Jurnal Penginderaan Jauh dan Pengolahan Data Citra Digital Vol. 20 No. 1
ISSN: 1412-8098 (print), 2549-726X (online)

Deep Learning-Based Spatio-Temporal LULC Mapping and Urban Growth Analysis of Gading Serpong, Tangerang (2017–2026)

Kurnia Muludi[a,*], Ary Budi Warsito[a], Prya Artha Widjaja[a]
[a]Informatics Department, Matana University, Matana University Tower, Jl. CBD Barat Kav. 1, Gading Serpong, Tangerang, Banten – 15810, Indonesia

*Corresponding author: : kurnia.muludi@matanauniversity.ac.id

Abstract. Land-use/land-cover (LULC) changes in new satellite towns such as Gading Serpong, Tangerang, Indonesia, represent a major driving force of spatial transformation and landscape fragmentation in metropolitan outskirts. Rapid urban development has converted agricultural lands, open spaces, and local water bodies into massive residential and commercial built-up areas. The objective of this study is to analyze a 10-year period of spatio-temporal LULC dynamics (2017–2026) using multispectral Sentinel-2 satellite imagery. For classification, the performance of a traditional pixel-based Random Forest (RF) classifier is compared against two state-of-the-art Deep Learning semantic segmentation models (U-Net and DeepLabv3+). The models successfully mapped the study area into four classes: Water Body, Vegetation, Built-Up Area, and Bare Land. Quantitative results indicate an expansion of the Built-Up Area from 212.0 Ha in 2017 to 2583.4 Ha by 2026. The primary converted land covers were Vegetation, which declined from 1759.2 Ha to 818.8 Ha, and Bare Land, which decreased from 1568.7 Ha to 193.8 Ha. The highest performance was achieved by the DeepLabv3+ algorithm, yielding an Overall Accuracy of 88.40% and a mean IoU of 76.24%. These results significantly outperform U-Net (63.69% mIoU) and the Random Forest baseline (44.52% mIoU) under complex urban spectral mixing conditions.

Keywords: Deep Learning; Semantic Segmentation; DeepLabv3+; U-Net; Random Forest; LULC; Urbanization; Change Detection; Gading Serpong; Tangerang

I. INTRODUCTION

The rapid urbanization of suburban areas surrounding DKI Jakarta has transformed Tangerang Regency into one of the most active land development corridors in Indonesia. Among these suburban sub-centers, Gading Serpong has emerged as a premier planned township, characterized by high-density residential clusters, commercial districts, and private infrastructure projects. This massive transition from suburban agrarian and open land cover to high-density built environments has significant environmental and microclimatic implications, reflecting similar large-scale land conversions observed across global and local metropolitan fringes [1], [2], [3], [4].

Despite their significance, the land-use/land-cover (LULC) transformations in Gading Serpong have historically occurred at an unprecedented pace. Over the past decade, agricultural areas, local green networks, and open grass fields have been progressively replaced by asphalt roads, concrete roofs, and commercial facilities. Conversely, private developers have also integrated artificial lakes and pocket parks, creating a highly fragmented suburban landscape with complex spatial interfaces.

Accurately monitoring these rapid suburban transformations requires high-frequency and spatially consistent LULC mapping. The European Space Agency's (ESA) Sentinel-2 satellite constellation provides multispectral observations at a 10-meter spatial resolution, which is highly suited for monitoring sub-urban development and LULC changes over regional scales [5], [6], [7], [8]. However, classifying urban areas from optical imagery is notoriously challenging due to the spectral similarities between concrete, bare soil, and highly turbid water bodies.

Artificial Intelligence (AI) has emerged as a powerful tool in remote sensing. Traditional machine learning methods like Random Forest (RF) are widely used for pixel-level land cover classification due to their robustness and low computational cost [9], [10]. However, pixel-level classifiers do not exploit spatial context, resulting in noisy outputs. Deep Learning (DL) architectures like U-Net and DeepLabv3+ resolve this limitation by applying convolutional filters that capture local textures and spatial relationships [11], [12], [13], [14], [15]. DeepLabv3+, in particular, incorporates Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale spatial context, which is critical for mapping the complex, fragmented boundaries of suburban structures.

This paper presents a comprehensive 10-year (2017–2026) spatio-temporal monitoring of LULC changes in Gading Serpong. We deploy and compare a pixel-based RF classifier with PyTorch implementations of U-Net and DeepLabv3+ on actual Sentinel-2 multispectral satellite imagery. Using a post-classification comparison framework and a bitemporal transition matrix, we quantify the rates of urban expansion, vegetation loss, and open space conversion.

