U-NET: SEMANTIC SEGMENTATION OF HIGH-RESOLUTION PADDY PARCELS IMAGERY USING ARCGIS PRO DEEP LEARNING ANALYST

Akhmalluddin Mohd Sujai*, Nurshafiza Zanal Abiddin, Nurul Aina Abdul Aziz, Bismawaty Hassan, Syazwani Basir, and Zuraimi Suleiman
Malaysian Space Agency (MYSA), No.13, Jalan Tun Ismail, 50480 Kuala Lumpur, Malaysia
Email: akhmal@mysa.gov.my*; nurshafiza@mysa.gov.my; aina@mysa.gov.my; bismawaty@mysa.gov.my; nursyazwani@mysa.gov.my; zuraimi@mysa.gov.my

KEY WORDS: Deep Learning, U-Net, Paddy, High Resolution Imagery, ArcGIS Pro

ABSTRACT: Malaysia Space Agency (MYSA) and the Department of Agriculture Malaysia (DOA) have collaborated to develop the Geospatial Information System (MakGeoPadi) to identify the exact areas of paddy cultivation throughout Malaysia to guarantee a sufficient national food supply. Statistics have shown that rice crop yields have decreased annually by 2,500 hectares, or an average of 1.2 percent since 2018. It is difficult to identify the current status of land use for paddy parcels every year in the 12 granary areas. Frequent land-use changes of paddy parcels into other crops or buildings contributed significantly to changes in paddy cultivation area changes. Deep convolutional neural network (DCNN) algorithms are gaining momentum, especially in pixel classification (image segmentation) without human intervention and U-Net is one of the most successful DCNN architectures. This study performed the entire U-Net deep learning workflow using ArcGIS Pro with the Deep Learning toolset. The objective of this study is to evaluate the U-Net model in classifying paddy cultivation area into i) active paddy parcel (PA) including four (4) major paddy-planting activities which are ploughing, irrigating, planting, and harvesting; ii) miscellaneous paddy parcel (PP); iii) permanent structures (SK); and iv) permanent crop (TK) with ResNet 34 as the backbone model in ArcGIS Pro vr. 2.9. A three-band (RGB) Pleiades satellite image of the Integrated Agriculture Development Area (IADA) Barat Laut Selangor was used to generate 16,000 samples of images and labels with a specified 224 x 224 pixel size. The training process took three hours and the Graphics Processing Unit (GPU) and NVIDIA RTX A2000 graphics card were used. The findings indicate that U-Net demonstrates strong performance across diverse image circumstances and generates precise and accurate segmentations. However, it can encounter difficulties when dealing with imbalanced classes, in which certain classes contain much fewer samples than others. To overcome this challenge, U-Net's segmentation performance can be improved by addressing class imbalance through techniques such as data augmentation, sample weighting, or the use of class-specific loss functions. Various aspects of training data preparation can be explored in the future to improve the deep learning semantic segmentation result.

1. INTRODUCTION

Oryza sativa L. or paddy rice, is one of the world's most important crops, with half of the world's population consuming rice daily (Rahman & Zhang 2023; Wahab et al., 2022; Omar et al., 2019). Asia produces more than 90% of the world's rice, with China being the greatest producer, followed by India, Indonesia, Bangladesh, and Vietnam (Bunawan et al., 2014; Brar et al., 2011). Malaysia has the smallest total area of paddy cultivating in the Southeast Asian region, with 689,268 hectares (Firdaus et al., 2020), and based on Annual Rice Consumption in Malaysia for 2019-2024, rice consumption in Malaysia has been increasing since 2019. However, paddy rice yield can vary due to numerous factors such as pests, diseases, climate change, and decreased paddy crop area, raising concerns about the sustainability of rice production and food security in the region (Omar et al., 2019).

Despite the widespread use of remote sensing studies to map paddy acreage, it is still challenging to map paddy fields in tropical areas where paddy is cultivated all year with a various of planting schedules and crop cycles. The biggest impediment to mapping yield potential is insufficient information on paddy-growing regions and yield potential. Nowadays, the manual annotation of high-resolution satellite images via visual digitization is used to determine the type of land utilisation in paddy fields, which is highly dependent on the competence of individuals to analyse the paddy field satellite images. However, this practice was found to be costly and time-consuming, hence an automated approach is needed. Several deep learning network topologies for pixel-wise image labelling, also known as semantic image segmentation, have been presented. Semantic image segmentation refers to the process of associating each individual pixel of an image with a predefined class label (Plath et al., 2009). Deep convolutional neural network (DCNN) algorithms have gained traction, particularly in non-human pixel classification (image segmentation). U-Net is one of the most successful DCNN architectures. Accurate information on paddy planting areas obtained from the data can provide a scientific foundation for agricultural production, forecasting and evaluating rice output, forecasting food prices, and arranging and planning national food production (Soh et al., 2023).

