

An artificial neural network approach to predict the future land use land cover of Great Malang region, Indonesia

Fatwa Ramdani^{1,*}, Budi Darma Setiawan¹, Alfi Nur Rusydi¹, Muhammad Tanzil Furqon¹

¹*Geoinformatics Research Group, Brawijaya University, Malang, Indonesia*

*e-mail for the corresponding author: fatwaramdani@ub.ac.id

Abstract

Great Malang region is developing rapidly with the population increase and inhabitant's activity, like migration and urbanization. Other activities like agricultural expansion as well as an uncontrolled residential development need to be monitored to avoid any negative impact in the future. The availability of free and open-source software, spatial high-resolution satellite imagery datasets, and powerful algorithms open the possibilities to map, monitor, and predict the future trend of land use land cover (LULC) changes. However, the accuracy and precision of this model is still in doubt, especially in the Great Malang region. Research is needed to provide a foundational basis and documentation on how the changes occur, where did the changes occur, and the accuracy of the predicted model. This study tries to answer those questions using the high spatial resolution of Sentinel-2 imageries. Combination of the fuzzy algorithm, artificial neural network, and cellular automata was utilized to process the datasets. We analysed four different scenarios of simulation and the result then compared. The different number of hidden layers and iteration was used and evaluated to understand the effect of different parameters in the prediction result. The best scenario was then used to predict future land use changes. This study has successfully produced the future LULC model of Great Malang region with high accuracy level (87%). The study also found that the land use transformation from agriculture to urban built-up area is relatively low, where changes of the built-up area over three periods of analysis are below than 5%. This is due to the physical condition of Great Malang region where mountainous areas are dominated.

Keywords: artificial neural network; fuzzy set and logic; cellular automata; Malang; Indonesia

I. INTRODUCTION

The world population will have more than 9 billion by 2050 (UN, 2019). According to the United Nations (UN), there are nine countries will have more than half of population growth between 2019 and 2050, and Indonesia is one of them (UN, 2019). The rapid population growth will produce challenges for sustainable development. This will increase to demand of housing, consumption as well as the pollution comes from anthropogenic source.

More than 50% world's population live in urban areas and United Nations (UN) projected that the number will be increased to almost 70% by year 2050 (UN, 2018). Most of the areas that will experience the population growth are located in Asia and Africa. Southeast Asia is one of the densest populations with Indonesia as the highest one and most of the population are living in the main cities.

The development of main cities in Indonesia is very rapid. One of the main

cities that have been growth rapidly is the Great Malang region. The Great Malang region consists of three administrative regions, which includes Malang City, Batu City, and Malang Regency. This region is a centre of education, tourism and agriculture which supports the majority of its population. With a total population of more than three and a half million people (BPS, 2018) and counting, there is rapid land conversion in the Great Malang region.

A study with high accuracy and precision is needed to monitor trends in land use and cover changes in the Great Malang region to avoid negative impacts on its population in the future. Changes in uncontrolled and unplanned land use and cover can cause disasters such as floods and landslides that can lead to casualties. A study is needed to determine the trend of changes in land use and cover in the Great Malang region to monitor, take inventory, and overcome the possible negative impacts in the future.

The combination of Fuzzy Algorithms (FA), Artificial Neural Network (ANN), Cellular Automata (CA), will be used to produce predictions of changes in land use and cover in the Great Malang region with high accuracy. FA is used to standardize the criteria for all geospatial data variables to be analysed. ANN is used as a land-use change transition model, and CA is used as a possible transition of land-use change that occur at the pixel level. In addition, analysis is needed on the effect of the number of hidden layers and iterations used in the accuracy and precision of the resulting land use maps. Hidden layers are layers in artificial neural networks that are located between the input layer and the output layer, where artificial neurons take a set of weighted inputs and produce output through the activation function. While the iteration is a repetition of the computational procedure applied to the results of the previous application, it is used as a means of obtaining a closer successive approach to solving the problem.

Until now, there is no current data and available information with a very high spatial resolution regarding changes in land use and cover in the Great Malang region. In addition, there is no scientific documentation that properly records changes in land use and cover in the Great Malang region.

1.1 Previous research

Some researchers have used FA in determining the best locations, for example, the most suitable locations from the industrial area (Junianto et al., 2018) and residential housing (Hadi et al., 2018). In these studies, the researchers used Digital Elevation Model (DEM) data, medium-spatial-resolution satellite imagery, and several factors such as road networks, agricultural areas, and river networks as constraints and determinants in land suitability analysis. The result states that accuracy with FA can reach ~ 80%.

Other researches also utilized fuzzy sets to determine the most suitable and safe areas to operate drones (Arihta et al., 2018). This study uses the function of sigmoidal membership to determine the appropriate location in operating the drone. The accuracy is more than 60%. While another research utilized FA to determine the suitable locations of apples plantation in Batu City (Lestantyo, 2019). The study achieved almost 93% accuracy. The studies related to fuzzy sets prove that regional suitability analysis can achieved good accuracy.

Furthermore, Ramdani, et al., (2019) conducted a study using the ANN algorithm to classify land use and cover in the Universitas Brawijaya area based on geospatial data with ultra-high spatial resolution acquired using drones. The study

revealed that the difference in the number of training data will greatly affect computational performance and the resulting accuracy. The more the training data, the higher the classification accuracy, however it will burden computing performance. The study used Radial Base Function Neural Network (RBFNN) and Multi-Layer Perceptron (MLP) algorithm. The researcher concludes that if the research focuses on accuracy, then it is better to use the RBFNN algorithm, whereas if research focuses on speed in producing classifications for alternative answers to decision making, then it is better to use MLP algorithms. This study proves that ANN algorithms can be used in the process of classifying land use and cover with good accuracy.

Another study used the Cellular Automation (CA) approach to predict changes in land use and cover in the downstream area of the Ci Tarum Watershed, Indonesia. Landsat satellite image data is used for extraction of land use and cover classification with the Supervised Maximum Likelihood classification algorithm. The results of this study stated that the driving factors that influence changes in land use and cover are the distance from the road and the city centre (Karsidi, 2004). Another studies using combination of CA and Markov Chain analysis to monitor and predict the land use changes in Iran, Morocco, and China (Ansari & Golabi, 2019; Hyandye & Martz, 2017; Jazouli et al., 2019; Liping et al., 2018). All of these studies employed Landsat family datasets for land use class extraction and using commercial IDRISI software for data processing.

Some researcher also introduced the SLEUTH (Slope, Land use, Exclusion, Urban extent, Transportation, and Hillshade) model (Clarke, 2008; Feng et al., 2012). It is a model for future urban growth prediction and land use change. The researcher suggested to execute subsequent simulations by testing and choosing different parameters to evaluate the performance of the SLEUTH model. Furthermore, uncertainty, class consistency, and mis-registration issues become more of a problem as more time periods are used in the SLEUTH model.

There are many studies have been implemented the combination of CA-Markov for land use modelling (Gharaibeh et al., 2020; Halmy et al., 2015; Mansour et al., 2020; Okwuashi & Ndehedehe, 2020; Rahnama, 2021; Wang et al., 2021; Wu et al., 2019; Zhou et al., 2020). However very limited studies using neural network combination with CA for modelling land-use dynamic. For instance, Almeida et al., (2008) conducted the simulation for Sao Paulo State. The researchers remarked that to achieve high accuracy, user intervention is still needed. Especially in order to define the training algorithm, the neural network architecture as well as the parameters. Furthermore, Du et al., (2018) compared the cellular automata with tree-based method for land use changes modelling of Greater Tokyo Area. In this study the result showed that the tree-based algorithm outperformed the Multi-Layer Perceptron (MLP) algorithm. However, the researcher also stress-out that it is depend on the size of the neighbourhood and land use type of the neighbourhood. Anthropogenic factors such as policy and socio-economic also play important roles in transition process.