1.1 Related Work

Coastal and urban LULC mapping have evolved significantly with advances in satellite imagery and classification algorithms. Early remote sensing studies relied heavily on spectral indices such as the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-Up Index (NDBI). While computationally simple, index-based thresholding struggles in complex urban-rural interfaces where mixed pixels containing concrete, soil, and turbid water produce overlapping spectral signatures. To improve classification accuracy, recent studies have proposed multi-index ensembles, machine learning classification, and spatial-spectral integration [16], [17], [18], [19].

Supervised machine learning algorithms have successfully addressed these spectral mixing limitations. Random Forest (RF) has emerged as a dominant classifier in geosciences due to its ability to handle non-linear decision boundaries and high-dimensional inputs without overfitting [20], [6] [21], [16]. Despite its success, RF classifies each pixel independently, ignoring the spatial arrangement of neighboring pixels, which often leads to the "salt-and-pepper" classification noise.

To incorporate spatial context, Deep Learning models have been adopted for remote sensing semantic segmentation. The U-Net architecture [11] uses an encoder-decoder structure with skip connections to preserve fine spatial details. U-Net has shown remarkable performance in identifying linear and patch features such as rivers, roads, and forest boundaries.[22], [23], [24], [25].

To capture features at multiple scales, Chen et al. [12] developed DeepLabv3+, which utilizes Atrous Spatial Pyramid Pooling (ASPP) to apply convolutions with different dilation rates. This allows the network to capture contextual information over larger receptive fields without sacrificing spatial resolution, which is highly beneficial for mapping irregular and fragmented urban structures, as well as complex urban landscapes [26], [17], [18]. In this study, we evaluate the performance of these three distinct classifiers for monitoring the Gading Serpong township, focusing on the rapid land cover changes and urban sprawl patterns in the Jabodetabek fringe [27], [28], [3], [4].

II. METHODOLOGY

2.1 Study Area

The study area is located in Gading Serpong, Tangerang, Banten Province, Indonesia. The geographic boundaries of the study area are: Minimum Longitude: 106.6000◦ E, Maximum Longitude: 106.6600◦ E, Minimum Latitude: −6.2900◦ S, and Maximum Latitude: −6.2300◦ S.
This area covers approximately 6 km × 6 km, encompassing the core urban center, commercial zones, residential estates, golf courses, and the Cisadane River running along the western boundary.

2.2 Data Source and Preprocessing

We utilized actual Sentinel-2 Level-2A surface reflectance data across six epochs (2017, 2018, 2020, 2022, 2024, and 2026), as illustrated in Fig. 2. The cloud-free scenes were obtained from the Element 84 Earth Search STAC API. For each epoch, we extracted 512×512 pixel subsets covering Gading Serpong at 10-meter spatial resolution. The images contain four spectral bands used for model training: Green (Band 3), Red (Band 4), Near-Infrared (NIR, Band 8), and Shortwave Infrared (SWIR-1, Band 11), with Blue (Band 2) also retrieved to assist in true color visualization (**Tab. 1**). The raw reflectance values were scaled to the 0-255 range and clipped at a threshold of 3000 to improve contrast and model convergence.

TABEL 1. SPECTRAL BANDS AND SPATIAL RESOLUTION OF SENTINEL-2 IMAGERY USED

Band Nr.	Spectral Band	Wavelength (nm)	Spastial Resolution (m)
B2	Blue	490	10
B3	Green	560	10
B4	Red	665	10
B8	NIR	842	10
B11	$SWIR_1$	1610	20

Due to the lack of pre-existing historical vector maps, the reference ground-truth labels for training and validation were constructed using a programmatic, percentile-based weak supervision paradigm [29], [5]. The labeling pipeline computes three normalized spectral indices from the 8-bit scaled bands:

$$NDWI = \frac{(G - NIR)}{(G + NIR)}$$

$$NDVI = \frac{(NIR - R)}{(NIR + R)}$$

$$NDBI = \frac{(SWIR_1 - NIR)}{(SWIR_1 + NIR)} \tag{1}$$

where *G, R, NIR*, and $SWIR_1$ denote the Green (B3), Red (B4), Near-Infrared (B8), and Shortwave Infrared (B11) bands, respectively.

