

# Urban Land Use Land Cover Extraction for Catchment Modelling Using Deep Learning Techniques

Siming Gong, James Ball, Nicholas Surawski

## ABSTRACT

Over the world, the likelihood of floods and managing the associated risk are a concern to many catchment managers and the population residing in those catchments. Catchment modelling is a popular approach to predicting the design flood quantiles of a catchment with complex spatial characteristics and limited monitoring data to obtain the necessary information for preparing the flood risk management plan. As an important indicator of urbanisation, land use land cover (LULC) plays a critical role in catchment parameterisation and modelling the rainfall-runoff process. Digitising LULC from remote sensing imagery of urban catchment is becoming increasingly difficult and time-consuming as the variability and diversity of land uses occurred during urban development. In recent years, Deep Learning Neural Networks (DNNs) have achieved remarkable image classification and segmentation outcomes with the powerful capacity to process complex workflow and features, learn sophisticated relationships, and produce superior results. This paper describes end-to-end data assimilation and processing path using Unet and DeepLabV3+, also proposes a novel approach integrated with the clustering algorithm MeanShift. These methods were developed to generate pixel-based LULC semantic segmentation from high-resolution satellite imagery of Alexandria Canal catchment, Sydney, Australia, and assess the applicability of their outputs as inputs to different catchment modelling systems. A significant innovation is using the MeanShift clustering algorithm to reduce the spatial noise in the raw image and propagate it to the deep learning network to improve prediction. All three methods achieved excellent classification performance, where the MeanShift+Unet has the highest accuracy and consistency on the test imagery. The final suitability assessment illustrates that all three methods are more suitable for the parameterisation of semi-distributed modelling systems rather than the fully-distributed modelling systems, where the MeanShift+Unet should be adopted for image-based impervious area extraction of urban catchment due to its superior prediction accuracy of 98.47%.

**Key words** | Catchment Modelling, Land use Land cover (LULC), Remote Sensing, Deep Learning

### Siming Gong

Faculty of Engineering and Information Technology,  
University of Technology Sydney,  
Sydney, Australia

### James Ball

Faculty of Engineering and Information Technology,

University of Technology Sydney,  
Sydney, Australia

**Nicholas Surawski**

Faculty of Engineering and Information Technology,  
University of Technology Sydney,  
Sydney, Australia

## INTRODUCTION

Floods are one of the most destructive and frequent natural disasters, causing many personnel casualties and property losses every year (Dawson et al., 2008; Jha et al., 2012; Willner et al., 2018; Zaman et al., 2012). The process of urbanisation has increased the flood risk to human society with the rapid growth of impervious areas such as buildings, roads and parking areas, which break the infiltration of stormwater to the deep soil and permanently change the natural rainfall-runoff process of the catchment (Arnold Jr & Gibbons, 1996; Ball et al., 2019; Jha et al., 2012). Unlike the nature catchment, the impervious surface degrades nature flood resilience by directly transferring rainfall to runoff, which increases the flood volume and leads to the earlier arrival of the flood peak (Holman - Dodds et al., 2003; Li et al., 2017). The weakening of catchment nature flood resilience results in the vulnerability of communities to flood when exposed to the storm burst beyond the capacity of the urban stormwater drainage system. In addition, other factors such as climate change, irresponsible and inappropriate city planning, and stormwater management strategies are growing trends that aggravate the devastation of floods in urban areas.

To understand the rainfall and runoff processes that interconnect the catchment hydrological components of a catchment and to further study the occurrence and consequence of flood events, catchment models were introduced to predict the catchment response to rainfall; limitations of measurement techniques and extremely complicated inner mechanism of catchment hydrology also encouraged the development of catchment models (Beven, 2012; Peel & McMahon, 2020). In the computer era, distributed catchment modelling systems have experienced significant progress with the development of remote sensing technology, geographic information system (GIS) and [increased computation power](#) (Salvadore et al., 2015); all of these developments allow the modeller to reproduce the catchment rainfall-runoff process at a scale suitable for the available data. These sophisticated catchment models require a mass of catchment data and the associated data digitalisation, assimilation, interpretation, parameterisation, and calibration to maintain the robustness of the model predictions. Meanwhile, data quality and quantity are more important than ever before; the adequacy and reliability of a dataset play a critical role in the model performance (Gan et al., 1997). However, identification and digitisation of the spatial characteristics of urban catchments such as LULC have become a significant problem that plagues many aspects (e.g. catchment modelling, urban planning) with the expansion of urban areas and the changing of land use (Hu et al., 2016; Liu et al., 2015). The importance of land use information for catchment modelling systems is reflected in determining the impervious ratio, infiltration rate and the associated surface roughness coefficient (e.g. Manning's n value) (Ball et al., 2019). Therefore, quickly providing accurate and reliable datasets to parameterise the catchment spatial characteristics for

modelling purposes has become a topic that needs to be solved in the modelling community (Abulohom et al., 2001; Ball et al., 2019; Peel & McMahon, 2020).

Breakthroughs in remote sensing techniques allow modellers to obtain numerous catchment spatial data with a visualisation form (Wulder et al., 2012). A typical representative of visualised catchment spatial data is the remote sensing imagery, which has become the indispensable dataset of catchment modelling systems (Beven, 2012; Klein & Barnett, 2003; Mackay et al., 2003). The interpretation of spatial information from the remote sensing imagery is required to estimate the parameters for catchment modelling purposes (Foody, 2004). This interpretation includes identification, delineation, and digitalisation of LULC (e.g. road, building, impervious/pervious area, vegetation and waterbody) (Boegh et al., 2004), which are time-consuming and require solid geographic knowledge (Niemi et al., 2019). In addition, the accuracy and reliability of the manual interpretation solely rely on the interpreter's ability and cannot be reproduced by other modelling groups or systems. This reduces the model's consistency when the catchment is experiencing LULC changes.

In recent years, the development of Deep Learning techniques has greatly facilitated the development of Computer Vision (CV) applications; the CV is a subbranch of machine learning with interdisciplinary knowledge involving how computers acquire information, process data, and gain advanced understanding from digital images or videos (Szeliski, 2010; Voulodimos et al., 2018). Examples of CV applications can be found in autonomous driving (Grigorescu et al., 2020), face recognition (Parkhi et al., 2015), image classification and segmentation (He et al., 2017; Ronneberger et al., 2015).

Meanwhile, Deep Learning has also achieved excellent performances in remote sensing applications. For example, (Rezaee et al., 2018) proposed a Deep Learning neural network (DNN) to classify wetland from remote sensing imagery. (Wenzhi Zhao et al., 2017) introduced an object-based convolutional neural network to discriminate building types and footprints in a catchment of Beijing and achieved over 90% accuracy. (Huang et al., 2018) modified the classic DNN architecture to adapt the multi-spectral data of high-resolution remote sensing imagery and achieved 91.25% precision in the land-use classification for the study area of Shenzhen and Hong Kong. (Yang & Li, 2015) used linear spectral unmixing method to conquer the effect of vegetation for extracting road from an urban remote sensing image. In summary, these works focus on extracting individual spatial features or classifying images with low LULC heterogeneity (e.g. rural catchments, urban catchments with low landscape coverage). [There is no universal methodology for modelling urban catchments, particularly in high heterogeneous LULC areas such as various landscape coverage and diverse building types \(Salvadore et al., 2015\); whether Deep Learning can achieve classification and segmentation of LULC for parameterising and modelling urban catchment, extensive research has not yet been undertaken.](#)

Inspired by related works, this research develops a novel approach utilising a Deep Learning-based CV tool. This approach aims to produce the LULC classification and segmentation of urban catchment remote sensing imagery with heterogeneous spatial features and present a reproducible workflow for any other Australian urban catchment under the same remote sensing database. The feasibility and practicability of this approach for catchment modelling systems are discussed also. The remainder of this paper is organised as follows. Following the introduction, the study catchment is described, proposed approach. The section

'Result and discussion' demonstrates the outputs and discusses their suitability for catchment modelling. Finally, the conclusions derived from the research is presented at the end.