The objective of this study is to calculate the extent of changes in land use and cover in Malang region during the period 2019 to 2021 and 2023 with a combination of FA, CA, and ANN using high-spatial-resolution Sentinel-2 satellite image data. The advantage of using satellite image data is due to its wide coverage area in one acquisition time so that it can be consistently maintained. In addition, the existence of

open and free satellite image processing software for the public makes this analysis easier and more accurate.

In addition, this study will monitor, map and take inventory of the current conditions of land use and cover with very high accuracy satellite image data. Geospatial data and information that has been successfully produced will enable stakeholders make better informed decisions to avoid the negative impacts of future development.

II. DATA & METHODOLOGY

The data used in this study is the optical sensor data from the acquisition of the Sentinel-2 satellite by the European Union's Copernicus Program in the Great Malang region, which includes Malang City, Batu City, and Malang Regency (Figure 1).

Sentinel-2 is an optical sensor that records objects on the surface of the earth with a satellite that orbits at an altitude of more than 780 km above the earth's surface. Sentinel-2 is part of the European Union's Copernicus Program which orbits more than 5 earth observation satellites. Sentinel-2 records using 13 different wavelengths, ranging from sensitive wavelengths for aerosols, vegetation, water, visible light wavelengths, to Short Wave Infra-Red (SWIR).

Sentinel-2 data that will be used is data with Level1-C and Level2-A, which is data with an area of more than 100 km² that has gone through a geometric correction process so that researchers no longer need to make these corrections. Sentinel-2 data will re-record the same area within five days with a 10-meter pixel spatial resolution (ESA, 2015).

Malang City is the second most populous city after the city of Surabaya in East Java. It is known to be a tourist and higher education destination for prospective university students in Indonesia. The population until 2018 reaches 866,118 (BPS, 2019a). Furthermore, Kota Batu is a fraction of Malang Regency, which has been officially established since 2001. It is currently a natural and family tourist destination for local and foreign tourists because of the presence of East Java Park 1, 2 and 3. The population of Batu City until 2018 is 205,788 (BPS, 2019b). Finally, Malang Regency is the largest administrative area compared to Malang City and Batu City, with the highest population. Until 2018, the population of Malang Regency was recorded at 2,591,795 (BPS, 2019c).

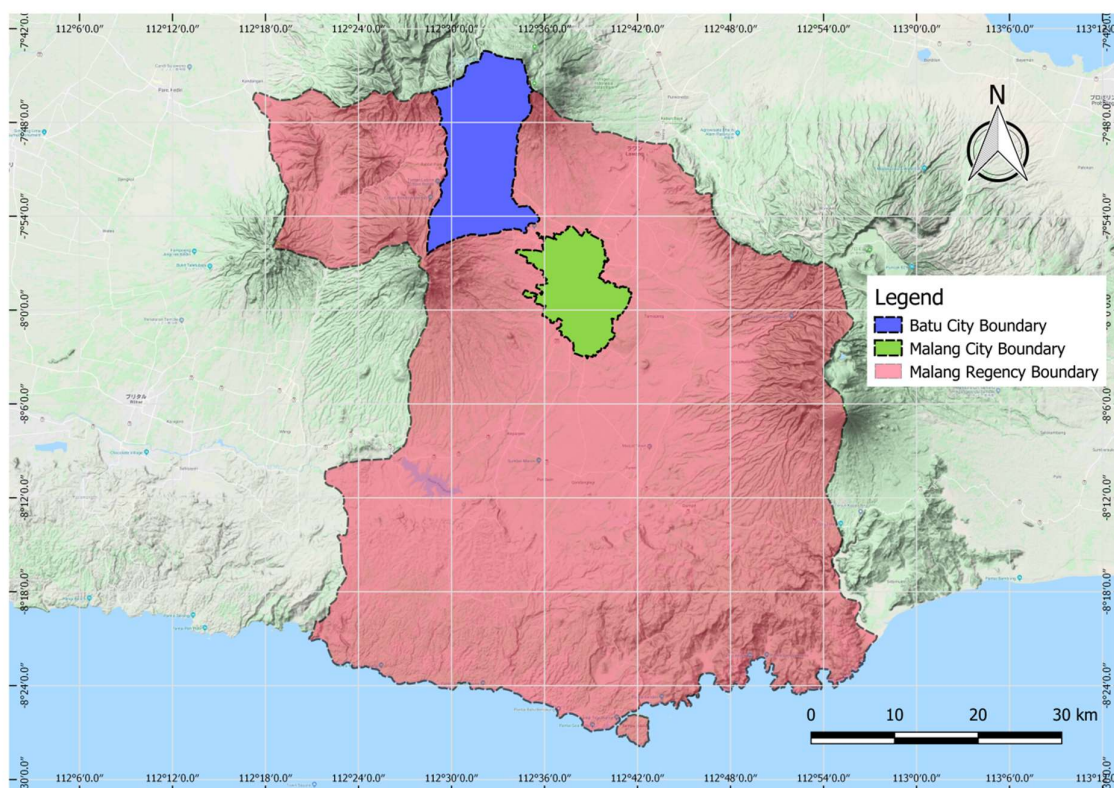


Figure 1. Study area of Great Malang region.

2.1 Procedure for data processing

To achieve the stated objective, in this research, a series of steps as illustrated in Figure 2 was developed. In general, the stages of research are divided into three; (A) the processing and classification stages of Sentinel-2 satellite image data (2015 and 2017) which will be used as input data in the next process; (B) the stages of data processing driving factors and constraints with FA, and the last step (C) is the stage of the prediction process using the ANN algorithm and CA.

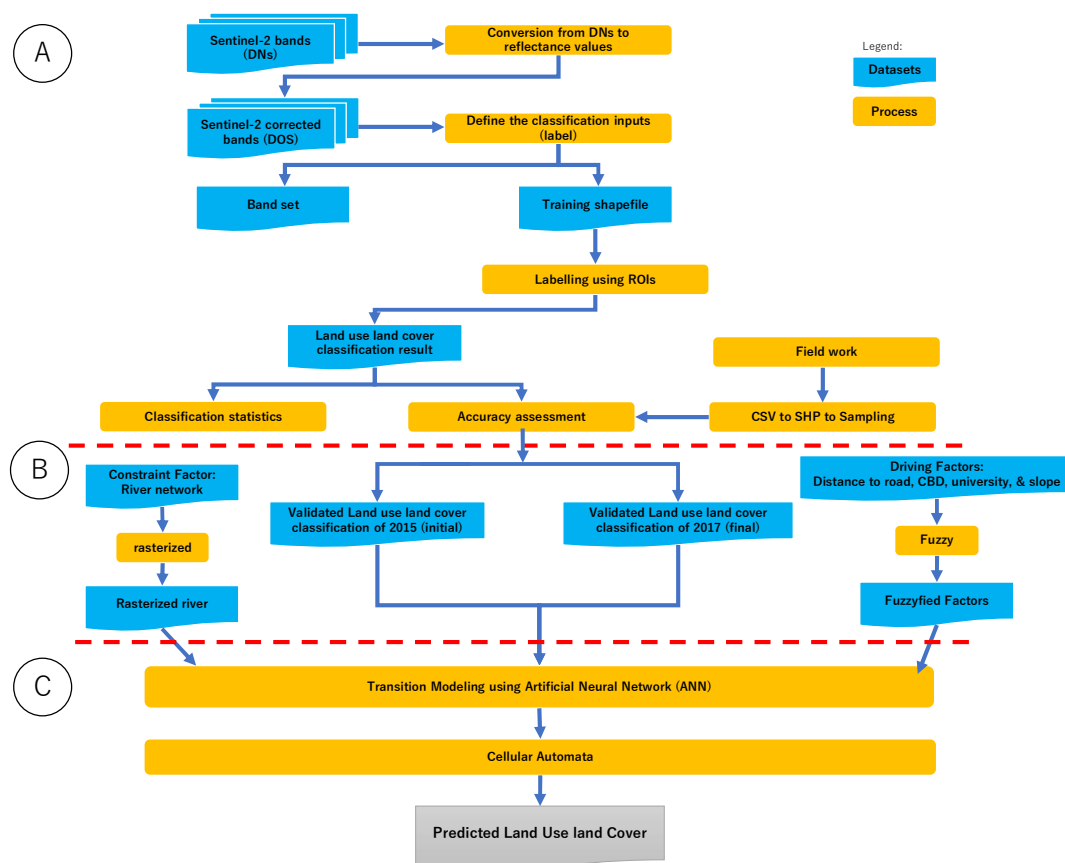


Figure 2. (A) Processing of Sentinel-2 Data, (B) Processing of Input (Driving and Constraint) Data, and (C) Prediction Process.

