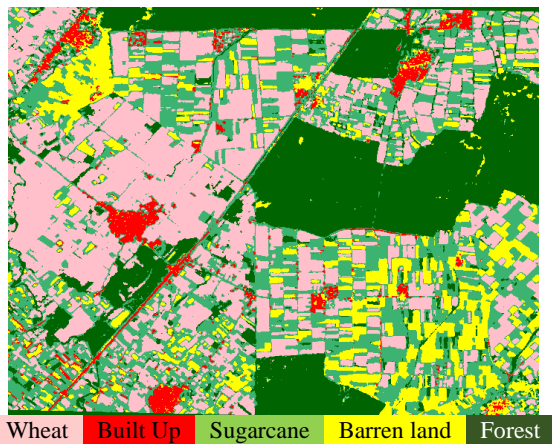
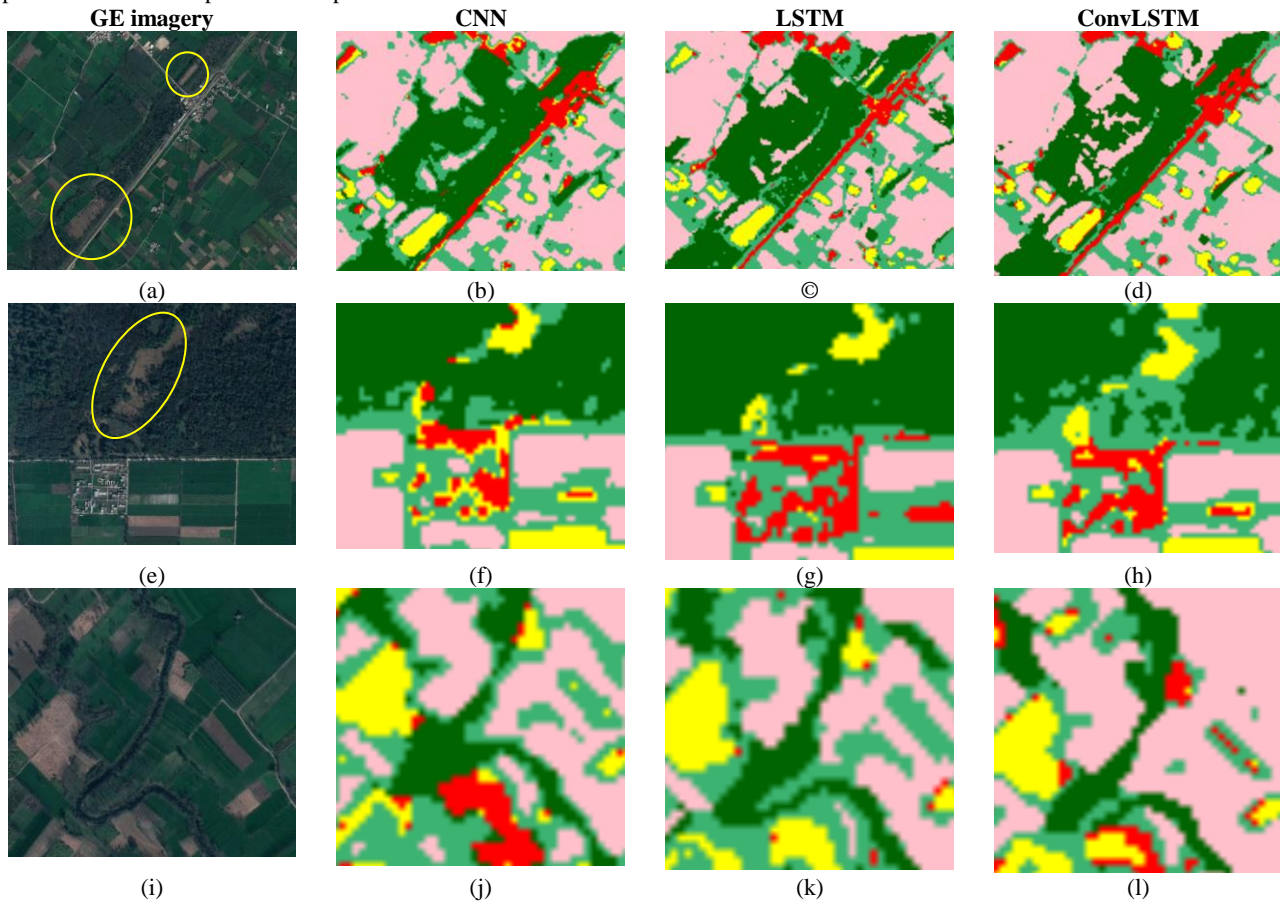


Figure 8 Classification image using LSTM model.