## **STUDY CATCHMENT**

The Alexandra Canal catchment was selected as the test area for this study. This catchment is located south of the great Sydney CBD area, Australia. The catchment area is 11.50 kilometres square and consists of four highly urbanised regions: Sheas Creek, Rosebery, Munni Street-Erskineville, and Alexandra Canal, with multiple land uses inclusive of residential (approx. 40%), industrial (approx. 25%), road (approx. 10%), parkland (approx. 22%) and water body (approx. 3%). The catchment land covers are varied and include single dwelling, terrace, dense apartment, industrial plants with the large impervious areas, and sizeable pervious areas of parkland and golf course. The drainage system of the Alexandra Canal catchment consists of subsurface pipes, pits, subsurface channels, open channels, culverts and flood mitigation structures. Shown in [Figure 1](#) (Esri, 2021) is the location of the catchment with respect to continental Australia and within the greater Sydney region.

A significant feature of the catchment is the high heterogenous LULC, such as the various roof colour and shapes, perennial vegetations, water bodies, railways, walking trails, roads, large pavement areas and buildings' shadows. A few ground features are difficult to identify on the imagery due to the obscuration of tree canopies. Additionally, in some regions of the catchment, the land zoning has changed from industrial land to commercial or residential land, leading to discontinuities in land cover between the expected and actual. The identification of these spatial features is essential for catchment model. For example, a critical inferred control parameter 'Impervious percentage' in Stormwater Management Model (SWMM) is highly related to the measured geometry of catchment LULC and associated roughness coefficient, which requires the modeller to identify and delineate all impermeable ground features (e.g. Roof and Road) (Choi & Ball, 2002). In summary, the LULC identification and classification of the catchment is a complex and very challenging task from the perspective of both computer vision and catchment modelling.



**Figure 1** | Location of Alexandria Canal catchment.

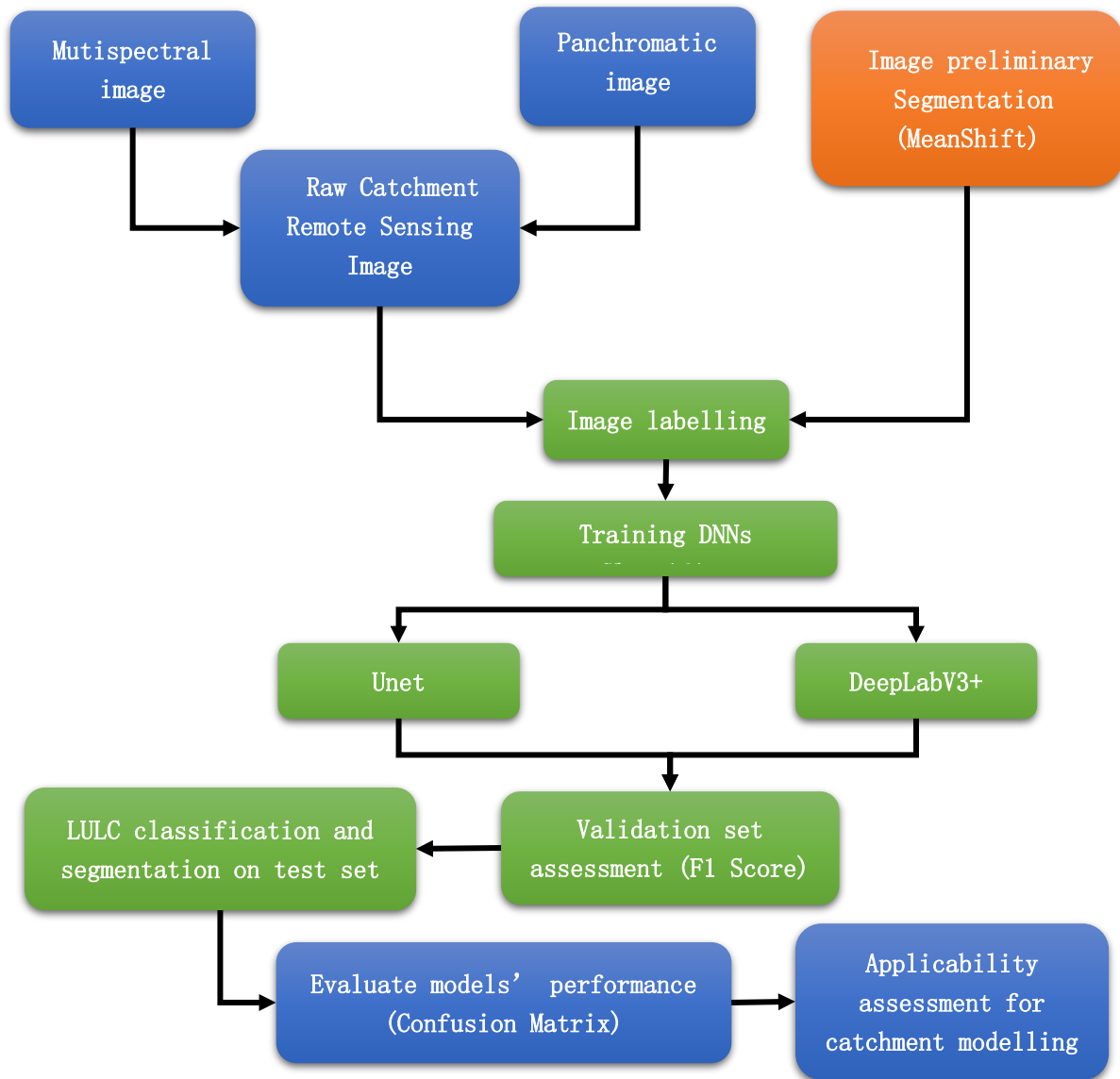
## **METHODOLOGY**

### **Outline of Methodology**

The proposed LULC classification and segmentation methods of the catchment imagery consists of three algorithms, which are the clustering algorithm MeanShift (Cheng, 1995), and the pixel classifiers Unet (Ronneberger et al., 2015) and DeepLabV3+ (Chen et al., 2018).

In the beginning, the catchment raw imagery was preprocessed to adapt to the Deep Learning environment. The MeanShift clustering algorithm was introduced to simplify the spectral information of the catchment imagery by producing a preliminary segmentation result. The segmented image was then used as the input of the pixel classifier Unet. Pure Unet and DeepLabV3+ classifiers were also implemented to classify and segment the LULC information using the catchment raw image as input for the control experiment. Then, the F1 score was adopted to evaluate the models' prediction accuracy on the validation set, and the associated images to demonstrate the bias between the ground truth and classification outputs. Finally, the performance of trained models on the test set was quantified using the confusion matrix to assess LULC accuracy associated with different catchment modelling systems. In machine learning, both the F1 score and confusion matrix are effective measures to analyse the accuracy of the model's prediction, where the F1 score is the harmonic mean of the precision and recall, the confusion matrix is an  $n \times n$  matrix by comparing the actual class of raw image and predicted class of classified image (Chicco & Jurman, 2020; Powers, 2011). Detailed

procedures are demonstrated in [Figure 2](#).



**Figure 2** | Flowchart of LULC extraction procedures proposed by this study.

### Imagery Pre-processing for GIS

Image preprocessing is the critical step in remote sensing applications. Many commercial software packages like ArcGIS and MapInfo have mature techniques and associated functions to process images for different industries and issues. The remote sensing data processing includes geometric referencing, image fusion, image mosaic, image cropping, cloud removal, shadow processing, and atmospheric correction (Camps-Valls et al., 2011; Yang & Li, 2015).

The focus of remote sensing image preprocessing is diverse as the different application requirements of various industries. For example, atmospheric correction is an essential process for remote sensing based agricultural management that reduces the impact of atmospheric scattering and clouds obstruction on images (Wu et al., 2005). Regarding catchment modelling, the focus is on LULC recognition and geometric accuracy, which requires high image resolution and geometric correction. Therefore, it is necessary to process the image fusion by

using the panchromatic image as the reference image to register the multi-spectral image to obtain the RGB remote sensing images with geographic coordinates and high resolution (Bai et al., 2015; Ghassemian, 2016; Thomas et al., 2008).

High-resolution remote sensing imagery for the catchment in 2019 was collected from MapServer (Spatial Services of New South Wales Government), with three bands (Red, Green and Blue) and a resolution of  $25344 \times 20992$  (0.29 meters per pixel width) (NSW Spatial Service, 2021). The coordinate system of source imagery is WGS 84 / UTM zone 56S. A Digital Surface Model (DSM) was utilised as the base layer for the treatment of multi-spectral image orthorectification and thereafter to geometrically correct the image to ensure it was planimetrically projected. The GIS platform ArcGIS Pro Version 2.8.2 was applied to delineate the catchment vector boundary. Following these steps, the catchment boundary was used to clip the remote sensing imagery and unify the geocoordinate system.