To address this issue, the Malaysian Space Agency (MYSA) and the Malaysian Department of Agriculture (DOA) collaborated to develop the Paddy Geospatial Information System (MakGeoPadi), which identifies the proper paddy planting regions throughout Malaysia. The MakGeoPadi system's primary role is to determine the area of 12 granaries throughout Malaysia. Meanwhile, precise paddy segmentation is crucial since land use changes frequently influence the actual area of paddy cultivation. The segmentation of paddy lots is divided into four categories, namely i) active paddy parcel (PA) including four (4) major paddy-planting activities which are ploughing, irrigating, planting, and harvesting; ii) miscellaneous paddy parcel (PP); iii) permanent structures (SK); and iv) permanent crop (TK) at lot level is significant due to the frequent change of land use yearly to meet the needs of the National Crop Cutting Survey (CCS), which must be reported annually and influences the amount of government subsidies given to the agricultural agency.

The objective of this study is to develop the U-Net paddy rice detection model to generate data that are qualified for reliable decision-making. To pursue higher accuracy, deep-learning applications based on the U-Net model were implemented into four categories: (i) active paddy parcel (PA), (ii) miscellaneous paddy parcel (PP), iii) permanent structures (SK) and iv) permanent crop (TK) with ResNet 34 as the backbone model in ArcGIS Pro vr. 2.9. The ArcGIS Pro with the Deep Learning Toolbox was used in this study to complete the entire U-Net deep learning workflow. A three-band (RGB) Pleiades satellite image of the Integrated Agriculture Development Area (IADA) Barat Laut Selangor were used to generate the images in this study.

2. METHODOLOGY

2.1 Study Area

IADA Barat Laut Selangor was selected as a research site due to being the largest producer of rice in Malaysia with a yield exceeding 5 t/ha and contributing significantly to the country's food security ambition of achieving a self-sufficiency level (SSL) of 70% for rice (Omar et al., 2019). This area is situated within Selangor State and is visually represented in Figure 1 through a series of multi-temporal satellite images. Covering an expansive 17,741 hectares (Sistem MakGeoPadi, 2023), this region is characterized by its level terrain and serves as a granary, hosting a diverse array of agricultural crops. Notably, paddy fields dominate the landscape, accompanied by oil palm, vegetables, fruits, and undeveloped land interwoven with residential and roadway infrastructure (Abiddin et al., 2020).

A notable feature of this location is its systematic irrigation network, which facilitates the cultivation of short-term crops in two cycles per year. Acknowledged as one of Malaysia's most prolific rice-producing areas, this fertile expanse played a pivotal role in generating an estimated 155,631 metric tonnes of rice in 2021 (Malaysian Department of Agriculture, 2022).