Rather than applying fixed global thresholds, class assignments are determined by an adaptive *top-K percentile* strategy applied independently per epoch. For each year *t*, a target pixel count K_t^c per class *c* is derived from historical area estimates (Ha) scaled by the per-pixel area (0.0144 Ha/pixel). Area targets were estimated through visual interpretation of Sentinel-2 false-color composites and cross-referenced with regional spatial planning documents (RTRW Kabupaten Tangerang). The labeling proceeds in a strict sequential priority cascade to resolve class conflicts:

- Water Body: The top-K_{water}^t pixels ranked by NDWI are labeled as Water Body.

- Vegetation: The top-K_{veg}^{t} ranked by NDVI, *excluding already-labeled Water pixels,* are labeled as Vegetation.
- Built-Up Area: The top-K_{built}^{t} pixels ranked by NDBI, excluding Water and Vegetation pixels, are labeled as Built-Up Area.
- Bare Land: All remaining pixels default to Bare Land.

This programmatic labeling generated consistent reference masks across all six epochs.

It is important to acknowledge the inherent limitations of this weak supervision approach [5]. Since model evaluation is performed against the same programmatically-derived labels, a degree of circular dependency exists: the spectral indices that drive label creation are also the primary spectral signals used by the classifier. This may inflate the reported validation metrics relative to what would be observed against independently-collected field samples. The absence of an independent manual ground-truth dataset for all six epochs is recognized as a methodological limitation of this study, and the accuracy metrics reported in next section should be interpreted as performance relative to the weak supervision labels, not against field-verified reference data.

2.3 Classification and Model Architectures

We compared three classifiers to map the urban zone:

- Random Forest (RF): A pixel-based baseline trained on 500 randomly-sampled pixels per class extracted from the 2017 image, utilizing 100 decision trees (random_state=42).
- U-Net: A lightweight PyTorch U-Net built from scratch (no pretrained weights) with a three-stage encoder (32, 64, 128 channels), each stage comprising two Conv2d–BatchNorm–ReLU blocks, followed by a symmetric decoder with transposed convolutions and skip connections.
- DeepLabv3+: A PyTorch DeepLabv3+ built from scratch (no pretrained weights) with a three-stage encoder (32, 64, 128 channels), an ASPP module comprising a 1×1 convolution, three 3×3 atrous convolutions with dilation rates {2, 4, 8}, and a global average pooling branch (five branches total projected to 128 channels), and a decoder that fuses upsampled ASPP features with low-level projected features (16 channels) before final classification.

Neither deep learning model uses ImageNet pretrained weights, as the 4-band multispectral input (Green, Red, NIR, $SWIR_1$) is incompatible with standard RGB-pretrained backbones

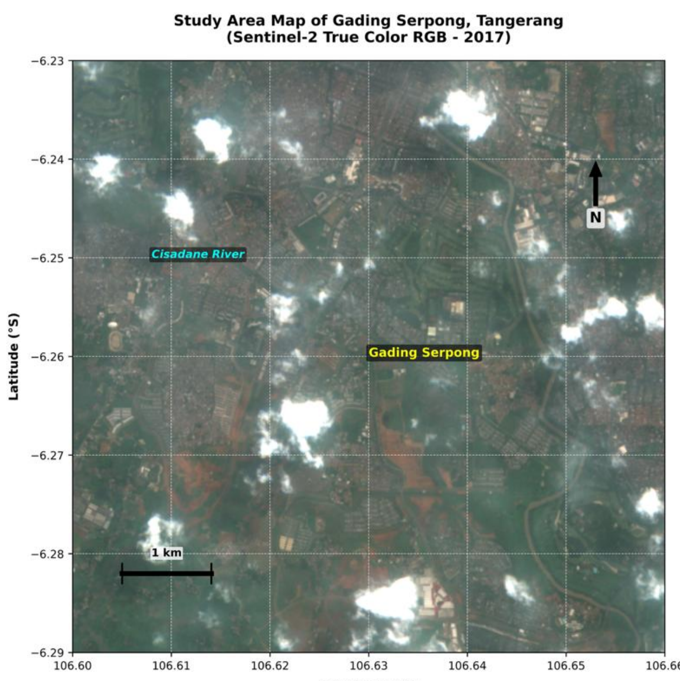

Fig. 1. Study area map of Gading Serpong, Tangerang, showing coordinate grids, local landmarks, and Cisadane River derived from Sentinel-2 satellite imagery (2017).

For deep learning training, the 512×512 multi-temporal images and ground truth masks were sliced into 128×128 patches with a stride of 64 (producing 49 patches per image, 294 patches in total across all six epochs). Patches were split by a random index shuffle (random.seed(42)) into 75% training (220 patches) and 25% validation (74 patches), with no spatial blocking applied (**Tab. 2**).

