A Multi-Feature Fusion Deep Convolutional Network based on A Coarse-Fine Structure for Cloud Detection

Dorcas Gicuku Mwigereri
School of Computing and Information Technology
Jomo Kenyatta university of Agriculture and Technology
Nairobi,Kenya
dorcausgicuku@gmail.com

Lawrence Nderu
School of Computing and Information Technology
Jomo Kenyatta university of Agriculture and Technology
Nairobi,Kenya
lawrence_nderu2@live.com

Tobias Mwalili
School of Computing and Information Technology
Jomo Kenyatta university of Agriculture and Technology
Nairobi,Kenya
mwalili@gmail.com

***Abstract*— Accurate detection of the cloud in remote sensing analytics is considered an essential task in remote sensing imagery with various spectral, temporal and spatial information. As a result we propose a multi-feature fusion deep convolutional neural networks for analysis of remote sensing satellite images so as to detect clouds, which are the region of interest. To ensure the algorithm was trained with data acquired from multiple satellites, the Landsat 7 ETM, Landsat 8 OLI/TIRS and Gaofen-1 wide field view datasets were used. The experimental results obtained showed that the proposed model gave accuracy, precision and recall measures of 95.2%, 89% and 89.9%respectively. The developed algorithm posted consistent and accurate results for cloud detection using satellite images that had clouds of different types and those obtained over different land surfaces that contain other objects in the images.**

Keywords*— *Deep convolutional neural networks, multi-feature fusion, remote sensing satellite imagery.

I . INTRODUCTION

Presence of clouds is seen to pose a challenge in the process of extracting information of the surface and/or atmosphere in remote sensing satellites in addition to affecting the amount of radiation a surface can receive[1]. Accurate detection of the cloud in remote sensing analytics has so far been seen to be a challenging task due to the various shapes the clouds may take in addition to having different ground objects captured in the satellite images[2]. To identify clouds given a satellite image, three approaches have so far been evaluated, that is, threshold based approaches which evaluate the reflectance and brightness across a given channel so as to detect presence of clouds, machine learning algorithms such as support vector machine(SVM)[3], artificial neural networks (ANN)[4] and random forest[5] that learn from handcrafted features so as to detect presence of clouds in an image and deep-learning technique[6] that automatically learn high complex features given a training dataset.

[7]reviewed literature on cloud detection using remote sensing satellite imagery from 2014 -2018.In their work, it was reported that most researchers explored a variety of cloud detection forms such as Cloud/No Cloud, Thin Cloud/ Thick Cloud and Snow/Cloud using threshold based techniques, deep learning algorithms and machine learning algorithms such as ANN, decision trees, random forests and Bayesian classification. Threshold based techniques were observed to perform differently for different climatic conditions or areas with different surface type thus making them have poor universality while machine learning algorithms that used handcrafted features were observed to be dependent on the feature selection process as different people have different understanding of clouds and their features. As a result, deep learning algorithms became more appealing due to their capability of automatically extracting highly complex features based on the spectral, temporal and spatial information provided in the training dataset.

Depending on the task being performed, deep learning techniques for cloud detection have been categorized to three major categories that is, patch based deep learning approach which uses an image patch as an input and gives a labelled output indicating whether the image is cloudy or not cloudy, region based deep learning techniques that segment an input image to different regions that are labelled using a pre-trained network and a pixel level deep learning technique that takes a fixed size input image and trains the model to output a pixel level labels that have the same size as the input image.

In this work, a multi-feature fusion deep convolutional neural network (MFF – DCNN) to predict presence of cloud given a remote sensing satellite image as an input is proposed. A single set feature, created by fusing the spectral, temporal and spatial features, is used to train the developed model for cloud detection.

II . RELATED WORK

[8] Developed a CNN algorithm for cloud detection based on a residual network (ReSNet) architecture which was crafted from the u-net architecture by adding a clipping layer, batch normalization and halving the depth of feature maps. The dataset obtained from the NASA Landsat 8 satellite consisted of five spectral band combinations that is the red/green/blue/infrared (RGBI) band, the red/green/blue (RGB) and the green band alone. In their work, they noted that the develop CNN model posted good results for semantic segmentation in addition to improving the performance levels and reducing the time it took to train the model by reducing the requirements during the preprocessing phase. To improve on the performance measure of the CNN algorithm for cloud detection, they stated the need for a method that would fully incorporate the spectral, spatial and temporal dimensions.