Stage one (A) is processing of Sentinel-2 data. Sentinel-2 data in the Great Malang region must first be converted into the actual reflectance value of the digital number (DN) value. The result is a corrected image from atmospheric and radiometric disturbances. Then the researcher will determine the land use classification label consisting of four main classes, (1) built-up area, which include all man-made structures, (2) sparse vegetation, which includes all low vegetation and agricultural land, (3) dense vegetation, include all high vegetation and tropical forest, and (4) waterbody. Next, in the classification process, all bands or channels must be combined together first, or known as bandset. Researchers will then choose Region of Interest (ROI) using pointer ROI so that the selected pixels are only pixels with homogeneous values. Each selected ROI must then be labelled according to the four main classes. The above steps must be carried out for all data in two different periods.

Then, researchers will use the Spectral Angle Mapper (SAM) algorithm to classify land use and cover. Finally, the extent of each land use and cover will be calculated and compared to the changes and validated with field survey data using Spectra GPS Handheld, a comparison matrix will be used to calculate the value of accuracy.

Then, stage two (B) is processing data of the driving factors, that is the distance from the road network, distance from the city centre, distance from the university, slope, and the constraint factor is the distance from the river network. Fuzzy sets will be applied to road network data with a limit value of 50 m. The closer it is to the road, the more likely it is that a class of land use and cover changes, whereas the farther it is from

the road network, the less likely there exist a class of land use and cover to change. While the distance from the city centre is set at 5 km, the distance from the university is 2 km, and slopes should not exceed 15%. River networks are used as constraints factor because it is assumed that the river may not experience any changes. Table 1 summarizes Fuzzy requirements for each driving and constraint factor. The value is based on the Indonesian National Standard (SNI-*Standart Nasional Indonesia*) No. 03-1733-2004 (BSN, 2004).

The last stage (C) is the determination of the probability change value through the ANN algorithm calculation process, and then changes in land use and cover will be simulated using the CA algorithm.

Table 1. Fuzzy set membership values.

No	Factors	Value	Factor Type
1	<i>Distance from the road network</i>	<i>50 m</i>	<i>Driving</i>
2	<i>Distance from the city centre</i>	<i>5 km</i>	<i>Driving</i>
3	<i>Distance from the university</i>	<i>2 km</i>	<i>Driving</i>
4	<i>Slope</i>	<i>Less than 15%</i>	<i>Driving</i>
5	<i>Distance from the river network</i>	<i>50 m</i>	<i>Constraint</i>

2.1.1 Pre-processing

At the stage of pre-processing, the data that has been obtained should pass through the stages of atmospheric and radiometric correction. In the Semi-automatic Classification (SCP) plugin of QGIS (Congedo, 2020), the researcher will process the correction automatically by activating the correction feature (Apply DOS1 atmospheric correction).

Furthermore, large satellite image data will be clipped using the administrative boundaries of Malang City, Batu City, and Malang Regency to facilitate the computational process. The classification results from two different periods (2015 and 2017) will then be used as input data in the analysis of predictions of changes in land use and cover in 2019.

The software that will be used in this study is QGIS with the Semi-automatic Classification plugin (Congedo, 2020). In addition, to carry out the process of predicting changes in land use and cover, researchers will use the MOLUSCE plugin with the CA and ANN features.

2.1.2 Processing

Cellular Automata (CA) method is a method commonly used in making predictions of land-use changes in a geographic information system (GIS). This method is based on the principle of cell/pixel that changes according to the conditions of its closest neighbours and various other spatial variables that influence.

Before being able to process data using the CA method, we must prepare several land use data in different periods, at least two periods. The data must be in a raster data format with the same georeferenced system so that it can be processed using the geocomputation method. In this case, the land use data used are 2015 and 2017 land use data from the Great Malang region (Malang City, Batu City, and Malang Regency)

which are processed from Sentinel-2 satellite image data using the semi-automatic classification method.

Supporting data such as road networks are used with the assumption that the area to be built in the future is an area close to the road network which is used as the main transportation access. This data is also known as a driving factor, which is a spatial variable that will affect the results of predictions. Fuzzy rules that are imposed are, assuming the area with proximity of the road network is as far as 50 m, the higher possibility of a change occurring in an area. The farther it is from 50 m, the lower the possibility of changes in land use and cover. The type of membership function used is sigmoid.

If the supporting data is available in vector format, it must be converted to a raster format. The approach is to use the “Rasterize” function in QGIS. In order to be processed in the CA method, all raster data must have the same geometry size, otherwise, the “Align” process must be done first. After all data has been prepared, we need to install the MOLUSCE plugin. To predict the land use of year 2019, the land use of year 2015 will be employed as “Initial” data, and the land use of year 2017 will be employed as “Final” data. This process will produce “land-use changes” dataset as the first input data in the input layer of neural network. While the driving and constraint factors will be set as the other input data in the input layer of neural network as shown in Figure 3.

In the training process, we use the ANN approach using 3, 5, 7, and 9 hidden layers with each of 500, 1000, and 2000 iterations as the “Stopping Criteria”. The scenario is used in order to get the best kappa value (close to 1). The example of architecture of ANN in this study is shown in Figure 3.

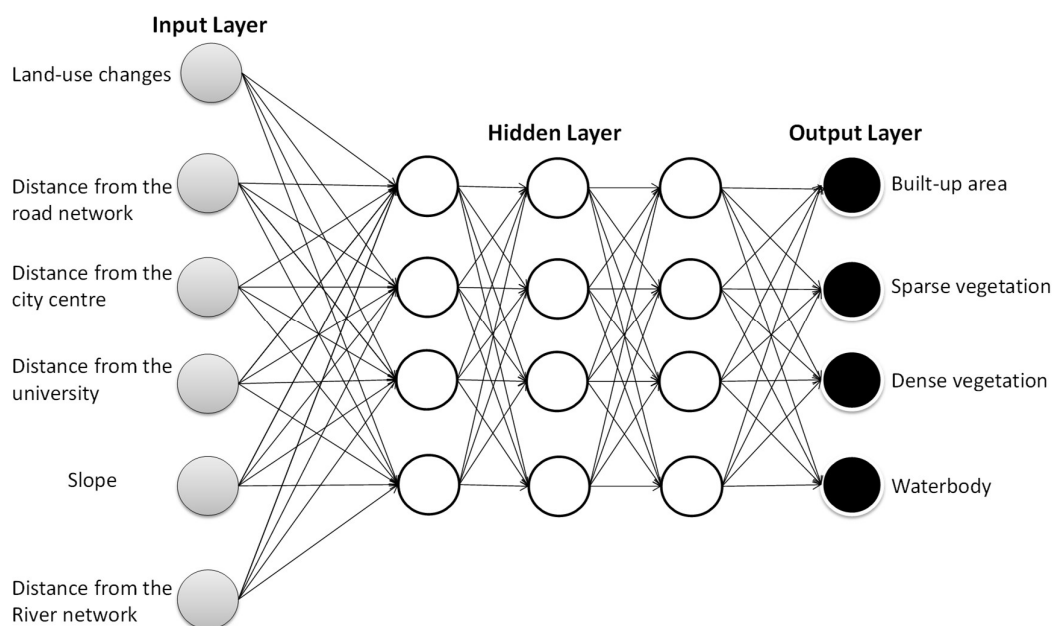


Figure 3. ANN architecture with three hidden layers.

