Cloud segmentation in Advanced Wide Field Sensor (AWiFS) data products using deep learning approach

Manaruchi Mohapatra, Prasun Kumar Gupta, Bhaskar Ramchandra Nikam*, Praveen Kumar Thakur
Indian Institute of Remote Sensing, Indian Space Research Organisation, Dehradun - 248 001, India
*Email: bhaskarnikam@iirs.gov.in

(Received: Nov 18, 2021; in final form Feb 28, 2022)

Abstract: Presence of cloud in optical remote sensing data hides the useful information and reduces the applicability of the data. Majority of operational techniques of extracting cloud cover from optical remote sensing data employ digital classification of individual pixels. These approaches ignore the spatio-temporal information about the cloud cover in the data and the fact that clouds are spatially continuous and highly dynamic entities. In traditional approaches, similar spectral properties of snow and cloud in shorter wavelength regions pose problems in accurate snow cover mapping and cloud masking. The present study proposes four encoder-decoder based convolutional neural networks (CNNs) for segmentation of Advanced Wide Field Sensor (AWiFS) optical data into four classes i.e. cloud, cloud shadow, snow and other features. The proposed CNNs have seven convolutional layers in encoding path and six convolutional layers in decoding path. Each CNN was tuned using simple grid search and trained with an average accuracy and loss of 0.96 and 0.02, respectively. The pixel-wise probability for each class was generated from the tuned CNNs using unseen data. The class assignment to each pixel was done by normalizing the probabilities from the CNN. For every pixel, the class having maximum normalized probability was said to be the class type of that particular pixel. The final output was compared with the outputs from a Random Forest (RF) Classifier and a self-digitized output. The deep learning model performed better than RF classifier, as the average accuracy values of 94% and 90% were achieved by the deep learning model and RF classifier, respectively. The proposed model can be used for cloud masking and snow cover mapping with higher accuracy and more robustness than other conventional methods over the AWiFS data.

Keywords: cloud segmentation, deep learning, encoder-decoder model, cloud masking, random forest classifier, convolutional neural networks

1. Introduction

Cloud along with cloud shadow in optical multispectral remote sensing images limits the applicability of the imagery causing problems in extraction of useful information and increase the error due to misclassification of features. Cloud cover causes the spatio-temporal discontinuity and hinders the application of time-series satellite images (Li et al., 2019). Cloud cover hinders the representation of actual surface features in the remote sensing data. In all applications of optical remote sensing, data with less percentage of cloud cover is preferred. Due to the high probability of presence of cloud in an optical data, automatic masking of cloud pixels in the data is considered as an important preprocessing step (Wu et al., 2018). Therefore, cloud detection and masking is one of the key problems in usage of optical images. Accurate identification and removal of clouds is necessary to reduce the negative impact of clouds on image applications. Another major problem is that cloud and snow has similar spectral reflectance, which makes segmentation of snow from cloud a difficult task. Snow and cloud have similar reflectance values in the lower wavelength region; but in higher wavelength regions (>1.5µm), cloud shows higher reflectance value as compared to snow. This property of cloud and snow has been traditionally exploited in order to separate snow and cloud in an optical imagery.

There has been growing interest in using Artificial Neural Networks, and specifically Convolutional Neural Networks, which forms the basis of Deep Learning models. These can perform efficient feature detection and are of much use in the field of Remote Sensing (Ma et al., 2019). Deep Learning models (networks) are composed of many layers that transform input data (e.g. images) to outputs (e.g., categories) while learning progressively the higher level features. The higher computational complexity that they involve is often ignored to achieve accurate results over large datasets. Image classification using deep learning began with AlexNet in 2012 and various advancements viz. ResNet, GoogleNet, etc. were available in open domain. Several experiments were conducted to use variants of these deep learning techniques for the purpose of cloud detection in remote sensing data sets, such as, Mateo-Garcia et al. (2017), Xie et al. (2017), Li et al. (2018), Tuia et al. (2018), Varshney et al. (2018), Zhang et al. (2018), Jeppesen et al. (2019) and Varshney et al. (2019). For creating image segments, a network is fed with an image and a corresponding set of pre-labeled pixels. Once the network learns attributes such as texture, tone and spatial correlation of the labeled pixels, it can classify the rest of the unlabeled pixels with this information. Such a trained network can then be used on an entirely new image, in order to classify it.

The current work aims to use the spectral information of visible, near infrared and shortwave infrared information of optical satellite image in order to segregate clouds from snow effectively. The purpose of this work is to build a robust neural network architecture especially designed for cloud, cloud shadow and snow detection and segmentation; in complex terrain and illumination conditions.

2. Data used

The dataset used in this study was of Advanced Wide Field Sensor (AWiFS) of Resourcesat – 2 satellite from the Indian Remote Sensing program. AWiFS acquire data in 4 wavelength ranges i.e. green (0.52 to 0.59 μm), red (0.62 to 0.68 μm), near infrared (0.77 to 0.86 μm) and short-wave infrared (1.55 to 1.7 μm) with spatial resolution 56m each (Table 1). The revisit period is 5 days and the radiometric resolution is 12 bits. The raster data of each scene comprised of around 17000 rows and 15000 columns. Due to the high spectral and temporal resolution, AWiFS data has been the pivot for various remote sensing applications such as land-use land-cover classification (Kandrika & Roy, 2008; Panigrahy et al., 2009; Haldar & Patnaik, 2010; Punia et al., 2011), domain specific studies in water resources (Kulkarni et al., 2006; Rajawat et al., 2007; Raju et al., 2008; Subramaniam et al., 2011; Karri et al., 2016) and disaster management (Bahuguna et al., 2008; Calle et al., 2008; Das et al., 2017). AWiFS data of northern part of India during mid-monsoon season was considered in the study as the presence of snow and cloud could be seen together. A representation of green band and SWIR band is shown in Figure 1 along with the marked portions of snow, cloud and other features to demonstrate the difference in spectral properties of the features in the two different bands.

The green band is useful to differentiate vegetation features from the snow and cloud features as snow and cloud features have high reflectance compared to other features. The short-wave infrared (SWIR) band is useful to differentiate snow from cloud and vegetation as snow has low reflectance in SWIR region. By using simple thresholding, the snow and cloud features can be separated from other features, heuristically, to get the primary mask for preparation of training data.

3. Methodology

The flowchart for general workflow of the proposed cloud segmentation method with three parts i.e., training data setup, deep learning model and model evaluation is visualized in Figure 2.

3.1 Training Data Setup

The accuracy of deep learning models depends upon the availability of good training samples. In this study, sample size of 512 × 512 is considered. The segmentation operation is carried out for four classes *viz.* cloud, cloud shadow, snow and other features. The image from the green channel (Band - 2) was classified into two different brightness value ranges separating cloud and snow from vegetation by visually considering the values of different features. A mask was generated by considering snow and cloud range as true and other range as false. The image from the SWIR channel (Band - 5) was classified into two different brightness value ranges separating snow from cloud and other features. A mask was generated by considering the snow range as true and the other as false. The common pixels from the two masks were extracted as cloud features. Snow and other feature masks were generated by removing the cloud pixels from the first and second masks, respectively. The above operations gave the preliminary segmentation of different features in the image. The preliminary segmentation output was then refined manually and the final segmentation images were generated. The segmentation mask had values 1, 2, 3 and 4 for features cloud, cloud shadow, snow and other features, respectively. The above process was performed for five different scenes of AWiFS.