[9] Proposed a technique of detecting clouds that was based on cloud segmentation by fusing multi-scale convolutional features (MSCF) with aim of improving on the accuracy of the convolutional neural network (CNN) for object detection especially when using a multispectral image that contains the visible and infrared bands only. The proposed deep learning technique was based on fully convolutional network (FCN) for pixel-to-pixel semantic segmentation and Semantic Segmentation (SegNet) architecture which was built as a convolutional encoder-decoder for semantic segmentation of the pixel values. To train and evaluate the proposed model, images from Gaofen-1 WFV satellite were used and the performance posted by the developed model compared with two other techniques, that is, a multi-feature combined method (MFC) and a deep convolutional network (DCN).The obtained results shows that the MSCN model performs better with an accuracy level of 97.85% compared to the MFC model that posted an accuracy of 96.80%. Additionally, the MSCF was seen to keep details of the cloud boundaries on the cloud mask produced. For future studies, they recommended use of cloud images obtained from different satellite imageries to investigate if their suggested model can be generalizable on the different datasets.

[10]proposed a two-step deep learning technique for cloud detection on remote sensing satellite imagery. The first process involved use of a feature concatenation network which would obtain the cloud probability map from deep the convolutional neural network while the second process involved extraction of multilevel structural features using a multi-window guided filtering so as to refine the cloud mask. To validate the proposed model, the 502 Gaefen-1 WFV cloud images collected from May 2013 to December 2016 and obtained from different global regions, was used. To evaluate the performance measure of the model, the accuracy, the Intersection-Over-Unions(IOU) , Hansens Kuipers Discriminant(HK), False Alarm Ratio(FAR) and Probability Of Detection(POD, metrics were used and then compared with traditional cloud detection methods such as a multi-feature combined (MFC)[11], Scene Learning for Cloud Detection on Remote-Sensing Images[12] and a progressive refinement scheme [13]. According to the quantitative results obtained the proposed model posted a better performance with an accuracy of 95.45%, POD of 89.09%, FAR of 2.67%, HK of 93.07% and an IOU of 85.38%. They further recommended on improvements of the computational efficiency of the proposed model for cloud detection.

III. METHODOLOGY

A. Dataset Description

To train and validate the model, satellite images obtained from different satellites and covering different land surfaces as illustrated in the fig 1 were used. A total of 90 Landsat 8 Operational Land Imager/Thermal Infrared Sensor (OLI/TIRS) satellite images with a resolution of 30m which are provided to the public by [14], 160 Landsat 7 Enhanced Thematic Mapper Plus(ETM+) with a resolution of 30m[15] and 100 Gaofen-1 wide field view(WFW) with a resolution of 16m[16] were used. The Landsat-8 images were broadly categorized to eight biomes, that is, barren, forest, grass/crop, shrub land, urban, water and are further divided to four classes, that is, cloud, thin cloud, cloud shadow and clear. According to [17], data used to train a model should be more as compared to training dataset with a preferred percentage split of 70/30 whereby 70% of the total dataset is used to train the model while the other 30% is used to evaluate its performance. In our study, to obtain the training and test dataset, percentage split was used to randomly split the datasets whereby 70% of the each dataset was used for training the model while the remaining 30% was used for testing the performance of the developed model for cloud detection given satellite images.

B. The Proposed Method

The proposed MFF – DCNN for cloud detection is composed of a deep coarse network for extraction of high level features and three deep fine extraction network for separation of the cloud pixels from other objects present in an image as illustrated in fig 2.A fully connected (FC) layer is used to flatten the outputs obtained from the coarse and the fine modules and the output obtained from this layer fed to a feature fusion layer that is used to fuse features obtained from the four network components. The output from the feature fusion layer used for classification.

1) Deep Coarse Network: The deep coarse network (DCN) is constituted of three convolutional layers for the purpose of extracting high level features. The first layer is structured such that it is comprised of 36 filters and a kernel size of 3*3. The rectified linear unit(ReLU) activation function is then be applied to the convolved patches and a 2*2/2 max-pool function applied to each response generated. The second and the third convolution layers are then modelled such that they are made of 64 filters with the ReLU activation function and a 5*5 kernel and a 2*2 max-pooling with stride 2.

