IJECBE (2024), 2, 3, 383–394
Received (1 June 2024) / Revised (28 June 2024)
Accepted (4 July 2024) / Published (30 September 2024)
https://doi.org/10.62146/ijecbe.v2i3.57
https://ijecbe.ui.ac.id
ISSN 3026-5258

International Journal of Electrical, Computer and Biomedical Engineering

RESEARCH ARTICLE

Implementation of U-Net for Paddy Field Mapping Using Very High-Resolution Satellite Imagery

Faiz Khairul Isbat,* Mia Rizkinia, and Dodi Sudiana

Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia
*Corresponding author. Email: faiz.khairul@ui.ac.id

Abstract

Mapping paddy fields using remote sensing is one method that can be used to determine the number of paddy fields, especially in Indonesia. Using this method can increase effectiveness in agricultural resource management. This research using one of very high resolution optical satellite imageries, namely pleiades which is capable for displaying data information on a larger scale. The paddy field classification model in this study uses U-net to classifier between paddy fields and non-paddy fields. This model has performance for rice field classification with an accuracy 0.58 and an F1 score of 0.35 (non-rice) and 0.694 (rice).

Keywords: Paddy Mapping, Remote Sensing, Pleiades, Very High Spatial Resolution, U-Net

1. Introduction

Paddy is one of the largest commodities as a staple food for people worldwide. Population growth and the diversion of paddy fields have caused the area of paddy fields to decrease. Proper and accurate monitoring of agricultural land area is crucial for estimating crop yields [1]. Mapping agricultural land using remote sensing can be a research method to determine estimates of agricultural yields to maintain food security [2] [3]. Pixel-based and object-based remote image sensing can be used to classify problems in agricultural land mapping [1].

Various methods are used to assist in land mapping. One method that is often used is Convolutional Neural Network (CNN). CNN is a deep learning method to extract spatial information for image classification or semantic segmentation. One of the developments of CNN is U-Net, which was initially developed for medical image segmentation [4]. The U-net use multi-scale encoders-decoders and skip connections [5]. In this research, the U-net method will be implemented to segment and classify paddy fields quickly and accurately [6].

Previous research related to methods in this field is still relatively small. Wang et al. [1] used the U-Net-based fully connected convolutional network (FCN) method to detect rice fields in China using CNES RGB and Sentinel 1 SAR satellite imagery with an IOU accuracy of 0.953. Meanwhile, Yong et al. [5] used the Mobile V2-Unet and RANSAC methods to detect field boundaries with an IOU accuracy of 0.908. The R2U-Net method was applied by Song et al. [6] using Sentinel-2 data morphological filtering and the Douglas-Peucker Algorithm (DPA) with an overall accuracy of 86.42% for R2U-Net and 84.15% for single temporal images. The next research was conducted by X.Huang, et. al. [2] regarding efforts to overcome food security that occurs in Rakhine, Myanmar. This research offers remote sensing using Very High Resolution satellites as an effort to support the characterization of failed, abandoned and cultivated rice classes from 2016 to 2018. Based on this research, overall accuracy results (%) were 86.5, 87.5 and 91.0 for 2016, 2017, 2018.

Remote sensing technology with artificial intelligence is one of the modern methods used to map raw rice fields abroad, making it easier to record land areas [7]. However, existing methods focus on datasets from rice fields located abroad. Therefore, a detection model specially trained with raw rice field data from Indonesia using artificial intelligence is needed.

This research develops a method for mapping paddy fields in subtropical areas using Very High Resolution (VHR) optical image data from the Airbus Pleiades satellite [8]. Satellite image data will be classified using the U-net method to obtain high-accuracy mapping results for small paddy fields.

2. Materials and Methods

2.1 Data Preparation

Klaten Regency is one of the districts located between Mount Merapi and the Thousand Mountains, with an altitude of between 75 to 160 meters above sea level and a geographic location of 110°26' – 110°47' E and 70°32' – 70°48' S [9]. Klaten Regency has many paddy fields, making it ideal for use as a research area for mapping paddy fields. This research uses Pleiades satellite data with the area of interest of Manisrenggo subdistrict located at 110°29'11.4"- 110°29'52.8"E and 7°42'45.1" - 7°43'21.2" S, as shown in Figure 1. This area has agricultural land, almost part of which is used as productive paddy fields. Based on data from the Klate Regency city government, in 2014 the area of rice fields in Manisrenggo sub-district was 3073 (Ha), with an area of harvested land of 3058 (Ha) [10]. In some seasons, the land will be used as "Tumpang Sari," which will be planted with crops other than paddy, such as corn, chilies, beans, and others, as shown in Figure 2.