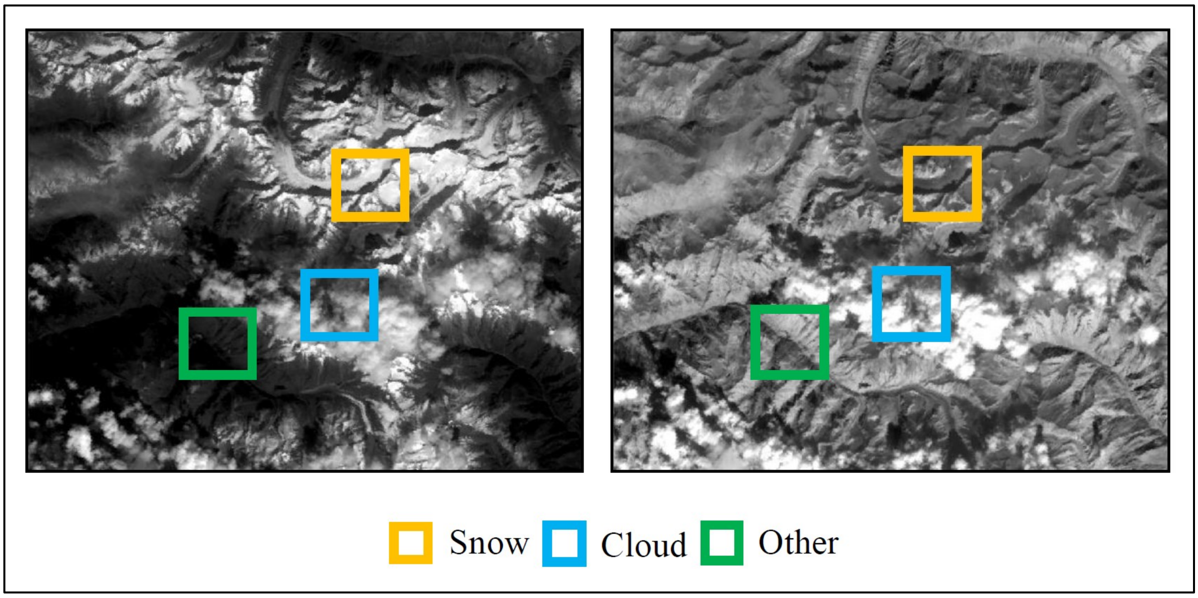

Figure 1. Green band (on left) and SWIR band (on right) with difference in spectral properties of snow, cloud and other features

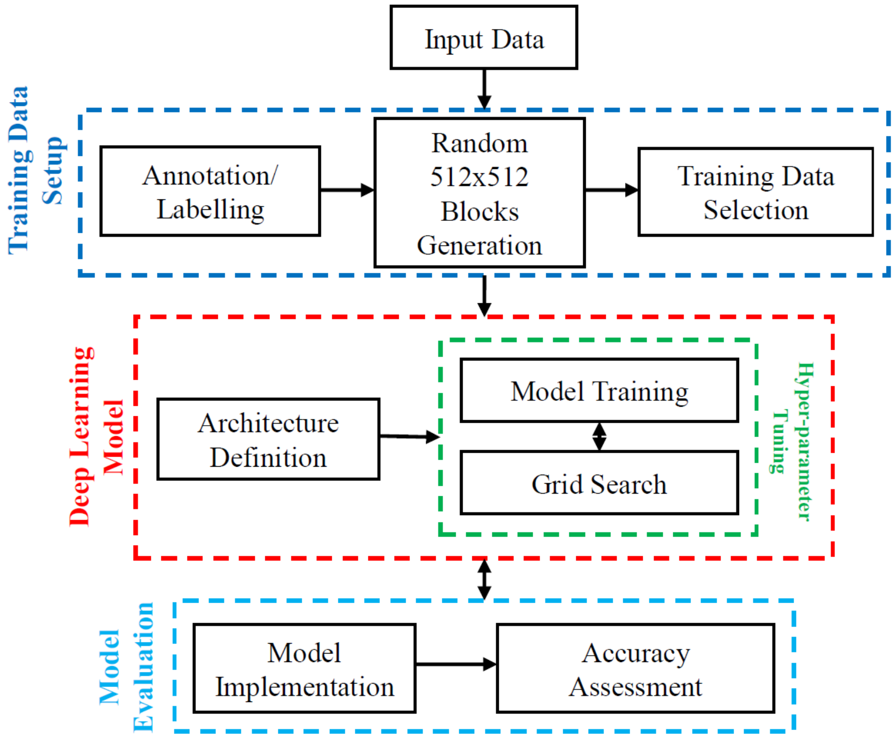

Figure 2. Flowchart of the methodology

Table 1. List of AWiFS images used in the present study

Sr. No	Satellite/Sensor	Row/Path	Date of Pass
1	Resourcesat-2 / AWiFS	97/48	09/10/2014
2	Resourcesat-2 / AWiFS	97/48	20/12/2014
3	Resourcesat-2 / AWiFS	97/48	06/02/2015
4	Resourcesat-2 / AWiFS	97/48	17/08/2015
5	Resourcesat-2 / AWiFS	97/48	28/10/2015

Random 512 × 512 pixel blocks (or samples) were clipped from all the images from the composite and the segmentation mask. Random samples having all the four classes were only considered for the training set. Finally, the training set contained 30% samples having maximum pixels as cloud, 30% samples having maximum pixels as snow, 20% samples having maximum pixels as cloud shadow and 20% samples having maximum pixels as other features. 154 numbers of training samples were hence selected, out of which 70% (108 blocks) were used for model training and remaining 30% (46 blocks) were used for accuracy assessment of the model outputs. Four different masks for four different classes were generated for each training sample by assigning the class value as true and the other class values as false. A sample training composite along with the masks is shown in Figure 3. The 4-band composite was used as the feature set and the four generated masks were used as the label dataset.

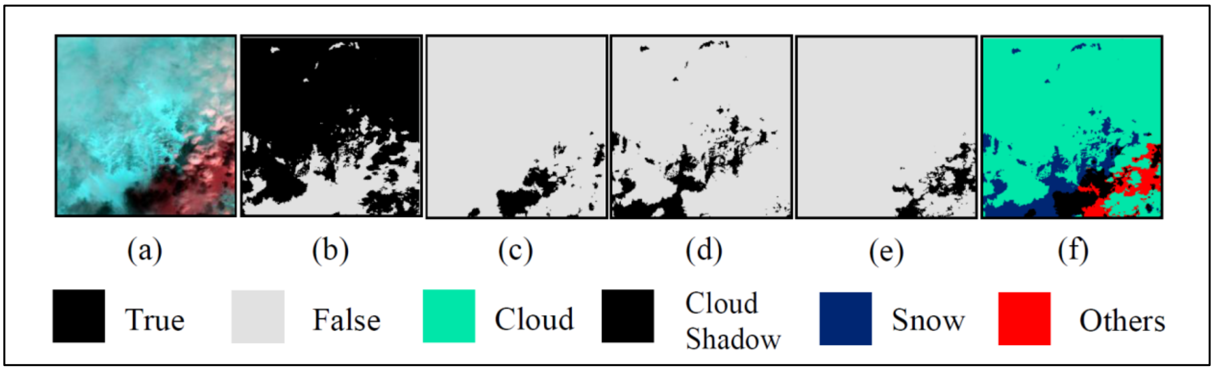

Figure 3. Sample training data set (a) FCC, (b) Mask for cloud, (c) Mask for cloud shadow, (d) Mask for snow, (e) Mask for other features, (f) Mask combining all classes