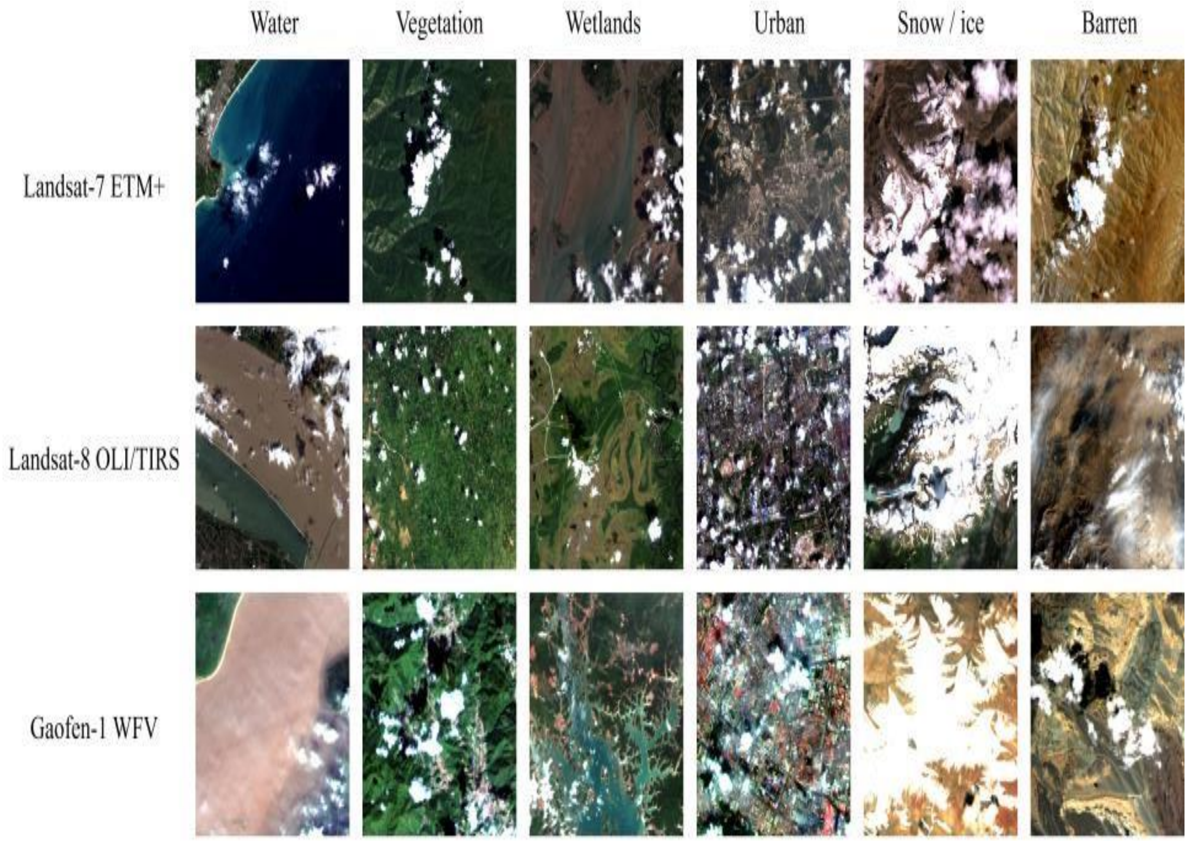

Fig 1: Display of the Obtained Datasets on various land surfaces Source: [18]

2) Deep Fine Extraction Network: Most of the images acquired from remote sensing satellites are rarely annotated and also lack bounding boxes to represent the most likely region of interest (ROI)[19]. As a result, this deep fine extraction network is developed to help in identification of the ROIs. The three deep fine extraction layer are built using the ResNet50 architecture which has so far proved to have better performance for object detection as compared to the other convolutional architectures in addition to helping to mitigate the vanishing gradient problem[20]. The areas marked by the bounding boxes are then extracted using ROI pooling layer so as to come up with a feature map to be fed to a fully connected layer that would compute the score for the input image for each class. The fig 3 illustrates the structure of the deep fine extraction module.