A summary of overall classification accuracy is provided in table 1. Although, from overall accuracy point of view, LSTM model shows the best performance but the models need to be evaluated rigorously. For this purpose, few patches are selected from the study area as reference patches which contain some unique feature. A comparative study on the classification of these patches by the three models is done in the next section.

4.1 Comparative Study

Patches are selected from the study region based on some unique features contained in it. The three models are analysed and evaluated on the basis of their performances in these patches. This comparison is performed to evaluate the



generalization capability of the three models as overall classification accuracy is not sufficient parameter in "overfitting" scenarios. Generalization on the other hand, is a qualitative measure of overfitting i.e. better the generalization, less the model is over-fitted.

	CNN	LSTM	ConvLSTM
Overall Accuracy (%)	95.02	96.8	93.6

Table 1 Summary of classification performance based on overall accuracy.

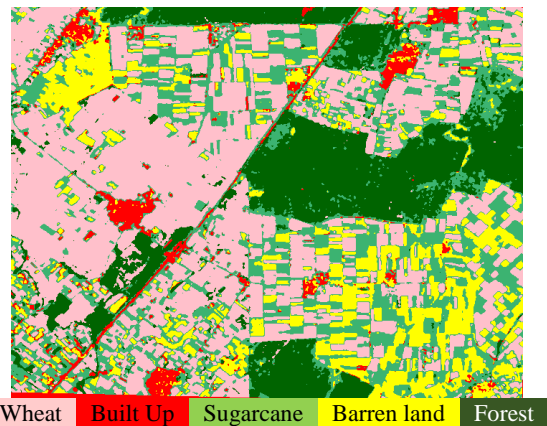


Figure 9 Classification image using ConvLSTM model

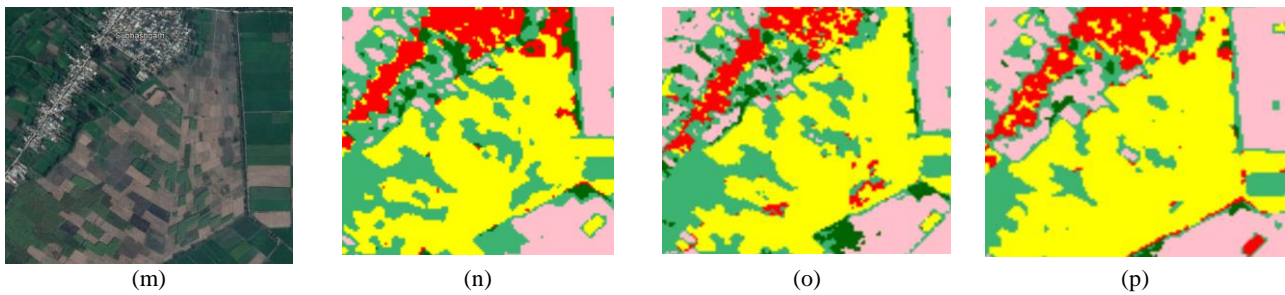


Table 2: Summary of model performances on reference patches

In table 2, figure a, e, i and, m are GE images of reference patches selected for comparison in the study region. In figure a, there are two barren fields marked in circles, a bigger one on the bottom left corner of the image and, a smaller one on the upper right side of the image. Observing the same features in the three classified images (see figure b, c and, d), it is clear that only LSTM model is able to classify the objects correctly. Both, CNN and ConvLSTM have correctly classified the bigger barren field but are unable to correctly classify the smaller one.

In figure e, there are barren areas in the forest (marked in ellipse). In CNN classification image (see figure f), these features are classified as barren, sugarcane and built up. In LSTM classification image (see figure g), these areas are classified as barren and sugarcane but with higher number of pixels correctly classified as barren. Whereas, in ConvLSTM classification image (see figure h), these areas are classified as barren and sugarcane but with higher number of pixels incorrectly classified as sugarcane.

Figure (i) shows agriculture fields and a tree line through the fields. In CNN and ConvLSTM classification images (see figure j and, l), few fallow fields are classified as built up. Whereas, in LSTM classification image (see figure k), they are correctly classified as barren.