2.1.3 Post-processing

Accuracy of classification result was calculated using a matrix table from the following URL: <http://vassarstats.net/kappa.html>. Where ground truth data has been obtained previously with field survey activities. Table 2 shows the data entry form field survey to produce the kappa values. Field surveys are conducted using a systematic unaligned sampling pattern. The field survey was conducted from July 8 to July 16, 2019. The survey was conducted using Spectra Precision GPS handhelds with a horizontal accuracy of 1 m in real-time.

Table 2. Data entry form to produce kappa values.

		Classified				
		Built-up area	Sparse vegetation	Dense vegetation	Waterbody	Totals
Ground truth	Built up area	----	----	----	----	----
	Sparse vegetation	----	----	----	----	----
	Dense vegetation	----	----	----	----	----
	Waterbody	----	----	----	----	----
Totals		----	----	----	----	----

III. RESULT

3.1 Fuzzified factors

The driving and constraint factors such as distance from the road network, distance from the city centre, distance from the university, slope, and river network that have been processed with FA are presented in Figure 4.

Figure 4 (A) to (D) are the driving factors, where blue colour visualizes a poor level of suitability, green colour is for an area with low level of suitability, yellow colour is for areas with fair level of suitability, orange colour is for the area with the good level of suitability, and red colour is for the area with excellent level of suitability. While Figure 4 (E) is a constraint factor, where the river area may not experience any changes.

Figure 4 (A) shows that along the road is an area that is suitable for land conversion, the farther away it is from the road network, the more inappropriate it will be. Meanwhile, Figure 4 (B) shows that a very suitable area is too limited. The farther it is from the city centre, the lesser its suitability, as is the case with Figure 4 (C), where the farther it is from the university location, the lesser its suitability. Furthermore, Figure 4 (D) clearly shows that the slope level in the central part of the Great Malang area covers very wide area and is suitable for conversion into an urbanized area. While Figure 4 (E) shows that the area around the river network is inappropriate and the farther it is from the river, the higher the level of suitability.

3.2 Simulation result

3.2.1 First scenario: 3 hidden layers with 500, 1000, and 2000 iteration

The simulation was carried out using 2015 and 2017 land-use class inputs from Sentinel data to predict land use in 2019. The results of the simulation process of predicting changes in land use using 3 hidden layers provide output with the lowest accuracy. Visually, this can be seen from Figure 5 (A) 500 iterations and (B) 1000 iterations that the built-up area actually decreases and is replaced with an agricultural or vegetation area. This is contrary to the theory of land use conversion; that it is not possible to change from a built-up area to a large scale agricultural or vegetation area. The factor of proximity to the city centre and the location of the university seem to greatly influence the outcome of predictions with this scenario. This is shown by Figures 5 (A) and (B), where land-use change appears to be circular following the factor of proximity to the city centre and university location.

Meanwhile, with the scenario of using 2000 iterations, the results are shown in Figure 5 (C), where the forest area was reduced rapidly in the southern part of the study area and turned into a large-scale agricultural area. This result is very different from the simulation results using other scenarios, where the forest area did not decrease rapidly in the southern part of the study area. This is also difficult because the southern slope of the study area is steep and varies so that it is technically very difficult to convert forest areas with steep slopes into agricultural areas.

3.2.2 Second scenario: 5 hidden layers with 500, 1000, and 2000 iteration

Simulations using scenarios of 5 hidden layers and 500 iterations, 1000 iterations, and 2000 iterations are shown in Figure 6 (A), (B), and (C) respectively. The simulation results show better results compared to using the hidden layer of scenario 3.

However, the simulation results with 500 iterations have the lowest accuracy with 1000 iterations and 2000 iterations. This is shown visually by the disappearance of forest land cover in the southern part of the study area which is a steep and varied slope. Figure 6 (A) shows the area of forest land cover that is converted to a massive agricultural area, while Figure 6 (B) and (C) shows the opposite. The area to be built constantly is expanding, but the forest area does not reduce significantly in the southern part of the study area.

3.2.3 Third scenario: 7 hidden layers with 500, 1000, and 2000 iteration

The next scenario is to use 7 hidden layers and 500, 1000, and 2000 iterations. The final results are shown sequentially in Figures 7 (A), (B), and (C). Scenarios provide better outcomes compared to the previous scenario, especially in 1000 and 2000 iterations, as shown in Figure 7 (B) and (C).

Figures 7 (B) and (C) visualize the simulation results which show that the built-up area continues to expand in all directions according to proximity to the road network, proximity to the city centre and university location, and sloping slopes. The forest area in the southern part of the study area did not decrease significantly due to steep slope factors. However, using a scenario with an iteration of 500 (Figure 7 (A)) gives less

accurate results, where the built-up area does not increase significantly and the forest area in the southern part of the study area actually disappears on a large scale and it is converted into an agricultural area.

3.2.4 Fourth scenario: 9 hidden layers with 500, 1000, and 2000 iteration

Visually, the scenario using 9 hidden layers gives the best results compared to other scenarios, especially using 1000 and 2000 iterations. Figure 8 (A), (B), and (C) visualize the simulation results using 9 hidden layers and 500, 1000, and 2000 iterations in sequence.

The simulation results using 500 iterations provide the least output because the built-up area increases by following only the proximity factor to the existence of the university location. This can be seen from Figure 8 (A) where the area is built in a circle depending on the proximity to the location of the city centre and the university. While using the number of 1000 iterations (Figure 8 (B)), the simulation results showed improvement with the increase in the built-up area that follows all factors. This increase in the built-up area resulted to a reduction in the area of agricultural land around the road network. The forest area in the south is maintained. Furthermore, an iteration of 2000 (Figure 8 (C)) gives an increasingly better outcome. The forest area in the southern part of the study area reduced but not significantly, while the built-up area or urban area became wider by transforming agricultural areas.

3.3 Accuracy assessment

The kappa value is obtained by comparing the class results of the classification with the results of the field survey. Table 3 shows the result of field survey, while Tables 4 to 5 show the results of kappa calculations. Using method 1 (Sheskin, 2004) and method 2 (Fleiss, 2003), the result of unweighted kappa produced observed kappa value of 0.83, while linear weighting produced observed kappa value of 0.9 for classification of LULC of year 2019.

