IOP Conference Series: Earth and Environmental Science

Modeling changes in land use using the integration of MLP-NN, CA-Markov models and GIS for settlement development in Tembalang District

To cite this article: S. Subiyanto *et al* 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **731** 012026

Modeling changes in land use using the integration of MLP-NN, CA-Markov models and GIS for settlement development in Tembalang District

S. Subiyanto[1*], Fauzi Janu Amarrohman[1] , Azizah Nur Rahmah[1]

[1]Geodesy Department, Faculty of Engineering, Diponegoro University, Indonesia
[*]sawitrisubiyanto@lecturer.undip.ac.id

Abstract. The effects of dynamic land use change can be modeled by Artificial Neural Network (ANN) models with Multi-layer Perceptron (MLP) network architecture and backpropagation algorithms. The setting process and predictive capabilities testing will be generated by the models with a combination of MLP-NN, CA-Markov methods and Geographic Iinformation System (GIS) from several previous studies so the results is accurate. This study aims to analisys land use change in Tembalang District in 2010 as first period models, 2014 as period models, and land use in 2018 as validation data, make a model of land use change with ANN methods and projection land use in Tembalang District in 2026. CA-Markov models are used for future projections. In modeling land use change, several driving force variables are used, namely distance to roads, rivers, settlements, and population density. This research are using maps for 2010, 2014 and 2018 from digitization process of hight resolution satellite imagery and validation land use in the field. In this research, data on land use change from 2010 to 2018 is dominated by land use changes from vacant land to settlements and housing. Settlement increased by 2,13%, housing increased by 102,69%, and vacant land already allocate increased by 47,32% and vacant land decreased by -61,18%. The results of modeling validation have a Kappa index of 0.959, an the root mean square value is 2.579 m that means this value is accepted, and 85% of the area between the prediction map and the digitization map are said to be appropriate, so this model is classified as having very good similarities with existing land use conditions in 2018. Overall prediction results show that land suitability is 70.52% and 29.48% of land is not in accordance with Semarang City RTRW map for 2011-2031.

Keywords : Land Use Change, HRSI, MLP, ANN, GIS

1. Introduction

The city of Semarang is the center of government in Central Java Province. As the capital of a fast-growing and dynamic province, the city of Semarang over bounded which indicates the development of the city towards the periphery or the countryside [1]. Tembalang District is one of the sub-districts located on the outskirts of Semarang with very rapid development. One of the implications is the many changes in land use that occur, including the construction of settlements and other physical buildings. The development of cities with land changes from one use to another land use is needed information and spatial analysis of these changes. [2]. Competition in meeting the increasing needs of land can cause land damage and decreased environmental quality of the city. These changes will continue along with the increasing number and activity of the population in carrying out economic, social and cultural life, which in turn will have both positive and negative impacts [3]. During the five year periods from

2010−2014, land use in Tembalang District experienced developments that led to the housing sector or land use for settlements, trade, industry and non-agricultural land use, which resulted in the conversion of land use functions mostly on land agriculture [4]. Therefore, monitoring and prediction regarding land use as an urban development are needed. Urban development is a process in which there is a change in the condition of an urban in different periods of time. The future land use of Tembalang District can be predicted using Geographic Information System (GIS) applications by integrating Artificial Neural Network (ANN) and Cellular Automata Markov (CA-Markov) models [5]. The ANN model is used to determine which locations or areas of land use are changing by looking at the matrix of land change opportunities. GIS is used in spatial aspects to build driving variables. The variables used as driving factors for land change in this research are accessibility factors in the form of distance to settlements, distance to main roads, and distance to toll roads, distance to rivers, and population density. The determination of that's driving factors is from analysis visualisation that land use changes occurs because they are related to the factors. This research also aims to see the direction of development of built land which is based on land use changes that occur, particularly industry and settlements. The results of this study are land use maps in 2010, 2014 and 2018, and maps of land use changes in 2010−2018, model validity test, and land use projection maps in 2018. The dynamics of land use change in East Ungaran District should be in line with the Spatial Plan (RTRW) of Semarang Regency, since RTRW is a space allocation guideline for the development of a region not beyond its carrying capacity [13]. To be able to predict future land use in spatial condition, spatial and dynamic based modeling is needed using Cellular Automata-Markov (CA-M) approach. Evaluate the synchronization of settlement land use in 2026 using Celluslar Automata-Markov (CA-M) approach to Spatial Plan District Tembalang 2011-2031 and see the potential problem. The predict future of land use changes in 2026 chosen because its closes to end of the periodic year Spatial Plan District Tembalang.