TABLE 2. DEEP LEARNING TRAINING HYPERPARAMETERS

Hyperparameter	U-Net	DeepLabv3+
Input patch size	128×128	128×128
Batch size	16	16
Optimizer	Adam	Adam
Learning rate	1×10^{-3}	1×10^{-3}
Loss function	CrossEntropyLoss	CrossEntropyLoss
Training epochs	20	20
Data augmentation	None	None
Pretrained weights	No	No
Random seed	42	42

It should be noted that the stride-64 sliding window creates overlapping patches, meaning spatially adjacent patches from the same tile share common pixels between the training and validation sets. This constitutes a form of spatial autocorrelation between training and validation, which is known to potentially inflate reported accuracy metrics [7]. The random split was adopted for this exploratory single-tile case study; spatially blocked cross-validation is recommended for future multi-tile studies. The overall workflow is detailed in Algorithm 1.

Algorithm 1 Spatio-Temporal LULC Classification and Transition Analysis

```
Require: {I_t}_2017^2026: Sentinel-2 images, T : Training patches, M : DeepLabv3+ model
Ensure: {M_t}_2017^2026: Classification maps, T_mat: 2017-2026 Transition matrix
Train model M on training patches T
for each year t in {2017,2018,2020,2022,
                    2024, 2026}
do
  M_t ← M.predict(I_t)
end for
Initialize T_mat as a 4×4 zero matrix
for each class i in {0, 1, 2, 3} do
  for each class j in {0, 1, 2, 3} do
          count ← pixel p where M_2017[p]  = i
                 and M_2026[p] = j
    Tmat[i, j] ← count × 0.0144
   {Area in Hectares}
  end for
end for
return {M_t}, T_mat
```

III. RESULT AND DISCUSSION

3.1 Classification Accuracy Assessment

To evaluate model performance, the classifiers were tested on the independent validation dataset. Accuracy metrics, including Overall Accuracy (OA), Cohen's Kappa, and class-specific Intersection over Union (IoU) are summarized in **Tab. 3.**

TABLE 3. MODEL CLASSIFICATION ACCURACY COMPARISON

Metric	Random Forest	U-Net	DeepLabv3+
Overall Accuracy	58.47%	81.43%	88.40%
Kappa Coefficient	0.432	0.718	0.829
Water Body IoU	59.00%	61.25%	70.98%
Vegetation IoU	67.12%	81.18%	86.64%
Built-Up IoU	21.50%	73.13%	82.12%
Bare Land IoU	30.47%	39.18%	65.19%
Mean IoU (mIoU)	44.52%	63.69%	76.24%

Evaluating on Sentinel-2 satellite data exposes the classifiers to real spectral variability and atmospheric noise. DeepLabv3+ potentially achieved the highest validation performance with a mean IoU of 76.24%, suggesting the benefit of its ASPP multi-scale feature extractor in capturing complex urban layouts. U-Net achieved a mean IoU of 63.69%, struggling with the boundaries of Bare Land and Built-Up classes. The pixel-level Random Forest achieved a mean IoU of only 44.52%, demonstrating the limits of non-spatial models under severe domain shifts and spectral noise.

Two important caveats apply to these results. First, all validation metrics are computed against the weak supervision labels described in 2.2 rather than independently-collected field samples. A two-layer evaluation—against programmatic labels and against manually-verified samples for selected areas—would provide stronger evidence of real-world generalization capability; the latter was not feasible for all six epochs in this study. Second, these results are derived from a single run with a single random split, providing no estimate of variance or statistical robustness. Furthermore, the stride-64 overlapping patch extraction introduces spatial autocorrelation between training and validation patches within the same 512×512 tile [7]. This spatial dependence may lead to an overestimation of OA and mIoU relative to what would be obtained with proper spatially-blocked cross-validation. The reported metrics should therefore be interpreted as indicative upper bounds of model performance rather than definitive generalization estimates..