The data using one of Very High Resolution (VHR) optical satellite imageries, namely Pleiades from the Badan Riset dan Inovasi Nasional (BRIN). This data was taken on June 02, 2022. Pleiades satellite data has several spectral bands such as Panchromatic: 450–820 nm, B0: 450–530 nm, B1: 510–590 nm, B2: 620–700 nm, B3:775–915 nm and has spatial resolution 30 cm for Panchromatic, 1.2 m for Multispectral bands [9] [11]. This data has a very high spatial resolution, so it can be used for mapping paddy fields per small area [12]. The dataset used has dimensions of 4,248×3,284, comprising 15,649,632 pixels.

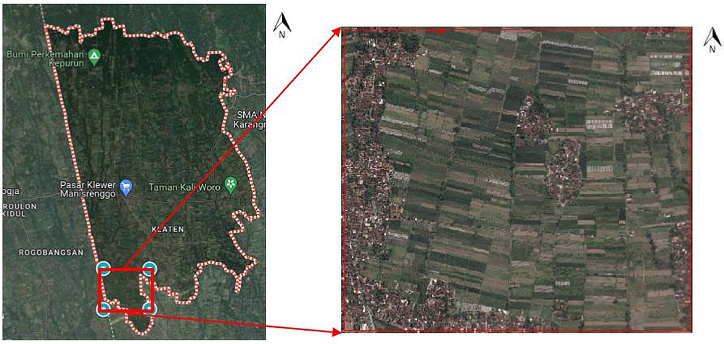

Figure 1. Area of interest: Manisrenggo Subdistrict, Klaten Regency.

Figure 2. An example of tumpang sari land.

A polygon digitization process first carries out optical satellite image data in the form of RGB to obtain the ground truth value for the paddy field area using ArcGIS software [7]. The ground truth will later be used as masking data as gray-scale values for reference in the machine-learning process [5]. The dataset used is divided into two classes. The first class comprises paddy fields, and the second non-paddy fields (houses, roads, mosques, schools, etc.). The data used is 288 satellite image datasets and 288 masking data with a size of 256×256 pixels. The dataset is divided into two classes, with the total amount of data shown in Table 1.

Table 1. Total images of paddy field and non-paddy field classes

Paddy Field Class	Non-Paddy Field Class
174 Satellite Images	114 Satellite Images
174 Masking Images	114 Masking Images

3. Methodology

The research workflow includes the data preparation stage, which includes creating a dataset, masking the data, pre-processing, and using the U-net model in the data training. The next step is testing the model, evaluating model performance, and predicting paddy fields. The processes are presented in a diagram, as shown in Figure 3.

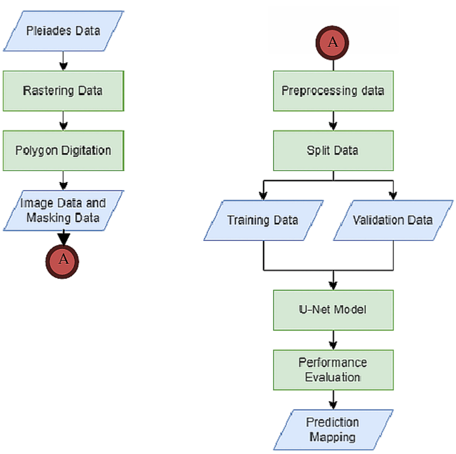

Figure 3. Block diagram of the research flow.

3.1 Data Acquisition

The dataset was created by digitizing polygons on images of rice fields to obtain ground truth values using ArcGIS Software, then digitizing the polygon as shown in Figure 4. Masking data created in the form of a grayscale image that represents the edge of the rice field [13]. The dataset consists of satellite optic RGB and masking data in reference data with 256×256 pixels, as shown in Figure 5.