3.2 Model Architecture

Deep Learning (Goodfellow et al., 2016) has been popular for CNN based image classification tasks. Recently, U-Net (Ronneberger et al. 2015) based remote sensing image classification applications have proved to be better than other algorithms. Current studies such as forest type mapping (Wagner et al., 2019), building non-building mapping (Huang et al., 2018) and cloud snow mapping (Varshney et al., 2019) have provided impetus on use of customised U-Net architecture in the field of image segmentation for mapping or identification of a specific type of feature in the remote sensing data. Image segmentation is a special deep learning problem as the dimensions of the input data is equal to the dimension of output segmented data with the depth being unity. Among many image segmentation algorithms, encoder-decoder (U-Net) based models have been widely used in the last decade. The primary aim of encoder-decoder model was to take an input and provide an output with only the important features preserved. The encoder part of the model divides the input into smaller chunks with only important features and the decoder model can recreate the original input using these chunks with high accuracy. Image segmentation is conceptually similar to encoder-decoder based models. The multi-band input images are reduced to smaller chunks using encoders and the segmentation map is recreated using the decoder part preserving only the important features that can identify a specific type of feature. The proposed cloud segmentation model is based on the U-Net, an encoder-decoder model. This model has been widely used for medical image segmentation due to its ability to provide better accuracy in relatively less training data as compared to other segmentation deep learning models (Wagner et al., 2019).

In the current case, as the AWiFS input image contains 4 spectral 4 bands and the segmented image would contain a single band, the input image must be reduced to represent information as in a single band image. The reduction process should be such that for each feature, the band which represents that feature relatively well should only be considered, as seen in case of methods such as Principle Component Analysis. The reduction process in this context would be similar to encoding where the input data is separated into smaller chunks with increased width by the application of successive convolutions using CNNs. The reduced chunks by the model can be compared to the training label in order to train the model by adjusting the weights and biases using a proper optimizer. In the process of encoding, the spatial information of the features is lost because of the successive reduction in dimension. To build the output segmented image with proper dimensions, a decoder model recreates the output using these smaller chunks in successive transposed convolutions using CNNs. A schematic diagram of the proposed model is shown in Figure 4.

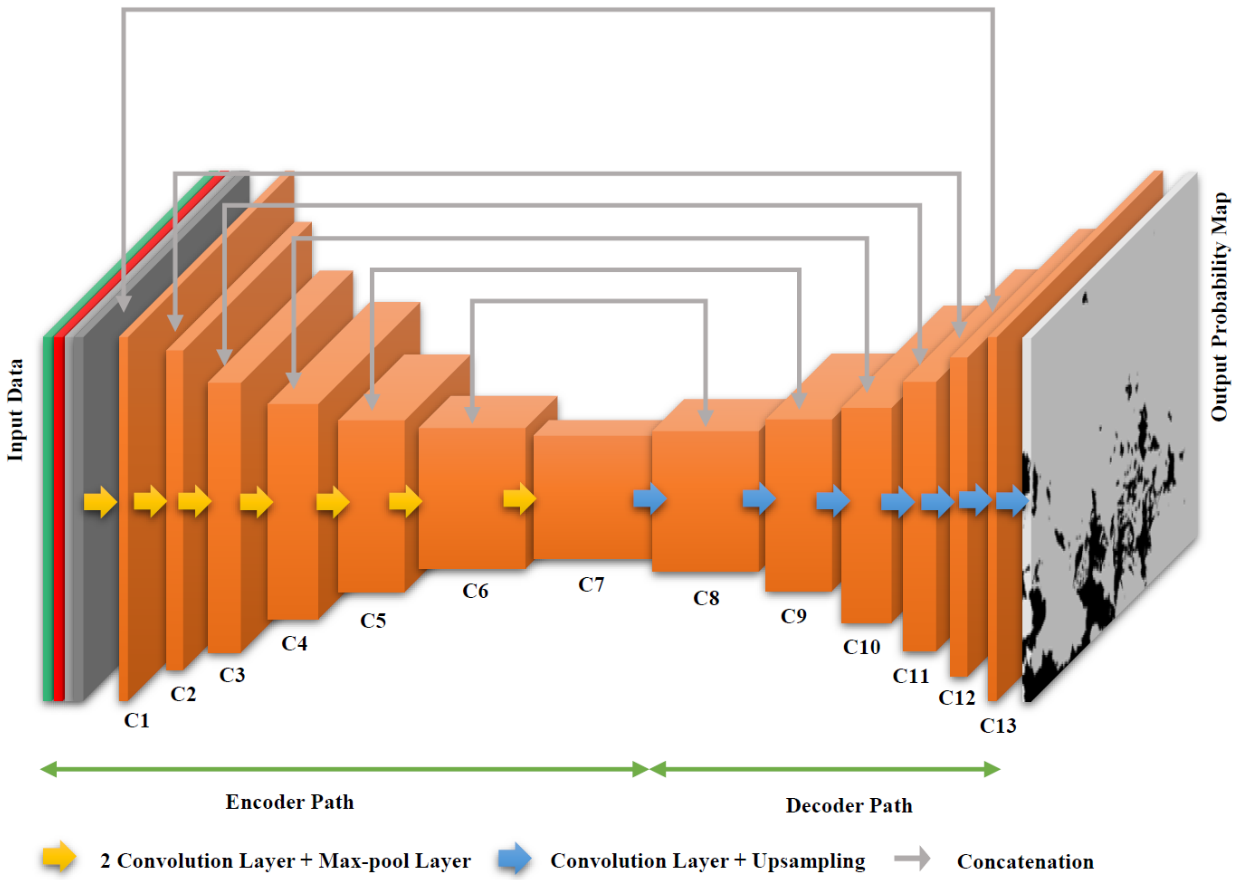

Figure 4. Schematic diagram of proposed network structure

The encoder path gradually reduced the size of the image while the depth was gradually increased starting from 512×512×4 to 8×8×1024. By using the encoder path, the network learnt the "WHAT" information in the image, however it lost the "WHERE" information. The decoder path gradually increased the size of the image while the depth was gradually reduced starting from 8×8×1024 to 512×512×1. By using the decoder path, the network recovered the "WHERE" information by gradual application of transposed convolutions and up-sampling. To get precise locations, at every step of the decoder path, skip connections were used by concatenating the output of the transposed convolution layers with the feature map from the encoder at the same level. After every concatenation, two consecutive regular convolutions were applied again so that the network could learn to assemble a more precise output. The output would be probabilities of each pixel belonging to a particular class. On a high level, the network has the relationship: Input (512×512×4) →Encoder → 8×8×1024 → Decoder → Output (512×512×1). The entire architecture, as represented in Figure 5 was written in Python 3.7 using Tensorflow version 1.1 which is the industry standard for deep learning models. With the following system configuration, training each model took ~18 minutes and prediction for each class took ~15 seconds each.