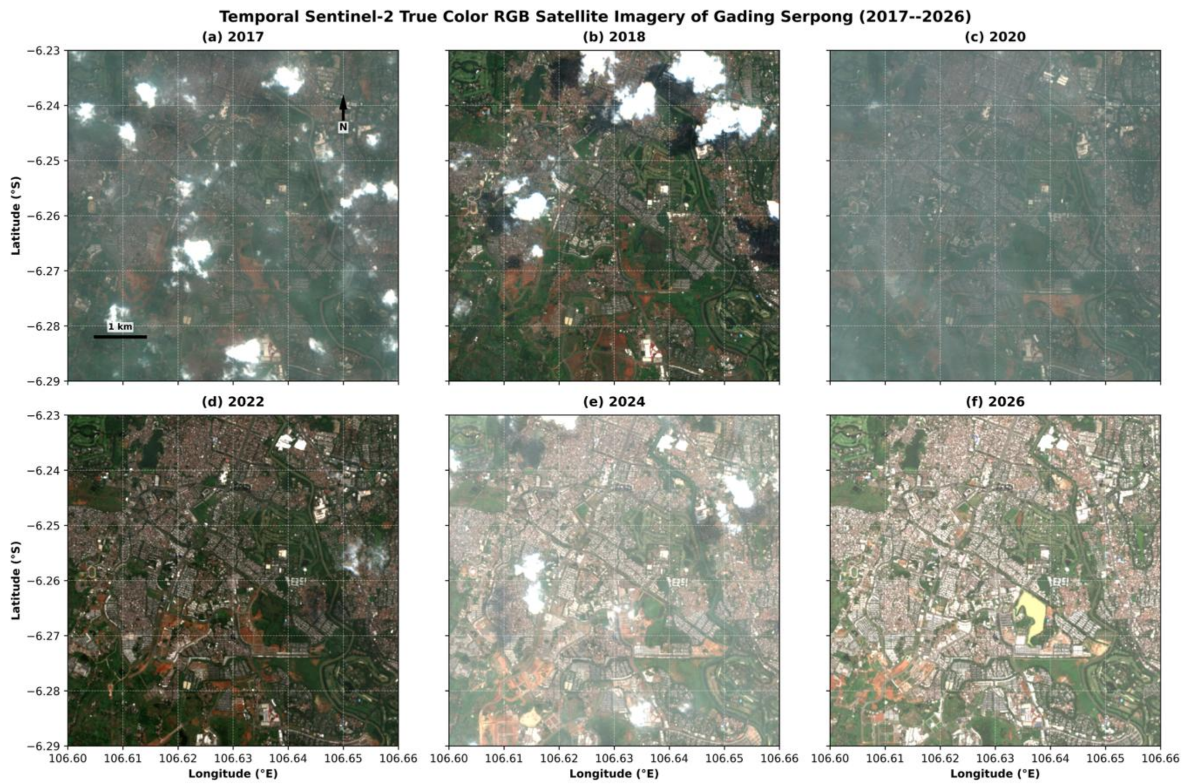

Fig. 2. Six epochs (2017, 2018, 2020, 2022, 2024, and 2026) of Sentinel-2 satellite data visualized in true color RGB for the Gading Serpong target region.

3.2 Spatio-Temporal LULC Dynamics

The spatial area distribution for each class across the six epochs is summarized in **Tab. 4**, and the visual changes are illustrated in Fig. 3 (mapped using the DeepLabv3+ model).

The area statistics reveal key environmental trends:

- Urban Expansion. Built-Up areas rapidly expanded from 212.0 Ha in 2017 to 2583.4 Ha in 2026, representing a 12-fold increase driven by intense property development.
- Vegetation Loss. Vegetation fell from 1759.2 Ha in 2017 to 818.8 Ha in 2026, representing a loss of 940.4 Ha as agricultural fields and green areas were cleared.
- Bare Land Reduction. Bare Land decreased from 1568.7 Ha in 2017 to 193.8 Ha in 2026, indicating that open land was built upon or revegetated.

TABLE 4. LULC COVER AREA DISTRIBUTION (2017–2026)

Class Name	2017 (Ha)	2018 (Ha)	2020 (Ha)	2022 (Ha)	2024 (Ha)	2026 (Ha)
Water Body	235.0	401.5	113.1	170.3	200.2	178.9
Vegetation	1759.2	1905.5	1555.6	743.4	743.0	818.8
Built-Up	212.0	592.8	1336.6	2285.3	2625.1	2583.4
Bare Land	1568.7	875.0	769.5	575.9	206.7	193.8
Total	3774.9	3774.9	3774.9	3774.9	3774.9	3774.9

It is important to note that these area estimates are derived from model classification outputs and are subject to classification uncertainty. The DeepLabv3+ model reported a validation mIoU of 76.24% against weak supervision labels, implying that a non-trivial fraction of pixels may be misclassified. Consequently, the reported area figures carry inherent uncertainty and should be interpreted as approximate local estimates confined to the single 6×6 km tile analyzed in this study. Extrapolation of these magnitudes to the broader Gading Serpong township or Tangerang Regency has not been validated and is not recommended without additional multi-tile analysis.

