This may be the author's version of a work that was submitted/accepted for publication in the following source:

Sarker, Chandrama, Mejias, Luis, Maire, Frederic D., & Woodley, Alan (2019)
Evaluation of the impact of image spatial resolution in designing a context-based fully convolution neural networks for flood mapping. In Digital Image Computing: Techniques and Applications (DICTA 2019), 2019-12-02 - 2019-12-04. (Unpublished)

This file was downloaded from: https://eprints.qut.edu.au/132714/

© 2019 The Author(s)

This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

Notice: Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.
Evaluation of the Impact of Image Spatial Resolution in Designing a Context-based Fully Convolutional Neural Networks for Flood Mapping

Chandrama Sarker, Luis Mejias, Frederic Maire, Alan Woodley
School of Electrical Engineering and Computer Science
Queensland University of Technology
Brisbane, Australia
{luis.mejias, frederic.maire, alan.woodley}@qut.edu.au; chandrama.sarker@hdr.qut.edu.au

Abstract—In this paper, our main aim is to investigate the context-based pixel-wise classification using a fully convolutional neural networks model for flood extent mapping from multispectral remote sensing images. Our approach helps to overcome the limitation of the conventional classification methods with low generalisation ability that used per-pixel spectral information for pixel-wise classification. In this study, a comparative analysis with conventional pixel-wise SVM classifier shows that our proposed model has higher generalisation ability for flooded area detection. By using remote sensing images with different spatial resolutions we also aim to investigate the relationship between imagesensor resolution and neighbourhood window size for context-based classification. Instead of fine-tuning a pre-established deep neural network model, we developed a preliminary base model with two convolutional layers. The model was tested on images with two different spatial resolutions of 3 meters (PlanetScope image) and 30 meters (Landsat-5 Thematic Mapper). During training phase we determined the structure of the convolutional layer as well as the appropriate size of the contextual neighbourhood for those two data types. Preliminary results showed that with increasing the scale of spatial resolutions the required neighbourhood size for training samples also increases. We tested different neighbourhood sized training samples to train the model and the analysis of the performance of those models showed that a 11 x 11 neighbourhood window for PlanetScope data and a 3 x 3 neighbourhood window for Landsat data were found to be the optimum size for classification. Insights from this work may be used to design efficient classifiers in scenarios where data with different resolutions are available.

Index Terms—remote sensing, convolutional neural network application, contextual classification, flood mapping

I. INTRODUCTION

Floods are one of the most frequently occurring natural disasters in Australia. Every year, floods are responsible for destruction of property, urban infrastructure and agricultural resources [1]. Spatial data of floods is the primary source of information used by various government and private disaster relief organizations. Those data help to obtain rapid assessment and mapping of flood extent which further help to plan the relief work more efficiently. Multispectral remote sensing image analysis have become an important tool for flood mapping replacing traditional manual ground truth collection during floods. Early classification and image analysis methods for flood mapping were primarily based on pixel-wise processing approaches [2]. Pixel-wise classification methods apply either rule based or supervised learning methods using per-pixel spectral information. Recently, with the advances in imaging sensors, more sophisticated machine learning methods have been adopted for flood extent mapping. Along with the single pixel-wise classification for flood mapping, researchers have started applying spectral unmixing models [3] for estimating the proportion of flood water. The high performance of all the above mentioned classification methods depend on the tuning of the model parameters for each new flooded area to be mapped. This is because the generalization ability of these methods is limited by the spatial heterogeneity of Earth’s surface and the varied nature of flooding. The spreading of flood waters on land causes inter-class (within different class types) spectral similarity and intra-class (within same class type) spectral heterogeneity. For example, turbid flood water reflectance resembles the reflectance of the concrete surface. Due to this complex nature of flood waters, the spectral responses from neighbourhood pixels or contextual information may provide substantial help over the pixel-based classification methods to distinguish flooded from other non-flooded areas with similar spectral properties and assign them appropriate class labels [4], [5]. Fundamentally, contextual classification is originated in the field of pattern recognition and computer vision and in remote sensing literature this method is gaining popularity with the advent of object based image segmentation approaches [6], [7].