GPU: 12GB GDDR5 K80
CPU: Single core Xeon Processors @2.3Ghz RAM: ~20 GB

The original scene having dimensions of around 17000×15000 were then divided into smaller images of dimension 512×512 and the deep learning model was used to generate a segmentation map. Every segmentation output was stitched back to the original dimension to obtain the final result. Accuracy Assessment was conducted in order to evaluate effectiveness and the capability of the proposed methodology to correctly classify different classes. The model architecture described above was deployed separately for four classes (cloud, cloud shadow, snow and other features); hereafter referred to as four models

The four models, each for each class were first structured and the training (70%) – testing (30%) dataset for each class was prepared. In order to regularize the model and to avoid the over-fitting condition, dropouts and batch normalization processes were added to the structure. In order to optimize the model and to get better accuracy, the hyper-parameters such as kernel size, activation function, optimization algorithm and dropout rate were tuned by implementing a simple grid search. Grid search works by implementing a defined set of hyper-parameter combinations and obtaining the parameter combination having maximum efficiency in an experimental setup.

The parameters after the tuning process were implemented on the deep learning models and were trained for 200 epochs each using the composite as feature data and the class mask as the label data. The accuracy and root-mean-square loss were calculated for each epoch taking 30% of the training set as validation set. The output probability map generated was then combined using pixel-by-pixel approach. For each pixel, the probability of four classes was normalized, and the class which had the maximum probability was assigned to the pixel. The model implementation workflow used in this work is shown in Figure 6.

.

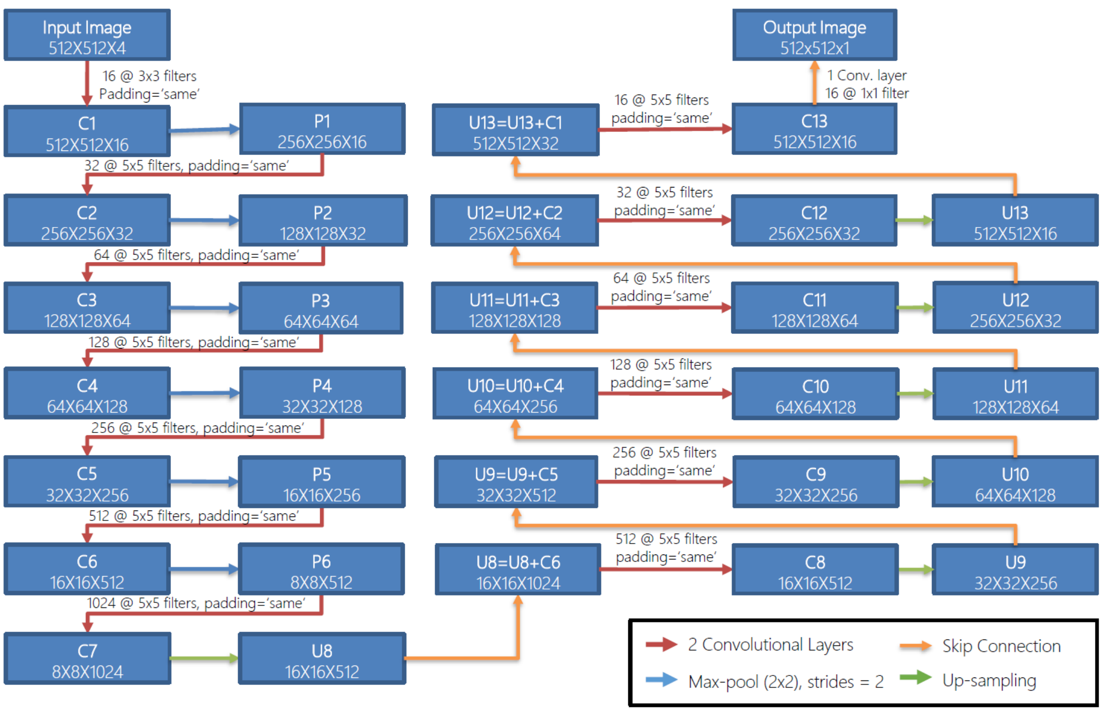

Figure 5. Final model architecture

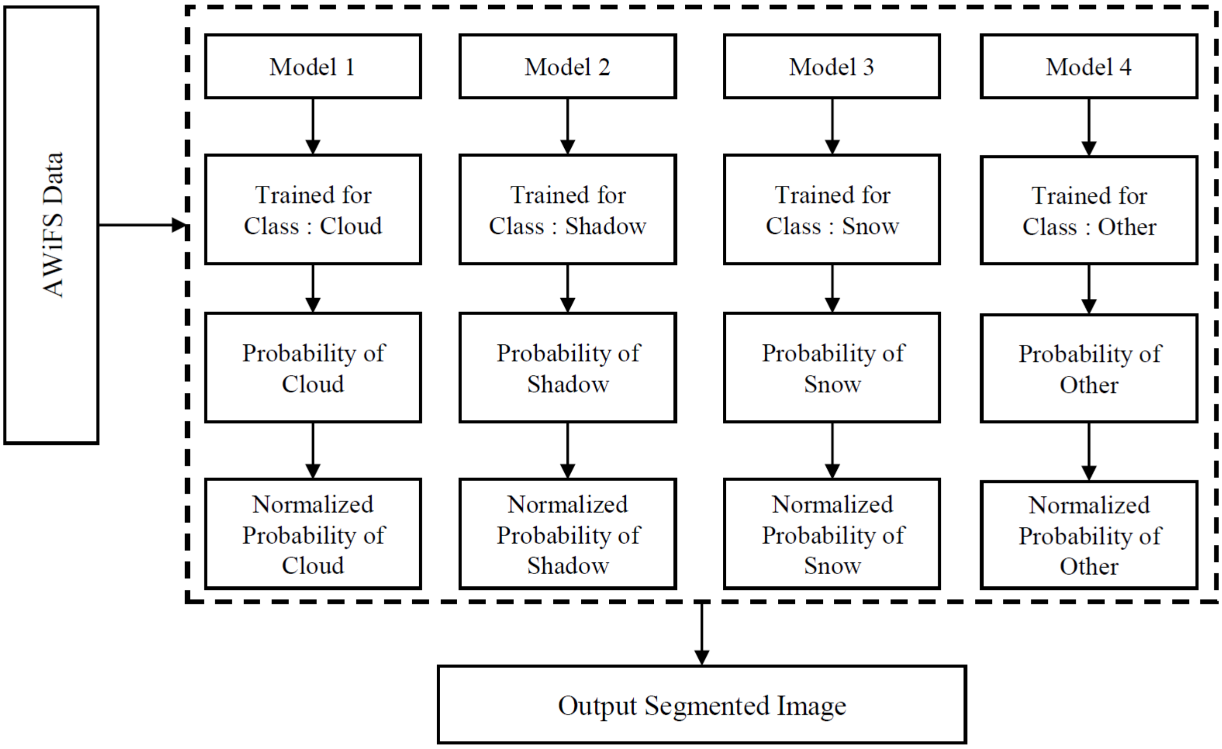

Figure 6. Model Implementation Workflow

3.3 Random Forest

The Random Forest (RF) based semantic image segmentation was also implemented for the sake of comparison with the proposed deep learning model. RF models are based on decision tree algorithms with improvements made to reduce the errors due to over-fitting. RF models introduce training time randomness into the trees and outputs of such randomized trees are combined into a single classifier. These randomness essentially works as a negative factor for model convergence and regularizes the model to provide accurate outputs. Schroff et al. (2008), Bosch et al. (2007) and Yin et al. (2007) demonstrated that RF generated lower test errors as compared to conventional decision trees and other image segmentation methods such as Support Vector Machines (SVMs). Moreover, RF models are considered straightforward and efficient owing to its sampling approach (Drönner et al., 2018).