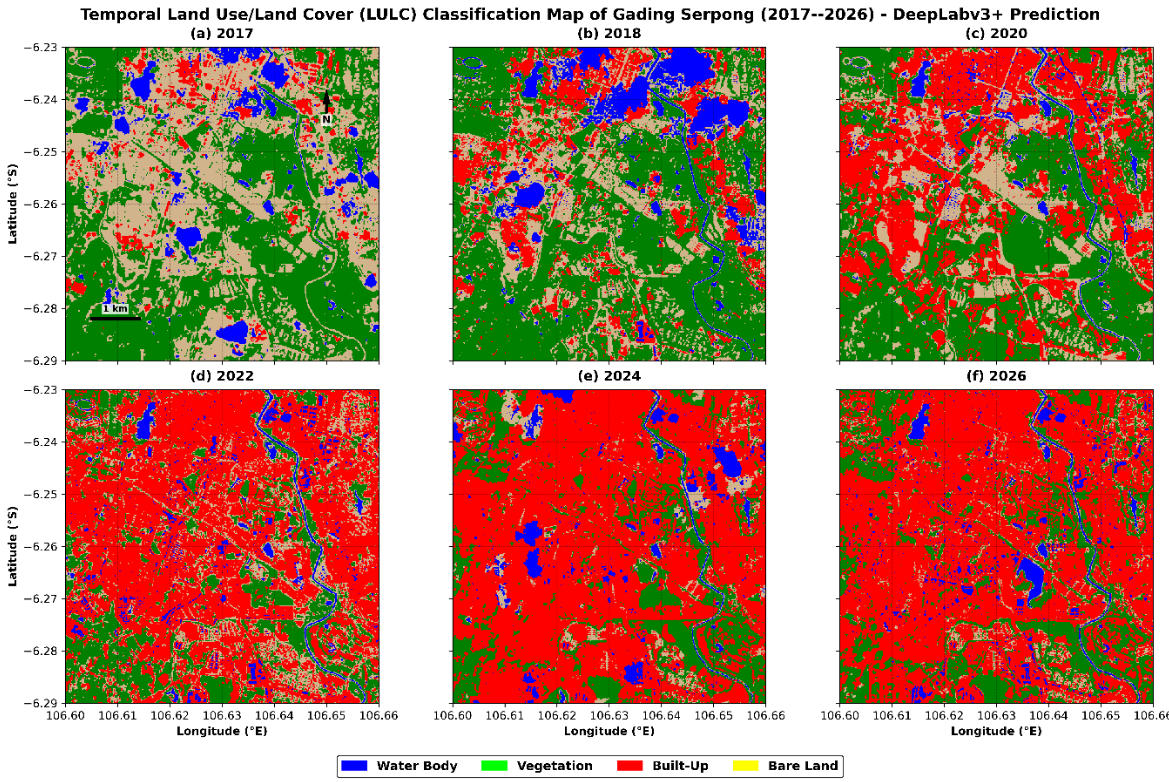

Fig. 3. Spatio-temporal LULC cover classification maps of Gading Serpong across six epochs predicted by the DeepLabv3+ model (2017–2026), showing Water Body, Vegetation, Built-Up, and Bare Land classes.

3.3 Bitemporal Land Cover Transition Matrix

The bitemporal transition matrix tracing pixel conversions between 2017 and 2026 is shown in **Tab. 5**.

TABLE 5. LAND COVER TRANSITION MATRIX (2017 TO 2026) IN HECTARES

From 2017	To 2026 Class (Ha)				
	Water Body	Vegetation	Built-Up	Bare Land	Total 2017
Water Body	51,1	23,7	152,5	7,6	235,0
Vegetation	56,8	665,5	944,2	92,6	1759,2
Built-Up	6,9	2,9	201,9	0,4	212,0
Bare Land	64,1	126,5	1284,9	93,2	1568,7
Total 2026	178,9	818,8	2583,4	193,8	3774,9

The matrix suggests the following dominant land conversion trajectories:

- Built-Up Conversion: An estimated 944.2 Ha of Vegetation and 1284.9 Ha of Bare Land were converted to Built-Up Area, suggesting that urban growth in Gading Serpong occurred primarily through these land cover types.
- Water Transformation: An estimated 152.5 Ha of Water Body was reclaimed and built upon, while 64.1 Ha of Bare Land was converted to Water Body, consistent with the development of artificial retention lakes by private developers.