### **Training Sample Generation**

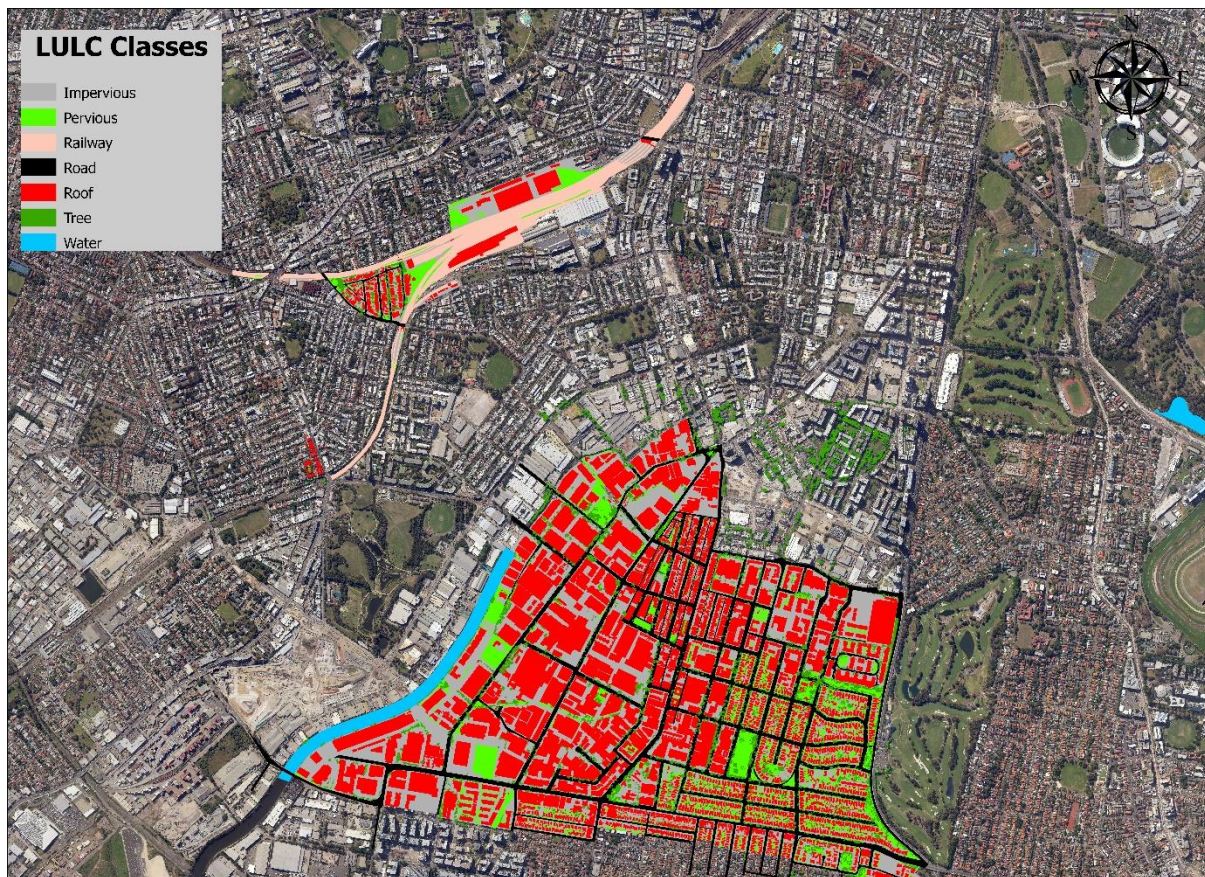
In machine learning, a common approach is to use a large number of task-related data sets to:

- Train the model,
- Adjust the model's parameters through continuous iteration of training error on the data sets, and
- Get the optimised model that best fits the model data set;

This is also referred to as supervised learning (Hastie et al., 2009). The construction of the training set requires a large number of labelled data (pictures, text files, videos, etc.) to provide a mapping channel for the model training. In the computer vision field, data labelling is used to identify the original images or videos and add one or more meaningful labels, which objectively establishes a mapping channel from data to results for the machine learning model to learn information from it (Cowie et al., 2011). For example, in the task of LULC recognition of remote sensing images, the label can be the binary representation of an object within an image, or it can be the semantic delineation of ground surface features at the pixel level. The starting point for data labelling requires humans to make judgments on designated unlabelled data. The trained model can be applied to make predictions on new data sets by learning hidden patterns within the human ascribed labels (Li et al., 2021). Labels are treated as objective criteria to train and evaluate the supervised machine learning model, often referred to as a "standard answer". The model training and accuracy of the prediction depends on the correctness of the 'standard answer', so it is essential to allocate sufficient time and resources for labelling to ensure the high accuracy of labels.

The ratio of the training and validation was set to 9:1. The trained model will be applied to make predictions on the entire catchment imagery (Figure 3) for testing the model's performance on the rest area of the imagery without labels. In catchment modelling systems, the determination of representative class schema depends on the comprehensive land cover conditions and the catchment spatial features that affect the catchment rainfall-runoff process. Therefore, I defined the catchment LULC class schema for this study as: roof, road, railway, water, impervious, pervious and tree by referring to the zoning map from the City of Sydney council and observable features of the catchment remote sensing imagery. The manual labelling approach was used to maintain the accuracy and consistency of label masks. It is worth mentioning that some ground surface features are occluded by the tree canopies, which leads to identification difficulties and incorrect labelling. To address this issue, I ignored the occluded ground features but delineated and labelled the tree canopies above them to ensure the stability

of the mapping channel. The tree can be seen as an uncertainty class, where its prediction should be analysed with other conditions to obtain the correct parameter for catchment modelling purposes. Approximately 25% of the image area was labelled. An image clipping tool was applied to grid the labelled area into 5636 small images with a size of  $256 \times 256$  and then rasterise them to generate the label masks by snapping to the pixel alignment. [Figure 3](#) shows the labelling area and class schema of the study.



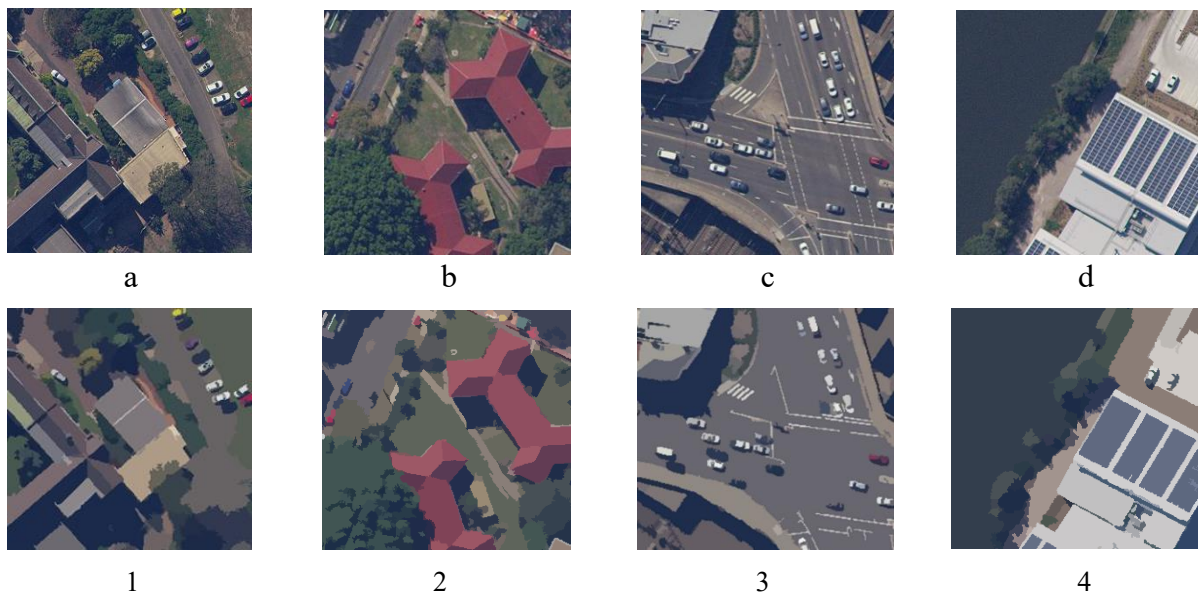
**Figure 3** | Class representatives and labels of the study catchment.