Recent studies have shown that convolutional neural networks learn the contextual information for classification process in the field of remote sensing [8] and it is also introduced in the field of mapping of flooded areas. For example, a recent work presented in [9] has applied convolutional neural networks (CNN) for flood extent mapping by learning the change detection of land features from a set of pre and post disaster aerial images with RGB channels. This study however, is limited to images with RGB channels and did not explore the spectral characteristics of flood water in the infra-red region of the electromagnetic spectrum. The work in [10] proposed a model architecture based on 4 deep networks consisting of dilated convolutions and deconvolutions layers to learn the
difference between flooded and non-flooded areas. Contextual information from remote sensing images of very high spatial resolutions were used in [10]. The study outperforms all baseline methods for detection of flooded areas. However, the model was only trained and tested on a single type of multispectral remote sensing image which did not demonstrate the generalisation ability of the model. Therefore, most of the present works are limited to using true-color images or only the visible channels and only applied to map a specific flooded region.

In this study, we designed a baseline configuration for a two-layered fully convolutional neural network for flood mapping from multispectral satellite images by utilising the spectral information of RGB channel as well as of the infrared channel. We adopted the idea of using pixel-patches (a block of pixels) as training samples instead of per-pixel (single pixel) training samples to take advantage of the contextual information for classification purpose. Each training pixel patch consists of a small block of pixels with the ‘target’ pixel is at the centre of the patch. With the purpose of utilising the neighbourhood spectral information of target pixels and not loosing the spectral information in the feature learning process, we kept the number of convolutional layers small. As our main aim is to obtain a flood extent classification map, we therefore used a convolutional layer at the classification stage to maintain the shape of the output image similar to the input image. The main contribution of the paper can be summarised as follows: First, to make the optimal use of visible and infra-red spectral information from neighbourhood pixels, we investigated the test and validation performance of the proposed fully convolutional neural network models using sets of sample-patches of different sizes and determined the optimal neighbourhood size for training samples. We tested the methodology on PlanetScope images (3 meters spatial resolutions) and Landsat-5 Thematic Mapper images (30 metre spatial resolutions) to determine the neighbourhood size of training samples for those two different types of images. This process also helped us to comment on the relationship exists between the optimum neighbourhood size of training samples and the spatial resolution of remote sensing images. Second, in the classification stage, we investigated the effect of context-based classification for flood extent mapping and evaluated the results by comparing with the conventional pixel-wise information based support vector machines (SVM) classifier. These findings form the basis to create deeper networks for detecting floods.

II. METHODOLOGY

We have proposed a fully convolutional neural network architecture to perform a supervised classification for detecting flooded and non-flooded class from multispectral satellite images.

A. Overview of the proposed network

The architecture of the proposed model is formed by stacking three convolutional layers. The last convolutional layer represents the classification layer. A softmax regression classification function was adopted in the last convolutional layer of the model. This function helps to estimate the probabilities of each class labels in the input feature map. The model was trained using categorical cross entropy loss function [5]. No pooling layers were used as our proposed model was designed to detect only two class types. No zero-padding was used because the models in this experiment were designed to perform a classification task based on spectral properties and hence, it is not necessary to preserve the spatial size of objects during training. The total number of epochs or iterations was set to 2,000 however, the training was stopped at 500 epochs, as the model started to overfit beyond 500 epochs. We adopted the initial learning rate at 0.001 and during the training process we fed the sample patches into the network in batches. Each batch consists of 128 sample patches. Stochastic gradient descent with Adam optimization algorithm was used by the model. Adam optimizer is based on adaptive moment estimation [11].