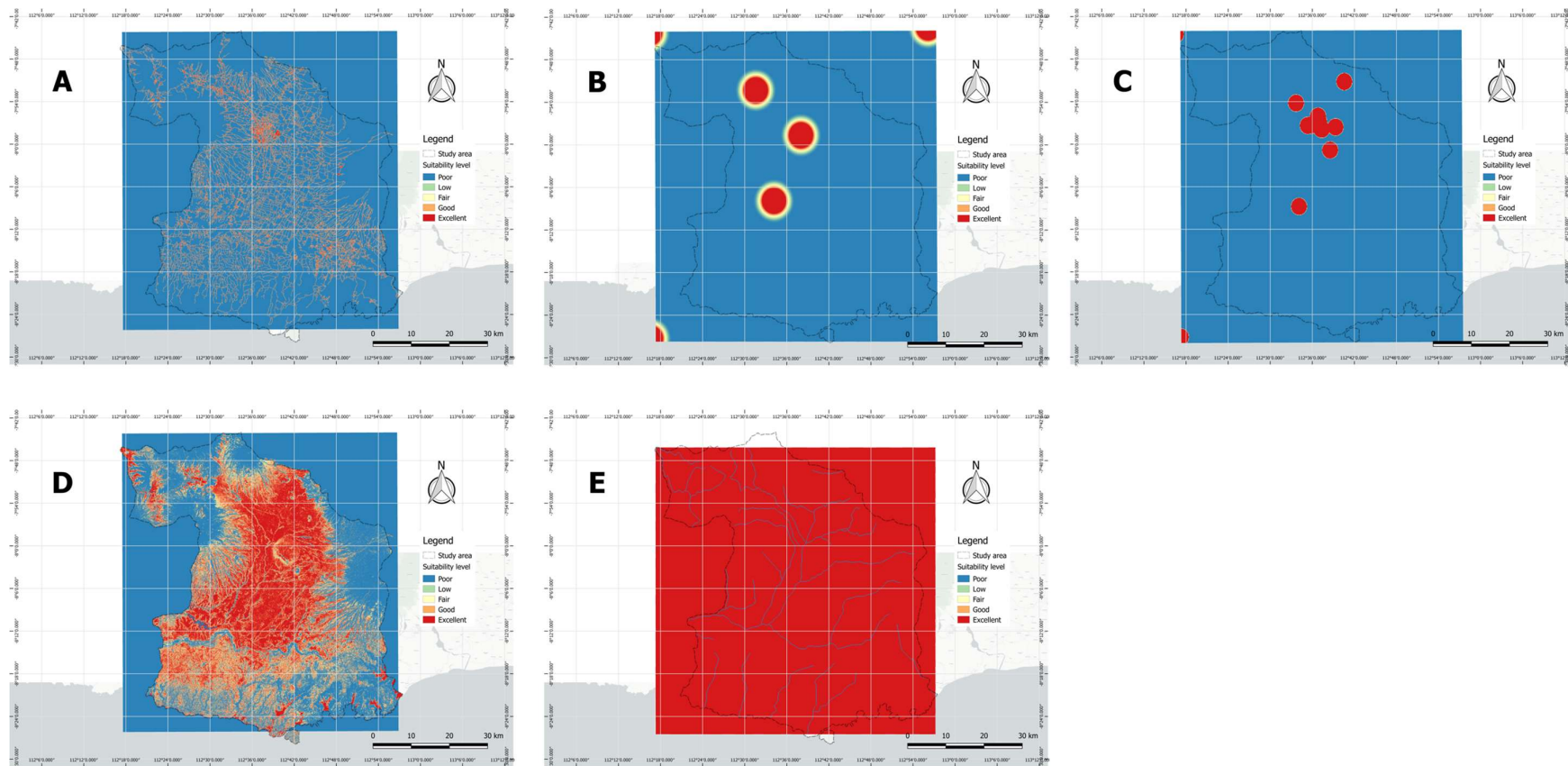


Figure 4. Fuzzified result using sigmoid membership function; (A) distance from road network, (B) distance from city centre, (C) distance from university, (D) slope, and (E) river network.

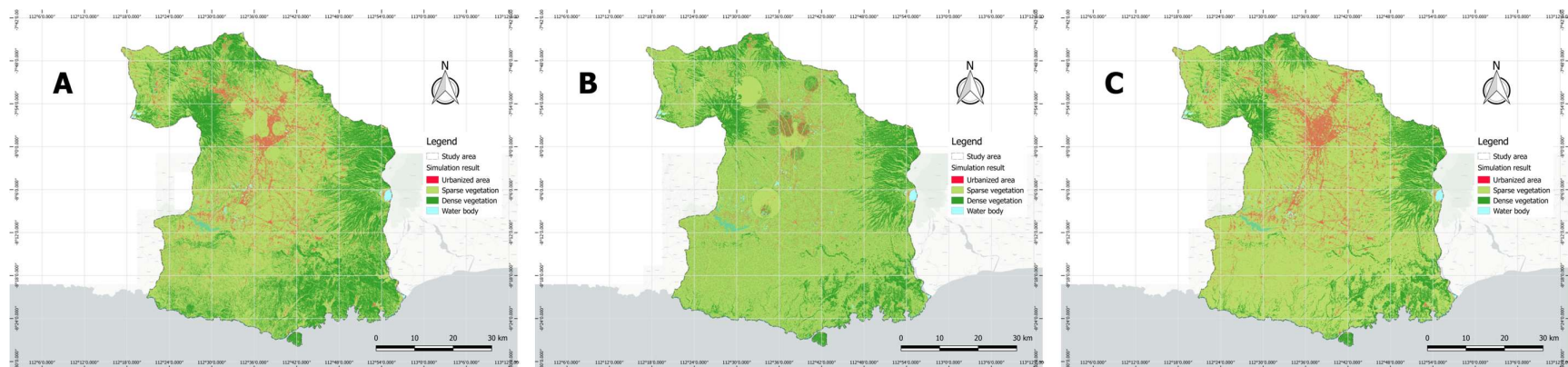


Figure 5. Simulation result of the first scenario: 3 hidden layers with (A) 500 iteration, (B) 1000 iteration, and (C) 2000 iteration.

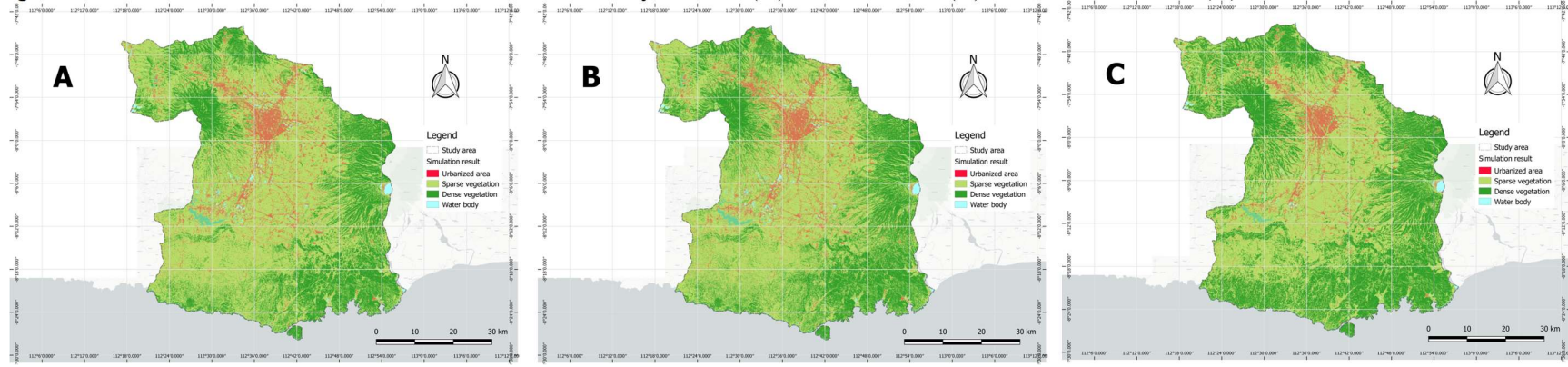


Figure 6. Simulation result of the second scenario: 5 hidden layers with (A) 500 iteration, (B) 1000 iteration, and (C) 2000 iteration.