### **MeanShift for image pre-segmentation**

The MeanShift image segmentation algorithm extracts general information from the complex features through coarsening and then segment the image (Comaniciu & Meer, 2002). The concept of MeanShift was first proposed by (Fukunaga & Hostetler, 1975). In unsupervised learning, MeanShift refers to a density-based non-parametric clustering algorithm that assumes the data of different clusters conform to the corresponding probability density distributions and finds the density vector of any sample points in which the highest vector means the direction of iterations. These sample points will eventually converge to the area with the maximum local density, and the points that converge to the same local maximum are considered members of the same cluster. (Comaniciu & Meer, 2002; Szeliski, 2010). In image segmentation, Meanshift can be used to segment image objects by clustering the neighbouring pixels with similar spectral and texture information.

Pixels of remote sensing imagery usually contain two types of information: coordinate information and colour information, which constitute the spatial features of the image (Foley et al., 1994). In the urbanised Alexandria Canal catchment, the regular geometric of artificial

structures and the similar colours of vegetation lay the foundation for the use of MeanShift for preliminary image segmentation. In addition, the spatial features of the imagery are very complex, and the associated data amount is enormous due to the high heterogeneity of the catchment LULC. Therefore, it is necessary to take preliminary segmentation measures to simplify the spatial features of the image. In this study, the MeanShift segmentation result was used as the input of the DNNs classifier to train the model and test set to make predictions of LULC. [Figure 4](#) shows the examples of segmentation by MeanShift.



**Figure 4** | Raw images (a, b, c, d) and their MeanShift segmentation (1, 2, 3, 4).

### Deep learning for image classification and segmentation

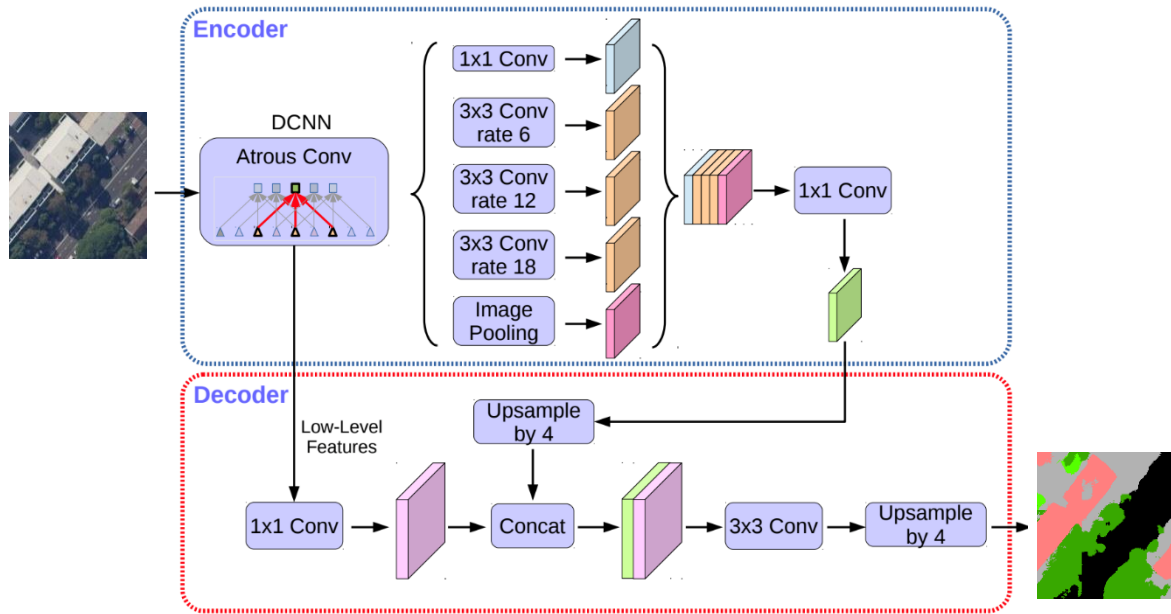
Two Deep neural network-based algorithms U-Net and DeepLabV3+, were selected to conduct the study catchment's LULC classification and semantic segmentation. In the distributed catchment modelling systems, any catchment internal spatial characteristics and distribution that affect the rainfall-runoff process should be considered and parameterised, where the spatial patterns are expressed by pixel in remote sensing imagery. Therefore, the pixel-based semantic segmentation methods were chosen instead of the object focused instance segmentation method.

**U-Net:** U-Net is one of the earliest algorithms for semantic segmentation using fully convolutional networks (Ronneberger et al., 2015). The proposed symmetrical U-shaped structure with innovative encoding and decoding paths has affected the development of many segmentation networks such as U-Net++ (Zhou et al., 2018) and SegNet (Badrinarayanan et al., 2017).

In this study, the pre-trained model Resnet-34 was applied as the backbone for enhancing the downsampling process. A schematic diagram of U-Net with the pre-trained backbone neural network ResNet-34 (He et al., 2016) is demonstrated in [Figure 5](#).

On the left side of the network is a series of downsampling operations composed of four convolution and max-pooling layers, and the right side of the network is called the expansive path with four upsampling layers to recover the feature map to its original resolution. The





**Figure 6** | Architecture of DeepLabV3+ (Chen et al., 2018).

The architecture of DeepLabV3+ is shown in [Figure 6](#). The main body Encoder part is a Deep Convolutional Neural Network (DCNN) with atrous convolution, the commonly used DCNN such as ResNet can be used, followed by the Atrous Spatial Pyramid Pooling module (ASPP) for introducing Multi-scale information. DeepLabv3+ introduced the Decoder part to improve the segmentation accuracy of class boundary by integrating the features of different levels.

### Accuracy assessment

A typical evaluation index of classification performance is accuracy, which can directly reflect the percentage of correct classification. However, in the semantic classification of urban remote sensing imagery, the percentages of each LULC class are not uniformly distributed, which lead to the evaluation bias to the big class with a large area if there is no adjustment to balance the data set. The Kappa coefficient (Stehman, 1996), an index that can punish the 'bias' of the model, is needed to replace a pure accuracy assessment. [A robust model appreciates a higher kappa coefficient, scilicet a balanced confusion matrix reflects a higher Kappa coefficient, vice versa.](#) The expression of Kappa coefficient is:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where  $p_o$  is the sum of all observed correct classification, and  $p_e$  is the hypothetical probability of each class, using the classified data to calculate the probabilities of each class from the view of users (Sim & Wright, 2005).

In this study, the Stratified Random method was applied to create 500 assessment points that are randomly distributed in each class. The number of points in each class is proportional to its class area. Then, the Kappa coefficient was calculated by confusion matrix to evaluate the prediction consistency.

## RESULTS AND DISCUSSION

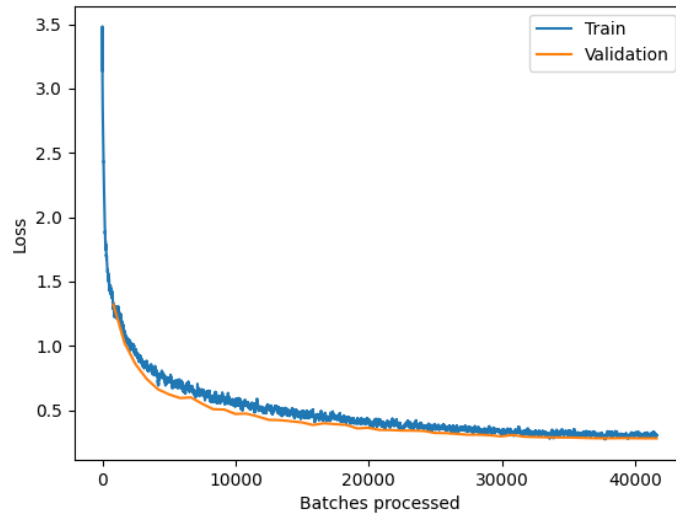
### Model training and prediction

In this study, three methods were used to predict LULC from remote sensing images of the study catchment, namely Unet, DeepLabV3+ and MeanShift+Unet, with the same test set, training set amounts and masks. The pre-trained backbone neural network and iteration parameter was set to Resnet-34 and 50 for all three methods. The difference is that the MeanShift+Unet method uses the preliminary segmented imagery as the base map instead of the raw images adopted by the other two methods. [Figure 7](#) shows the training and validation loss curves. The training process of the three models was smooth, and the fast convergence of the loss was achieved, which indicate that the gradient is continuously decreasing and the selected learning rate is appropriate. Unet has the best training performance, followed by MeanShift+Unet, and the last one is DeepLabV3+. A vibration occurred on the verification curve of DeepLabV3+, and the fitting degree between the verification and training is not as good as the former two. [Figure 8](#) shows the performance of the three methods on the verification set.

