Comparison of Built-Up Expansion Prediction Models for Magelang City and Its Peri Urban Areas: CA-Logistic Regression, CA-SVM, CA-MLP, and CA-RF

Balqis, H. R.,[1] Hidayati, I. N.[2*] and Susilo, B.[2]
[1]Master Program of Remote Sensing, Universitas Gadjah Mada, Yogyakarta, Indonesia
E-mail: hasna.rafifa.balqis@mail.ugm.ac.id
[2]Department of Geographic Information Science, Universitas Gadjah Mada, Yogyakarta, Indonesia
E-mail: iswari@ugm.ac.id, bowosusilo@ugm.ac.id
**Corresponding Author*

DOI: https://doi.org/10.52939/ijg.v22i5.4982

Abstract
Monitoring and prediction of built-up land expansion is necessary to support spatial planning. This study predicts built-up land expansion using CA-Markov integrated with four machine learning models, including logistic regression, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Random Forest (RF). These models were applied to Magelang City and its peri-urban areas from 2013 to 2025. During 2013 to 2025, Magelang City and its peri-urban areas experienced LULC changes. Residential buildings, non-residential buildings, agriculture land, and agroforestry land increased in area over time. Meanwhile, bare land and shrubland decreased in area over time. The results also show that there was an expansion of built-up land in Magelang City and its peri-urban areas from 2013 to 2025, from 2,674.35 ha to 3,483.81 ha. The 2013 and 2019 LULC maps were used as input in predicting built-up land expansion in 2025. The model was trained using LULC change in 2013 to 2019 and selected driving factors. The driving factor with the most significant influence was the distance to existing built-up land. The results showed that the SVM model achieved the highest accuracy, with an overall accuracy of 98.4% and a Figure of Merit (FOM) of 15.3%. This indicates that the SVM model is a most suitable model for predicting built-up land expansion in Magelang City and its peri-urban areas.

Keywords: Built-Up, Expansion, Cellular Automata, Machine Learning, Prediction

1. Introduction

Land conversion is the process of changing land from one use to another, whereby this phenomenon tends to convert non-built up land into built-up land and agricultural land into non-agricultural land [1]. This issue needs to be considered in regional development because it has a direct impact on environmental sustainability, food security, and spatial planning [2]. Land conversion often occurs on built-up land because built-up land is the type of land use that experiences the most rapid change [3]. Built-up land can be defined as land that has undergone a process of hardening or development [4]. The conversion of non-built-up land to built-up land has occurred in several cities in Indonesia, one of which is Magelang City. According to data from the Environment Agency, there has been a reduction in agricultural land during 2013 to 2023 from 210.7 ha to 105.54 ha. This phenomenon can occur due to the conversion of agricultural land into built-up land [5].

The conversion of non built-up land into built-up land in Magelang City may be influenced by its position as a growth centre. This reflects Magelang City's role as the Purwomanggung Regional Activity Centre in accordance with Central Java Provincial Regulation No. 6 of 2010. The Purwomanggung area is a development zone consisting of Purworejo Regency, Wonosobo Regency, Magelang Regency, Magelang City and Temanggung Regency. Magelang City's position as a growth centre will make it an area that experiences rapid development compared to the surrounding areas [6]. This is also supported by the strategic location of Magelang City, which is at the intersection of transportation and economic routes between Semarang-Magelang-Yogyakarta and Purworejo, as well as at the intersection of local and regional tourist routes between Kopeng-Borobudur Yogyakarta and Dieng [7]. This strategic location will trigger increased

economic growth in Magelang City. The potential for economic growth will trigger an increase in population and the expansion of built-up land, as it can encourage people from other regions to move to Magelang City [8].