The first two convolutional layers consist of learnable filters which help the layers to learn patterns of the input training data of different class types. The filters are of size 3x3. We experimented 6 different neighbourhood sizes for experiment (N) and these are- 3x3 (N-3), 7x7 (N-7), 9x9 (N-7), 11x11 (N-11), 13x13 (N-13) and 23x23 (N-23). We also investigated the effect of different numbers of filters used in each of the first two convolutional layers during hyper-parameter tuning. The number of filters in each layer varies from 2 to 256 in powers of 2, and used 8 combinations between layer -1 (L1) and layer-2 (L2) by keeping the number of filters same to make the process time permissible. The best performing model was selected out of (6 different training sets * 8 different filter numbers =) 48 different models based on the validation and test performances [5]. Figure 1 represents a fully convolutional model architecture using 32 filters of (3x3) size and each sample patch in the input data represents a (11x11x4) multidimensional matrix. The model networks were written in Python 3.5 using Keras API. We also utilized MATLAB 2018a and JetBrains PyCharm community Ed. 2017.3.

B. Available Remote Sensing Data set

Landsat-5 Thematic Mapper (TM) and PlanetScope remote sensing images were used for training and testing the classification model. The Landsat data consists of 14 images representing flooded areas during 2011 Queensland and NSW flooding events. Landsat-5 TM image has 6 reflective bands (band 1-5 and band 7) with 30 meters spatial resolution and 1 thermal band (band 6) with 120-meters spatial resolution which is resampled to 30 meters to the downloadable version. The details of the Landsat-5 TM data are listed in Table I. The PlanetScope data set is a collection of 325 images with 3 meters spatial resolution that were collected from 6 locations across Australia covering 8 flood events between 1st June, 2016 and 1st May, 2017. The Planet data was provided by Planet as well as the organisers of the 2017 Multimedia Satellite Task at the MediaEval Benchmarking Initiative for
Fig. 1. Overview of the proposed Fully Convolutional Neural Network architecture. The neural network is designed by stacking 3 convolutional layers. First two convolutional layers contain 32 filters of size 3 × 3 each. The last convolutional layer contains 3 filters with size 1 × 1 and a softmax regression function to obtain class label probability for each pixel.

| Spectral Bands | Wavelength (μm) | Spatial Resolution (m) |
|----------------|----------------|------------------------|
| 1              | 0.45-0.52 (blue) | 30                     |
| 2              | 0.52-0.60 (green) | 30                     |
| 3              | 0.63-0.69 (red) | 30                     |
| 4              | 0.76-0.90 (near IR) | 30                   |
| 5              | 1.55-1.75 (mid IR) | 30                   |
| 6              | 10.4-12.5 (thermal) | 30                |
| 7              | 2.08-2.35 (mid IR) | 30                |

Source: [13].

| Spectral Band | Wavelength (μm) | Spatial Resolution (m) |
|----------------|----------------|------------------------|
| 1              | 0.455-0.515 (blue) | 3                     |
| 2              | 0.5-0.59 (green) | 3                     |
| 3              | 0.59-0.67 (red) | 3                     |
| 4              | 0.78-0.86 (near IR) | 3                   |

Source: [14]

C. Reference Data

During flood most areas are inaccessible for ground truth data collection. It is also difficult to obtain a precise map of flooding extent as the extent of flood water during a flood event may changes daily. Therefore, for the evaluation process we decided to use published data as the approximated representation of the extent of flooded areas on ground that help to validate our experiment. In a recent study conducted by Geoscience Australia, a comprehensive mapping of surface water for the Australia has been made possible by the analysis of temporal Landsat data covering the period of 1987 to 2014. The study has provided a series of confidence level maps of water presence covering Australia. The aim of this web service is to develop a better understanding about where the water is present throughout the years (permanent water features), where is infrequently observed (intermittent water features), where flooding has been occasionally noticed and where there is no-water presence at all. The work achieves an overall accuracy of 97% with 93% correct identification of areas containing water [15]. The confidence map showing flood-water and no-water pixels in the water summary database was utilized to get the labels for flood-water and non-water pixels.

For PlanetScope data a set of labelled images have been provided in which a class label of 1 for showing evidence of flooding and 0 for showing no evidence of flooding [12]. The labelled images are generated by segmentation process to mask out the flooded areas by human annotators. These mask images were considered as the reference data for PlanetScope training samples.