2. Material and Methode

2.1. Research Location

The location of this research is in the Tembalang District, Semarang city, Central Java. Tembalang District consists of 12 sub-districts and approximately 300 meters on average height from sea level. The research location can be seen in **Figure 1(a).**

2.2. Research Limits

The research limitations in this study are as follows.

1. The research location is in Tembalang District, Semarang City, Central Java.
2. Research data used for land use maps are obtained from the digitization of very high resolution satellite imagery namely Quickbird with a spatial resolution of 60 cm in 2010, Wordlview-2 with a spatial resolution of 60 cm in 2014 and the Aerial Phothography Medium format with a spatial resolution of 10 cm a year 2018.
3. Administrative boundary maps, main road map and main river maps, and population density data of Tembalang District, Semarang City from Development Planning Agency Semarang City.
4. Software used for processing CA and ANN models is QGIS Desktop 2.18.10 by adding the plugin molusce.
5. ArcGIS software is used for spatial analysis.
6. The method used in this study is.
    1) Classify land use with the digitization method to create land use maps in 2010, 2014 and 2018.
    2) The intersect overlay method is to get the area of land use change per district.
    3) An artificial neural network approach to modeling land use change.
    4) Celllular Automata method for predicting land use changes in 2026 and testing the validity of predictive models.
    5) Model validation using kappa method and spatial overlay analysis (differences in centroid points and area).

7. Changes in land use analyzed in this study are settlement, housing, empty land already allocated, vacant land and non-plot vacant land.
8. Sampling for land use validation in the field as many as 120 points, approximately 68% of the 176 polygons of data resulting from land use classification in 2018. The sampling number has met the minimum validation requirements which are limited to 20% of the total polygon data based on the 2012 BPN NSPK. The location of the validation can be seen in the figurecan be seen in **Figure 1 (b).**
9. Variables that drive changes in land use are used euclidean distance to settlements, to main roads, to rivers, and population density can be seen in **Figure 2.**

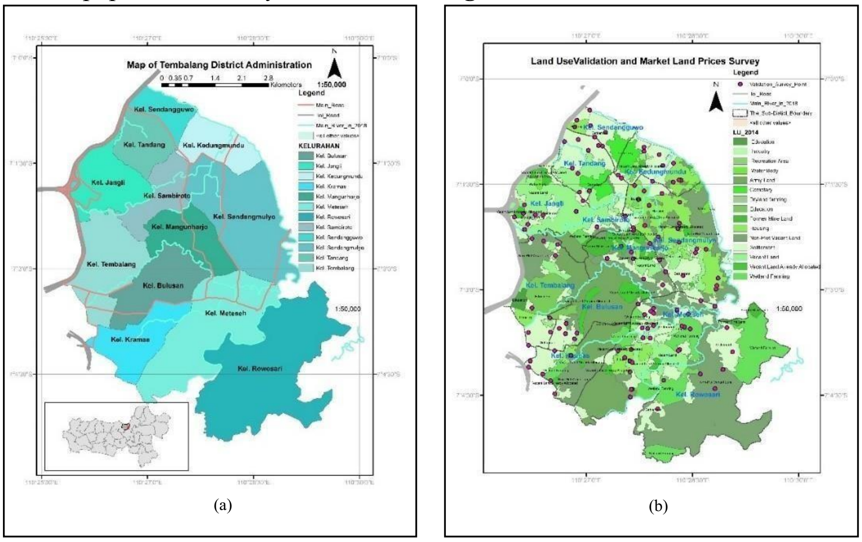

Figure 1. (a) Map Administration of the Tembalang District and (b) Overlay of settlement and a distribution of 120 sample points for field check land used for field surveys.