Land conversion in Magelang City has led to increasingly limited land availability in the city centre, prompting the community to expand their settlements to the peri-urban areas [9]. Peri-urban areas often experience rapid development as part of urban expansion [10]. This indicates that peri-urban areas of Magelang City also have the potential to experience built-up land development as a result of the urban expansion occurring in Magelang City [11]. Administratively, Magelang City directly borders several sub-districts in Magelang Regency, including Mertoyudan Sub-district, Secang Sub-district, Tegalrejo Sub-district, Bandongan Sub-district, and Candimulyo Sub-district. Urban expansion in these subdistricts was marked by a decline in agricultural land area over time, with a reduction of 518 ha from 2009 to 2011 [12]. The expansion of built-up land to the peri-urban areas can also be driven by factors such as good accessibility, the availability of existing built-up land, and the planning of new infrastructure[13]. One major infrastructure project that could have an impact is the Yogyakarta-Bawen Toll Road that can encourage the development of built-up land centred on toll exit points [14]. This is because the toll road will improve accessibility for the community, thereby impacting changes in the structure and pattern of land use in the surrounding area and triggering the conversion of land use from non-built up to built up [15].

Urban expansion in Magelang City and its peri-urban areas has created the potential for uncontrolled built up land development, which needs to be monitored regularly [16]. In addition, this can also be supported by predictions of future built-up land expansion [17]. Monitoring and prediction can be carried out using remote sensing data to observe historical land use change patterns[18]. One type of data that can be used is Landsat 8 OLI and Landsat 9 OLI-2 imagery with medium spatial resolution, which has several advantages, including high temporal resolution, free availability, and easy access [19]. Simulation of built-up land expansion can be carried out using cellular automata (CA), a simulation method that can determine the behavioural change process of a system. CA has five constituent elements, including cellular space, the state of each cell, neighbouring cells that can influence the status of a cell, rules that can determine cell status changes, and time [20]. The CA model is a popular model that can capture the urbanisation process from the past to the future [17].

Cellular automata requires driving factors for urban expansion, both spatial and social factors, so that the prediction results can match the dynamics of society [21]. The influence of driving factors will shape the probability of land use change, which can be analysed using machine learning models. Machine learning models that can be used include logistic regression, support vector machine (SVM), multilayer perceptron (MLP), and random forest (RF). Four machine learning models were selected to represent different modelling approaches commonly used in land-use change prediction. Logistic Regression was included as a baseline statistical model to capture linear relationships between driving factors and built-up expansion. SVM and MLP were employed to represent non-linear machine learning approaches capable of modelling complex relationships in spatial data. RF was selected as an ensemble learning method known for its strong predictive performance and robustness in land-use change modelling.

Several previous studies have compared machine learning models integrated with CA in predicting land use change [22][23][24] and [25]. Each model will produce different predictions and accuracy, so a comparison is needed to determine the most suitable model for predicting urban expansion in Magelang City and its peri-urban areas. This study aims to (1) monitor built-up land expansion, (2) compare machine learning models integrated with CA for predicting built-up land expansion, and (3) determine the most suitable machine learning model for predicting built-up land expansion in Magelang City and its peri-urban areas.

2. Material and Method

2.1 Study Area

This research was conducted in the administrative area of Magelang City and other subdistricts in Magelang Regency adjacent to the city. Administratively, Magelang City is adjacent to the north by Secang District, to the south by Mertoyudan District, to the east by Tegalrejo and Candimulyo Districts, and to the west by Bandongan District. Magelang City has an area of 18.56 km^2, which is divided into 3 urban villages and 17 rural villages [26]. Then, the subdistricts adjacent to Magelang City, including Mertoyudan, Bandongan, Secang, Tegalrejo, and Candimulyo, have areas of 46.25 km^2, 48.56 km^2, 51.39 km^2, 38.26 km^2, and 49.28 km^2, respectively [27]. The total area of the entire study area is 252.29 km^2 that divided into 8 sub-districts. The study area map is presented in Figure 1.