D. Data Pre-processing

Landsat-5 TM images were required to be registered with their corresponding reference images. The Matlab Image Processing toolbox was used to perform the image registration process. The MATLAB registration process helps to align each reference image with its corresponding satellite image at pixel level which is required to generate the training samples. The image registration process involves the second order polynomial geometric mapping transformation followed by control-point mapping function [16]. In case of PlanetScope data, the images and their corresponding reference image are already co-registered. Image co-registration process is followed by a normalization of the input images before adding them to the network. Normalization is a prerequisite to remove any location bias present in images. In this study we performed Minmax normalization. Minmax normalization is often called feature scaling by which the values of a numeric attribute of feature data (in this study the numeric property is the intensity values of the image pixels) are reduced to a scale between 0 and 1 [17]. The value of Z which is the normalized value of an attribute from the set of observed attribute X can be obtained by the following formula (Equation 1):

\[ Z = \frac{X - minX}{maxX - minX} \] (1)

The next step was to generate training, validation and test sets from input images. In our study instead of using single pixels as training samples, we used small patches of pixels as samples which were utilised by the model during training.
process to understand the underlying pattern of the class types [18]. The total number of input images were divided in two sets with a 90%-10% ratio and using them for generating training and test sample-patches. Further, the training set was subdivided into train and validation sets with a 80%-20% ratio. The three sets further used to generate different sets of sample-patches of different sizes. The final selection of training and test samples from their corresponding sample-patches also followed a random choice method with no duplication. Each training set covers sample patches consisting of a total of 2.94 million pixels (1.47 million pixels per class). The class label of the pixel located at the centre of the corresponding image patch obtained from the reference data was referred the class type for the training sample patch. Pixels surrounding the centre pixel were formed the neighbouring area for information extraction.

E. Experiment

We performed two separate experiments of model training using inputs from PlanetScope and Landsat data respectively. The model’s performance was evaluated on 8 different neighbourhood sizes. We preferred to apply odd numbers for the neighbourhood size to obtain a specific centre pixel location. PlanetScope images have 4-channels (Table II) and therefore, similar channels of Landsat-5 TM images were considered while generating training samples. This makes the training sample sets to only differ in spatial resolutions. Evaluation of validation and test performances of the model determined which neighbourhood window size and what number of learnable filters help the model to perform optimally for PlanetScope and Landsat remote sensing data respectively. The validation set was required in this process to assess the generalisation ability of the trained model. If the validation performance is deteriorating while the training performance is improving over the epochs, the model assumes to be not well trained and has a very low generalisation ability. In deep learning terms it is called overfitting [18].

III. Results and Discussions

In this section we present the experiment results in two separate subsections - first, we present the results and related discussions on the training, validation and test performance of the proposed model and the selection of optimally performing model for each of the two different type of satellite images. The second part evaluates classification results achieved by applying the selected best model on the PlanetScope and Landsat test dataset which were not included into the model’s training process.

Assessment of classification performance of the models over different PlanetScope and Landsat-5 TM images was the final stage of the experiment. We have calculated and analysed different coefficients from the confusion matrix to evaluate the performance of the model for flood extent mapping. The coefficients include precision (P) and recall (R) and F1-score and intersection over union metric or Jaccard index.

To calculate the P and R from each classification result, we calculated the true positive (refers to pixels correctly classified as flood water), the false positive (refers to non-flooded pixels erroneously classified as flood water) and the false negative (refers to flood water pixels that are missed) [19] from each classification output. The calculation of P and R are shown in Equation 2 and Equation 3 respectively.

\[
P = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}
\]

\[
R = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}
\]

F1-score estimates the harmonic average of the precision (P) and recall (R) and ranges between [0,1]. The F1-score near 1 indicates best classification performance with perfect precision and recall. The score can be calculated by implementing the following formula:

\[
F1 = 2 \times \frac{P \times R}{P + R}
\]