2.3. Modeling Changes in Land Use
Model is a simple form of a very complex system performance. The reciprocal or causal relationship between utilization for land management and land use change can be studied and analyzed in the form of pemodella [3]. Uses Modeling land use change is to understand the dynamics of land change caused by various factors driving change, including socio-economic factors [8], predicting the economic and environmental impacts of these changes and evaluating government policies in determining the direction of land use and land use [9].

2.4. Artificial Neural Network (ANN) method
Artificial Neural Network (ANN) is a technique or approach to information processing that is inspired by the workings of the biological nervous system, especially on human brain cells in processing information. The ANN method, which is applied to modeling changes in land cover / use, works in four stages, namely (1) determining input and network architecture, (2) creating a network using a portion of pixels from the input, (3) testing the network using all pixels from inputs, and (4) uses information that has been generated by the network to predict future land use changes [10]. Multilayer Neural Network or better known as Multilayer Perceptron (MLP) is a neural network consisting of input layer, hidden layer and output layer network [10].

2.5. Projections using CA-Markov
One of the advanced capabilities of GIS technology can be used in the preparation of Cellular Automata models. CA-Markov (Cellular Automata Markov) are simple models of spatial distributed processes in GIS [11]. One of the strengths of CA-Markov is that it can be integrated with other models based on

visuals, statistics, and artificial intelligence [12]. According to [13] the traditional form of CA-Markov consists of five main components namely cell, state, rules, neighboard, and time. A matrix of transition opportunities will be generated in this process and be the basis for future land use projections.

2.6. Model Validation Test

This modeling validation aims to determine the accuracy of the resulting land use prediction map. One common method used for model validation is the kappa statistical method. Kappa can be calculated from a contingency table between two data sets. The value for Kappa ranges from 1 which means it shows the perfect agreement, until −1 indicates there is no agreement at all [14]. Another method that can be used for visual model validation is the overlay method. This method takes into account the centroid points between the two land use maps to get the Root Mean Square (RMS) value and the transformation or shift value. The overlay method also takes into account the vast difference between the two land uses to validate the modeling. Modeling land use change has several uses, among others, to explore a variety of activities where the occurrence of a change in land use is driven by socio-economic factors.

2.7. Prediction Map Making Stage

This ANN model is run by using the molusce plugin which is available in QGIS 2.18.10 software. The land use map used is only two point years, namely the land use map in 2010 and 2014. The stages of making predictive maps are as follows.

1. Evaluating Correlation Pearson
   All four driving factors were tested with Pearson correlation. Pearson correlation measures the relationship between one variable with each land use with a range of values 0−1, where 0 indicates no relationship, while a value of 1 indicates a close relationship between these variables with land use.
2. Area Changes
   At this stage a table of addition and reduction of area for each land use is produced. At this stage, a transition matrix of land use change is also produced.
3. Transition Potensial Modelling
   This study uses an artificial neural network model by conducting several network simulations by setting parameter values. The parameters used in the simulation to find the best RMS values are 1 px neighborhood, learning rates 0.010 and 0.001, momentum iterations 100 and 1000, and hidden layer 1-5. The model will stop when it has reached the specified conditions.
4. Cellular Automata Simulations
   In molusce, the length of the (automatic) prediction time is t1+(t1-t0), where t1 is the final year and t0 is the initial year, then the prediction produced in this study is 2013+ (2013-2008) = 2018. If you want more from 2018, then add the number of iterations (multiplied by 3 / (t1-t0)).
5. Validations Model
   According to [14] (Altman, 1991 in Kubangun dkk., 2016) kappa value of 0.81-1.00 shows a very good agreement, 0.61-0.80 is good, 0.41-0.60 is moderate, 0, 21-0.40 is less than moderate, and a value <0.21 is said to be bad.
6. The land use prediction in 2026
   The modeling process is carried out in several stages to produce a prediction of land use in 2018, which will then be validated against the existing land use map. The results of the model are processed so that the resulting land use prediction in 2026 with Cellular Automata Markov method.
7. Conformity Analysis of Prediction Results of Land Use in 2026 with Semarang City RTRW
   The land use prediction in 2026 is then analyzed using an intersect overlay to find out the classification of land uses that are appropriate or not in accordance with the Semarang City Spatial Planning Map 2011-2031 from the aspects of usage, area and location.