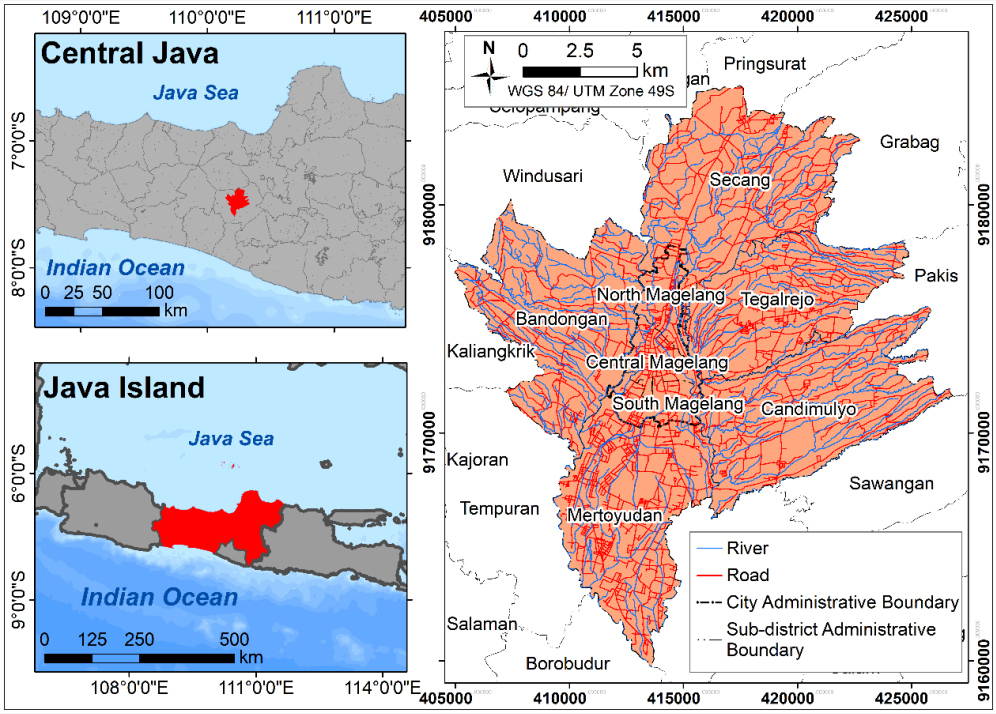

Figure 1: Study area of Magelang City and its surrounding districts: Mertoyudan, Bandongan, Secang, Tegalrejo, and Candimulyo

Magelang City is strategically located in the centre of Java Island. In addition, Magelang City is also located at the intersection of economic and transportation routes between Yogyakarta-Magelang Semarang and Temanggung-Purworejo. Magelang City is also located at the intersection of regional and local routes between Yogyakarta-Borobudur Kopeng and the Dieng Plateau. In terms of topography, Magelang City is classified as highland because it is located at an altitude of ± 370 metres above sea level with a slope gradient ranging from 5^0 to 45^0[28]. Then, Mertoyudan District, which borders Magelang City to the south, is the residential centre of Magelang Regency [29]. Meanwhile, Bandongan District has hilly topography and cool air, known as an agricultural area and a rapidly developing residential area [30]. Secang Subdistrict is located in the north and serves as a strategic route connecting Magelang City and the northern part of Central Java, with dynamic economic and trade activities [31]. Tegalrejo Subdistrict is located in the west of Magelang City and has productive agricultural land as well as small-scale industrial and service activities growing around the residential area [32]. Meanwhile, Candimulyo Subdistrict is located in the southwest with a relatively flat to undulating landscape, dominated by rice farming and horticulture activities [33]. The methodologhy in this study is presented in Figure 2.

2.2 Data Collection

In this study, the data collected included remote sensing imagery data to extract LULC classes and data on driving factors of built-up land expansion. The imagery data used were Landsat 8 OLI images from 2013 and 2019 and Landsat 9 OLI-2 images from 2025. The images used were L1TP level images that had been geometrically corrected. Other data used were data on the driving factors of built-up land expansion. The data used came from various sources. Some data was obtained from agencies in Magelang City and Magelang Regency. The data obtained from agencies included the distribution of educational facilities, health facilities, industrial areas, commercial facilities, roads, rivers, and population density. This data was obtained in vector format. Then, the Euclidean distance process was carried out to measure the distance to these objects, and the driving factor data was produced in raster format. Other data used was DEMNAS, which was obtained from tanahairindonesia.go.id. The DEMNAS used had a spatial resolution of 8 metres. DEMNAS was used to extract elevation data. Next, DEMNAS was processed to obtain slope data. All data was then resampled to have the same spatial resolution of 15 metres with the UTM zone 49S projection system with WGS 1984 datum. The specifications of the data used in the study are presented in Table 1.