Following the same methods for training of the deep learning model, same training and testing data were used for the RF model. The outputs of the RF model were subjected to evaluation in order to tune the model.

3.4 Accuracy Assessment

In order to compare the model performance with the existing methods of classification, a random forest model was trained using the same training data which was used to train the deep learning model. To assess the accuracy of the model, both random forest and deep learning model was implemented to generate classification outputs on an unseen data. The unseen data was then hand digitized and was compared with the outputs from both random forest and deep learning models. Overall accuracy was calculated for both the cases.

4. Results

The initial model was iteratively trained using different combinations of defined hyper-parameters and the combination that showed highest performance efficiency is chosen. The different categories of hyper-parameters, which were explored and their efficiency is shown in Figure 7.

The results from grid search were considered and the existing models were fine tuned to form final models which were trained separately for each class. During the training, all the models were evaluated and the performance of the model was monitored using the Tensorboard interface. In this study, model accuracy and loss were used as the performance indices for training the deep learning models which are depicted in Figure 8. The final models were implemented on the test dataset to find probability map of each class. The probability values were normalized and the class having maximum probability value was selected in the final segmentation map. Figure 9 depicts the output obtained by using four models on the given input data. Accuracy assessment was performed using unseen data samples where each sample had thin and thick clouds over snow and land features so as to see the model performance to detect the clouds in the extreme conditions possible. The results of the assessment is shown in Table - 2. The comparative analysis of results of deep learning model with self-digitized reference data and with the output of RF are shown in Figure 10 and Table-3.

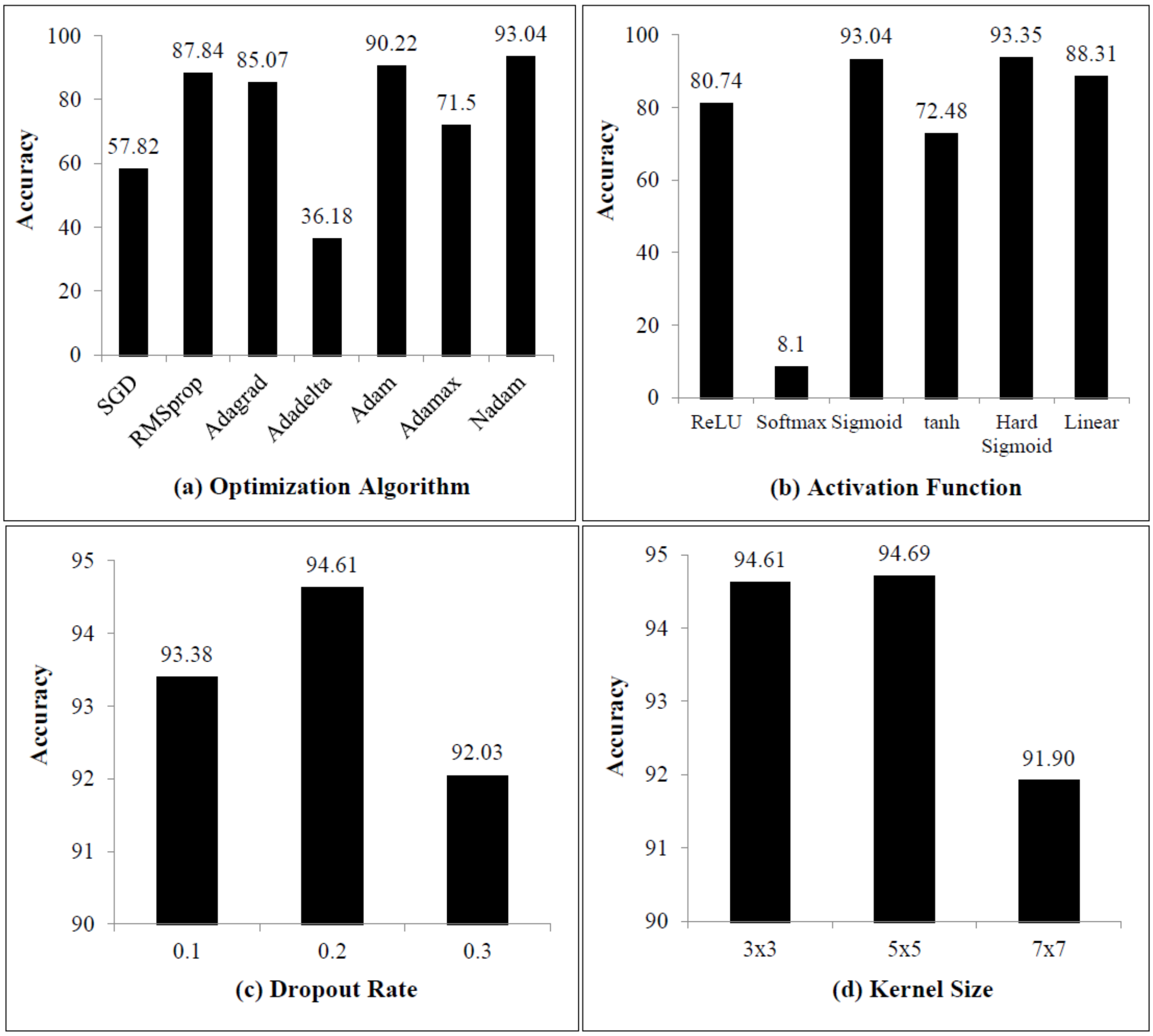

Figure 7. Simple grid search results for finding optimum (a) optimization algorithm, (b) activation function, (c) dropout rate and (d) kernel size