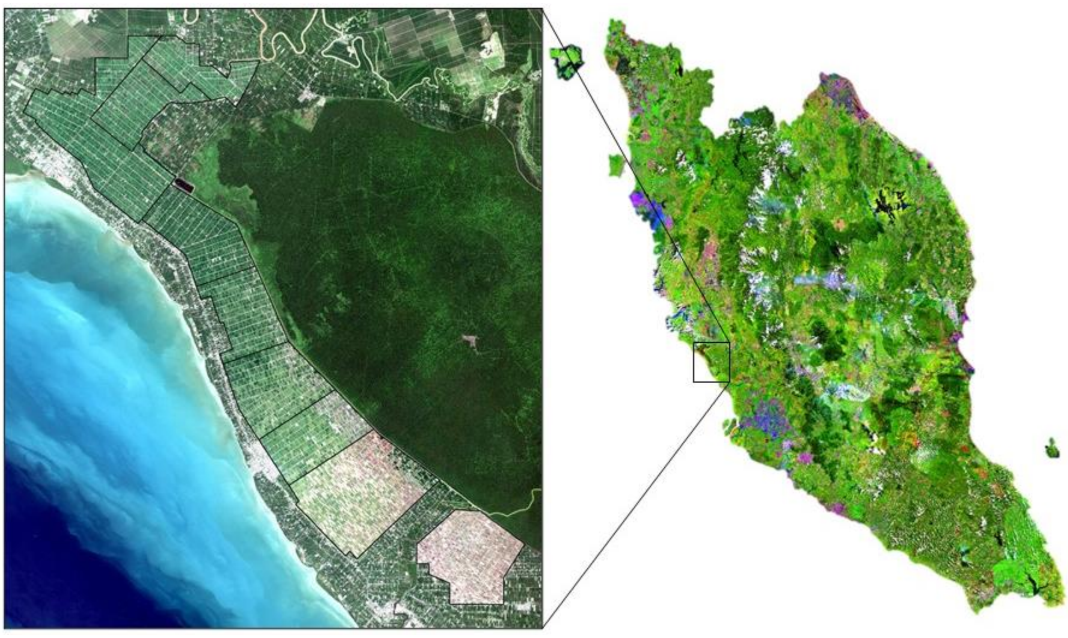


Figure 1. IADA Barat Laut Selangor Granary Area

2.2 Dataset Preparation

2.2.1 Satellite Image Processing: Multi-temporal Pleiades satellite images were acquired that consist of i) cultivated paddy areas that include four (4) major paddy-planting activities, which are ploughing, irrigating, planting and harvesting; and ii) uncultivated paddy areas that include roadways, houses, and other land usages. The three-band, true colour (RGB) satellite imagery underwent enhancement processes to enhance colour presentation and contrast. The improved images were then subset to exclude non-granary areas of the image scene.

2.2.2 Ground Truth Preparation: The National Digital Cadastral Database (NDCDB) lots were sourced from the Department of Survey and Mapping Malaysia (JUPEM), the authoritative entity in land surveying. These lots were segmented manually into four (4) categories: i) active paddy parcel (PA) including four (4) main paddy-planting activities which are ploughing, irrigating, planting, and harvesting; ii) miscellaneous paddy parcel (PP); iii) permanent structures (SK); and iv) permanent crop (TK) by overlaying the cadastral lot with the multi-temporal Pleiades satellite images. The segmentation output was prepared in a standard shapefile GIS format in order to produce ground truth samples for deep learning training. The precision of the segmentation was verified through an on-site verification programme by the Integrated Agriculture Development Area (IADA) Barat Laut Selangor as the authorized department. Figure 2 shows a sample of the Pleiades image and the ground truth.