3) Feature Fusion layer: The feature fusion layer is introduced to the architecture so as to ensure that the features extracted from the deep-coarse and deep-fine networks are combined to form a single feature set for classification.To avoid overfitting and improve on the generalization of the model, cross entropy loss function[21] was used to regularize the feature fusion process and the softmax regression function[22], defined in equation (1) used for classification.

$$f_{softmax}(x) = \left[\frac{\exp(x)}{\Sigma_i^n \exp(x)}\right]^T \qquad (1)$$

where x is the filter i score obtained from the previous layer and f_{sotmax} is the corresponding output.

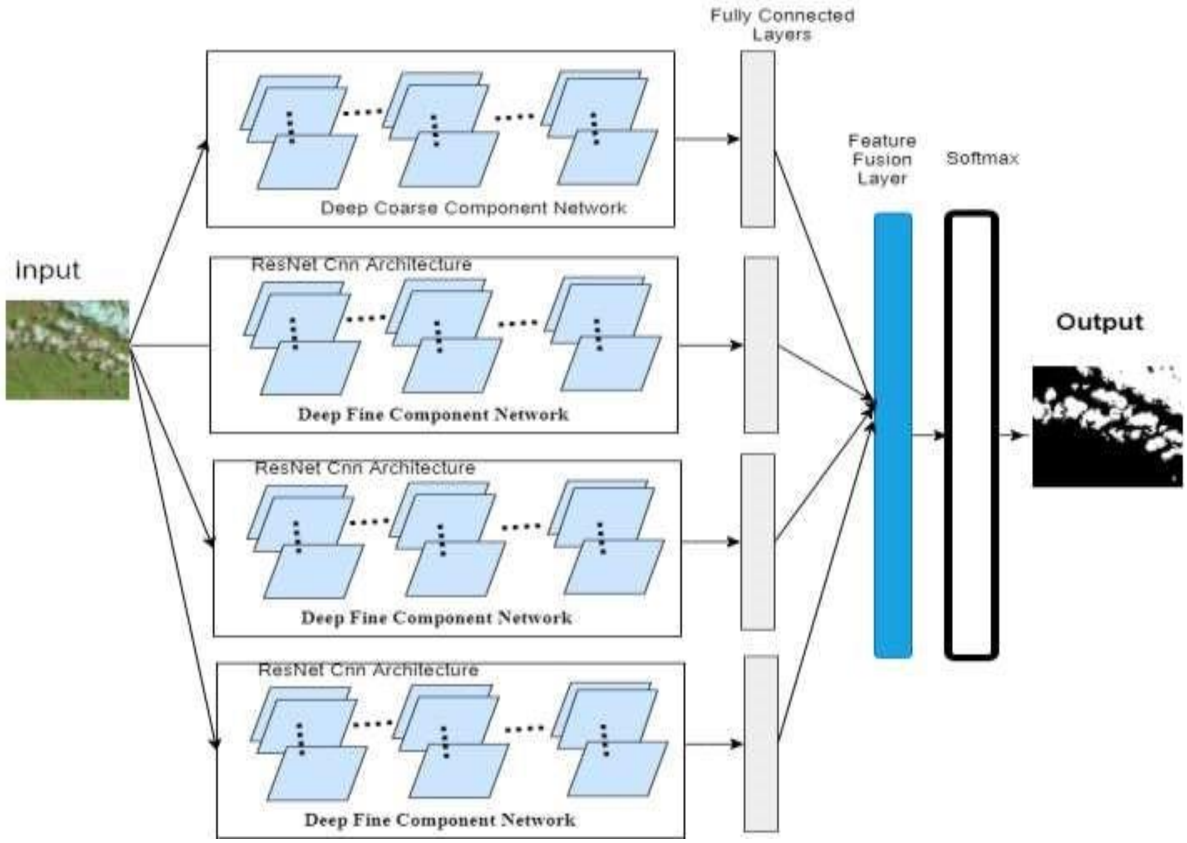

Fig 2: The Proposed Multi-Feature Fusion Deep Convolutional Neural Network for cloud detection