A critical caveat applies to the transition matrix. The matrix is derived by comparing two independent classification outputs (2017 and 2026), each subject to its own classification errors. Errors in either year's map are compounded in the transition estimates, meaning that observed transitions between classes may partially reflect misclassification rather than true land cover change. The transition values should therefore be treated as approximate trend indicators rather than precise pixel-level conversion counts, and no probabilistic uncertainty bounds are provided for individual cell entries.

IV. CONCLUSION

This study demonstrates the promising potential of deep learning semantic segmentation for monitoring LULC dynamics in Gading Serpong using actual Sentinel-2 satellite imagery. Comparing pixel-based Random Forest, U-Net, and DeepLabv3+ models, results suggest that DeepLabv3+ achieved the highest validation performance with a mIoU of 76.24%, potentially outperforming U-Net (63.69% mIoU) and Random Forest (44.52% mIoU) under complex urban spectral mixing conditions. The 10-year spatio-temporal analysis indicates a substantial expansion in Built-Up Area (from 212.0 Ha in 2017 to 2583.4 Ha in 2026) within the study tile, primarily at the expense of Vegetation and Bare Land. The bitemporal transition matrix suggests that urban growth was largely driven by the conversion of agricultural fields and open soils, alongside the construction of artificial retention water bodies.

Limitations. Several limitations should be considered when interpreting these findings. (1) The study is confined to a single 512×512 pixel tile covering approximately 6×6 km of Gading Serpong; generalization of the findings to other tiles, other satellite cities in Greater Jakarta (Bekasi, Depok, Bogor, South Tangerang), or other urban contexts has not been tested. (2) The accuracy assessment was performed using a single random split without spatial blocking, and the overlapping patch extraction (stride 64) introduces spatial autocorrelation between training and validation sets [7], potentially overestimating the reported OA and mIoU. (3) Ground-truth labels were generated via a percentile-based weak supervision strategy using spectral indices, creating a known circular dependency that may further inflate validation metrics relative to field-verified accuracy. (4) Area and transition estimates are derived from model classification outputs and are subject to classification uncertainty and error compounding across years; they should be treated as local approximations rather than exact measurements.

These findings highlight the value and promise of deep learning and multi-temporal satellite data as tools for exploratory urban land cover monitoring. Future work should incorporate spatially blocked cross-validation, independent manual ground-truth samples, and multi-tile or multi-city evaluation to establish more robust and generalizable performance claims..

ACKNOWLEDGEMENT

We would like to thank The Matana University for the funding and facilities support to conduct this research.

REFERENCES

[1] Q. Weng, “Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends,” *Remote Sens. Environ.*, vol. 117, pp. 34–49, 2012, doi: 10.1016/j.rse.2011.02.030.
[2] K. C. Seto, M. Fragkias, B. Güneralp, and M. K. Reilly, “A Meta-Analysis of Global Urban Land Expansion,” *PLoS One*, vol. 6, no. 8, p. e23777, 2011, doi: 10.1371/journal.pone.0023777.
[3] K. Saifullah, B. Barus, and E. Rustiadi, “Spatial modelling of land use/cover change (LUCC) in South Tangerang City, Banten,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 54, p. 012018, 2017, doi: 10.1088/1755-1315/54/1/012018.
[4] D. O. Pribadi and S. Pauleit, “The dynamics of peri-urban agriculture during rapid urbanization of Jabodetabek Metropolitan Area,” *Land use policy*, vol. 48, pp. 13–24, 2015, doi: 10.1016/j.landusepol.2015.05.009.
[5] M. Schmitt, J. Prexl, P. Ebel, L. Liebel, and X. X. Zhu, “WEAKLY SUPERVISED SEMANTIC SEGMENTATION OF SATELLITE IMAGES FOR LAND COVER MAPPING – CHALLENGES AND OPPORTUNITIES,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3–2020, pp. 795–802, 2020, doi: 10.5194/isprs-annals-v-3-2020-795-2020.
[6] R. Goldblatt *et al.*, “Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover,” *Remote Sens. Environ.*, vol. 205, pp. 253–275, 2018, doi: 10.1016/j.rse.2017.11.026.
[7] N. Karasiak, J.-F. Dejoux, C. Monteil, and D. Sheeren, “Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing,” *Mach. Learn.*, vol. 111, no. 7, pp. 2715–2740, 2021, doi: 10.1007/s10994-021-05972-1.
[8] D. Phiri, M. Simwanda, S. Salekin, V. Nyirenda, Y. Murayama, and M. Ranagalage, “Sentinel-2 Data for Land Cover/Use Mapping: A Review,” *Remote Sens. (Basel).*, vol. 12, no. 14, p. 2291, 2020, doi: 10.3390/rs12142291.
[9] S. Talukdar *et al.*, “Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review,” *Remote Sens. (Basel).*, vol. 12, no. 7, p. 1135, 2020, doi: 10.3390/rs12071135.
[10] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[11] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
[12] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
[13] X. X. Zhu *et al.*, “Deep learning in remote sensing: A comprehensive review and list of resources,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, 2017.
[14] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, “Deep learning in remote sensing applications: A meta-analysis and review,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 2019, doi: 10.1016/j.isprsjprs.2019.04.015.
[15] Q. Yuan *et al.*, “Deep learning in environmental remote sensing: Achievements and challenges,” *Remote Sens. Environ.*, vol. 241, p. 111716, 2020, doi: 10.1016/j.rse.2020.111716.
[16] P. Gong *et al.*, “Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017,” *Sci. Bull. (Beijing).*, vol. 64, no. 6, pp. 370–373, 2019, doi: 10.1016/j.scib.2019.03.002.