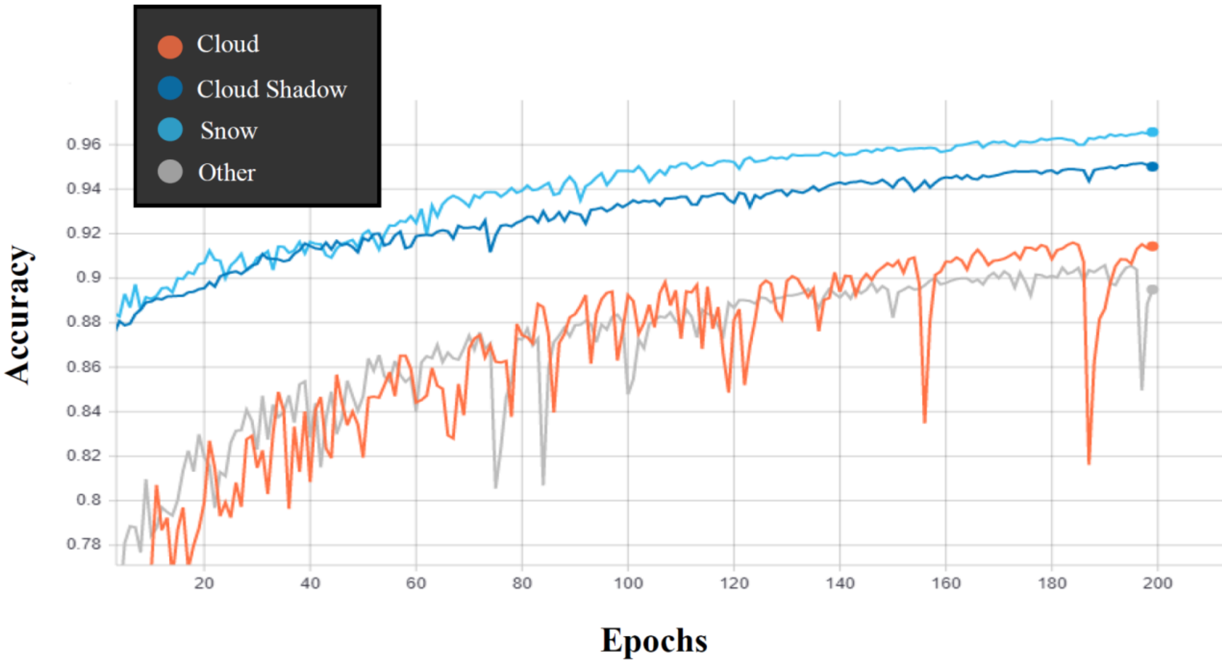

Figure 8. Model Evaluation Indices: Accuracy (on top), Loss (on bottom)

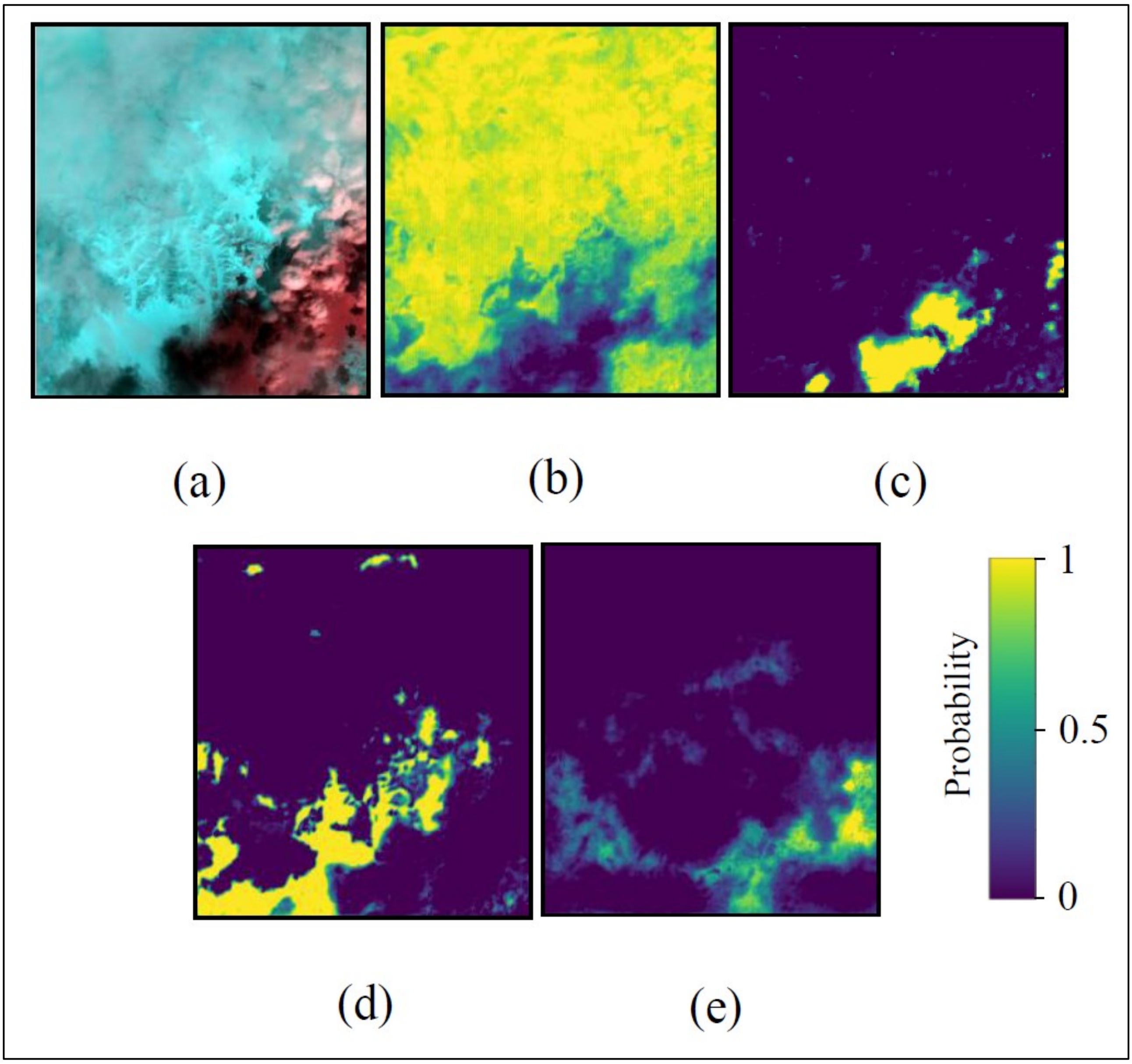

Figure 9. (a) Input test data, (b) to (e) Output probability map of cloud, cloud shadow, snow and other features respectively (brighter color represents higher value)

Table 2. Confusion matrix for the Deep Learning Model

	Cloud	**Cloud Shadow**	**Snow**	**Other**
Cloud	**200078**	56	1425	914
Cloud Shadow	1244	**12797**	1139	1142
Snow	4602	273	**24963**	6
Other	4371	515	15	**8604**