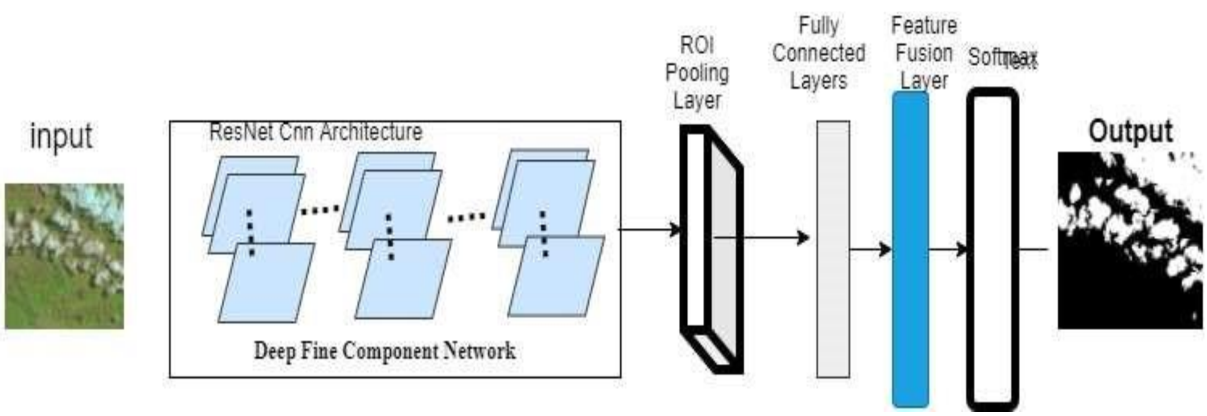

Fig 3: The Architecture of the Deep Fine Extraction Network

IV. RESULTS

A. Performance metrics

The developed model was trained to detect clouds given a remote satellite image dataset that captures different land surfaces in addition to containing spatial, temporal and spectral information. To evaluate the performance levels for extracting and fusing multiple features for cloud detection using the proposed model, the accuracy, precision and recall measures as specified in equation (2),(3) and (4)respectively were used.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Where TP is the true positives, TN is the true negatives, FP is the false positives and FN is the false negatives.

B. *Results and Discussion*

In our work the performance measure of the FCN network proposed by [9], multi-feature fusion of point and block feature using SVM classifier proposed by [2] and a multilevel feature cloud detection FCN proposed by [10] are compared with our proposed model. These models were evaluated using the test datasets that contained a total 57 Landsat-8 images that were different from the images used to train the model and the accuracy results obtained are as summarizer in table 1.These results showed that our model performed the best with an accuracy level of 99.06 which was 1.21% higher than that of the fully convolutional network implemented using the Segnet architecture. Which performed as the second best model for cloud detection. Of the four compared models, the model composed of the feature concatenation and window guided filtering gave the lowest accuracy level of 92.92%.

TABLE 1: QUANTITATIVE ANALYSIS OF THE OBTAINED RESULTS

Author	Architecture	Accuracy (%)
[9]	Fully Convolutional layers + Segnet	97.85
[2]	Multi-feature fusion of point feature and block feature	95.36
[10]	Feature Concatenation + Window guided Filtering	92.92
Ours	MFF – DCNN	99.06

Furthermore, the recall and precision measure obtained when testing our model were 89.9% and 89% respectively and an increase in the number of epochs to about 25 epochs was seen to lead to an increase in both the training and validation accuracy of the proposed model but this was seen to stabilize. Further increase of the number of epochs led to an unstable increase and decrease in the training and validation accuracy as illustrated in the fig 4.

The ability of the proposed MFF – DCNN to effectively detect both thick and thin clouds is mainly been attributed to the use of ResNet50 skip connections in the DFCN that enables the model utilize all available feature for cloud detection.

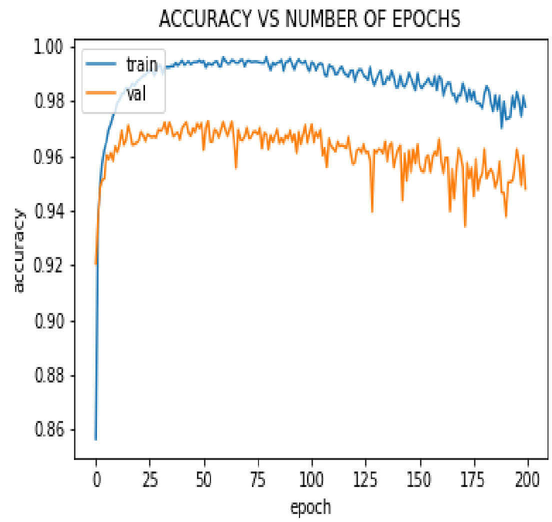

Fig 4: Proposed Model training and validation accuracy against the number of epochs