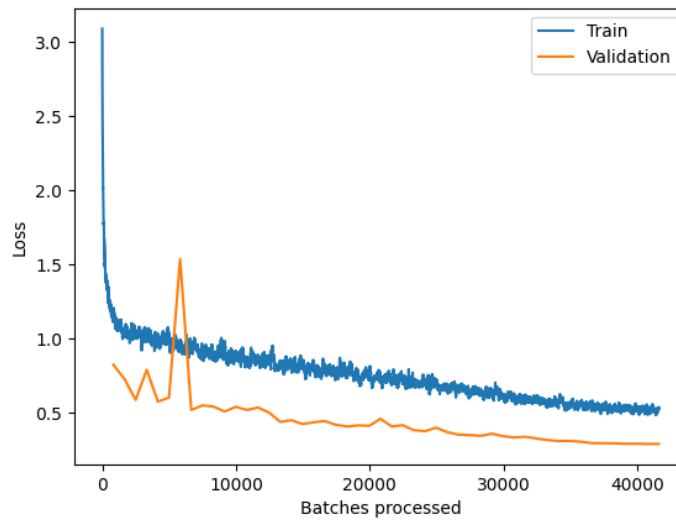
The F1 score is involved to evaluate the model's performance in terms of prediction accuracy. F1 score is an index used in statistics to measure the accuracy of a two-class model considering both the precision and recall of the classification model, which can be regarded as a harmonic average of the model's precision and recall (Powers, 2011; Tharwat, 2020). The equation of F1 score is:

$$F_1 = \frac{2}{recall^{-1} + precision^{-1}}$$

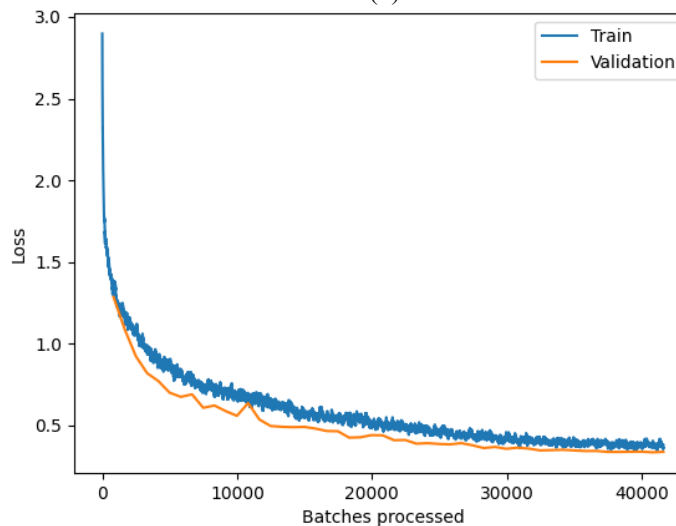
The advantage of the F1 score is avoiding the unstable recall and precision due to the excessive number of positive samples. [Figure 9](#) shows the F1 score of three models in each class. The three methods have minor differences in the F1 scores of each LULC class. Unet performs better in major classes (e.g. roof, impervious), and DeeplabV3+ performs better in minor classes (e.g. tree, road), which may be powered by its excellent boundary segmentation capability. The overall performance of MeanShift+Unet is lower than the other two methods, but the prediction is more balanced, which means the better consistency between input and output.



(a)



(b)



(c)

**Figure 7** | Training and validation loss curve of (a) Unet, (b) DeepLabV3+ and (c) MeanShift+Unet.

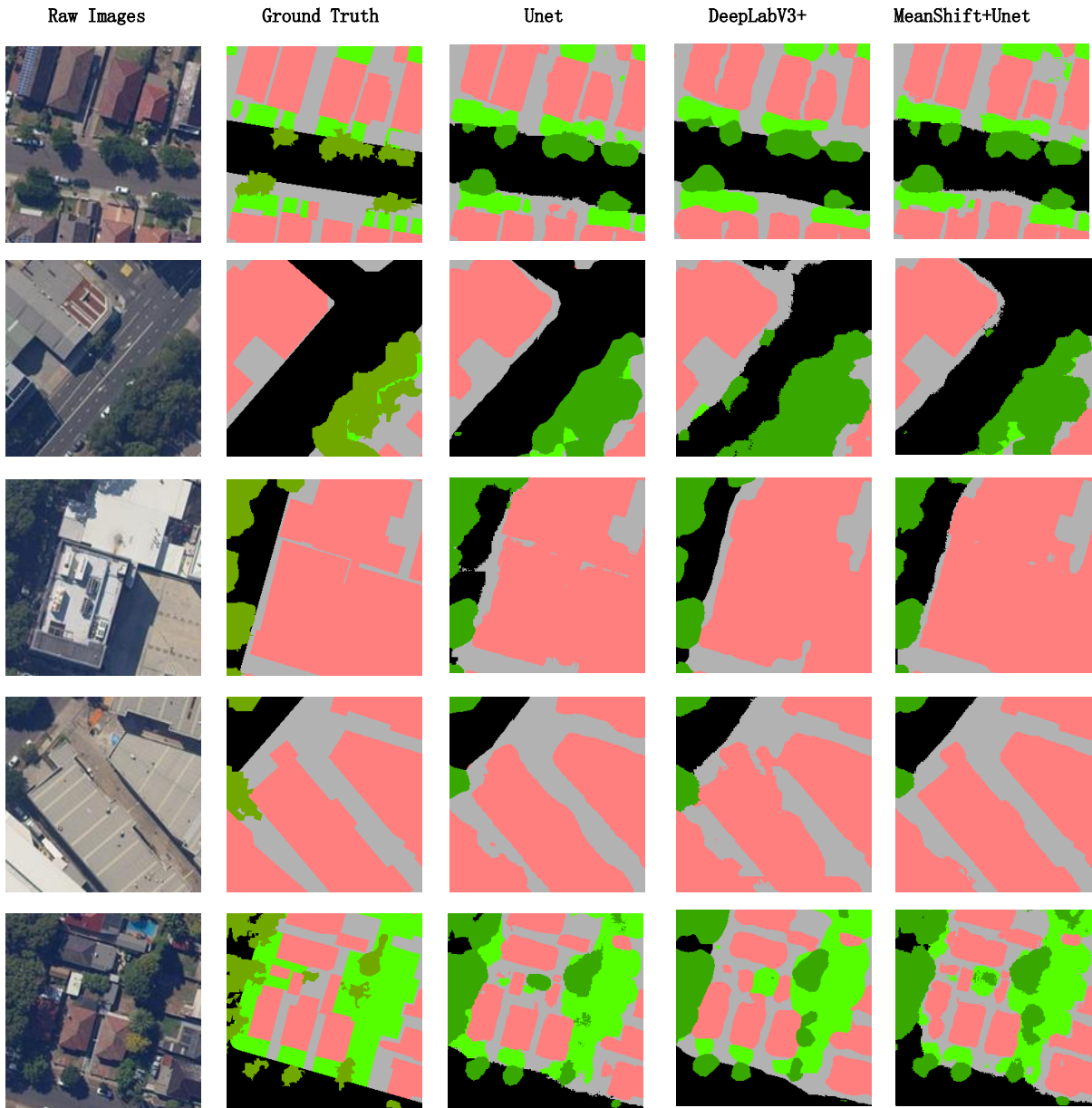
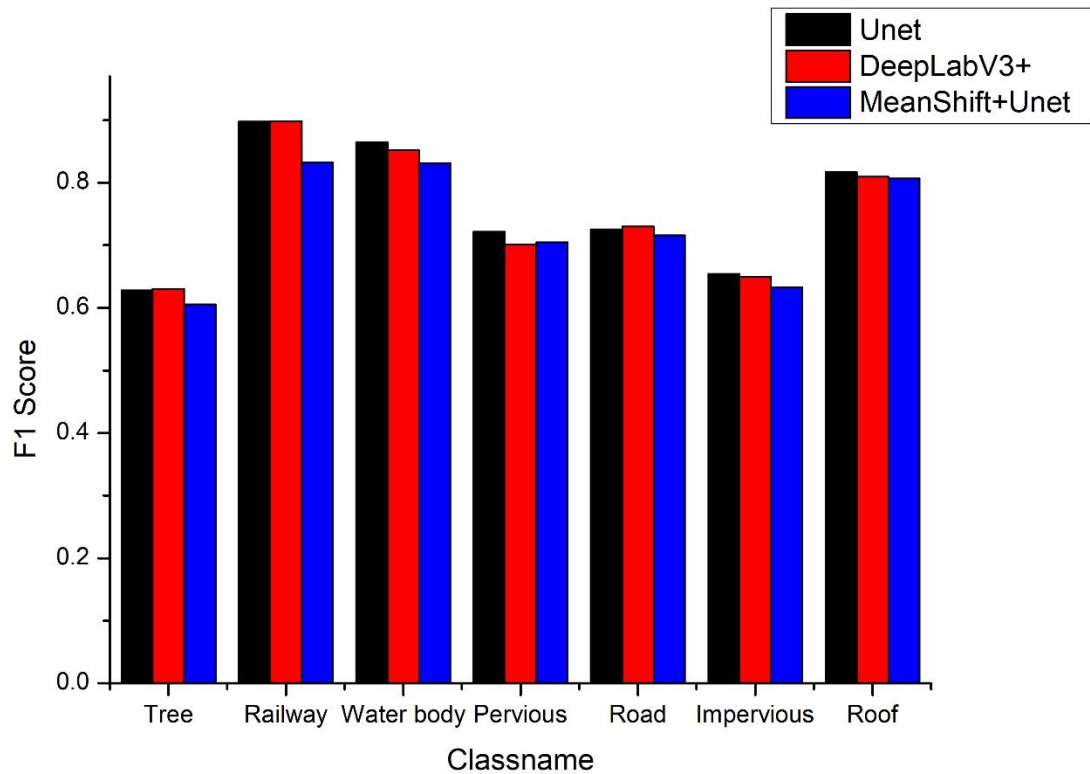