In order to evaluate the ability of the classifiers for correct detection of class labels, we used the Intersection over Union (IOU) metric or Jacard Index. IOU is a measure of pixel-wise classification accuracy that ranges from 0% to 100% [10]. The higher the percentages, the more similar the predicted classification to the true labels. The IOU metric is computed as

\[
IOU = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive} + \text{False Negative}}
\]

A. Performance evaluation of the model networks

The test accuracy and performance loss of models trained on N-9, N-13 and N-23 training samples resulted low accuracy both PlanetScope and Landsat data and this performance outcome made us discard those three neighbourhood sizes for further investigations. Table III listed the highest test accuracy and corresponding loss obtained by using different neighbourhood sized training samples for both the PlanetScope and Landsat data set.

| Size of Receptive Field | Test Performance |
|------------------------|------------------|
|                        | PlanetScope Data | Landsat Data  |
|                        | Accuracy (%)     | Log Loss      | Accuracy (%)     | Log Loss      |
| 3                      | 87               | 0.34          | 91               | 0.22          |
| 7                      | 87               | 0.33          | 91               | 0.24          |
| 9                      | 81               | 0.47          | 89               | 0.35          |
| 11                     | 89               | 0.31          | 87               | 0.31          |
| 13                     | 84               | 0.24          | 86               | 0.38          |
| 23                     | 66               | 1.05          | 85               | 0.40          |

Representation of the performance of 48 models is also beyond the scope of this paper due to page restrictions. Therefore we listed only the best performing models in Figure 2 and
Considering the accuracy and loss curves in Figure 3 it can be outlined that the model performed optimally trained on N-3 samples with a combination of 32 filters in convolutional layers (the blue and yellow curves) for Landsat data. The model achieves an accuracy of 91% and 0.20 loss rates. The accuracy and loss rates obtained from the model’s performance on test set for Landsat data also shows that model using N-3 samples achieves overall good performance with highest accuracy values of 91% and loss value of 0.22 with a combination of 32 learnable filters (Table V).

**TABLE V**

**MODEL’S PERFORMANCE ON TEST SET FOR LANDSAT DATA**

| Number of Filters | Size of Neighbourhood Window |
|-------------------|------------------------------|
|                   | N3                          | N7                          | N11                         |
|                   | Accuracy (%)Log Loss         | Accuracy (%)Log Loss         | Accuracy (%)Log Loss         |
| N3                | 2 78 0.46 85 0.36 71 0.22 | 64 91 0.22 87 0.24 85 0.37 | 128 91 0.22 87 0.24 85 0.33 |
| N7                | 4 89 0.32 85 0.32 82 0.42  | 8 88 0.30 87 0.30 85 0.36  |
| N11               | 8 88 0.30 87 0.30 85 0.36  | 16 90 0.25 90 0.24 85 0.37  |
|                   | 32 91 0.22 90 0.36 86 0.54  | 64 91 0.22 90 0.24 87 0.51  |
|                   | 128 91 0.22 91 0.24 87 0.53  | 256 91 0.22 91 0.24 87 0.52  |

The final step of experiment was to evaluate the classification performance of the best performing model across different flood images of PlanetScope and Landsat data. A set of test images along with their corresponding reference images of PlanetScope data were provided by the Planet as well as the organisers of the 2017 Multimedia Satellite Task. Two test images were randomly selected from that set to evaluate the classification performance of the model for this study. For a comparative analysis of the performance of neighbourhood spectral information for flood extent mapping we compared the results of our CNN model with of a conventional SVM classifier which uses per-pixel spectral information for the classification. Figure 4 lists the images, their corresponding true label images and the classification results.

For a comparable visualisation of classification results of those test images, the accuracy was calculated based on precision, recall, F1-score and Intersection over Union (IOU)
Fig. 3. Training and validation accuracy (A<sub>plots</sub>) and loss (L<sub>plots</sub>) plots over 500 epochs of the model using 32 learnable filters and different set of training samples for Landsat data. The number of learnable filters is specified as 'F-number' in every sub-caption, e.g. 'F-32' refers to 32 filters used in the model. TA and VA denote training and validation accuracy. TL and VL denotes training and validation loss.