Table 1: Specifications of the data used in the research

No.	Data	Type	Source
1	Landsat 8 OLI	Raster	USGS, EarthExplorer
2	Landsat 9 OLI-2	Raster	USGS, EarthExplorer
3	Educational Facilities	Vector	Public Works and Spatial Planning Agency
4	Health Facilities	Vector	Public Works and Spatial Planning Agency
5	Industrial Areas	Vector	Public Works and Spatial Planning Agency
6	Commercial Facilities	Vector	Public Works and Spatial Planning Agency
7	Road	Vector	Public Works and Spatial Planning Agency
8	River	Vector	Public Works and Spatial Planning Agency
9	Population Density	Vector	Central Statistics Agency
10	DEMNAS	Raster	tanahair.indonesia.go.id

2.3 LULC Classification

LULC classification was carried out using Landsat 8 and Landsat 9 imagery with L1TP product level. Classification was performed on Landsat 8 OLI imagery from 2013 and 2019 and Landsat 9 OLI-2 imagery from 2025. The date of acquisition of the data for each image used is presented in Table 2. Image pre-processing was carried out with radiometric correction to TOA reflectance level using Equation 1:

$$\rho_\lambda = M_p \cdot Q_{cal} + A_p$$

Equation 1

Where :

ρ_λ = TOA *reflectance*
M_p = Spectral reflection multiplication scale
Q_{cal} = *digital number* (DN)
A_p = Spectral reflectance addition scale

Table 2: Date of image data acquisition

No.	Data	Acquisition Date
1	Landsat 8 OLI	12 September 2013
2	Landsat 8 OLI	13 September 2019
3	Landsat 9 OLI-2	19 July 2025

The next step is to perform pan-sharpening, which is the process of combining high-resolution panchromatic band with low-resolution multispectral bands to produce colour images with high spatial resolution[19]. Landsat 8 OLI and Landsat 9 OLI-2 images were used, which have multispectral bands with a spatial resolution of 30 m and panchromatic bands with a spatial resolution of 15 m. Pan-sharpening was performed using the Gram-Schmidt method, which combines multispectral and panchromatic band images by applying vector orthogonal transformations. This method was chosen because it can increase the spatial resolution of the image while maintaining the original spectrum of the multispectral bands so that it does not change much after pan-sharpening [34]. This process produces a multispectral image with a spatial resolution of 15 m, and all LULC classifications in the analysis were generated at this spatial resolution. The image bands used in the LULC classification consist of red, green, and blue (RGB) bands, as these bands provide sufficient spectral information to distinguish major land-cover types such as vegetation, built-up areas, and water bodies [35]. Since the main objective of this study is to analyse urban expansion, the visible bands are adequate for identifying built-up areas and other dominant land-cover classes. In addition, the pan-sharpening process improves the spatial resolution of the imagery, which enhances the detection of urban features [19].

LULC classification was performed using a random forest model on Google Earth Engine (GEE). Classification with random forest was performed by forming trees, beginning with the creation of subsets of data equal to the number of iterations performed. Each tree is trained with a random subset of the training data and has a random subset of features to perform splits. The reflectance values of objects in each band of the training data are entered into the tree. Then, each tree produces its own class decision, and majority voting is performed to determine the LULC class of a pixel [36]. LULC classification is divided into seven classes, namely residential buildings, non-residential buildings, bare land, agriculture land, agroforestry land, shrubland, and water body. The samples used in LULC classification were determined using stratified random sampling, where the sample size was determined using the Slovin method [37] using Equation 2:

$$n = \frac{N}{1 + N(e)^2}$$

Equation 2

Where :

n = sample size
N = population size
e = margin of error (0.05)

The samples were divided into training samples to build the model and testing samples to test the LULC classification results at a ratio of 70:30. The testing sample is used to validate the LULC classification against the ground truth. The LULC classification for 2025 was validated through field surveys conducted in the study area. Meanwhile, the validation of the 2013 and 2019 LULC classifications was carried out using high-resolution imagery available on the Google Earth platform. The accuracy of the LULC classification results was tested using a confusion matrix, which is a square matrix representing the number of samples classified in each class and the actual class according to the field survey results. The confusion matrix shows the overall accuracy, user's accuracy, producer's accuracy, omission error, and commission error values [38]. Then, the seven land use classes will be generalised into built-up and non-built-up land classes as shown in Table 3.