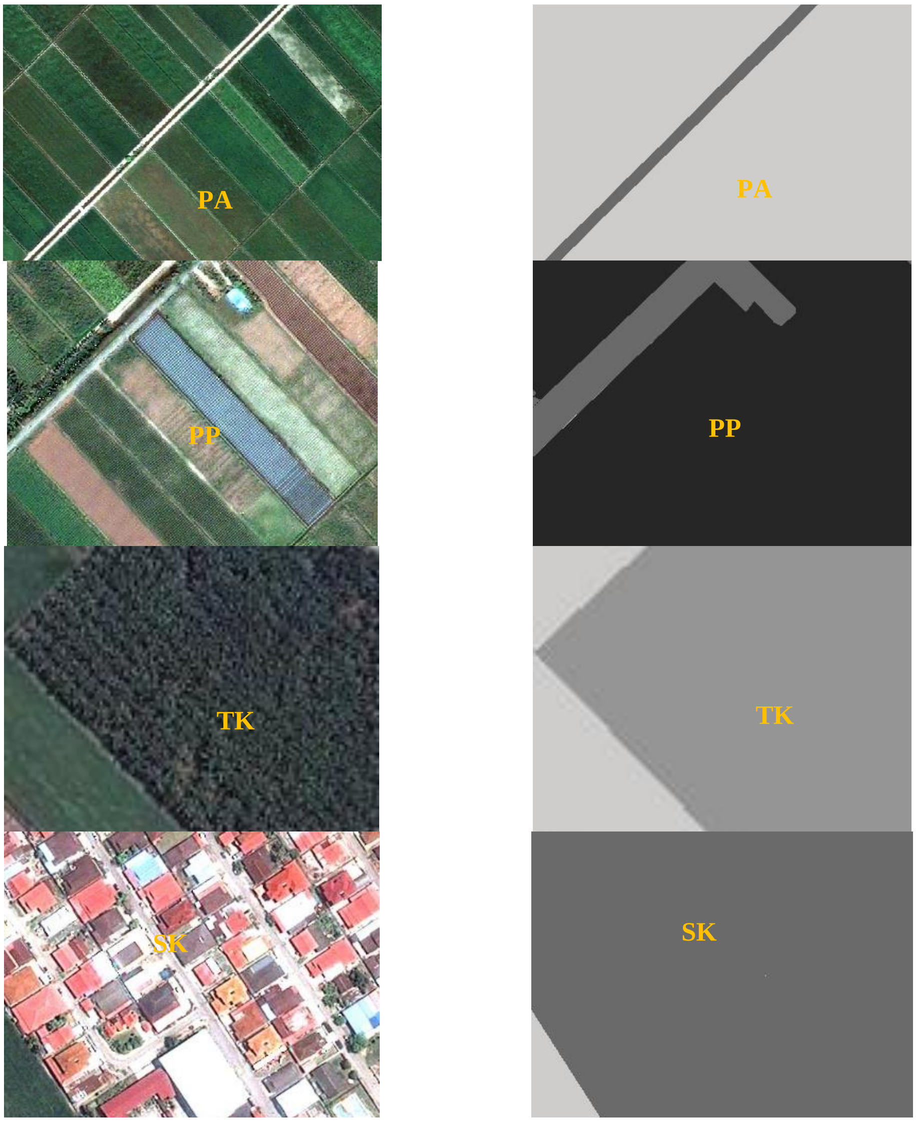


Figure 2. Some samples of cropped satellite image and theirs ground truth labels.

2.3 Training Dataset Extraction

The training dataset was generated by utilizing an unsigned 8-bit satellite image in RGB TIF format along with a ground truth polygon shapefile. This combination enabled the dataset to be created. The segmented polygon was labelled with four different class labels: PA (1), PP (2), SK (3), and TK (4). Both the image and ground truth data were divided into pairs of tiles that corresponded to the same geographic region, as seen in Figure 3. The RGB values of the input features were represented by one tile, and the class number assigned to each pixel based on the training polygons was represented by the second tile. A total of 16,000 samples consisting of images and corresponding ground truths with dimensions of 224 x 224 pixels were generated for the training of the U-Net Deep Learning Network. The entire process, from extraction to analysis was carried out using ESRI's ArcGIS Pro software under the Train Deep Learning Model tools.

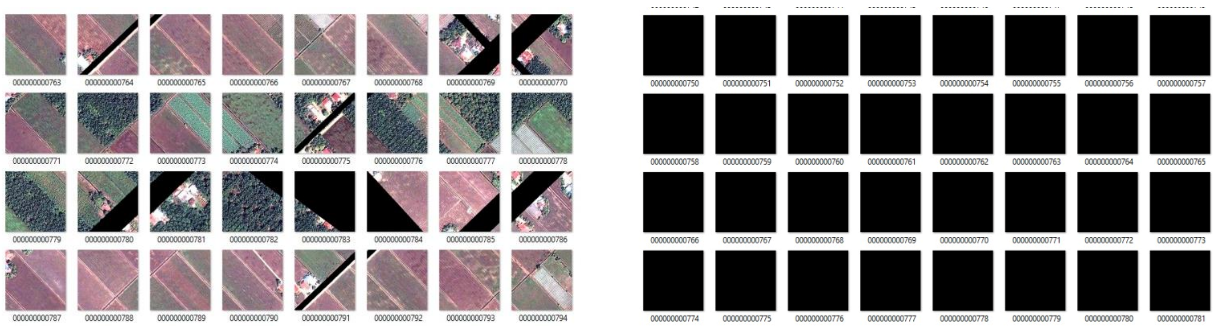


Figure 3. Image and ground truth tiles

2.4 U-Net Deep Learning Network Training

Semantic segmentation involves categorizing individual image pixels into specific classes. Semantic segmentation holds a significant position in enhancing image comprehension and is indispensable for various image analysis assignments, offering both accurate and expedited segmentation outcomes (Pashaei et al., 2020). U-Net is a well-known image segmentation algorithm that was originally invented and first used for biomedical image segmentation (Ronneberger et al., 2015). The U-Net architecture utilizes an autoencoder concept to generate a compact latent representation for segmentation. It consists of two main parts: the encoder captures image context through convolutional and pooling layers, while the decoder facilitates localization using transposed convolutions (Kunapuli et al., 2021). The structure of U-Net is shown in Figure 4. In this study, the U-Net model architecture is proposed with the ResNet-34 backbond model. Training lasted three hours and was conducted in ArcGIS Pro software with a graphics processing unit (GPU) and an NVIDIA RTX A2000 graphics card. The training process produces 95.5% accuracy, as shown in Figure 5 with an EMD file of the trained model which, will be used in image classification.