Figure 4. Polygon Digitation

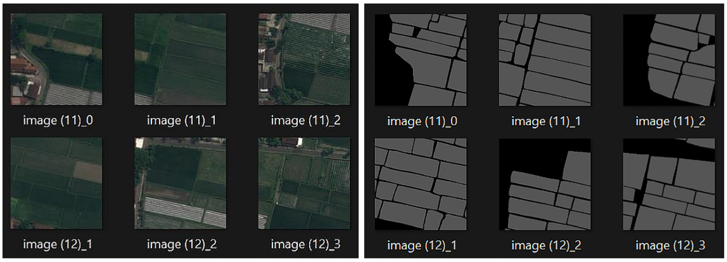

Figure 5. Examples of image and masking data.

3.2 *Preprocessing*

In the preprocessing stage, several processes are carried out on the satellite image dataset and masking image. The dataset was then subjected to band normalization pre-processing to normalize the value of all images in the dataset. The next step was selecting bands of the satellite images. In this study, the input images consist of 3 channels, i.e., red, green, and blue. Furthermore, the dataset is divided into 70% training data and 30% validation.

3.3 *U-Net*

In this study, the neural network used for segmentation is U-Net, where this architecture was previously used for biomedical image segmentation [14]. The U-net architecture has an encoder decoder structure in the deep learning process, where this structure can speed up model operation and reduce the number of unnecessary model parameters [5]. In this research, we used Encoder Structure (256, 128, 64, 32, 16) with Maxpooling layers to reduce dimensions by maintaining tensor parameters, then

Batch normalization layer, which functions to normalize tensors, then Dropout layer as regularization to reduce overfitting. The Up-sampling layer is used in the Decoder process. The Decoder structure (16, 32, 64, 128, 256) has a Transpose Convolution layer to duplicate the feature map and a Concatenate layer to combine the feature map with the previous encoder output via Skip Connections. The Skip Connection layer functions as backpropagation by directly connecting the previous layer in the network to deeper layers, as show in Figure 6.

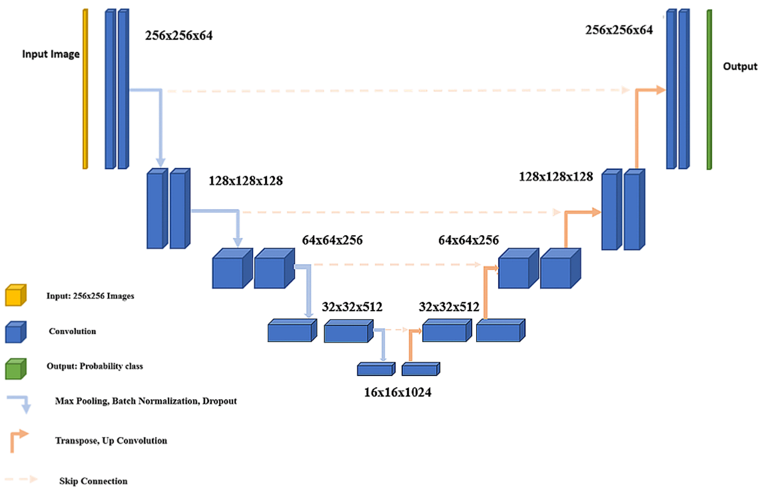

Figure 6. U-net architecture.

The image data and masking data that have been divided into training and validation set will be used as input data on the conv1 layer. The input size in the model is (256,256,3) which means that the value represents the size and dimensions to be processed in the U-net model [15]. In conv1, image data and masking will be carried out in a 2D convolution process with 64 kernel filters measuring 3x3. After the convolution process is complete, then ReLu activation is applied which functions to introduce non-linearity to the model. The results of the convolution will then be normalized using batch normalization. The result of conv1 is in the form of feature map extraction with max poolling. Next, conv2 will be carried out using a 128 kernel filter with a size of 3x3 and perform the same process as conv1. In conv3, we use a 256 kernel filter with a size of 3x3 and continue the same process as with the previous layer [16]. In conv4, the layer dropout process is applied first before the feature map is carried out using max pooling. Next, a convolution was carried out with a 3x3 kernel 512 filter, and the same process was continued as before. In conv5, drpout is carried out again as in the conv4 layer, then a 1024 kernel filter with a size of 3x3 is carried out. After carrying out the encoder process, up-sampling is carried out using a concatenate layer which functions to connect between previous layers that have the same size by establishing a skip connection. In conv6, conv7, conv8 and conv9 the same convolution process is carried out but uses different kernel sizes. On conv10 it will process the conv9 input as output using activation softmax [4].