IOP Conf. Series: Earth and Environmental Science **731** (2021) 012026 doi:10.1088/1755-1315/731/1/012026

3. Results and discussion

3.1. The physical and social variables

Before modelling, driving factors are needed, namely distance to road, distance to river, distance to settlements and population density per village. Supporting data was obtained using the Euclidean distance tool in ArcGIS 10.4 and exported in the form of raster, as well as the Land Use Maps in 2010, 2014 and 2018. The driving factor can be seen in Figure 2.

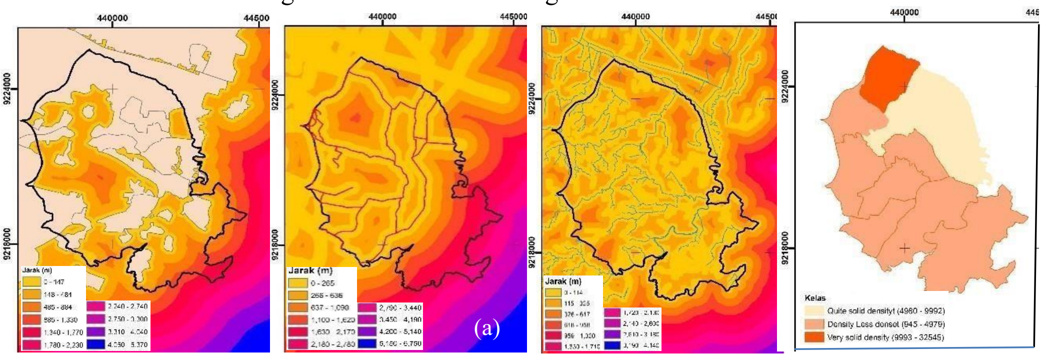

Figure 2. The sample data of driving variables : (a) distance to settlements, (b) distance to the road, (c) distance to the river, and (d) population density

3.2. Analysis of Changes in Land Use in 2010-2014-2018

Analysis of changes in land use begins with population growth has led to an increase in housing, employment and social facilities. In the case of Tembalang District, changes in the use of agricultural land that can be utilized in rural areas are increasingly narrowing. The change was followed by an increase in the amount of land as housing, settlements, vacant land that had been allocated for industrial development, educational facilities, and so on which had previously been agricultural areas. Land use changes in 2010 - 2014 and 2014 - 2018 can be seen in **Table 1**.

Table 1 Changes in Area of Land Use (Ha).

LU Name	Land Use in 2010 (Ha)	Land Use in 2014 (Ha)	Land Use in 2018 (Ha)	Merge Land Use in 2010 and 2014 (Ha)	Merge Land Use in 2010 and 2014 (%)	Merge Land Use in 2014 and 2018 (Ha)	Merge Land Use in 2014 and 2018 (%)
Army area	498,07	498,07	498,07	0,00	0,00%	0,00	0,00%
Cemetary	89,82	89,82	89,82	0,00	0,00%	0,00	0,00%
Dryland farming	75,61	61,00	53,00	-14,61	-19,33%	-8,00	-13,11%
Education	85,06	92,68	92,68	7,62	8,95%	0,00	0,00%
Former Mine Land	15,40	18,89	18,89	3,49	22,63%	0,00	0,01%
Housing	425,91	663,29	863,29	237,38	55,73%	200,00	30,15%
Industry	0,26	0,46	0,54	0,20	77,40%	0,08	16,41%
Non-Plot Vacant Land	764,48	733,49	704,00	-30,99	-4,05%	-29,49	-4,02%
Recreation Area	3,73	3,73	4,73	0,00	0,00%	1,00	26,79%
Settlement	984,05	995,90	1.005,00	11,85	1,20%	9,10	0,91%
Vacant Land	659,44	435,49	256,00	-223,95	-33,96%	-179,49	-41,22%