Figure 8 | Illustration of LULC segmentation on the validation set.



**Figure 9** | F1 score of Unet, DeepLabV3+ and MeanShift+Unet in each class.

### Confusion matrix and kappa coefficient

In the performance evaluation strategy, the confusion matrix and kappa coefficient were applied to assess the performance of the three trained models in predicting the LULC of the study catchment.

Five hundred random points were created on the catchment imagery as assessment points, where the number of points in each class is proportional to its corresponding area percentage.

The kappa coefficient for all three trained models exceeded 0.75; the Kappa coefficient values are 0.7929, 0.7593 and 0.7955 for Unet, DeepLabV3+ and MeanShift+Unet, respectively. These kappa coefficients indicate that the classification results of the three trained models are strongly consistent, and there is no bias towards the major classes with large areas. The U\_Accuracy and P\_Accuracy represent the User Accuracy and Prediction Accuracy for reflecting the model's accuracy from different observation channels. For example, in the tree class of Table 1, 16 points lost the correctness among the total 72 assessing points, and 56 points were counted to the correct class within 65 classified samples so that the P\_Accuracy and U\_Accuracy are 0.7778 and 0.8615. [Table 1](#), [Table 2](#) and [Table 3](#) show the confusion matrixes and kappa coefficient of the three trained models.

**Table 1** | Confusion matrix of the classification results achieved by Unet.

Classname	Tree	Railway	Water Body	Pervious	Road	Impervious	Roof	Total	U_Accuracy
Tree	56	0	0	7	0	1	1	65	0.8615
Railway	0	7	0	0	2	1	0	10	0.7000
Water Body	2	0	8	0	0	0	0	10	0.8000
Pervious	7	0	1	66	0	3	0	77	0.8571
Road	3	0	0	0	57	3	0	63	0.9048
Impervious	4	0	0	24	8	79	11	126	0.6270
Roof	0	0	0	0	0	5	150	155	0.9677
<b>Total</b>	72	7	9	97	67	92	162	506	
<b>P_Accuracy</b>	0.7778	1	0.8889	0.6804	0.8507	0.8587	0.9259		0.8360

**P\_Accuracy=Prediction Accuracy | U\_Accuracy = User Accuracy**  
**Overall Accuracy = 0.8360 | Kappa = 0.7929**

**Table 2** | Confusion matrix of the classification results achieved by DeepLabV3+.

Classname	Tree	Railway	Water Body	Pervious	Road	Impervious	Roof	Total	U_Accuracy
Tree	56	0	0	3	4	2	0	65	0.8615
Railway	0	6	0	0	0	1	0	7	0.8571
Water Body	2	0	9	1	0	0	0	12	0.7500
Pervious	6	0	0	67	0	5	1	79	0.8481
Road	1	0	0	2	55	1	0	59	0.9322
Impervious	7	0	0	21	4	71	15	118	0.6017
Roof	0	1	0	3	4	12	146	166	0.8795
<b>Total</b>	72	7	9	97	67	92	162	506	0
<b>P_Accuracy</b>	0.7778	0.8571	1	0.6907	0.8209	0.7717	0.9012		0.8103

**P\_Accuracy=Prediction Accuracy | U\_Accuracy = User Accuracy**  
**Overall Accuracy = 0.8103 | Kappa = 0.7593**

**Table 3** | Confusion matrix of the classification results achieved by MeanShift+Unet.

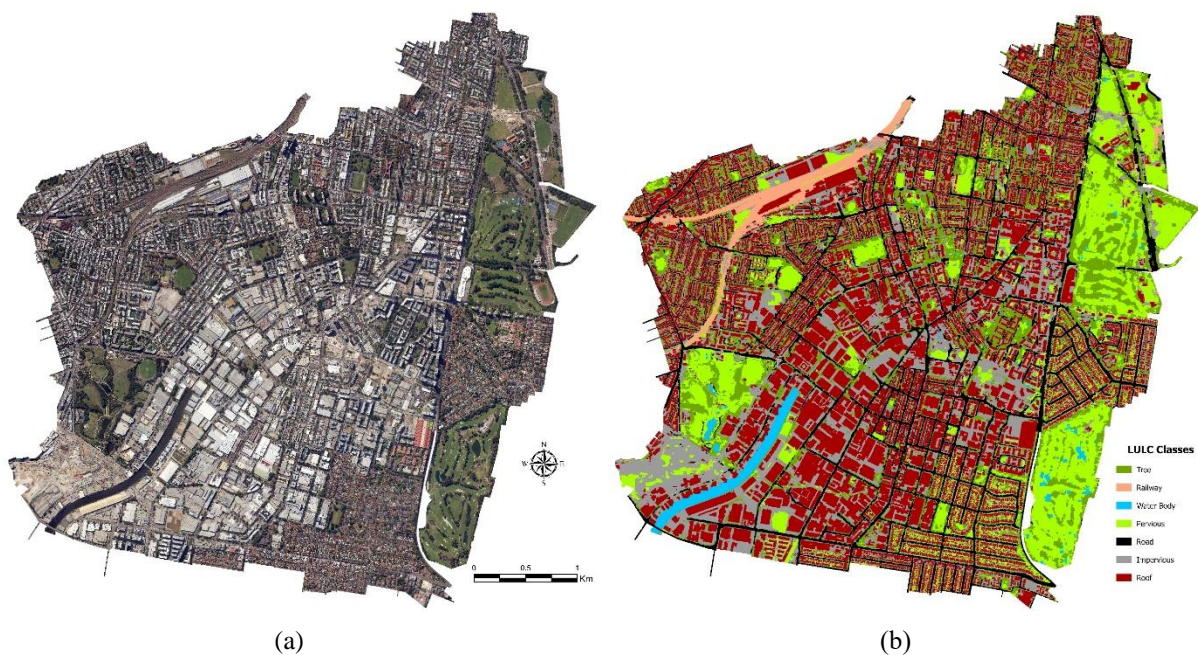
Classname	Tree	Railway	Water Body	Pervious	Road	Impervious	Roof	Total	U_Accuracy
Tree	56	0	0	7	0	1	1	65	0.8615
Railway	0	7	0	0	2	1	0	10	0.700
Water Body	2	0	8	0	0	0	0	10	0.800
Pervious	7	0	1	66	0	3	0	77	0.8571
Road	3	0	0	0	57	3	0	63	0.9048
Impervious	4	0	0	24	8	80	11	127	0.6299
Roof	0	0	0	0	0	4	150	154	0.9740
<b>Total</b>	72	7	9	97	67	92	162	506	
<b>P_Accuracy</b>	0.7778	1	0.8889	0.6804	0.8507	0.8695	0.9259		0.8379

**P\_Accuracy=Prediction Accuracy | U\_Accuracy = User Accuracy**  
**Overall Accuracy = 0.8379 | Kappa = 0.7955**

## Discussion for catchment modelling

This section will discuss the suitability of the test results of the above three methods in two types of catchment modelling systems (Fully-distributed and Semi-distributed modelling systems). Both fully-distributed and semi-distributed models involve considering catchment spatial patterns distribution when determining the runoff generation and routing, where the remote sensing datasets such as imagery and DEM are widely applied in obtaining spatial features for modelling purposes. (Ball et al., 2019; Bell et al., 2009).