3.4 Performance Evaluation

Model evaluation will be carried out using confussion matrix, *accuracy*, *precision*, *recall*, and $F1$ *score* values are obtained from the U-net model designed in this study. The method of calculating the value is shown in equations (1), (2), (3), (4).

$$accuracy = \frac{TP + TN}{ntotal} \tag{1}$$

$$accuracy = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\ score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{4}$$

The *ntotal* is the sum of all the data used in the evaluation process (in this case, the amount of image data used for testing). True Positive (TP) is the amount of data that is predicted to be true as positive (in this case rice fields), True Negative (TN) is the amount of data that is correctly predicted as negative (in this case non-rice fields), False Positive (FP) is data that is actually negative, but predicted by the model as positive and False Negative (FN) is data that is actually positive but predicted by the model as negative.

4. Result and Discussion

This research usesd the Pleiades RGB image data with greyscale masking images used as the reference data. The U-net was trained to learn from the data training in classifying paddy and non-paddy fields by extracting features [16]. Several experimental schemes were conducted to test the model and obtain the best performance. The testing schemes and settings of parameters are described in Table 2.

Table 2. Experiment schemes

Scheme	Parameter	
Tuning Epoch	25	35
Tuning Learning Rate	1×10^{-4}	1×10^{-5}
Splitting Dataset	70:30	80:20
Drop Out layer	0.4	0.5

Each experimental scheme was carried out to obtain model performance suitable for solving cases in mapping rice fields. The model scheme includes a tuning epoch to test how the performance of model. The following scheme aimed to fine tune the learning rate parameter, where it determines how the model changes in response based on the weights given. Tuning the splitting ratio of the dataset was carried out to see how the amount of training and validation data influenced the learning process.

Table 3. Training accuracy results of the experimental schemes.

Scheme	Parameter Value	Training Accuracy	Validation Accuracy
Tuning Epoch	25	96%	94%
	35	95%	93%
Tuning Learning Rate	1×10^{-4}	96%	80%
	1×10^{-5}	96%	78%
Splitting Dataset	70:30	96%	94%
	80:20	80%	90%
Drop Out layer	0.4	80%	80%
	0.5	96%	94%

Drop-out ratio settings are carried out to reduce overfitting in the model-learning proces [15]. The model performance is shown in Table 3 based on each scheme.

The initial experiment was carried out on the number of epoch schemes for the model training process. In the epoch parameter tuning experiment, using data splitting 70:30, learning rate 1×10^{-4}, drop out ratio 0.5 for each training. The results obtained for the epoch of 25 were 1% higher compared to the epoch of 35, so in the next experiment, the number of epochs that would be used would be 25. In the learning rate tuning experiment, using data splitting 70:30, epoch 25, drop out ratio 0.5 for each training. The learning rate tuning scheme produces the same accuracy, so the following parameter used a learning rate of 1×10^{-4}. In the dataset splitting using epoch 25 learning rate 1×10^{-4}, drop out ratio 0.5, for each trial The dataset splitting scheme has quite significant differences in accuracy results, whereas splitting with a ratio of 70:30 has yielded higher results. This happeneds because the larger the training data used, the more complex the learning process and the more overfitting occurs in the model. In the drop out layer scheme using 25 epochs, learning rate of 1×10^{-4}, splitting the dataset 70:30. The drop out layer of 0.5 is better than the drop-out layer of 0.4, with training accuracy of 96%.

The U-net training accuracy performance results obtained 96%, as shown in Figure 7. These results are quite good for the classification of paddy and non-paddy fields. The hyperparameter set consists of an epoch of 25, a learning rate of 1×10-4, dropout layer 0.5, with a splitting dataset of 70:30, batch size 4, early stopping, activation ReLu, and Adam optimizer.

The next stage is the performance evaluation of the model on the testing result. The model using a testing image with a size of 256x256 pixels to determine the values of accuracy, precision, recall and F1 score. Based on the testing stage, the results of accuracy were obtained of 0.585, precision 0.4782 (non-paddy) and 0.6205 (paddy), recall 0.2878 (non-paddy) and 0.7875 (paddy), F1 score 0.3594 (non-paddy) and 0.6941 (paddy).