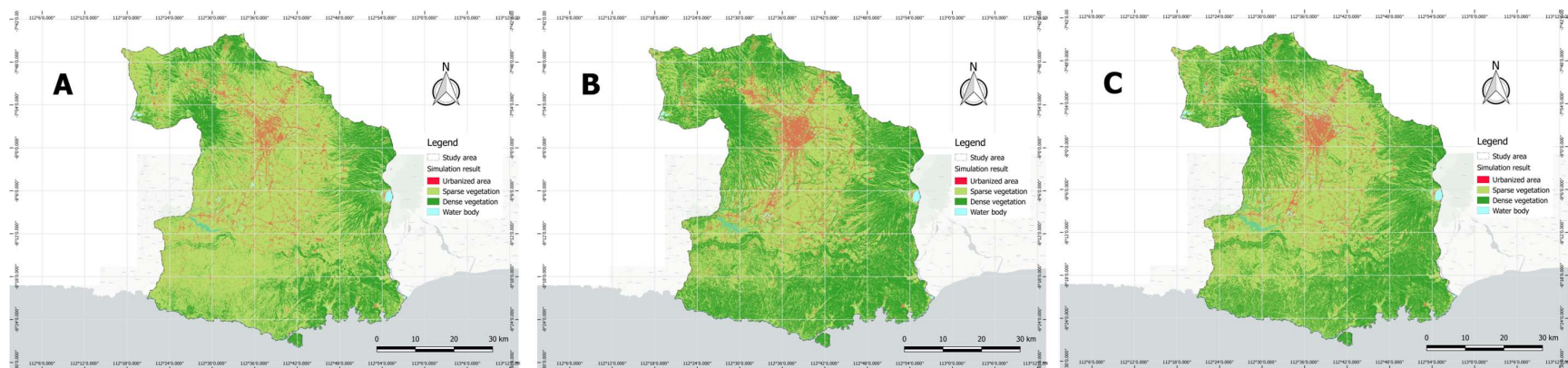


Figure 7. Simulation result of the third scenario: 7 hidden layers with (A) 500 iteration, (B) 1000 iteration, and (C) 2000 iteration.

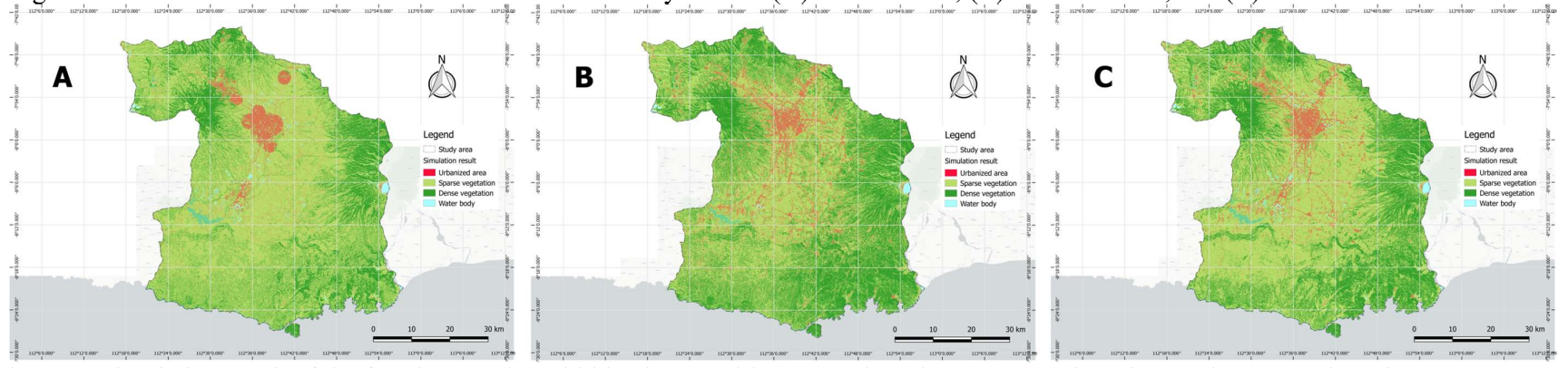


Figure 8. Simulation result of the fourth scenario: 9 hidden layers with (A) 500 iteration, (B) 1000 iteration, and (C) 2000 iteration.

Table 3. The field survey data sets

		Predicted				Totals
		Built-up area	Sparse vegetation	Dense vegetation	Waterbody	
Ground truth	Built up area	12	0	0	0	12
	Sparse vegetation	0	9	3	0	12
	Dense vegetation	0	3	9	0	12
	Waterbody	0	0	0	12	12
Totals		12	12	12	12	48

Table 4. The result of unweighted kappa.

<i>Observed Kappa</i>	<i>Standard Error</i>	<i>.95 Confidence Interval</i>	
0.833		<i>Lower Limit</i>	<i>Upper Limit</i>
<i>Method 1</i>	0.0636	0.7086	0.958
<i>Method 2</i>	0.0637	0.7085	0.9581

Table 5. The result of kappa with Linear Weighting.

<i>Observed Kappa</i>	<i>Standard Error</i>	<i>.95 Confidence Interval</i>	
		<i>Lower Limit</i>	<i>Upper Limit</i>
0.9	0.0407	0.8202	0.9798

The results of the analysis in Table 6 showed that the best accuracy was obtained using 7 hidden layers and 2000 iterations with a kappa value of 0.393. The second rank was obtained using 9 hidden layers and 2000 iterations with a kappa value of 0.392, and the last rank was obtained using the scenario of 9 hidden layers and 1000 iterations with a kappa value of 0.390.

Furthermore, analysis was also carried out based on different land uses. Of the four land uses analysed, the built-up area had the highest accuracy, 86.39% for producer accuracy and 75.81% for user accuracy compared to other land use classes. Table 7 shows the results of accuracy based on a single land-use class.

Table 6. kappa value

		<i>Number of iteration</i>		
<i>Number of hidden layer</i>		500	1,000	2,000
3		0.229	0.241	0.217
5		0.266	0.360	0.360
7		0.364	0.364	0.393
9		0.380	0.390	0.392

Table 7. Accuracy based on the single land-use class

<i>Class</i>	<i>Producer accuracy (%)</i>	<i>User accuracy (%)</i>
<i>Built-up area</i>	86.39	75.81
<i>Sparse vegetation</i>	39.18	54.71
<i>Dense vegetation</i>	31.23	20.49
<i>Waterbody</i>	30.01	31.75

The results of the accuracy analysis of year 2019 can be used as a reference that the next prediction data will also have the same accuracy, especially in the built-up area. Prediction results are only done for the built-up area because only this class has the best accuracy compared to other land use classes. While sparse vegetation (low vegetation and agricultural areas) and dense vegetation (high vegetation and tropical forests), and waterbodies were excluded from analysis, since they had low accuracy, 39%, 31%, and 30% respectively. The built-up area has an accuracy of almost 87% in 2019, so it is assumed that the area built in 2021 and 2023 also has the same accuracy.

3.4 Statistical analysis

Statistical analysis is performed to determine the relationship between the number of hidden layers and iterations with the accuracy produced. RStudio software is used to calculate the correlation between two independent variables (hidden layer and iteration) and one dependent variable (accuracy). The input data is shown in Table 6 and the calculation results are shown below:

Residuals:

Min	1Q	Median	3Q	Max
-0.049518	-0.018286	-0.009241	0.020781	0.059304

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.529e-01	3.447e-02	4.435	0.00164 **
HL	2.600e-02	4.507e-03	5.769	0.00027 ***
IT	1.782e-05	1.616e-05	1.103	0.29873

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03491 on 9 degrees of freedom

Multiple R-squared: 0.7931, Adjusted R-squared: 0.7471

F-statistic: 17.25 on 2 and 9 DF, p-value: 0.0008337

The results of statistical calculations show a positive correlation with a strong relationship between the number of hidden layers and iterations with the accuracy produced, this is indicated by the value of the adjusted R-squared 0.75. Furthermore, statistical analysis is also conducted separately to determine the relationship between one independent variable and one dependent variable with an exploratory approach. The relationship between the number of hidden layers with accuracy and the number of iterations with accuracy is presented by Figures 9 and 10, respectively.