Fig. 4. Classification performance of the convolutional model and SVM classifier on different PlanetScope images showing floods.

Fig. 5. Classification performance of the convolutional model and SVM classifier on different Landsat-5 TM images showing floods.
metric rate for each class type. The performance of the CNN and SVM models for test images are represented in Tables VI and VII respectively. Compared to the precision and recall rate of both the classifiers it is observed that the CNN model proves to be more robust to the detection of flooded areas from test images. The CNN model achieves high F1-score and IOU percentages of flood water class for both the PlanetScope Test data sets (Table VI). This proves that the CNN classifier can detect flooded areas with high precision and the robustness of the model to classify Test data sets which were not included during training. While, the accuracy components in Table VII show that the SVM classifier is not able to distinguish the two class types. It is evident from the classification output of the second Test image (the fourth column in Figure 4) that the shallow cloud cover was misclassified as flood water pixels by the SVM classifier. The classification accuracy of the second Test set is also very low compared to the CNN classification result using the second Test set. Therefore, from the classification results of PlanetScope Test images it is obvious that the SVM classifier did not prove to be as robust to the unseen Test images (the Test images were not included during the training process) as the CNN classification model.

Similarly, for the Landsat data we selected two image subsets showing flooding occurrences in the mapped area and applied the convolutional model and the SVM model to evaluate the classification performance. The test images, their reference images and corresponding classification results are listed in Figure 5. The accuracy measures for CNN and SVM classification results are listed in Tables VIII and IX respectively. The precision and recall of flood water class for CNN and SVM classification results of the second Landsat Test image did not show improvements by using context-based information. However, the classification accuracy increased for the first Test image. Overall, the application of context-based CNN classifier generated more accurate classification results than conventional SVM classification but, it is observed that using PlanetScope data set the context-based CNN classification method produced comparatively high precise (precision and recall both over 90%) flooded extent maps. With a coarser (comparatively low) spatial resolution it can be expected that Landsat images do not contain the finer spectral details of land features as PlanetScope. Moreover, the use of limited number of channels (we used only Red, Green, Blue and Near Infrared channels of Landsat images in this study) led to less information capture which could be a reason that the model may not be able to learn about the spectral difference of class types optimally.

### IV. CONCLUSION

In this work, we proposed and investigated an automated flood mapping method by using a fully convolutional neural networks trained on context-based or neighbourhood information. We experimented using PlanetScope flood images with a spatial resolution of 3 meters and Landsat-5 Thematic Mapper flood images with a spatial resolution of 30 meters. This experiment helped us to verify the previous research that the size of neighbourhood window of training samples and the spatial resolution of remote sensing images are inversely co-related. With the increase in the spatial resolution, a pixel in a satellite image contains more detail information but for a small area (3 meters per pixel) and because of this with a large neighbourhood size sample patches the model performed better and attained its optimal performance with 11 x 11 neighbourhood size samples for PlanetScope data. Conversely, for Landsat images the

### TABLE VI
**MEASURE OF ACCURACY FOR CNN MODEL: PLANETSCOPE TEST DATA**

| Accuracy Measures | Class Types | Test-1 | Test-2 |
|-------------------|-------------|--------|--------|
|                   | Flood Water | Non-flood Water | Flood Water | Non-flood Water |
| Precision (%)     | 98.48       | 92.59   | 92.55   | 96.00   |
| Recall (%)        | 93.90       | 98.14   | 96.47   | 91.60   |
| F1                | 0.96        | 0.98    | 0.94    | 0.95    |
| IOU (%)           | 92.57       | 91.00   | 89.52   | 88.24   |