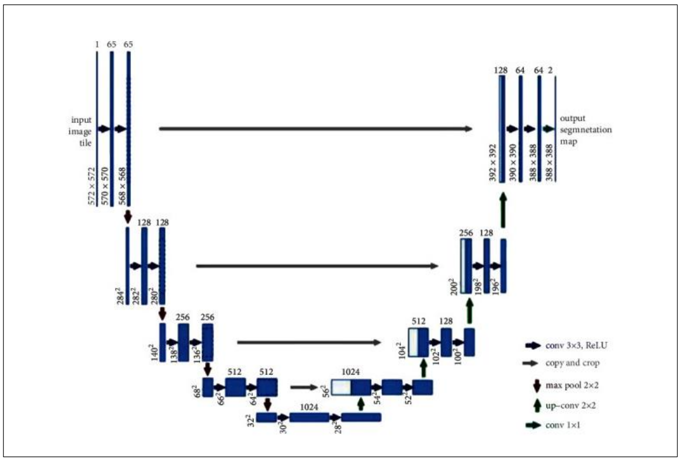
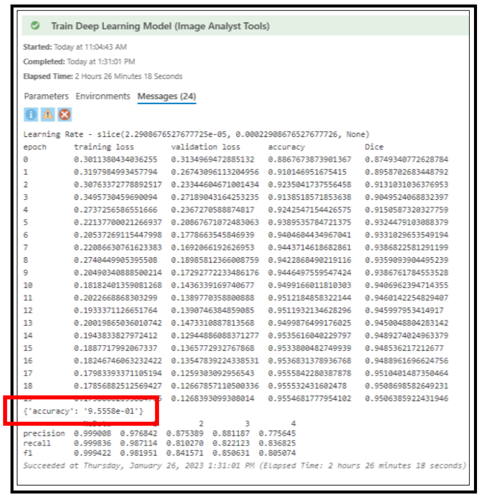


Figure 4. Illustration of U-Net convolution network structure. The left side of the U-shape is the encoding stage, also called contraction path with each layer consisting of two 3 ∗ 3 convolutions with ReLu activation and a 2 ∗ 2 maximum pooling layer. The right side of the U-shape, also called expansion part, consists of the decoding stage and the upsampling process that is realized via 2 ∗ 2 deconvolution to reduce the quantity of input channels by half (Ronneberger et al., 2015).



Train Deep Learning Model (Image Analyst Tools)

Started: Today at 11:04:43 AM

Completed: Today at 1:31:01 PM

Elapsed Time: 2 Hours 26 Minutes 18 Seconds

Parameters Environments Messages (24)

Learning Rate - slice(2.2908676527677725e-05, 0.00022908676527677726, None)

epoch	training loss	validation loss	accuracy	Dice
0	0.3011380434036255	0.3134969472885132	0.8867673873901367	0.8749340772628784
1	0.3197984993457794	0.26743096113204956	0.910146951675415	0.8958702683448792
2	0.30763372778892517	0.23344604671001434	0.9235041737556458	0.9131031036376953
3	0.3495730459690094	0.27189043164253235	0.9138518571853638	0.9049524068832397
4	0.2737256586551666	0.2367270588874817	0.9242547154426575	0.9150587320327759
5	0.22137700021266937	0.20867671072483063	0.9389535784721375	0.9324479103088379
6	0.20537269115447998	0.1778663545846939	0.9404604434967041	0.9331029653549194
7	0.22086630761623383	0.1692066192626953	0.9443714618682861	0.9386822581291199
8	0.2740449905395508	0.18985812366008759	0.9422868490219116	0.9359093904495239
9	0.20490340888500214	0.17292772233486176	0.9446497559547424	0.9386761784553528
10	0.18182401359081268	0.1436339169740677	0.9499166011810303	0.9406962394714355
11	0.2022668868303299	0.1389770358800888	0.9512184858322144	0.9460142254829407
12	0.1933371126651764	0.1390746384859085	0.9511932134628296	0.945997953414917
13	0.20019865036010742	0.1473310887813568	0.9499876499176025	0.9450048804283142
14	0.1943833827972412	0.12944886088371277	0.9535616040229797	0.9489274024963379
15	0.1887717992067337	0.1365772932767868	0.9533800482749939	0.948536217212677
16	0.18246746063232422	0.13547839224338531	0.9536831378936768	0.9488961696624756
17	0.17983393371105194	0.1259303092956543	0.9555842280387878	0.9510401487350464
18	0.17856882512569427	0.12667857110500336	0.955532431602478	0.9508698582649231
[illegible]	[illegible]	0.1268393099308014	0.9554681777954102	0.9506385922431946