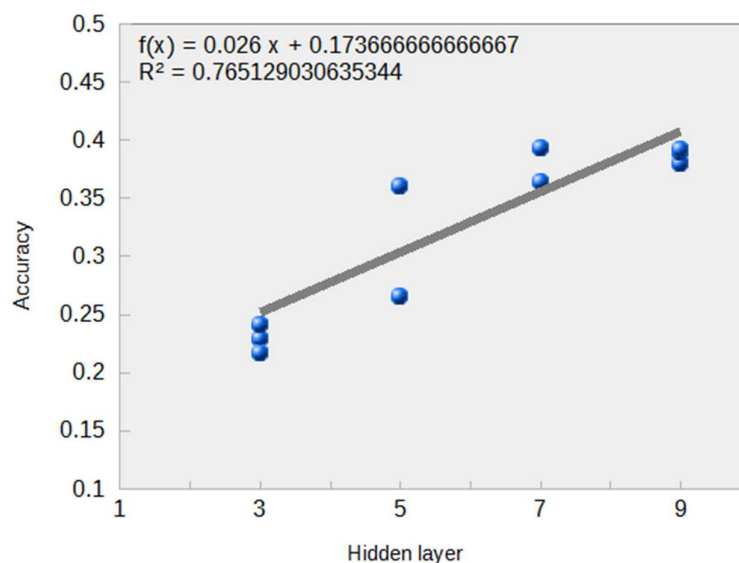


Figure 9. Relationship between numbers of the hidden layer and accuracy.

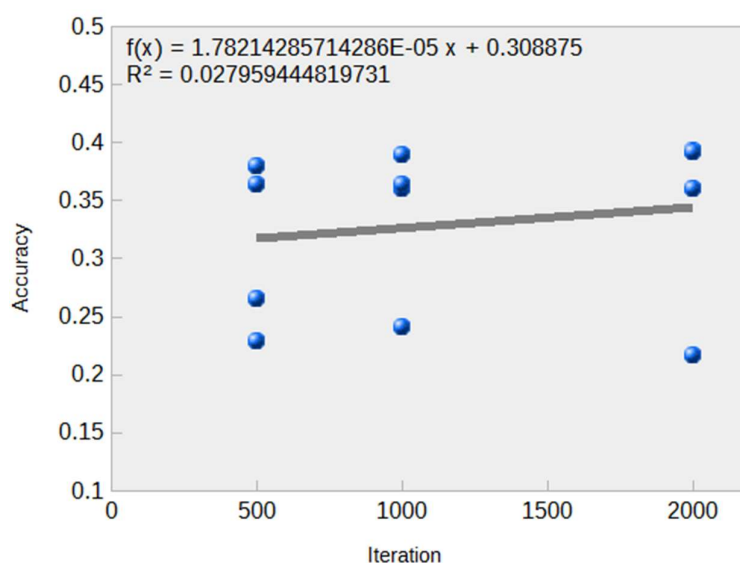


Figure 10. Relationship between numbers of iterations and accuracy.

Figures 9 and 10 show that the number of hidden layers has a stronger relationship than the number of iterations to the accuracy produced. This is indicated by the position of the trend line on the number of hidden layers that are tilted compared to the number of iterations. In addition, the R2-value of the relationship between the numbers of hidden layers with accuracy is 0.765, which means strong. While the R2-value of the relationship between the number of iterations and accuracy is only 0.027, which means very weak.

3.5 Built-up area prediction in the future

Prediction analysis of built-up area expansion was carried out using neural network

methods and CA with the best scenarios, using 7 hidden layers and 2000 iterations. The analysis was carried out for the periods of 2021 and 2023, where the input data used is year 2017 and 2019 for predictions of 2021, and year 2019 and 2021 for predictions of 2023.

The highest change in the built-up area occurs within the period of 2017 to 2019, while there has been a drastic decline between the 2019 to 2021 period, as well as the period of 2021 to 2023. However, even though the increase in the built-up area has decreased, generally, the area built up continues to grow along with demand and population growth. The increase in built-up area from 2017 to 2023 is presented in Table 8.

Table 8. The changes of the built-up area over three periods of analysis.

<i>Class</i>	<i>Period</i>			
	<i>2017 (initial)</i>	<i>2019 (predicted)</i>	<i>2021 (predicted)</i>	<i>2023 (predicted)</i>
<i>Built up area</i>	<i>185,452.20</i>	<i>202,467.15</i>	<i>205,316.64</i>	<i>209,279.61</i>
<i>Changes (%)</i>	<i>-</i>	<i>4.5</i>	<i>1.39</i>	<i>1.89</i>

IV. DISCUSSION AND CONCLUSION

This study successfully proved that the number of hidden layers and iterations will affect the outcome of the land-use model in the Great Malang region. However, that does not mean that the more number of hidden layers will produce a model with better accuracy.

In this study, we tested various scenarios with a different number of hidden layers and iterations. The scenario with 7 hidden layers gives the most accurate land-use model output with a kappa value close to 0.4. Furthermore, the higher number of iterations, the better and consistent the results, where the number of iterations 2000 predicts the expansion of the built-up area with an accuracy of more than 85%.

The results of the statistical analysis also show that there is a positive and strong relationship between the number of hidden layers and iterations used with the results of accuracy obtained where the adjusted R-squared value is close to 0.80. Separately, the results of exploratory statistical analysis show that the number of hidden layers has a stronger relationship compared to the number of iterations to the accuracy of the land use prediction model.

The predicted increase in the built-up area shows a low trend from the 2019 period to 2021 and the period 2021 to 2023. In the 2019 period to 2021, the built-up area only increased by 1.39%, while between the periods 2021 to 2023, the area built only increased by 1.89%.

The factors that influence land-use change in this study have been able to predict the expansion of the built-up area with fairly good accuracy so that in the future it is expected that the growth of unplanned well-built up areas can be minimized.

Data availability

Ramdani, Fatwa (2020), "Land Use Modeling of Malang", Mendeley Data, V2, doi: 10.17632/47scmwsp3t.2