### TABLE VII
**MEASURE OF ACCURACY FOR SVM MODEL: PLANETSCOPE TEST DATA**

| Accuracy Measures | Class Types | Test-1 | Test-2 |
|-------------------|-------------|--------|--------|
|                   | Flood Water | Non-flood Water | Flood Water | Non-flood Water |
| Precision (%)     | 87.50       | 77.74   | 43.05   | 23.51   |
| Recall (%)        | 70.18       | 98.60   | 46.19   | 21.31   |
| F1                | 0.87        | 0.77    | 0.44    | 0.22    |
| IOU (%)           | 69.41       | 72.04   | 28.67   | 12.58   |

### TABLE VIII
**MEASURE OF ACCURACY FOR CNN MODEL: LANDSAT TEST DATA**

| Accuracy Measures | Class Types | Test-1 | Test-2 |
|-------------------|-------------|--------|--------|
|                   | Flood Water | Non-flood Water | Flood Water | Non-flood Water |
| Precision (%)     | 66.16       | 39.12   | 85.01   | 41.52   |
| Recall (%)        | 45.19       | 60.37   | 58.80   | 73.83   |
| F1                | 0.53        | 0.50    | 0.69    | 0.53    |
| IOU (%)           | 36.70       | 31.13   | 53.27   | 36.19   |

### TABLE IX
**MEASURE OF ACCURACY FOR SVM MODEL: LANDSAT TEST DATA**

| Accuracy Measures | Class Types | Test-1 | Test-2 |
|-------------------|-------------|--------|--------|
|                   | Flood Water | Non-flood Water | Flood Water | Non-flood Water |
| Precision (%)     | 63.37       | 36.94   | 77.42   | 47.57   |
| Recall (%)        | 31.98       | 68.31   | 70.36   | 56.72   |
| F1                | 0.42        | 0.47    | 0.73    | 0.51    |
| IOU (%)           | 26.99       | 31.54   | 58.38   | 34.90   |
required size of the neighbourhood window is as small as 3 x 3 pixels because each pixel in Landsat data contains information of a large area (30 meters per pixel). We also investigated and determined the optimal number of learnable filters to use in convolutional layers for the remote sensing dataset during the training phase of the proposed model. The model with 64 learnable filters preformed optimally using PlanetScope image samples and the model with 32 learnable filters performed optimally using Landsat images. Finally, we tested model’s classification performance on satellite images covering different flood events. We also used conventional SVM classification on the same test images to compare the classification performance using context-based information with conventional pixel-based classification performance. The test classification outlined that our proposed classification method outperforms the conventional SVM classification method. The context-based CNN classification model works best on high resolution PlanetScope images compared to its performance on low resolution Landsat images. Moreover, in doing a comparative analysis between the two classification methods, the context-based CNN classification model was able to detect flooding extent with very high precision (98.48%) and recall (93.90%) rates from unseen data that proves its potentiality to apply on real-time flood images for flood mapping. The conventional SVM classification model on the contrary, was not robust on unseen data set. We therefore, can consider this as the main limitation of the conventional classifier which uses per-pixel spectral information is its low generalisation ability that makes it unsuitable to use for flood mapping from unseen (the data which is not used to train the model) remote sensing data.

Moreover, to use the full potential of the spectral information of Landsat images in our future work, we will focus on the use or all 6 spectral bands (red, green, blue, near infrared, middle infrared and shortwave infrared) which may enhance the CNN model effectiveness in learning the different class types. As a part of future work, we will also introduce the third class type ‘permanent water’ into the classification process. This will be beneficial to enhance accuracy of the detection of flooded areas by distinguishing flood water from permanent water areas.

ACKNOWLEDGMENT

The authors would like to thank Mr. Norman Mueller (Earth Observation Scientist, Geosciences Australia) for providing us the access to Water Summary data set. All Landsat-5 TM data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at USGS/EROS, Sioux Falls, SD. http://lpdaac.usgs.gov. The High Performance Computing (HPC) lab of Queensland University of Technology has been utilized to run the experimental work on training data preparation and the model training.