The remote sensing-based catchment modelling systems have rigorous requirements for the geometric accuracy of spatial features in estimating the model's parameters. Therefore, the selected index of suitability assessment for catchment modelling should be the P\_Accuracy (Prediction Accuracy) values of the CV model on the test set, which can reflect the classification and segmentation correctness of the CV model on each LULC class. [Table 1](#), [Table 2](#) and [Table 3](#) illustrate that the three methods' overall prediction accuracy have all exceeded 0.8, yielding 0.8360, 0.8103 and 0.8379 for Unet, DeepLabV3+ and MeanShift+Unet, respectively. In addition, the Kappa coefficient of MeanShift+Unet is the highest among the three methods, which means MeanShift+Unet has the best consistency in predicting the LULC of the catchment. The advantage of consistency will be gradually expanded with the increasing of the prediction area, which is beneficial to maintaining the stability and robustness of the parameter estimation model under the changing catchment conditions. Therefore, MeanShift+Unet method has the best performance in extracting LULC information at pixel scale among the three with the overall P\_accuracy of 0.8379. [Figure 10](#) shows the MeanShift+Unet LULC prediction of the entire catchment.



**Figure 10** | Alexandria Canal catchment raw image (a) and MeanShift+Unet prediction of LULC (b).

MeanShift+Unet prediction is also tested at the subcatchment scale under the schema of semi-distributed modelling systems, which have a pre-determined flow direction and the requirement of fewer parameters. The spatial distribution is ignored within the subcatchment scale of the semi-distributed model (Sitterson et al., 2018). For example, (Choi & Ball, 2002) categorized the control parameters of the SWMM into measured and inferred parameters, where the inferred parameters such as impervious ratio, subcatchments length/width, depression storage, infiltration and manning's roughness coefficients are highly related to the classification result and segmentation geometry of catchment LULC expected to be obtained from the MeanShift+Unet prediction. Therefore, the original four sub-classes (railway, road, impervious and roof) treated as the impervious feature were merged into one class (impervious) in the confusion matrix. The new class schema is 'Tree', 'Pervious', 'Waterbody' and 'Impervious'. The purpose of this simplified class schema is to assess the suitability of the semi-distributed modelling systems by ignoring the classification errors among the sub-classes with impervious rainfall responses. [Table 4](#) and [Table 5](#) shows the confusion matrixes of the DeepLabV3+, Unet and MeanShift+Unet after merging classes. One table (Table 5) is used to show the simplified confusion matrix for both Unet and MeanShift+Unet due to their same accuracy distribution among the 500 assessing points. The P\_Accuracy value of 'Impervious' was significantly increased in the three methods under the simplified class schema, in which the 'Impervious' P\_Accuracy occurred from Unet, MeanShift+Unet, and DeepLabV3+ are 0.9847, 0.9847 and 0.9634, respectively. The kappa coefficient of DeepLabV3+ has been improved but is still lower than the other two methods.

Considering the highest Kappa coefficient achieved by MeanShift+Unet in the original class schema, MeanShift+Unet is recommended for estimating the initial parameters of semi-distributed modelling systems. Reasons are as follows: First of all, the semi-distributed model does not require high geometric accuracy of the spatial features within the subcatchment scale, which reduce the probability of models' failure by downgrading the resolution of prediction error. Secondly, the estimated parameters for the semi-distributed model are more accurate as the impervious area prediction accuracy (0.9847) of MeanShift+Unet is very close to the accuracy of manual delineation. In addition, the parameter uncertainty caused by the CV model error is objectively reduced since the number of parameters of the semi-distributed model is much less than that of the fully-distributed model. Finally, the higher Kappa coefficient of MeanShift+Unet method illustrates that its better classification consistency could maintain a stable prediction accuracy on the sizeable urban catchment with high heterogeneity of LULC. However, the number of subcatchments in a metropolis could be very large due to the complex LULC and drainage system controlled land divisions. In the specific implementation of MeanShift+Unet on catchment parameterisation, it is necessary to simplify the large number of subcatchments to a certain extent to reduce the impact of the error caused by the high resolution of spatial features.

Nevertheless, this method still has some significant drawbacks as

- There is a significant gap in achieving accurate geometric parameters of the catchment spatial features at pixel scale by referring to the overall P\_Accuracy. This error will be

gradually enlarged, with the expansion of the study catchment area, and cause the collapse of the entire model eventually.

- The occlusion of tree canopy and buildings' shadows could lead to incorrect determinations of the underlying ground features. Further study about mitigation approach is necessary to reduce their negative effect.
- The spectral similarity between the bare soil and the impervious area, resulting in a large portion of bare soil with pervious rainfall response was classified as impervious feature, which will further lead to the uncertainty of the parameters and the instability of the catchment model.
- Errors and misprediction could occur when using these approaches on other remote sensing databases such as Google and Quickbird due to the pixel spectral variation caused by different image processing methods.

It is recommended to train the deep learning models on the associated remote sensing dataset before applying the approach to a practical catchment model. The LULC class schema used in this study is not fixed. Modellers should set up suitable LULC classes to represent catchment spatial features and correlate to the catchment rainfall-runoff process.

**Table 4** | Confusion matrix of DeepLabV3+ after merging classes.

ClassValue	Tree	Water Body	Pervious	Impervious	Total	U_Accuracy
Tree	56	0	3	6	65	0.8615
Water Body	2	9	1	0	12	0.75
Pervious	6	0	67	6	79	0.8481
Impervious	8	0	26	316	350	0.9028
Total	72	9	97	328	506	
P_Accuracy	0.7778	1	0.6907	0.9634		

P\_Accuracy=Prediction Accuracy | U\_Accuracy = User Accuracy

Overall Accuracy = 0.8854 | Kappa = 0.7721

**Table 5** | Confusion matrix of Unet and MeanShift+Unet after merging classes.

ClassValue	Tree	Water Body	Pervious	Impervious	Total	U_Accuracy
Tree	56	0	7	2	65	0.8615
Water Body	2	8	0	0	10	0.8000
Pervious	7	1	66	3	77	0.8571
Impervious	7	0	24	323	354	0.9124
Total	72	9	97	328	506	
P_Accuracy	0.7778	0.8889	0.6804	0.9847		

P\_Accuracy=Prediction Accuracy | U\_Accuracy = User Accuracy

Overall Accuracy =0.8953 | Kappa = 0.7900

## CONCLUSION

This paper proposes three methods for LULC classification on Alexandria Canal catchment

remote sensing imagery. The study imagery was preprocessed, and the MeanShift algorithm was applied to achieve preliminary segmentation of the catchment LULC objects with similar spatial features. The raw catchment image samples were trained using Unet and DeepLabV3+, and the MeanShift+Unet was trained by the pre-segmented catchment image of MeanShift. Then, the Unet, DeepLabV3+ and MeanShift+Unet worked as the classifier on the study catchment imagery. Finally, the confusion matrix was applied to assessing the suitability of the prediction for catchment modelling systems.

In contrast with the pure Unet and DeepLabV3+, MeanShift+Unet has achieved the best prediction accuracy and excellent consistency in the study catchment. Both original and simplified confusion matrices indicate that the proposed methods are more applicable to estimating spatial parameters for the semi-distributed catchment modelling systems, where the MeanShift+Unet is recommended due to its high accuracy (0.9847) in classifying impervious features. In addition, it is necessary to explore the performance of MeanShift+Unet on multi-sources remote sensing dataset and statistic the balance between the accuracy of the MeanShift+Unet and the catchment scale further, as well as the influence of image noise such as tree canopy and buildings' shadow on the classification accuracy, so as to improve the practicality of the CV model in the catchment modelling field.