According to[20] the residual block of the ResNet architecture as illustrated in the fig 4 enables a model train deeper neural networks by optimally tuning the number of layers while training the model. Consequently, it is associated with its high capabilities of addressing the vanishing gradient problem that is frequent especially when more layers are added to a deep learning model. Fusing both the high level features obtained from the deep-coarse network and the low level features from the DFCN enables the model take into consideration all features available in remote sensing satellite imagery, that is, the spectral, textural and spatial information during the training process. Thus the model is capable of learning more information and as a result enable it improve on its predictive predictability.

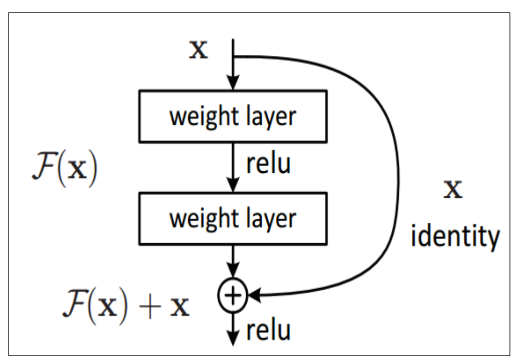

Fig 5: ResNet Residual Block

V. CONCLUSION

In this work, we propose a multi-feature fusion extraction based on deep convolutional neural network for cloud detection in remote sensing analytics given a satellite image that consists of spectral, spatial and textural information. The proposed MFF – DCNN model architecture consisted of a deep-coarse network for extraction of high level features and deep-fine network for extraction of low level features. For identification of the region of interest, feature fusion layer made of a fully connected layer was then used to fuse features identified in the deep-coarse and deep-fine networks and the result fed to a softmax regression function for classification. The cross entropy loss function was then used to regularize the outputs. The quantitative results obtained showed that the proposed model was capable of performing well given datasets that have different clouds types with varying cloud size and density and as a result, it was concluded that this model can also be replicated to different scenarios for accurate and reliable cloud detection. The proposed model is thus seen to be useful in providing insights about clouds in remote sensing analytics and as a results it has proved to be useful in tasks such as prediction of the amount of solar irradiance a given surface can receive given the locations' atmospheric and cloud conditions.

For future works, we recommend evaluation of the proposed model on images obtained from other satellites such as Sentinel-2, SPOT-5 and the Moderate Resolution Imaging Spectroradiometer (MODIS) data so as to evaluate generalizability of the proposed model on different datasets. Additionally, more research should be made on how to improve on the computational efficiency of the proposed model.

ACKNOWLEDGEMENT

The authors of this paper would like to thank the AFRICA-ai JAPAN for funding this project under the AFRICA-ai-JAPAN project innovation research grants (2020/2021).

REFERENCES

[1] T. Bai, D. Li, K. Sun, Y. Chen, and W. Li, "Cloud Detection for High-Resolution Satellite Imagery Using Machine Learning and Multi-Feature Fusion," *Remote Sens.*, vol. 8, no. 715, pp. 1–21, 2016, doi: 10.3390/rs8090715.

[2] Z. X. Zhang Bing , Wu Wei, Shi Qin, Yuan Chenhzong, "Cloud Detection of Remote Sensing Image Based on Multi Feature Fusion," in *5th IEEE International Conference on Big Data Analytics Cloud*, 2020, pp. 298–303.

[3] H. Ishida, Y. Oishi, K. Morita, K. Moriwaki, and T. Y. Nakajima, "Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions," *Remote Sens. Environ.*, vol. 205, pp. 390–407, 2018, doi: https://doi.org/10.1016/j.rse.2017.11.003.

[4] J. Jang, A. A. Viau, F. Anctil, and E. Bartholomé, "Neural network application for cloud detection in SPOT VEGETATION images," *Int. J. Remote Sens.*, vol. 27, no. 4, pp. 719–736, 2006, doi: 10.1080/01431160500106892.