In this research, using the best scenario, an accuracy result of 58.5% was obtained for the classification of rice fields in Indonesia. When compared with the research of X.Huang, et. al. [2], the overall accuracy (%) values were 86.5, 87.5 and 91.0 for 2016, 2017, 2018. The accuracy levels obtained tended to be lower compared to

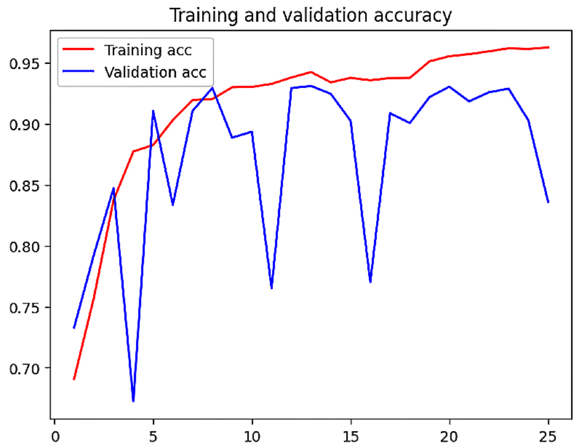

Figure 7. Result of the training and validation accuracy of the U-net model

previous research. This happeneds because agricultural land in Indonesia tends to be heterogeneous with various kinds of plants. One of the things that influences the performance of the model is the type of rice field used as "*tumpang sari*" land, so that the model created is still unable to differentiate between "*tumpang sari*" land.

From this model, predictions were designed using testing data with an image size of 1367×668 pixels. Test the model using larger data to show the performance of the model that has been created. The data used for predictions is image data with a landscape of paddy fields and buildings, as shown in Figure 8.

Figure 8. The image data that is used for prediction.

The U-net model was able to differentiate between paddy fields and non-paddy fields in the prediction data, as shown in Figure 9. The prediction results show that the division of paddy fields in each small area is visible.

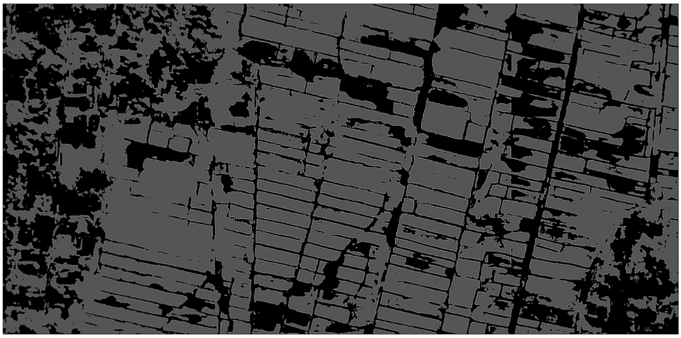

Figure 9. Prediction result (black area: non-paddy fields, grey area: paddy fields).

Based on the prediction results of the model, there are still several errors in predicting the paddy field and non-paddy field classes, as shown in Figure 10. These errors occur because the model could not detect the non-paddy field class on a large scale, so errors occur in the pixel predictions of the non-paddy classes. The complexity of the model and variations in the non paddy class cause errors in predicting this class. Paddy land in Indonesia planted as "tumpang sari" was not detected well in this model, this is because the color of the paddy land is different from other paddy fields. However, the model created to predict paddy percil performed well in predicting paddy land on a large scale.

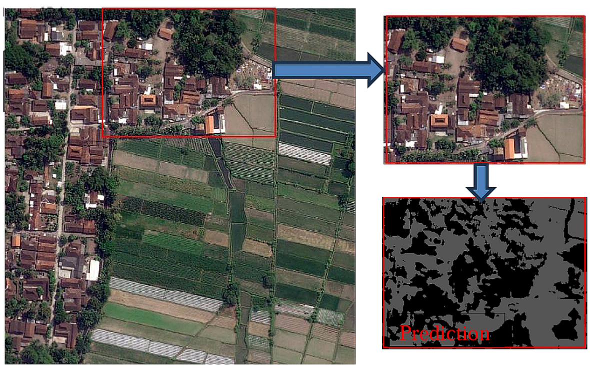

Figure 10. Prediction errors in larger scale.

Prediction errors occur in areas with non-paddy classes such as buildings and roads. This happeneds because the size of the buildings and roads in the non-paddy class was small, so the model could not detect in detail. Overall, the model could distinguish rice fields with smaller sizes, but there were still missclassification when detecting land with non-paddy classes. Based on the experiment, the U-net model using VHR data is capable to detect percil rice fields. However, the model should be improved in detecting non-rice classes due to the small size. Further research is needed to improve the U-net model to predict rice fields on a large scale.