Table 3: LULC Classification

No.	Class	LULC Classification
1	Built-up land	Residential buildings
		Non-residential buildings
2	Non built-up land	Bare land
		Agriculture land
		Agroforestry land
		Shrubland
		Water Body

2.4 LULC Transition and Driving Factors Map

From the LULC map, a transition map of LULC changes from 2013 to 2019 was then created. This transition map has two classes, namely changed and unchanged. Then, maps of the driving factors of built-up land expansion were created, consisting of 10 factors, including distance to roads, distance to rivers, distance to educational facilities, distance to health facilities, distance to industrial areas, distance to commercial facilities, distance to existing built-up land, population density, elevation, and slope. The driving factors used in this study were selected based on their relevance to the urban expansion process as reported in previous land-use change studies [23] [24] and [25], as well as the availability and spatial consistency of data for the study area. These factors represent physical and accessibility-related characteristics that may influence the likelihood of land conversion into built-up areas. Topographic variables such as slope and elevation were included because terrain conditions affect the feasibility of urban development. Accessibility-related variables, including distance to roads and distance to facilities, were selected because areas with better accessibility tend to attract more development. In addition, population density was considered to represent socio-economic pressure that may drive urban expansion.

The driving factors were analysed using Cramer's V test, which can describe the strength of the relationship between the driving factors and land use transition based on the Chi-square (χ^2) statistical value of the independence test. The Cramer's V value ranges from 0 to 1, and the relationship between the two variables is stronger if the value is closer to 1 [39]. If the Cramer's V values range between 0.15 and 0.4, the variables are considered to have a moderate influence on land-use changes and are deemed acceptable. Values greater than 0.4 indicate a strong association between land-use change patterns and driving variables, whereas variables with values below 0.15 are excluded from the model due to weak relationships [40].

The selected driving factors were tested for multicollinearity to determine the correlation between the factors used. Multicollinearity occurs when two or more independent variables are highly correlated, which can affect the stability and reliability of regression coefficients. To detect multicollinearity among the driving factors, the Variance Inflation Factor (VIF) and tolerance values were calculated. VIF is derived from the coefficient of determination (R^2) obtained by regressing each explanatory variable against the remaining variables. A VIF value close to 1 indicates no correlation among variables, while values greater than 5 indicate potential multicollinearity. The tolerance value, defined as the reciprocal of VIF, indicates multicollinearity when it is lower than 0.1–0.2 [41].

2.5 Transition Potential Map

A transition potential map is a map that shows the probability of each pixel changing from non-built-up land to built-up land. Transition potential maps have pixel values ranging from 0 to 1, where pixels with values closer to 1 have a higher probability of change. Transition potential maps are created using pixel values from driving factor maps and LULC transition map. In this analysis, sample points were determined using stratified random sampling. Transition potential maps were created using four machine learning models, including logistic regression, SVM, MLP, and RF. Transition potential maps were generated using the R programming language in the RStudio environment, which was used to implement the machine learning models and perform spatial data processing.

2.5.1 Logistic regression

Logistic regression is a model that analyses the relationship between several independent variables and several dependent variables that are categorical.