{'accuracy': '9.5558e-01'}

	[illegible]	[illegible]	2	3	4
precision	0.999008	0.976842	0.875389	0.881187	0.775645
recall	0.999836	0.987114	0.810270	0.822123	0.836825
f1	0.999422	0.981951	0.841571	0.850631	0.805074

Succeeded at Thursday, January 26, 2023 1:31:01 PM (Elapsed Time: 2 hours 26 minutes 18 seconds)

Figure 5. Trained model accuracy

2.5 Validation of trained U-Net Model

The validation of the trained model is carried out to evaluate the model's performance and ensure that the model can generalize successfully with the new or unseen dataset. The study employs a train-test dataset with a ratio of 70:30. Under the Classify Pixels Using Deep Learning tools, the EMD file created in the previous step is used to test the model. At this stage, we can compare the testing accuracy to the training accuracy.

3. RESULT AND DISCUSSION

3.1 U-Net Semantic Segmentation Output

The total area of 17,00 hectares of segmented images was successfully produced from the tested dataset as illustrated in Figure 6. The pixel classification process took about 30 minutes, which is significantly shorter than the 3 hours required for training. Based on the observations, major classes such as PA and artificial features like permanent structure (SK) are well-defined. On the other hand, class PP which has almost similar features to PA and TK tends to be misclassified with certain different conditions. Figure 7 highlights several classification issues.

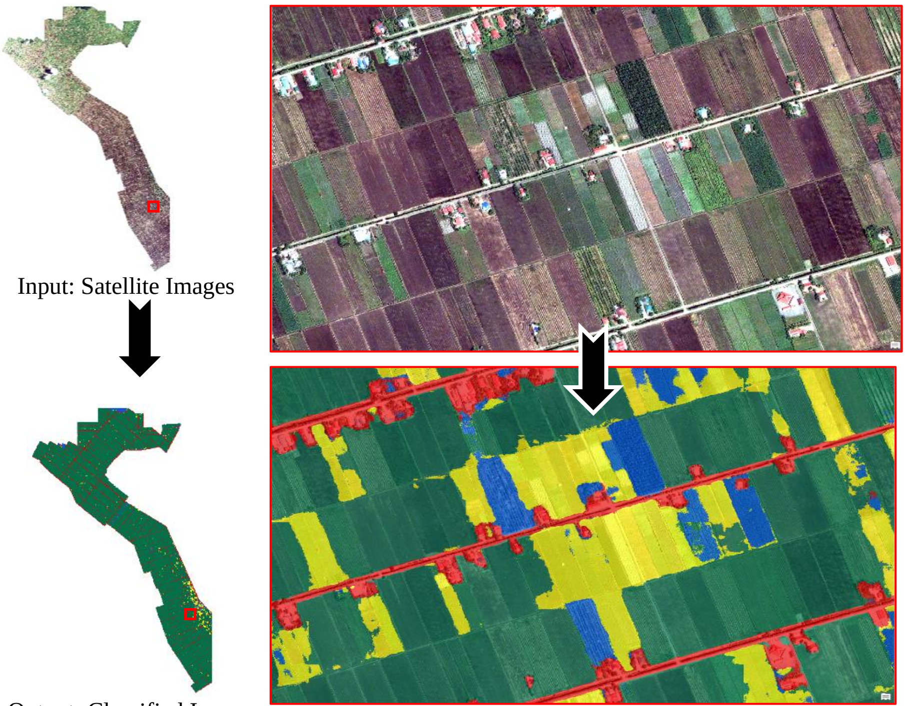
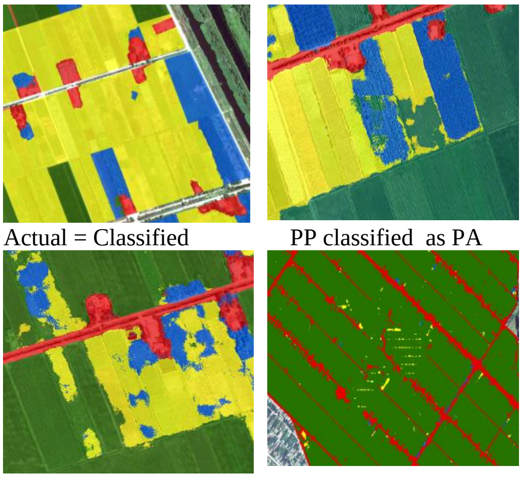