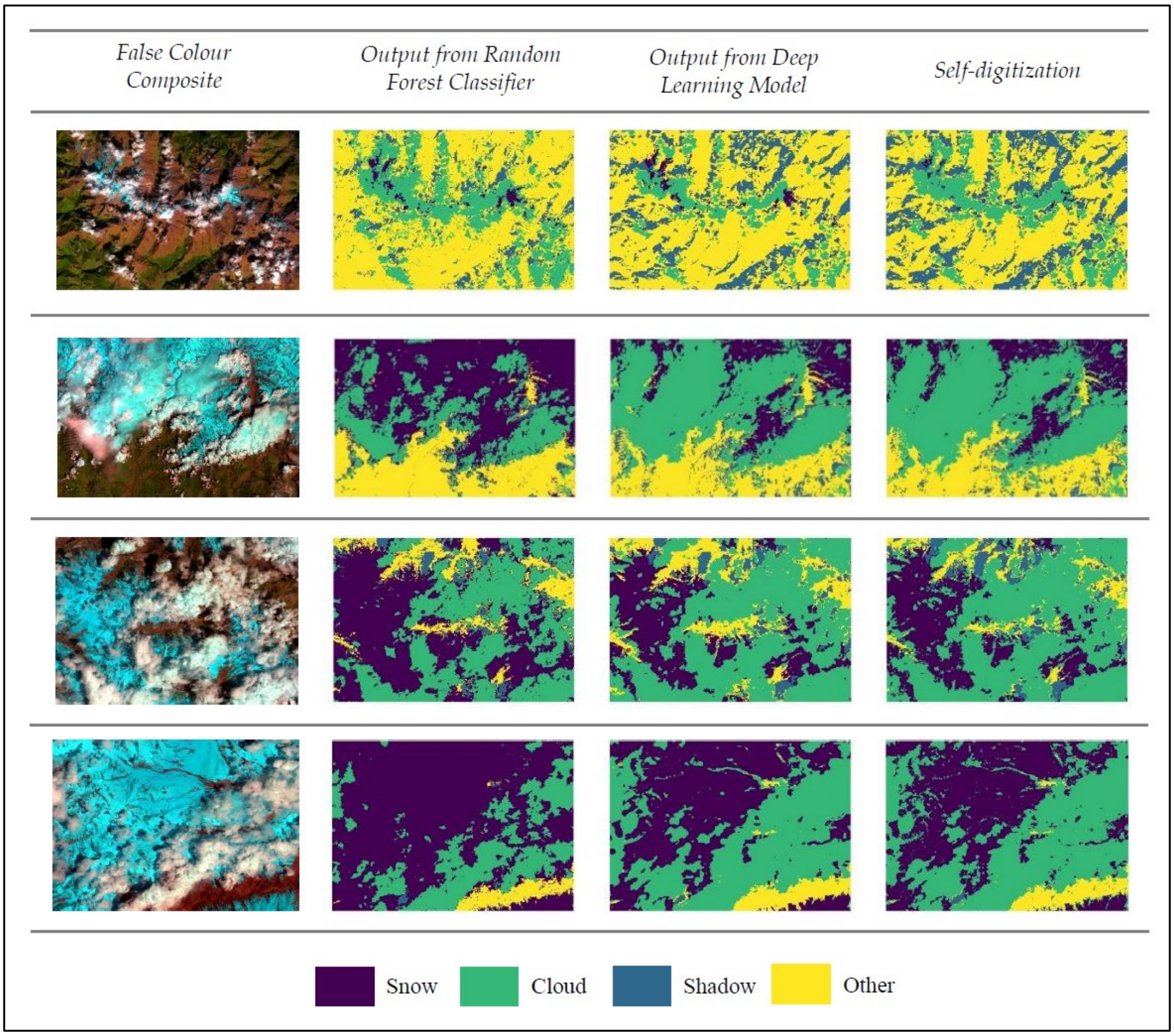

Figure 10. Sample accuracy assessment results

Table 3. Overall Accuracy of Random Forest Classifier and Deep Learning Model

Class	Overall Accuracy (in percentage)	
	Random Forest Classifier	Deep learning Model
Snow	87.49	93.06
Cloud	88.95	93.29
Cloud shadow	93.46	94.41
Other features	93.16	95.00

5. Discussions

The proposed deep learning model was implemented using Tensorflow 2.x and the input data processing was performed using Python scripts. The training and testing dataset were annotated using AWiFS multispectral data by utilizing the spectral differences between the required feature classes. Even though such thresholding approach could separate the snow and cloud cover features, in most cases the approach falls short due to similar spectral behavior of snow and cloud cover and requires human intervention for accurate segmentation. The current work aimed towards development of a deep learning image segmentation model based on the U-Net architecture which could automate the segmentation process with optimum accuracy. U-Net architecture was considered due to its capability to perform optimally even with smaller training datasets. The model was structured and tuned for different combinations of hyper-parameters so as to prepare a model which was better fit to the given dataset. The tuning ensured that the best parameter set was used to structure the model. However, measures were considered while training of the model to prevent model overfitting such as regularization layers and comparison of test and train accuracy. Four different models were trained for the required features and the output of each model was combined using a simple probability normalization approach where the prior probability of the pixels being any particular class was multiplied to the likelihood to calculate the posterior probability. The final model output showed an overall accuracy of around 94%. As a model for comparison, a random forest model was trained using the same training dataset and the same samples were used for testing the model performance. For the same unseen samples, the comparison of the overall accuracies from the deep learning model and RF model showed that the deep learning model was able to perform image segmentation with higher accuracy. The accuracy of the deep learning model was also found to perform consistently well with different case scenarios. The deep learning models were found to overcome the need of the manual interventions required in heuristic approaches. Along with providing a cloud mask, this methodology can also be helpful to the organizations that provide earth observation data to estimate and specify the percentage cloud cover present in the metadata.

As the deep learning model was implemented in Tensorflow, the model was scalable for larger training datasets. However, one drawback of the model can be the fixed size of input files. Due to the fixed input size, the original AWiFS scenes has to be divided into smaller arrays with dimensions as required for the model to be passed through the model. The outputs generated from the model has to be stitched to the original dimension. This could significantly increase the time required for segmentation of an entire scene. However, the time could be reduced by changing the input file dimension of the model to a larger size. The time consumption can be further reduced by using a system with higher computational power such as High Performance Clusters (HPCs). The number of blocks/training datasets used in the present study, were found to be sufficient for the current study area, however, for global application of this model the number of training samples may need to be increased.

In the present study the four single class classifiers have been used; however the effect of using a single multi-class classifier for task can be explored as a future study. The accuracy assessment of the present model is done using self-digitized reference datasets and classified outputs of RF model. The performance of present deep learning model was observed to be satisfactory with respect of RF. However, cross-validation can also be performed using classified outputs of different ML algorithms and other techniques such as semantic segmentation, object-based image classification, textural classification, etc. implemented on the same study area.

6. Conclusions

The study proposed a deep learning network model for multi-class segmentation of AWiFS scenes into four classes i.e. cloud, cloud shadow, snow and other features in order to replace the manual, conventional processes of cloud masking. The models were structured and the best hyper-parameter combinations were chosen using simple grid search. The accuracy assessment of the combined segmentation results generated by the four models showed that the deep learning model performs better than the random forest classifier trained on the same dataset with an overall accuracy of around 94%. The proposed model could better identify the cloud features from snow features. In multiple cases, Random forest classifier failed to detect the thin clouds whereas the deep learning model could correctly detect cloud in such cases. Each model took around 18 minutes to train and around 15 seconds to predict on an unseen data. The output from the model can be further used to generate cloud masks and snow products. The deep learning model can also be extended to be used in data from other sensors such as Linear Imaging Self Scanning (LISS) – III. Although the deep learning model lacked in some measures, such as misclassifying thin clouds, this could be rectified by increasing the training sample, deeper encoder-decoder network and higher computational power.

References

Bahuguna A., S. Nayak and D. Roy (2008). Impact of the tsunami and earthquake of 26th December 2004 on the vital coastal ecosystems of the Andaman and Nicobar Islands assessed using RESOURCESAT AWiFS data. *International Journal of Applied Earth Observation and Geoinformation*, *10*(2), 229–237. https://doi.org/10.1016/j.jag.2008.02.010

Bosch A., A. Zisserman and X. Munoz, (2007). "Image Classification using Random Forests and Ferns," *IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, 2007, pp. 1-8, doi: 10.1109/ICCV.2007.4409066.