This model can estimate the probability of an event occurring by entering training data into a logistic curve [42]. A transition potential map is created with the driving factor value as the independent variable and LULC transition as the dependent variable. The logistic regression used is binary logistic regression because the dependent variable used only consists of two categories, namely changed and unchanged. Logistic regression produces coefficient values that indicate the magnitude of each driving factor's influence on LULC transition and the probability of LULC transition for each pixel [43]. The probability of transition is calculated using Equation 3:

$$p = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p$$

Equation 3

Where:

p = probability of a grid cell having a LULC transition
α = constant value
X = influencing factor
β = coefficient of driving factor X

2.5.2 Support Vector Machine (SVM)

Support Vector Machine is a supervised learning method that maps input vectors to a higher-dimensional space and then separates the data using a hyperplane based on maximum margin [44]. In creating a transition potential map using SVM, driving factors are used as predictor variables and LULC transitions are used as target variables. The basic principle of SVM is to project data into a high-dimensional space through a kernel to separate classes with maximum distance or margin using a hyperplane. The SVM used in this study uses a radial basis function kernel. The kernel in SVM is a way of calculating the similarity between data in a high-dimensional space without calculating the coordinates of that space [45]. The RBF kernel calculates the proximity of data using Equation 4:

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right)$$

Equation 4

Where:

x_i, x_j = two data points
$\| xi - xj \|^2$ = euclidean squared distance between two points
γ = parameters that regulate the width of influence

The RBF kernel works by calculating the K value for each pair of data points that indicate their proximity in space. Then, the kernel value projects the data into infinite-dimensional space where linear separation occurs. SVM then uses the kernel value to determine the hyperplane that can maximise the distance between classes. The SVM model was optimized through hyperparameter tuning using 5-fold cross-validation. The tuning process evaluated the regularization parameter (*C*) and the kernel parameter (*sigma*), which control the trade-off between model complexity and classification error as well as the influence range of the kernel function. The optimal parameter combination obtained from the tuning process was C = 0.5 and *sigma* = 0.252574, which produced the best model performance and was therefore used in the final SVM model. The trained SVM model is then used to calculate the probability value of change in each pixel in the image. SVM calculates probabilities by converting margins into probabilities, a process known as plat scaling [46]. This method adjusts the logistic function to the margin results using Equation 5:

$$P\left(y = 1 \mid x\right) = \frac{1}{1 + \exp(Af(x) + B)}$$

Equation 5

Where:

$f(x)$ = margin value of SVM
A and B = parameters estimated from validation data

2.5.3 Multilayer Perceptron (MLP)

MLP is a type of feedforward neural network that has several layers, consisting of an input layer, a hidden layer, and an output layer [47]. In creating a transition potential map using the MLP model, driving factors are used as predictor variables and LULC transitions as target variables. MLP analyses data using a simple neural network with an input layer consisting of driving factors of built-up land expansion. The process is carried out by randomly initialising the weights for each connection between neurons. Then, forward propagation is carried out with a sigmoid activation function. Next, the error is calculated, followed by back propagation by calculating the error derivative against the weights and updating the weights so that the error decreases [48]. Iteration is carried out until the minimum error is achieved. The MLP model was optimized using a grid search approach with 5-fold cross-validation to determine the optimal hidden layer size. Several candidate values (1, 3, 5, and 7) were evaluated during the tuning process. The results indicated that a hidden layer size (*size*) of 1 produced the best model performance and was therefore selected for the final MLP model. The trained MLP model is then used to calculate the probability value of change in

each pixel in the image. Probability calculations are performed using the sigmoid function, which will produce a probability value of change ranging from 0 to 1 [49]. The sigmoid function changes the values produced by the output layer using Equation 6:

$$y = \frac{1}{1+e^{-z}}$$

Equation 6

Where:

y = probability of change
e = Euler number (~2.71828)
z = result of the output layer

2.5.4 Random Forest (RF)

RF is an ensemble method based on decision trees, in which this method constructs many trees consisting of random subsets of samples and features. The trees are constructed by creating subsets of training samples through a bagging approach [50]. Each tree will be trained with a different subset of data and produce its own decision. RF model parameters were optimized by tuning the *mtry* parameter using a grid search approach with 5-fold cross-validation. Several candidate values were evaluated, and the optimal parameter was selected based on the highest classification accuracy, resulting in *mtry* 5 for the final model. The number of trees (*ntree*) was set to 500 to ensure stable and robust prediction performance. Decision making is based on majority voting, where each tree will give a decision and the class that receives the most votes will be selected as the final decision [51]. In creating a transition potential map using the RF model, driving factors are used as predictor variables and LULC transitions as target variables. The driving factor values of each pixel are entered into each tree in the forest, then each tree will predict whether the class will change or unchanged. Next, the final probability value is calculated by dividing the number of trees that predict change by the total number of trees [52]. The resulting probability value will range from 0 to 1.