Figure 6. U-Net Classification Output



Actual = Classified

PP classified as PA

PP classified as TK

Cloud classified as PA

	1	PA
	2	PP
	3	SK
	4	TK

Figure 7. Analysis of Classification Output

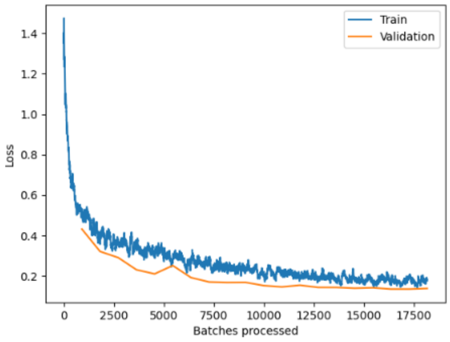
3.2 Accuracy Assessment

The accuracy assessment is performed using the ArcGIS Pro Compute Confusion Matrix tool to compare the accuracy of the test dataset against the training dataset. There are 200 randomly selected points, with 50 points allotted to each class and two fields titled classified and ground truth. Figure 8 presents the confusion matrix resulting from the accuracy assessment tool. Miscellaneous paddy parcel (C_2) has the lowest user accuracy (82%), with 8 samples misclassified as paddy parcel (C_1), while active paddy parcel (C_1) shows 100% true classification. In terms of producer accuracy, permanent structure (C_4) shows 100% accuracy, followed by permanent crop (C_3), which has 96% accuracy and just 2 incorrectly classified samples out of 51 samples. Again, active paddy parcel (C_1) obviously contributes to high false negatives (83%), with 8 samples out of 60 from (C_2) wrongly classified as (C_1). The overall accuracy assessment achieved in this demonstration was 92% with a Kappa of 0.89, indicating nearly perfect agreement.

OBJECTID *	ClassValue	C_1	C_2	C_3	C_4	Total	U_Accuracy	Kappa
1	C_1	50	0	0	0	50	1	0
2	C_2	8	41	1	0	50	0.82	0
3	C_3	1	0	49	0	50	0.98	0
4	C_4	1	4	1	44	50	0.88	0
5	Total	60	45	51	44	200	0	0
6	P_Accuracy	0.833333	0.911111	0.960784	1	0	0.92	0
7	Kappa	0	0	0	0	0	0	0.893333
Click to add new row.								

Figure 8. Confusion matrix result

Based on the observation, the Miscellaneous paddy parcel (C_2) category is imbalanced, which can be detrimental to the learning process since it is biased. In addition, the number of training samples having the expected segmentation labels should be large and evenly distributed for each label. Ideally, each labelled categories should have an equal number of observations (Johnson & Khoshgoftaar 2019). The tested dataset, which included noise such as clouds also will also pose a challenge to the model. The fact that, the accuracy of test dataset is somewhat slightly lower than the accuracy of training dataset (95%) indicates that the model is overfitting as shown in Figure 9. There are several methods for avoiding overfitting in the model, including employing the cross-validation method, early stopping the training and dropping out (Cogswell et.al, 2015).



Analysis of the model

Per class metrics:

	NoData	1	2	3	4
precision	0.998284	0.977517	0.838602	0.859958	0.758593
recall	0.999379	0.982600	0.792452	0.838209	0.797957
f1	0.998831	0.980052	0.814874	0.848944	0.777777

Figure 9. Graph shows overfitting model

3.3 Post-Processing and Integration to MakGeoPadi System

The raw classification output from U-Net must go through several post-processing steps especially when related to paddy cultivation activities that are managed with fixed lot boundaries. Thus, the ArcGIS Pro spatial analyst tool is critical for raster-to-vector conversion. Several strategies, such as polygon intersection, elimination, smooth polygon, layer organization and database are used to meet the requirements of the existing system. The final output of the paddy parcel that has been published in MakGeoPadi is shown in Figure 10.