Vacant Land Already Allocated	132,73	159,67	195,53	26,94	20,30%	35,87	22,46%
Water Body	10,66	10,66	10,66	0,00	0,00%	0,00	0,00%
Wetland Farming	399,98	382,06	353,00	-17,92	-4,48%	-29,06	-7,61%
	4.145,22	4.145,22	4.145,22				

Based on Table 1, it can be seen that the Tembalang District land use conditions in 2010 were the most dominant settlement land, which is an area of 984,05 ha (23,74%). The most dominant land use condition in Tembalang District in 2014 was settlement land, which was 995,89 ha (24,3%). The most dominant land use condition in Tembalang District in 2018 is agricultural land, which is an area of 1005 ha (24,24%). In the Tembalang district, in addition to being dominated by settlement areas as well as housing/residential areas and Non-Plot Vacant Land can be seen in **Figure 3.**

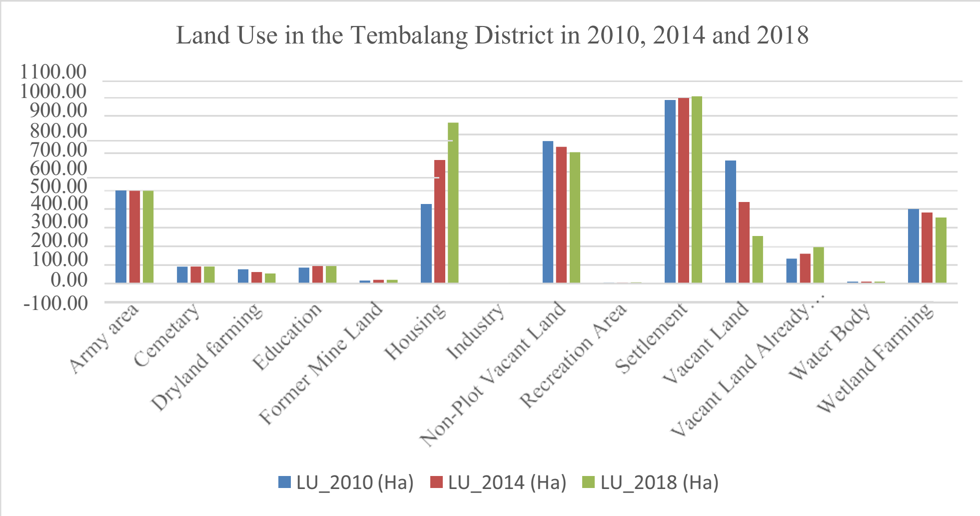

Figure 3. Chart Land Use in the Tembalang District in 2010, 2014 and 2018

In the period of four years, namely between 2010 and 2014, the condition of land use in the Tembalang district area has changed a lot. The use of land as an Vacant Land Already Allocated area has increased by 20,30%. Housing land experienced an increase in land area of 55,73% ha. Settlement land land has a slight increase by 11,85% and the area that has decreased by a large area Vacant Land by 33,96%. Whereas in the period of 4 years from 2014 to 2018 in Tembalang District the land changes were almost the same as the period 2010 to 2014 and can be seen in **Figure 4.** Confusion matrix method is used to test the accuracy of categorical data using samples obtained from the field. Confusion matrix results obtained an average value of the accuracy of the maker of 95%, an average value of 96% of user accuracy, overall accuracy or overall accuracy of 93% and kappa accuracy of 0.91