2.6 Transition Probability Matrix

The transition probability matrix for built-up and non-built-up land classes was created using a Markov chain. Markov Chain is a statistical method used to calculate the probability of transition of land cover classes from one type to another based on historical data [53]. Markov chain can calculate the probability of one land use class changing to another within a certain period of time based on two land use maps from different times [54]. In this study, a Markov chain was used to calculate the probability of non-built-up land changing to built-up land in 2025 based on LULC maps from 2013 and 2019. The Markov chain calculated the transition probability by counting the number of pixels in each class in the first and second years. Then, the probability is calculated by dividing the number of pixels that changed from the initial class to the final class by the total initial pixels of the initial class. Each probability value is then used to construct a transition probability matrix [55]. Probability values range from 0 to 1, where the closer the probability value is to 1, the greater the likelihood of change [56]. The probability matrix represents the likelihood of land use change over a 6-year period, from 2013 to 2019, and is used to model changes over the next 6 years, from 2019 to 2025. Furthermore, the probability matrix and the number of pixels in each class in 2019 are used to predict the number of pixels in each class in 2025 using Equation 7:

$$S_{t+1} = S_t \cdot P$$

Equation 7

Where :

$S_{(t+1)}$ = prediction of the number of pixels in each class in the future, i.e. t+1
P = transition probability matrix
S_t = number of pixels in each class at time t

2.7 LULC Change Simulation Using CA-Markov

Simulation of LULC changes was carried out using a cellular automata model. Cellular automata (CA) is a dynamic system, where space is represented by a grid of uniform cells, each of which can be in one of a limited number of states. The state of each cell evolves over time according to a set of transition rules that depend on the states of its neighbouring cells [57]. The basic components of cellular automata consist of cell space, cell state, neighbourhood, transition rules, and time steps [58]. Simulations were conducted to predict LULC changes in 2025 based on LULC maps from 2013 and 2019. In this study, a raster of LULC classification results with a cell size of 15x15 m was used. The neighbourhood filter used was the Moore neighbourhood filter, which is a 3x3 neighbourhood filter that considers 8 neighbouring pixels around the central pixel. Simulation of LULC changes using CA was conducted using two main inputs: a transition probability matrix derived from a Markov Chain and transition potential maps generated using four machine learning models, namely logistic regression, SVM, MLP, and RF. The transition probability matrix estimates the quantity of land-use changes between classes over time, while the machine learning models generate transition potential maps representing the probability of each cell being converted into built-up areas based on the analyzed driving factors.