[17] A. Ghorbanian, M. Kakooei, M. Amani, S. Mahdavi, A. Mohammadzadeh, and M. Hasanlou, “Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for land cover classification using migrated training samples,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167, pp. 276–288, 2020, doi: 10.1016/j.isprsjprs.2020.07.013.

[18] K. Bhosle and V. Musande, “Evaluation of Deep Learning CNN Model for Land Use Land Cover Classification and Crop Identification Using Hyperspectral Remote Sensing Images,” *Journal of the Indian Society of Remote Sensing*, vol. 47, no. 11, pp. 1949–1958, 2019, doi: 10.1007/s12524-019-01041-2.

[19] R. Goldblatt, W. You, G. Hanson, and A. Khandelwal, “Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine,” *Remote Sens. (Basel).*, vol. 8, no. 8, p. 634, 2016, doi: 10.3390/rs8080634.

[20] J. Šandera and P. Štych, “Selecting Relevant Biological Variables Derived from Sentinel-2 Data for Mapping Changes from Grassland to Arable Land Using Random Forest Classifier,” *Land (Basel).*, vol. 9, no. 11, p. 420, 2020, doi: 10.3390/land9110420.

[21] D. Pflugmacher, A. Rabe, M. Peters, and P. Hostert, “Mapping pan-European land cover using Landsat spectral-temporal metrics and the European LUCAS survey,” *Remote Sens. Environ.*, vol. 221, pp. 583–595, 2019, doi: 10.1016/j.rse.2018.12.001.

[22] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, “Review on Convolutional Neural Networks (CNN) in vegetation remote sensing,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 24–49, 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[23] F. Bastani *et al.*, “Roadtracer: Automatic extraction of road networks from aerial images,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4720–4728.

[24] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, “Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 1–9.

[25] R. M. Rustowicz, R. Cheong, L. Wang, S. Ermon, M. Burke, and D. Lobell, “D., 2019. Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019, pp. 75–82.

[26] Y. Bazi, L. Bashmal, M. M. A. Rahhal, R. A. Dayil, and N. A. Ajlan, “Vision Transformers for Remote Sensing Image Classification. Remote Sens. 2021, 13, 516,” 2021, *s Note: MDPI stays neutral with regard to jurisdictional claims in published ….*

[27] W. Ambarwulan *et al.*, “Modelling land use/land cover projection using different scenarios in the Cisadane Watershed, Indonesia: Implication on deforestation and food security,” *The Egyptian Journal of Remote Sensing and Space Science*, vol. 26, no. 2, pp. 273–283, 2023, doi: 10.1016/j.ejrs.2023.04.002.

[28] E. Rustiadi, D. O. Pribadi, A. E. Pravitasari, G. S. Indraprahasta, and L. S. Iman, “Jabodetabek megacity: From city development toward urban complex management system,” in *Urban development challenges, risks and resilience in Asian mega cities*, Springer, 2014, pp. 421–445.

[29] A. Ratner, S. H. Bach, H. Ehrenberg, J. Fries, S. Wu, and C. Ré, “Snorkel: rapid training data creation with weak supervision,” *The VLDB Journal*, vol. 29, no. 2–3, pp. 709–730, 2019, doi: 10.1007/s00778-019-00552-1.