Calle A., F. Gonzalez-Alonso and S. Merino de Miguel (2008). Validation of active forest fires detected by MSG-SEVIRI by means of MODIS hot spots and AWiFS images. *International Journal of Remote Sensing*, *29*(12), 3407–3415. https://doi.org/10.1080/01431160701596164

Das P. K., B. Sahay, M. V. R. Seshasai and D. Dutta (2017). Generation of improved surface moisture information using angle-based drought index derived from Resourcesat-2 AWiFS for Haryana state, India. *Geomatics, Natural Hazards and Risk*, *8*(2), 271–281. https://doi.org/10.1080/19475705.2016.1201149

Drönner J., N. Korfhage, S. Egli, M. Mühling, B. Thies, J. Bendix, B. Freisleben and B. Seeger (2018). Fast cloud segmentation using convolutional neural networks. *Remote Sensing*, *10*(11), 1–24. https://doi.org/10.3390/rs10111782

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

Haldar D. and C. Patnaik (2010). Synergistic use of multi-temporal Radarsat SAR and AWiFS data for Rabi rice identification. *Journal of the Indian Society of*

Remote Sensing, *38*(1), 153–160. https://doi.org/10.1007/s12524-010-0006-x

Huang B., K. Lu, N. Audeberr, A. Khalel, Y. Tarabalka, J. Malof, A. Boulch, B. Le Saux, L. Collins, K. Bradbury and S. Lefèvre (2018). Large-scale semantic classification: outcome of the first year of inria aerial image labeling benchmark. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS),* 6947-6950. https://doi.org/10.1109/IGARSS.2018.8518525

Jeppesen J. H., R. H. Jacobsen, F. Inceoglu and T.S. Toftegaard (2019). A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*, *229*(May), 247–259. https://doi.org/10.1016/j.rse.2019.03.039

Kandrika S. and P. S. Roy (2008). Land use land cover classification of Orissa using multi-temporal IRS-P6 awifs data: A decision tree approach. *International Journal of Applied Earth Observation and Geoinformation*, *10*(2), 186–193. https://doi.org/10.1016/j.jag.2007.10.003

Karri S., B. Gharai, S. V. S. Sai Krishna and P.V. N. Rao (2016). Impact of AWiFS derived land use land cover on simulation of heavy rainfall. *Remote Sensing and Modeling of the Atmosphere, Oceans, and Interactions VI*, *9882*, 98821M. https://doi.org/10.1117/12.2223627

Kulkarni A. V., S. K. Singh, P. Mathur and V. D. Mishra (2006). Algorithm to monitor snow cover using AWiFS data of RESOURCESAT-1 for the Himalayan region. *International Journal of Remote Sensing*, *27*(12), 2449–2457. https://doi.org/10.1080/01431160500497820

Li Z., H. Shen, Q. Cheng, Y. Liu, S. You and Z. He (2018). Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 197-212.

Li Z., H. Shen, Q. Cheng, W. Li and L. Zhang (2019).Thick Cloud Removal in High-Resolution Satellite Images Using Stepwise Radiometric Adjustment and Residual Correction. Remote Sens. 2019, 11, 1925.

Ma L., Y. Liu, X. Zhang, Y. Ye, G. Yin and B. A. Johnson (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *152*(November 2018), 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Mateo-Garcia G., L. Gomez-Chova and G. Camps-Valls (2017). Convolutional neural networks for multispectral image cloud masking. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2017-July*, 2255–2258. https://doi.org/10.1109/IGARSS.2017.8127438

Panigrahy R. K., S. S. Ray and S. Panigrahy (2009). Study on the utility of irs-p6 AWiFS SWIR Band for crop discrimination and classification. *Journal of the Indian Society of Remote Sensing*, *37*(2), 325–333. https://doi.org/10.1007/s12524-009-0026-6

Punia M., P. K . Joshi and M.C. Porwal (2011). Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data. *Expert Systems with Applications*, *38*(5), 5577–5583. https://doi.org/10.1016/j.eswa.2010.10.078

Rajawat A. S., M. Gupta, B. C. Acharya and S. Nayak (2007). Impact of new mouth opening on morphology and water quality of the Chilika Lagoon - A study based on Resourcesat-1 LISS-III and AWiFS and IRS-1D LISS-III data. *International Journal of Remote Sensing*, *28*(5), 905–923. https://doi.org/10.1080/01431160600904949

Raju P. V., M. V. R. Sesha Sai and P. S. Roy (2008). In-season time series analysis of Resourcesat-1 AWiFS data for estimating irrigation water requirement. *International Journal of Applied Earth Observation and Geoinformation*, *10*(2), 220–228. https://doi.org/10.1016/j.jag.2008.01.004

Ronneberger O., P. Fischer and T. Brox (2015). U-Net: convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, http://arxiv.org/abs/1505.04597

Schroff F., A. Criminisi and A. Zisserman (2008). Object class segmentation using random forests. *BMVC 2008 - Proceedings of the British Machine Vision Conference 2008*. https://doi.org/10.5244/C.22.54

Subramaniam S., A. V. Suresh Babu and P. S. Roy (2011). Automated Water Spread Mapping Using ResourceSat-1 AWiFS Data for Water Bodies Information System. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *4*(1), 205–215. https://doi.org/10.1109/JSTARS.2010.2085032

Tuia D., B. Kellenberger, A Pérez-Suay and G. Camps-Valls (2018). A deep network approach to multitemporal cloud detection. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2018-July*, 4351–4354. https://doi.org/10.1109/IGARSS.2018.8517312

Varshney D., P. K. Gupta, C. Persello and B.R. Nikam (2018). Snow and Cloud Discrimination using Convolutional Neural Networks. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. IV-5. 59-63. 10.5194/isprs-annals-IV-5-59-2018.

Varshney D., P.K. Gupta, C. Persello and B.R. Nikam (2019). Deep convolutional networks for cloud detection using resourcesat-2 data. *IEEE International Geoscience and Remote Sensing Symposium* (*IGARSS*), 9851-9854. https://doi.org/10.1109/IGARSS.2019.8898628

Wagner F. H., A. Sanchez, Y. Tarabalka, R. G. Lotte, M. P. Ferreira, M. P. M. Aidar, E. Gloor, O. L. Phillips and L. E. O. C. Aragão (2019). Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation*, 1–16. https://doi.org/10.1002/rse2.111

Wu W., J. Luo, X. Hu, H. Yang and Y. Yang (2018). A Thin-Cloud Mask Method for Remote Sensing Images Based on Sparse Dark Pixel Region Detection. *Remote Sensing*, *10*(4). https://doi.org/10.3390/rs10040617

Xie F., M. Shi, Z. Shi, J. Yin and D. Zhao (2017). Multilevel cloud detection in remote sensing images based on deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(8), 3631–3640. https://doi.org/10.1109/JSTARS.2017.2686488

Yin P., A. Criminisi, J. Winn and I. Essa (2007). Tree-based Classifiers for Bilayer Video Segmentation. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. https://doi.org/10.1109/CVPR.2007.383008

Zhang Z., A. Iwasaki, G. Xu and J. Song (2018). Small Satellite Cloud Detection Based On Deep Learning and Image Compression. *Preprints*, *February*, 1–12. https://doi.org/10.20944/preprints201802.0103.v1