5. Conclusion

The small paddy field classification method is proposed using very high-resolution Pleiades satellite data and the U-net classification method. The U-net method utilizes an encoder to speed up the computational process of training models and uses a decoder to multiply features to get optimal model results. The use of skip connections helps reduce the model's complexity, thereby speeding up the data training process. The results of the model show a training accuracy performance of 96%, with an accuracy 0.58, F1 score of 0.35 (non-rice) and 0.694 (rice). This shows that the U-net classifier feature works in small detail to differentiate between paddy and non-paddy fields. For future research, tuning hyperparameters such as the use of regularization, adding attention layers, augmentation in data preprocessing or combined backbone U-net with other architecture to improve accuracy and performance in the classification of paddy field mapping.

Acknowledgement

We thanked Badan Riset dan Inovasi Nasional (BRIN) for providing the Pleiades image dataset.

References

[1] M. Wang et al. "Agricultural Field Boundary Delineation with Satellite Image Segmentation for High-Resolution Crop Mapping: A Case Study of Rice Paddy". In: *Agronomy* 12.10 (2022). DOI: 10.3390/agronomy12102342.

[2] X. Huang, B. Ziniti, and N. Torbick. "Communication assessing conflict driven food security in rakhine, myanmar with multisource imagery". In: *Land* 8.6 (2019). DOI: 10.3390/LAND8060095.

[3] U. Nation. *Department of Economic and Social Affairs Sustainable Development*. https://sdgs.un.org/goals. Accessed: 2024-07-06.

[4] F. Waldner and F. I. Diakogiannis. "Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network". In: *Remote Sens. Environ.* 245.October 2019 (2020), p. 111741. DOI: 10.1016/j.rse.2020.111741.

[5] Y. He et al. "Automated detection of boundary line in paddy field using MobileV2-UNet and RANSAC". In: *Comput. Electron. Agric.* 194.December 2021 (2022), p. 106697. DOI: 10.1016/j.compag.2022.106697.

[6] W. Song et al. "Hierarchical extraction of cropland boundaries using Sentinel-2 time series data in fragmented agricultural landscapes". In: *Comput. Electron. Agric.* 212.August (2023), p. 108097. DOI: 10.1016/j.compag.2023.108097.

[7] L. Xu et al. "Paddy rice mapping in thailand using time-series sentinel-1 data and deep learning model". In: *Remote Sens.* 13.19 (2021). DOI: 10.3390/rs13193994.

[8] R. Nurtyawan and N. Fiscarina. "Assessment of the Accuracy of Dem From Panchromatic Pleiades Imagery (Case Study: Bandung City. West Java)". In: *Int. J. Remote Sens. Earth Sci.* 17.1 (2020), p. 34. DOI: 10.30536/j.ijreses.2020.v17.a3329.

[9] European Space Agency. *Pleiades*. https://earth.esa.int/eogateway/missions/pleiades. Accessed: 2024-07-06.

[10] Klatenkab.go.id. *Pertanian dan Ketahanan Pangan*. https://klatenkab.go.id. Accessed: 2024-07-06.

[11] Satellite Imaging Corporation. *Pleiades-1A Satellite Sensor (0.5m)*. https://www.satimagingcorp.com/satellite-sensors/pleiades-1/. Accessed: 2024-07-06.

[12] C. Yan et al. "High-resolution mapping of paddy rice fields from unmanned airborne vehicle images using enhanced-TransUnet". In: *Comput. Electron. Agric.* 210.November 2022 (2023), p. 107867. DOI: 10.1016/j.compag.2023.107867.

[13] W. Xiao, S. Xu, and T. He. "Object-Based Algorithm — A Implementation in Hangjiahu Plain in China Using GEE Platform". In: *Remote Sens.* 13 (2021), pp. 1–20.

[14] N. Siddique et al. "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications". In: *IEEE Access* 9 (2021), pp. 82031–82057. DOI: 10.1109/ACCESS.2021.3086020.

[15] I. Salehin and D. K. Kang. "A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain". In: *Electron.* 12.14 (2023). DOI: 10.3390/electronics12143106.

[16] M. Li and A. Stein. "Mapping land use from high resolution satellite images by exploiting the spatial arrangement of land cover objects". In: *Remote Sens.* 12.24 (2020), pp. 1–21. DOI: 10.3390/rs12244158.