REFERENCES

[1] S. Chandrama, M. Luis, and W. Alan, “Integrating recursive bayesian estimation with support vector machine to map probability of flooding from multispectral landsat data,” in 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, Dec 2016, pp. 1–8.
[2] J. Zhao, Y. Zhong, H. Shu, and L. Zhang, “High-resolution image classification integrating spectral-spatial-location cues by conditional random fields,” IEEE Trans. Image Process., vol. 25, no. 9, pp. 4035–4045, 2016.
[3] K. Aziz, A. Rahman, G. Fang, and S. Shrestha, “Application of artificial neural networks in regional flood frequency analysis: a case study of australia,” Stochastic Environmental Research and Risk Assessment, vol. 28, no. 3, pp. 541–554, 2014.
[4] C. M. Gurney, “The use of contextual information in the classification of remotely sensed data,” Photogrammetric Engineering and Remote Sensing, vol. 49, no. 1, pp. 155–64, 1983.
[5] C. Sarker, L. Mejias, P. Maire, and A. Woodley, “Flood mapping with convolutional neural networks using spatio-contextual pixel information,” unpublished.
[6] P. Dao and Y. Liou, “Object-oriented approach of landsat imagery for flood mapping,” International Journal of Computer Applications, vol. 7, no. 5, pp. 5077–5079, 2015.
[7] L. Pulvirenti, M. Chini, N. Pierdicca, L. Guerriero, and P. Ferrazzoli, “Flood monitoring using multi-temporal cosmo-skymed data: image segmentation and signature interpretation,” Remote Sensing of Environment, vol. 115, no. 4, 2011.
[8] E. Moggiyori, Y. Tarabalka, G. Charpiat, and P. Alliez, “Convolutional neural networks for large-scale remote sensing image classification,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 2, pp. 645–657, 2017.
[9] S. Amit and Y. Aoki, “Disaster detection from aerial imagery with convolutional neural network,” in International Electronics Symposium on Knowledge Creation and Intelligent Computing,(IES-KCIC), Surabaya, Indonesia, Sep 2017, p. 1–7.
[10] K. Nogueira, G. Fadel, Samuel, I. C. Dourado, d. o. Wernecke, Rafael, A. Munoz, Javier, O. A. Penatti, T. C. Rodrigo, Li, J. d. Santos, and R. d. S. Torres, “Exploiting convnet diversity for flooding identification,” IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 9, pp. 1446–1450, 2018.
[11] D. Kingma and J. Lei Ba, “Adam: a method for stochastic optimization,” in 3rd International Conference for Learning Representations, 2015.
[12] G. Guillaume, B. Benjamin, D. Claire-Hélène, Z. Maia, R. Michael, D. Emmanuel, B. Dmtry, S. Richard, G. J. Jones, and L. Martha, “Mediaeval 2017 multimedia benchmark workshop,” in Working Notes Proceedings of the MediaEval 2017 Workshop, 2017.
[13] A. Ahmad, “Classification simulation of razaksat satellite,” vol. 53, 2013, p. 472 482.
[14] P. L. Inc., “Planet imagery product specifications,” Planet, Product Specifications, 2018.
[15] N. Mueller, A. Lewis, D. Roberts, S. Ring, R. Melrose, J. Sixsmith, L. Lynburner, A. McIntyre, P. Tan, S. Cumow, and S. Ip, “Water observations from space: mapping surface water from 25 years of landsat imagery across Australia,” Remote Sensing of Environment, vol. 174, pp. 341 – 352, november 2016.
[16] Mathworks. (2019) Control point registration. [Online]. Available: https://au.mathworks.com/help/images/control-point62registration.htm
[17] C. Saranya and G. Manikandan, “A study of normalization techniques for privacy preserving data mining,” International Journal of Engineering and Technology, vol. 5, pp. 2701–2704, 2013.
[18] D. Rosebrooke, Understanding convolutions. PyImageSearch.com, 2018.
[19] D. Notti, D. Giordan, F. Cal, A. Pepe, F. Zucca, and J. Pedro Galve, “Potential and limitations of open satellite data for flood mapping,” Remote Sensing, vol. 10, no. 11, pp. 1–30, 2018. [Online]. Available: https://www.mdpi.com/2072-4292/10/11/1673