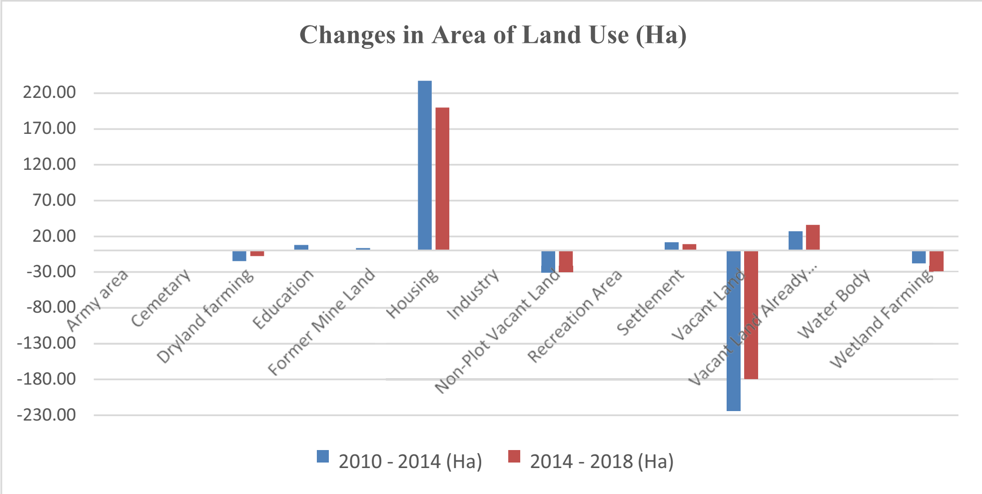

Figure 4. Chart Land Use in the Tembalang District in 2010, 2014 and 2018

3.3. Model Validation Level

3.3.1. Input Model

The map used as input is the 2010 land use map as the initial and the 2014 land use map as final. The driving factors used in this study are distance to the road, distance to the settlement, distance to the river, and population density. Maps of distances to roads, settlements, rivers, and population density maps are shown in **Figures 2 (a), (b), (c), (d).**

3.3.2. Evaluating Correlation

The second stage is evaluating correlation, calculating the relationship between driving factors or variables using the Pearson's correlation method. Based on the **Table 2**, the correlation values range from -1, 0 to 1. A value of -1 indicates a perfect negative correlation, a value of 0 indicates no correlation and a value of 1 indicates a perfect positive correlation. The correlation between roads and settlements is very strong because it approaches 1, which is 0.899. Similarly, the correlation between roads with rivers, and settlements with rivers both have positive values and are close to 1. While the population density factor with the other three factors has a negative correlation.

Table 2. Pearson Correlation Results

	River	Settlement	Main Road	Population density
River	-	0,844	0,803	-0,053
Settlement		-	0,899	-0,337
Main Road			-	-0,246
Population density				-

Pearson's correlation only explains the strength of the relationship without regard to causality, which is affected and which influences.

3.3.3. Area Changes

At this stage a table of changes in land use area in 2010-2014 was produced and a transition matrix that shows the opportunities for one land use change to another*Model Validation Level. the area change results have been described in section 3.2.*

3.3.4. Transition Potensial Modelling

The method used to calculate the transition potential is the Artificial Neural Network (ANN), where this method uses input data and driving factors to calibrate and model land use changes. The network topology used is Multi Layer Perception with a 4-3-2 structure, which is 4 nodes at the input layer, 3 nodes at the hidden layer and 2 nodes at the output layer. Each nodes in the layer will be related to other nodes, where the connection or connection path contains a weight (W) in the form of a matrix whose size depends on the number of input nodes, hidden nodes and output nodes. The results of modeling using ANN can be seen in Table 3.

Table 3. Modeling Accuracy Results

Neighbourhood	1 px
Learning rate	0,010
Maximum iterations	100
Hidden layer	5
Momentum	0,050
Δ Overall accuracy	-0,00057
Min validation overall accuracy	0,00377
Current validation kappa	0,97710

3.3.5. Cellular Automata Simulations

This study uses a repetition value from 1000 to 100, and at a repetition value of 100 a good RMS result is obtained. Adding the number of iterations to the optimum iteration limit will increase the RMS value. Adding an iteration value that exceeds the optimum limit will cause a decrease in the RMS value. A large iteration value (1000) does not always provide good accuracy, and a learning rate that is too low will cause the algorithm to reach convergence longer. Momentum determines the amount of weight change from a training. Momentum with a value of 0.050 is considered to have the best performance. The ANN modeling results curve is presented in **Figure 5.**