References

- Almeida, C. M., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S. (2008). Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science*, 22(9), 943–963. <https://doi.org/10.1080/13658810701731168>
- Ansari, A., & Golabi, M. H. (2019). Prediction of spatial land use changes based on LCM in a GIS environment for Desert Wetlands – A case study: Meighan Wetland, Iran. *International Soil and Water Conservation Research*, 7(1), 64–70. <https://doi.org/10.1016/j.iswcr.2018.10.001>
- Arihta, L. S., Ramdani, F., & Saputra, M. C. (2018). Pengembangan Aplikasi Safety Flying Drone Map (Syndrone Map) pada Perangkat Mobile Android dengan Menggunakan Metode Spatial Multi- Criteria Evaluation. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(10), 3515–3524.
- BPS. (2018). *Proyeksi Penduduk Menurut Kabupaten/Kota dan Jenis Kelamin di Jawa Timur 2018*. BPS Provinsi Jawa Timur. <https://jatim.bps.go.id/dynamictable/2018/02/05/314/proyeksi-penduduk-menurut-kabupaten-kota-dan-jenis-kelamin-di-jawa-timur-2018.html>
- BPS. (2019a). *Jumlah Penduduk di Kota Malang Menurut Kecamatan dan Jenis Kelamin, 2011-2020*. BPS Kota Malang. <https://malangkota.bps.go.id/dynamictable/2019/05/15/19/jumlah-penduduk-di-kota-malang-menurut-kecamatan-dan-jenis-kelamin-2011-2020.html>
- BPS. (2019b). *Jumlah Penduduk Menurut Jenis Kelamin dan Kecamatan di Kota Batu, 2018*. BPS Kota Batu. <https://batukota.bps.go.id/statictable/2019/10/03/458/jumlah-penduduk-menurut-jenis-kelamin-dan-kecamatan-di-kota-batu-2018.html>
- BPS. (2019c). *Penduduk, Luas Wilayah dan Kepadatan Penduduk Menurut Kecamatan, Tahun 2015-2018*. BPS Kabupaten Malang. <https://malangkab.bps.go.id/statictable/2015/03/14/403/penduduk-luas-wilayah-dan-kepadatan-penduduk-menurut-kecamatan-tahun-2015-2018.html>
- BSN. (2004). *Standar Nasional Indonesia (SNI) No. 03-1733-2004 about the Procedures for urban housing planning*. BSN Indonesia. <http://sni.litbang.pu.go.id/index.php?r=/sni/new/sni/detail/id/694>
- Clarke, K. C. (2008). Mapping and modelling land use change: an application of the SLEUTH model. *Lecture Notes in Geoinformation and Cartography*, 9783540691679, 353–366. https://doi.org/10.1007/978-3-540-69168-6_17
- Congedo, L. (2020). *Semi-Automatic Classification Plugin Documentation Release 7.0.0.1 Luca Congedo*. <https://doi.org/http://dx.doi.org/10.13140/RG.2.2.29474.02242/1>
- Du, G., Shin, K. J., Yuan, L., & Managi, S. (2018). A comparative approach to modelling multiple urban land use changes using tree-based methods and cellular automata: the case of Greater Tokyo Area. *International Journal of Geographical Information Science*, 32(4), 757–782. <https://doi.org/10.1080/13658816.2017.1410550>

- ESA. (2015). *Colour Vision for Copernicus – The Story of Sentinel-2*.
https://esamultimedia.esa.int/docs/EarthObservation/Sentinel-2_ESA_Bulletin161.pdf
- Feng, H. H., Liu, H. P., & Lü, Y. (2012). Scenario Prediction and Analysis of Urban Growth Using SLEUTH Model. *Pedosphere*, 22(2), 206–216.
[https://doi.org/10.1016/S1002-0160\(12\)60007-1](https://doi.org/10.1016/S1002-0160(12)60007-1)
- Fleiss, J. L. (2003). *Statistical Methods for Rates and Proportions*. John Wiley & Sons.
- Gharaibeh, A., Shaamala, A., Obeidat, R., & Al-Kofahi, S. (2020). Improving land-use change modeling by integrating ANN with Cellular Automata-Markov Chain model. *Heliyon*, 6(9), e05092. <https://doi.org/10.1016/j.heliyon.2020.e05092>
- Hadi, M., Hariyanto, S., Ramdani, F., & Saputra, M. C. (2018). Sistem Informasi Geografis Kesesuaian Lahan Perumahan di Kota Malang menggunakan Metode MCE. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(1), 263–272.
- Halmy, M. W. A., Gessler, P. E., Hicke, J. A., & Salem, B. B. (2015). Land use/land cover change detection and prediction in the north-western coastal desert of Egypt using Markov-CA. *Applied Geography*, 63, 101–112.
<https://doi.org/10.1016/j.apgeog.2015.06.015>
- Hyandye, C., & Martz, L. W. (2017). A Markovian and cellular automata land-use change predictive model of the Usangu Catchment. *International Journal of Remote Sensing*, 38(1), 64–81. <https://doi.org/10.1080/01431161.2016.1259675>
- Jazouli, A. El, Barakat, A., Khellouk, R., Rais, J., & Baghdadi, M. El. (2019). Remote sensing and GIS techniques for prediction of land use land cover change effects on soil erosion in the high basin of the Oum Er Rbia River (Morocco). *Remote Sensing Applications: Society and Environment*, 13(May 2018), 361–374.
<https://doi.org/10.1016/j.rsase.2018.12.004>
- Junianto, D. D., Ramdani, F., & Pramono, D. (2018). Sistem Informasi Penentuan Lokasi Pembangunan Kawasan Industri di Kabupaten Mojokerto Menggunakan Metode Multi-Criteria Evaluation. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(9), 2476–2484.
- Karsidi, A. (2004). Spatial Analysis of Land Use / Land Cover Change Dynamics Using Remote Sensing and Geographic Information Systems : A Case Study in the down stream and surroundings of the Ci Tarum watershed [The University of Adelaide]. In *Doctoral Thesis* (Issue May).
<https://digital.library.adelaide.edu.au/dspace/handle/2440/22077>
- Lestantyo, P. (2019). Utilization of Current Data for Geospatial Analysis of the Appropriateness of Apple Plantation Land Based on Fuzzy Inference Systems. *Journal of Information Technology and Computer Science*, 4(1), 64.
<https://doi.org/10.25126/jitecs.20194196>
- Liping, C., Yujun, S., & Saeed, S. (2018). Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China. *PLoS ONE*, 13(7), 1–23.
<https://doi.org/10.1371/journal.pone.0200493>
- Mansour, S., Al-Belushi, M., & Al-Awadhi, T. (2020). Monitoring land use and land cover changes in the mountainous cities of Oman using GIS and CA-Markov modelling techniques. *Land Use Policy*, 91(December 2019), 104414.
<https://doi.org/10.1016/j.landusepol.2019.104414>
- Okwuashi, O., & Ndehedehe, C. E. (2020). Integrating machine learning with Markov chain and cellular automata models for modelling urban land use change. *Remote*

- Sensing Applications: Society and Environment*, 21(December 2020), 100461.
<https://doi.org/10.1016/j.rsase.2020.100461>
- Rahnema, M. R. (2021). Forecasting land-use changes in Mashhad Metropolitan area using Cellular Automata and Markov chain model for 2016-2030. *Sustainable Cities and Society*, 64(November 2019), 102548.
<https://doi.org/10.1016/j.scs.2020.102548>
- Ramdani, F., Furqon, M. T., Setiawan, B. D., & Nur, A. (2019). Analysis of the application of an advanced classifier algorithm to ultra-high resolution unmanned aerial aircraft imagery – a neural network approach. *International Journal of Remote Sensing*, 00(00), 1–21. <https://doi.org/10.1080/01431161.2019.1688413>
- Sheskin, D. (2004). *Handbook of Parametric and Nonparametric Statistical Procedures* (3rd ed.). CRM Press.
- UN. (2018). World Urbanization Prospects. In *Demographic Research* (Vol. 12).
<https://doi.org/10.4054/demres.2005.12.9>
- UN. (2019). World Population Prospects 2019: Highlights. *United Nations Publication*, 141, 1. <https://population.un.org/wpp>
- Wang, S. W., Munkhnasan, L., & Lee, W.-K. (2021). Land use and land cover change detection and prediction in Bhutan's high altitude city of Thimphu, using cellular automata and Markov chain. *Environmental Challenges*, 2(November 2020), 100017. <https://doi.org/10.1016/j.envc.2020.100017>
- Wu, H., Li, Z., Clarke, K. C., Shi, W., Fang, L., Lin, A., & Zhou, J. (2019). Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. *International Journal of Geographical Information Science*, 33(5), 1040–1061. <https://doi.org/10.1080/13658816.2019.1568441>
- Zhou, L., Dang, X., Sun, Q., & Wang, S. (2020). Multi-scenario simulation of urban land change in Shanghai by random forest and CA-Markov model. *Sustainable Cities and Society*, 55(July 2019), 102045.
<https://doi.org/10.1016/j.scs.2020.102045>