Figure (m) shows sugarcane fields, fallow fields and built up. The sugarcane fields are correctly classified by the CNN model but fallow fields are classified as built up (see figure n). The LSTM model successfully maps the sugarcane fields, fallow lands and built up areas (see figure o) with very few fallow fields misclassified as built up. The ConvLSTM model is unable to identify sugarcane fields efficiently. Also, fallow fields are classified as built up by this model.

Overall, from the discussion in the previous sections, it is clear that LSTM based DNN model shows the best generalization performance. The reason for LSTMs better performance over CNNs is the power of LSTMs to memorize patterns which helps them in classifying time series data. Although ConvLSTMs have also LSTMs in their model but the convolution process overshadows LSTM's performance. This study is a preliminary study, hence more studies are suggested to pin point the advantages and disadvantages of the considered models.

5. CONCLUSION

With free and timely availability of multi sensor data, crop monitoring is easier than ever. SAR and optical data can be used synergically for crop classification. Here, Sentinel 1 and Sentinel 2 data are processed and used together for crop classification. Various deep neural network models are utilized for crop classification using SAR and optical data. Three

dimensional convolutional neural network (3D CNN) models, long short-term memory neural network (LSTM) models and, convolutional long short-term memory neural networks (ConvLSTM) models are designed, utilized and evaluated in crop classification task. Performance evaluation is done on the basis of both classification accuracy and generalization performance. LSTM based model have shown superior performance than 3D CNNs and ConvLSTMs in both measures. This study also suggests for more comparative studies as there are numerous internal parameters at play and are changing as DNNs are evolving. Hence, more elaborative studies are required.

REFERENCES

- Anon. n.d. 'How to Use the SNAP API from Python - SNAP - Confluence'. Retrieved 15 August 2018a (<https://senbox.atlassian.net/wiki/spaces/SNAP/pages/19300362/How+to+use+the+SNAP+API+from+Python>).
- Anon. n.d. 'Keras Documentation'. Retrieved 15 August 2018b (<https://keras.io/>).
- Blaes, Xavier, Laurent Vanhalle, and Pierre Defourny. 2005. 'Efficiency of Crop Identification Based on Optical and SAR Image Time Series'. *Remote Sensing of Environment* 96(3–4):352–65. Retrieved 11 August 2018 (<https://www.sciencedirect.com/science/article/pii/S0034425705001045>).
- Gu, Jiuxiang et al. n.d. *Recent Advances in Convolutional Neural Networks*. Retrieved 15 August 2018 (<https://arxiv.org/pdf/1512.07108.pdf>).
- Löw, F., U. Michel, S. Dech, and C. Conrad. 2013. 'Impact of Feature Selection on the Accuracy and Spatial Uncertainty of Per-Field Crop Classification Using Support Vector Machines'. *ISPRS Journal of Photogrammetry and Remote Sensing* 85:102–19. Retrieved 11 August 2018 (<https://www.sciencedirect.com/science/article/pii/S0924271613001974>).
- Sainath, Tara N., Oriol Vinyals, Andrew Senior, and Hasim Sak. 2015. 'Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks'. Pp. 4580–84 in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. Retrieved 15 August 2018 (<http://ieeexplore.ieee.org/document/7178838/>).
- Wang, X. Y., Y. G. Guo, J. He, and L. T. Du. 2016. 'Fusion of HJ1B and ALOS PALSAR Data for Land Cover Classification Using Machine Learning Methods'. *International Journal of Applied Earth Observation and Geoinformation* 52:192–203.

Retrieved 11 August 2018
(<https://www.sciencedirect.com/science/article/pii/S0303243416300964>).

Xie, Chengjun et al. 2018. 'Multi-Level Learning Features for Automatic Classification of Field Crop Pests'. *Computers and Electronics in Agriculture* 152:233–41. Retrieved 11 August 2018
(<https://www.sciencedirect.com/science/article/pii/S0168169916308833>).

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. 'Deep Learning'. *Nature*. Retrieved
(<http://arxiv.org/abs/1603.05691>).

Zhu, Xiao Xiang et al. 2017. 'Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources'. *IEEE Geoscience and Remote Sensing Magazine* 5(4):8–36. Retrieved 13 August 2018
(<http://ieeexplore.ieee.org/document/8113128/>).