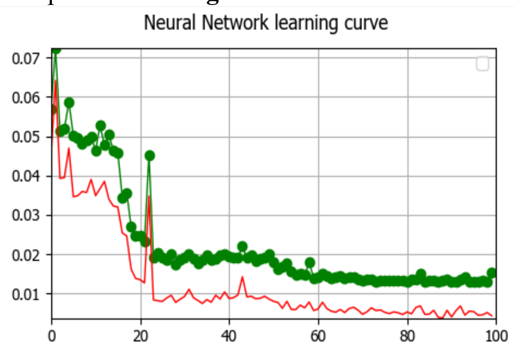

Figure 5. Modeling Curve

3.3.6. Validations

The validation results using the overlay method resulted in an RMS of 2.579 m and 85% of the area between the map digitized and the map that was predicted accordingly. Based on **Figure 5** shows that the results of the validation of this projection indicate that the Kappa value. The prediction model that has been made is then validated. Validation results the Kappa method is 0.959. According [15] the Kappa value of 0.81-1.00 indicates a very good agreement. In other words, the 2018 prediction results can be used for modeling and predicting land use in 2026.

3.4. Land Use Area Prediction in 2026

Field validation for land use surveys is carried out simultaneously and can been show in **Figure 1 (b)**. The availability of vacant land that reduced is largely due to the transition of vacant land to settlements. One example of changing vacant land into housing use is shown in **Figure 6.**

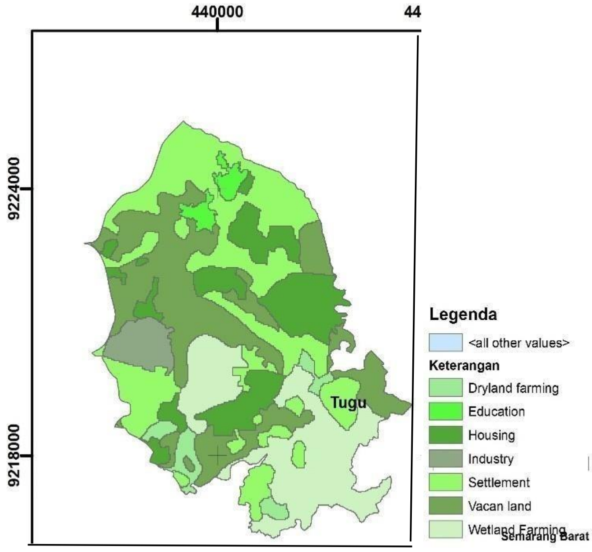

Figure 6. Prediction of land use area in 2016

3.5. Conformity Analysis of Prediction Results of Land Use in 2026 with Semarang City RTRW

Analyzing the suitability of the results of the Tembalang district land use prediction in 2026 with the Semarang City RTRW map is done by overlay intersect. Declared appropriate if there are similarities in class, area and location between the results of modeling predictions with the Semarang City RTRW map. The 2026 land use prediction itself is based on changes that occurred in 2010-2014 and the 2018 prediction using Cellular Automata Simulation. This modeling method is based on pixel changes in the previous year, namely 2018. This prediction is also influenced by the driving factors used such as distance to the road, distance to the river, distance to settlements and population density. Prediction results indicate the possibility of the growth of the City of Semarang will be covered by residential land both irregular settlements and regular settlements, plantations, industry and vacant land. While the Semarang City RTRW map.

4. Conclusion

Based on the results and description of the discussion of the research that has been done, the following conclusions can be drawn:

1. Changes in land use in the Tembalang district in the 2010-2018 period were dominated by the reduction in the area of vacant land and non plot vacant land, as well as the increase in Vacant Land Already Allocated, settlements and housing.
2. Modeling changes in land use using the MLP-NN approach in 2010 and 2014 with the factors driving show excellent model accuracy results (kappa value of 0, 95 or 95%). Model validation resulting from the 2018 land use prediction map with the existing map yields an RMS value of 2.579 and 85% of the area between the two maps is appropriate. The land use class that has a great opportunity to change to another land use is a plantation with a high enough value (0.67).
3. The results of the compatibility of the Tembalang distric land use prediction in 2026 against the Semarang City Spatial Plan are stated to be good and appropriate. Overall prediction results show

land suitability of 69.30% is suitable and 30.70% is not in accordance with Semarang City RTRW map for 2011-2031.