Figure 10. MakGeoPadi interface with updated classification of paddy parcel area

4. CONCLUSION

In conclusion, a deep convolutional neural network (DCNN) using U-Net segmentation is able to semantically segment the paddy cultivation area into four (4) categories, which are: i) active paddy parcel (PA) includes four (4) main paddy-planting activities which are ploughing, irrigating, planting and harvesting; ii) miscellaneous paddy parcel (PP); iii) permanent structures (SK); and iv) permanent crop (TK) with an accuracy of 92%. Various aspects of preparing training data can be investigated in the future to improve the deep learning semantic segmentation results.

5. ACKNOWLEDGEMENT

The authors would like to acknowledge DOA as a strategic partnership in MakGeoPadi System development and IADA Barat Laut Selangor in the operationalization of MakGeoPadi System.

6. REFERENCES

Abiddin, N.Z., Hasan, Z.A., Zulkifley, M.A., Abdani, S.R., Hassan, B., Azmi, A., & Aziz, N.A.A. 2020. Deep Learning Segmentation Technique for Mapping and Updating Paddy Cultivations at Plot Level Using High Resolution Satellite Data. *Asean Conference on Remote Sensing*, pp. 1-8.

Annual rice consumption in Malaysia 2019-2024. Statista Research Department, May 25, 2023; https://www.statista.com/statistics/1370840/malaysia-annual-rice-consumption/

Brar, D.S. & Singh, K. 2011. Oryza. In Wild crop relatives: genomic breeding and resources: cereals; Kole, C., Eds; Springer-Verlag, Berlin, Heidelberg, pp. 321-366.

Bunawan, H., Dusik, L., Bunawan, S.N., & Amin, N.M. 2014. Rice tungro disease: From identification to disease control. *World Applied Sciences Journal*, Vol. 31(6), pp.1221-1226.

Cogswell, M., Ahmed, F., Girshick, R., Zitnick, L., & Batra, D. 2015. Reducing overfitting in deep networks by decorrelating representations. *arXiv preprint arXiv:*Vol 4, pp. 1-10

Firdaus, R.R., Leong Tan, M., Rahmat, S.R., & Senevi Gunaratne, M. 2020. Paddy, rice and food security in Malaysia: A review of climate change impacts. *Cogent Social Sciences*, Vol. 6(1), pp. 1818373

Johnson, J. M., & Khoshgoftaar, T M. 2019. Survey on deep learning with class imbalance. *Journal of Big Data*, Vol. 6(1), pp. 1-4.

Kunapuli, S.S., & Bhallamudi, P.C. 2021. A review of deep learning models for medical diagnosis. In *Machine Learning, Big Data, and IoT for Medical Informatics*, Elsevier, Oxford, pp. 389-404.

Omar, S.C., Shaharudin, A., & Tumin, S.A. 2019. The status of the paddy and rice industry in Malaysia. Khazanah Research Institute. Kuala Lumpur. Malaysia.

Pashaei, M., Starek, M.J., Kamangir, H., & Berryhill, J. 2020. Deep learning-based single image super-resolution: an investigation for dense scene reconstruction with UAS photogrammetry. *Remote Sensing*, Vol. 12(11), pp. 1757.

Plath, N., Toussaint, M., & Nakajima, S. 2009. Multi-class image segmentation using conditional random fields and global classification. In *Proceedings of the 26th annual international conference on machine learning*, Vol. 1, pp. 817-824).

Rahman, A.R., & Zhang, J. 2023. Trends in rice research: 2030 and beyond. *Food and Energy Security*, Vol. 12(2), pp.390.

Ronneberger, O., Fischer, P., & Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, Vol. 18, pp. 234-241.

Soh, N.C., Shah, R.M., Giap, S.G.E., Setiawan, B.I., & Minasny, B. 2023. Automated near-real-time mapping and monitoring of rice growth extent and stages in Selangor Malaysia. *Remote Sensing Applications: Society and Environment*, Vol. 1(1), pp. 100993.

Sistem Maklumat Geospatial Tanaman Padi, Retrieved August 18, 2023; http://makgeopadi.mysa.gov.my/jelapang/_main

Statistik Tanaman (Sub-Sektor Tanaman Makanan). Unit Geospatial Pertanian dan Statistik Bahagian Perancangan Strategik Jabatan Pertanian Semenanjung Malaysia, 2022. Malaysia, pp. 56.

Wahab, W.A., Talip, N., Basir, S., Akbar, M.A., Saad, M.F.M., & Bunawan, H. 2022. Disease Development and Discovery of Anatomically Resistant Features towards Bacterial Leaf Streak in Rice. *Agriculture*, Vol. 12(5), pp. 629.