## ACKNOWLEDGMENTS

The land zoning map, drainage network and cadastral map of Alexandria Canal catchment are provided by the City of Sydney Council, Sydney, Australia. Sincere thanks are given for the support of the City of Sydney Council and their flood risk management team.

## REFERENCES

- Abulohom, M., Shah, S., & Ghumman, A. (2001). Development of a rainfall-runoff model, its calibration and validation. *Water Resources Management*, 15(3), 149-163.
- Arnold Jr, C. L., & Gibbons, C. J. (1996). Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American planning Association*, 62(2), 243-258.
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
- Bai, L., Xu, C., & Wang, C. (2015). A review of fusion methods of multi-spectral image. *Optik*, 126(24), 4804-4807.
- Ball, J., Babister, M., Nathan, R., Weinmann, P., Weeks, W., Retallick, M., & Testoni, I. (2019). Australian Rainfall and Runoff-A guide to flood estimation. In: Commonwealth of Australia (Geoscience Australia).
- Bell, V., Kay, A., Jones, R., Moore, R., & Reynard, N. (2009). Use of soil data in a grid-based hydrological model to estimate spatial variation in changing flood risk across the UK. *Journal of Hydrology*, 377(3-4), 335-350.
- Beven, K. J. (2012). *Rainfall-runoff modelling : the primer* (Second ed.): Wiley-BlackWell.
- Boegh, E., Thorsen, M., Butts, M., Hansen, S., Christiansen, J., Abrahamsen, P., Hasager,

- C., Jensen, N. O., van der Keur, P., & Refsgaard, J. C. (2004). Incorporating remote sensing data in physically based distributed agro-hydrological modelling. *Journal of Hydrology*, 287(1-4), 279-299.
- Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., & Malo, J. (2011). Remote sensing image processing. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 5(1), 1-192.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). *Encoder-decoder with atrous separable convolution for semantic image segmentation*. Paper presented at the Proceedings of the European conference on computer vision (ECCV).
- Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on pattern analysis and machine intelligence*, 17(8), 790-799.
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1), 1-13.
- Choi, K.-s., & Ball, J. E. (2002). Parameter estimation for urban runoff modelling. *Urban water*, 4(1), 31-41.
- Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5), 603-619.
- Cowie, R., Cox, C., Martin, J.-C., Batliner, A., Heylen, D., & Karpouzis, K. (2011). Issues in data labelling. In *Emotion-oriented systems* (pp. 213-241): Springer.
- Dawson, R., Speight, L., Hall, J., Djordjevic, S., Savic, D., & Leandro, J. (2008). Attribution of flood risk in urban areas. *Journal of Hydroinformatics*, 10(4), 275-288.
- Esri. (2021). World Imagery. In *DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community*.
- Foley, J. D., Van Dam, A., Feiner, S. K., Hughes, J. F., & Phillips, R. L. (1994). *Introduction to computer graphics* (Vol. 55): Addison-Wesley Reading.
- Foody, G. M. (2004). Mapping land cover from remotely sensed imagery for input to hydrological models. In *Neural Networks for Hydrological Modelling* (pp. 269-289): Swets and Zeitlinger Lisse.
- Fukunaga, K., & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1), 32-40.
- Gan, T. Y., Dlamini, E. M., & Biftu, G. F. (1997). Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology*, 192(1-4), 81-103.
- Ghassemian, H. (2016). A review of remote sensing image fusion methods. *Information Fusion*, 32, 75-89.
- Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, 37(3), 362-386.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Overview of supervised learning. In *The elements of statistical learning* (pp. 9-41): Springer.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask r-cnn*. Paper presented at the Proceedings of the IEEE international conference on computer vision.

- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Identity mappings in deep residual networks*. Paper presented at the European conference on computer vision.
- Holman -Dodds, J. K., Bradley, A. A., & Potter, K. W. (2003). Evaluation of hydrologic benefits of infiltration based urban storm water management 1. *JAWRA Journal of the American Water Resources Association*, 39(1), 205-215.
- Hu, T., Yang, J., Li, X., & Gong, P. (2016). Mapping urban land use by using landsat images and open social data. *Remote sensing*, 8(2), 151.
- Huang, B., Zhao, B., & Song, Y. J. R. S. o. E. (2018). Urban land-use mapping using a deep convolutional neural network with high spatial resolution multi-spectral remote sensing imagery. *214*, 73-86.
- Jha, A. K., Bloch, R., & Lamond, J. (2012). *Cities and flooding: a guide to integrated urban flood risk management for the 21st century*: The World Bank.
- Klein, A. G., & Barnett, A. C. (2003). Validation of daily MODIS snow cover maps of the Upper Rio Grande River Basin for the 2000 -2001 snow year. *Remote Sensing of Environment*, 86(2), 162-176.
- Li, H., Ding, L., Ren, M., Li, C., & Wang, H. (2017). Sponge city construction in China: A survey of the challenges and opportunities. *Water*, 9(9), 594.
- Li, K., Li, G., Wang, Y., Huang, Y., Liu, Z., & Wu, Z. (2021). *CrowdRL: An End-to-End Reinforcement Learning Framework for Data Labelling*. Paper presented at the 2021 IEEE 37th International Conference on Data Engineering (ICDE).
- Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3), 512-530.
- Mackay, D. S., Samanta, S., Nemani, R. R., & Band, L. E. (2003). Multi-objective parameter estimation for simulating canopy transpiration in forested watersheds. *Journal of Hydrology*, 277(3-4), 230-247.
- Niemi, T. J., Kokkonen, T., Sillanpää, N., Setälä, H., & Koivusalo, H. (2019). Automated Urban Rainfall - Runoff Model Generation with Detailed Land Cover and Flow Routing. *Journal of Hydrologic Engineering*, 24(5), 04019011.
- NSW Government Spatial Service. (2021). Historical, Aerial and Satellite Imagery. Retrieved from [https://www.spatial.nsw.gov.au/products\\_and\\_services/aerial\\_and\\_historical\\_imagery](https://www.spatial.nsw.gov.au/products_and_services/aerial_and_historical_imagery)
- Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.
- Peel, M. C., & McMahon, T. A. (2020). Historical development of rainfall - runoff modeling. *Wiley Interdisciplinary Reviews: Water*, 7(5), e1471.
- Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63.
- Rezaee, M., Mahdianpari, M., Zhang, Y., Salehi, B. J. I. J. o. S. T. i. A. E. O., & Sensing, R. (2018). Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. *11(9)*, 3030-3039.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation*. Paper presented at the International Conference on Medical image computing and computer-assisted intervention.

- Salvadore, E., Bronders, J., & Batelaan, O. (2015). Hydrological modelling of urbanized catchments: A review and future directions. *Journal of Hydrology*, *529*, 62–81.
- Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, *85*(3), 257–268. doi:10.1093/ptj/85.3.257
- Sitterson, J., Knightes, C., Parmar, R., Wolfe, K., Avant, B., & Muche, M. (2018). An overview of rainfall–runoff model types.
- Stehman, S. (1996). Estimating the kappa coefficient and its variance under stratified random sampling. *Photogrammetric Engineering and Remote Sensing*, *62*(4), 401–407.
- Szeliski, R. (2010). *Computer vision: algorithms and applications*: Springer Science & Business Media.
- Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*.
- Thomas, C., Ranchin, T., Wald, L., & Chanussot, J. (2008). Synthesis of multi-spectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics. *IEEE Transactions on GeoScience and Remote sensing*, *46*(5), 1301–1312.
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, *2018*.
- Wenzhi Zhao, Shihong Du, Emery, & J, W. (2017). Object-Based Convolutional Neural Network for High-Resolution Imagery Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), 3386–3396. doi:https://doi.org/10.1109/JSTARS.2017.2680324
- Willner, S. N., Otto, C., & Levermann, A. (2018). Global economic response to river floods. *Nature Climate Change*, *8*(7), 594–598.
- Wu, J., Wang, D., & Bauer, M. E. (2005). Image-based atmospheric correction of QuickBird imagery of Minnesota cropland. *Remote Sensing of Environment*, *99*(3), 315–325.
- Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, *122*, 2–10.
- Yang, J., & Li, P. (2015). Impervious surface extraction in urban areas from high spatial resolution imagery using linear spectral unmixing. *Remote Sensing Applications: Society and Environment*, *1*, 61–71.
- Zaman, M. A., Rahman, A., & Haddad, K. (2012). Regional flood frequency analysis in arid regions: A case study for Australia. *Journal of Hydrology*, *475*, 74–83.
- Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3–11): Springer.