Acknowledgement
This research was financially suported by The Faculty of Engineering, Diponegoro University, Indonesia trough Strategic Research Grant 2020.

References

[1] Subiyanto, S and Fadilla, L. 2017. *Monitoring land use change and urban sprawl based on spatial structure to prioritize specific regulations in Semarang, Indonesia*. The 3rd Internasional Symposium for Sustainable Landscape Development2017. Earth Environ. Sci. 179 012029.
[2] Quigley, J.M. and Berkeley Larry, B. 2005. *The Effects of Land Use Regulation on the Price of Housing: What Do We Know? What Can We Learn?.* A Journal of Policy Development and Research.Volume 8, Number 1. U.S. Department of Housing and Urban Development. University of California, Berkeley.
[3] Tasha, K. (2012). *Pemodelan Perubahan Penggunaan Lahan Dengan Pendekatan Artificial Neural Network*. Fakultas Pertanian: Institut Pertanian Bogor, (May 2014), 75.
[4] Tusianto, A., Taryono, & Sigit, A. A. (2015). *Evaluasi Kesesuaian Penggunaan Lahan Kota Salatiga Tahun 2010-2014 Terhadap Rencana Tata Ruang Wilayah Kota Salatiga Tahun 2010-2030*. Fakultas Geografi, Universitas Muhammadiyah Surakarta, 1–15.
[5] Pijanowski, B.C. and Tayyebi. 2009. *Urban Expansion Simulation Using Geospati Information System and Artificial Neural Networks*. International Journal of Environmental Research, Vol. 3, No. 4, 2009, pp. 493-502
[6] Batty, M and P. A. Longley. 1994. *Urban Modelling in Computer Graphic and Geographic Information System Environments.* Environment and Planning. Vol. 19, p. 663-688
[7] Bockstael, N. et al. 1995. *Ecological Economic Modelling and Valuation of Ecosystems*. Ecological Economics. Vol. 14, p. 143-159
[8] King, A. W., A. R. Johnson, R. V. O'Neill and D. L. De Angelis. 1989. *Using Ecosystem Models to Predict Regional CO2 Exchange Between The Atmosphere and The Terrestrial Biosphere*. Global Biogeochemical Cycles. Vol. 3, p: 337-361
[9] Wu, Q. et al. 2006. *Monitoring and Predicting Land Use Change in Beijing Using Remote Sensing.* Landscape and Urban Planning. Vol. 78, p: 322-333
[10] Atkinson, P. and A. Tatnall. 1997. *Neural Network in Remote Sensing*. International Journal of Remote Sensing. Vol. 18(4), p. 699-709
[11] Baja, S. (2012). *Perencanaan Tata Guna Lahan Dalam Pengembangan Wilayah*. Yogyakarta: CV. Andi Offset.
[12] Wijaya, M. S., & Umam, N. (2015). *Pemodelan Spasial Perkembangan Fisik Perkotaan Yogyakarta Menggunakan Model Cellular Automata dan Regresi Logistik Biner*. Majalah Ilmiah Globë, 17(2), 165–172.
[13] Parasdyo, M. M., & Susilo, B. (2012). *Komparasi Akurasi Model Cellular Automata untuk Simulasi Perkembangan Lahan Terbangun dari Berbagai Variasi Matriks Probabilitas Transisi.* Masterplanning Futures, 238–274.
[14] Hagen-zanker, A. (2014). *Multi-method assessment of map similarity Multi-method assessment of map similarity Advanced use of Kappa statistics*. (June).
[15] Kubangun, S.H. and K. G. 2016. Model Perubahan Penutupan / Penggunaan Lahan Untuk Identifikasi Lahan Kritis Di Kabupaten Bogor, Kabupaten Cianjur , Dan Kabupaten Sukabumi. *Majalah Ilmiah Globe*, *18*, 21–32.