[5] N. Ghasemian and M. Akhoondzadeh, "Introducing two Random Forest based methods for cloud detection in remote sensing images," *Adv. Sp. Res.*, vol. 62, no. 2, pp. 288–303, 2018, doi: https://doi.org/10.1016/j.asr.2018.04.030.

[6] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. S. Toftegaard, "A cloud detection algorithm for satellite imagery based on deep learning," *Remote Sens. Environ.*, vol. 229, pp. 247–259, 2019, doi: https://doi.org/10.1016/j.rse.2019.03.039.

[7] S. Mahajan and B. Fataniya, "Cloud detection methodologies: variants and development—a review," *Complex Intell. Syst.*, p. 11, 2019, doi: 10.1007/s40747-019-00128-0.

[8] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. Skjødeberg, "Remote Sensing of Environment A cloud detection algorithm for satellite imagery based on deep learning," *Remote Sens. Environ.*, vol. 229, no. May, pp. 247–259, 2019, doi: 10.1016/j.rse.2019.03.039.

[9] Z. Li, H. Shen, Y. Wei, Q. Cheng, and Q. Yuan, "Cloud detection by fusing multi-scale convolutional features," in *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci*, 2018, vol. IV, pp. 7–10.

[10] X. Wu and Z. Shi, "Utilizing Multilevel Features for Cloud Detection on Satellite Imagery," *Remote Sens.*, vol. 10, p. 1853, 2018, doi: 10.3390/rs10111853.

[11] Z. Li, H. Shen, H. Li, G. Xia, P. Gamba, and L. Zhang, "Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," *Remote Sens. Environ.*, vol. 191, pp. 342–358, 2017, doi: https://doi.org/10.1016/j.rse.2017.01.026.

[12] Z. An and Z. Shi, "Scene Learning for Cloud Detection on Remote-Sensing Images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 8, pp. 4206–4222, 2015, doi: 10.1109/JSTARS.2015.2438015.

[13] Q. Zhang and C. Xiao, "Cloud Detection of RGB Color Aerial Photographs by Progressive Refinement Scheme," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7264–7275, Nov. 2014, doi: 10.1109/TGRS.2014.2310240.

[14] L. Missions, "Landsat 8," *USGS Science for changing world.*, 2017. https://www.usgs.gov/land-resources/nli/landsat/landsat-8?qt-science_support_page_related_con=0#qt-science_support_page_related_con.

[15] E. O. System, "Landsat 7 Earth Observing System," *Earth Observing System*, 2017. https://eos.com/landsat-%0A7/ (accessed Jun. 04, 2020).

[16] Eoportal, "Gaofen-1 Satellite Missions - eoPortal Directory.," *Satellite Missions*, 2020. https://directory.eoportal.org/web/eoportal/satellite-missions/g/gaofen-1#:~:text=Gaofen-1 - Satellite Missions - eoPortal Directory&text=Gaofen-1 (gao fen %3D,Administration)%2C Beijing%2C China. (accessed Jul. 05, 2020).

[17] J. P. Romano and C. Diciccio, "By Cyrus DiCiccio Technical Report No . 2019-03 April 2019 Department of Statistics," California, USA, 2019.

[18] Z. Li, H. Shen, Q. Cheng, and Y. Liu, "Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors."

[19] W. S. Yongjie Zhan, Jian Wang, Jianping Shi, Guangaliang Cheng, Lele Yao, "Distinguishing Cloud and Snow in Satellite Images via Deep Convolutional Network," *IEEE Geosci. Remote Sens. Lett.*, no. 14, pp. 1785–1789., 2017, doi: https://doi.org/10.1109/LGRS.2017.2735801.

[20] S. Saha, K. M. Khabir, S. S. Abir, and A. Islam, "A newly proposed object detection method using Faster

R-CNN inception with ResNet based on Tensorflow," in *Real-Time Image Processing and Deep Learning 2019*, 2019, vol. 10996, pp. 246–256, doi: 10.1117/12.2523930.

[21] Y. Upadhyay, "Regularization techniques for Neural Networks," *towards data science*, 2019. https://towardsdatascience.com/regularization-techniques-for-neural-networks-e55f295f2866 (accessed Feb. 10, 2020).

[22] H. Mahmood, "The Softmax Function, Simplified," *towards data science*, 2018. https://towardsdatascience.com/softmax-function-simplified-714068bf8156 (accessed Jul. 12, 2020).