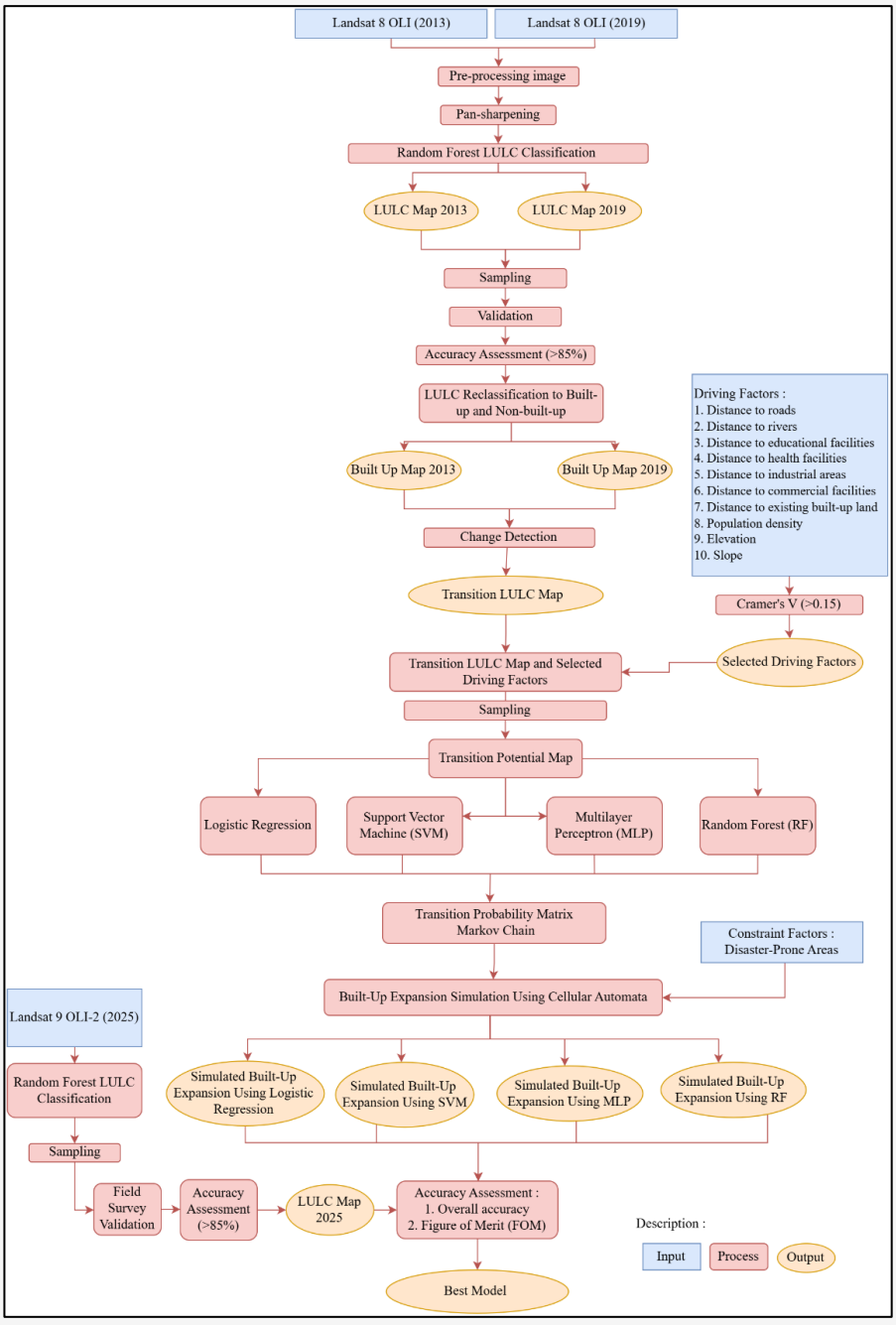

Figure 2: Flow chart on model comparison in predicting built-up land expansion

The transition potential maps were subsequently integrated into the CA–Markov framework to guide the spatial allocation of land-use changes. In this process, the Markov chain determines the expected amount of land-use transition, whereas the CA component allocates these changes spatially based on neighborhood rules and transition suitability derived from the machine learning models. Cells with higher transition potential values have a greater likelihood of being converted during the simulation process. Through this integration, the model can simulate both the quantity and spatial distribution of future LULC changes. Each machine learning model produced a separate transition potential map, which was then used within the CA–Markov simulation to evaluate the effectiveness of different models in predicting spatial patterns of LULC change.

The combination of transition probability matrix from Markov Chain and transition potential maps from four machine learning models is then used by cellular automata to predict built-up land expansion in Magelang City and its peri urban areas in 2025. This simulation also uses disaster-prone areas as a constraint area, consisting of flood-prone and landslide-prone areas. The use of disaster-prone areas as a constraint means that built-up land development will only be simulated in areas that are not prone to disasters. The use of constraints aims to ensure that predictions form an ideal pattern and support sustainable development.

2.8 Accuracy Assessment

Accuracy assessment of LULC change simulation results was conducted to determine the extent to which the model built could represent actual conditions in the field. Accuracy testing was conducted to determine the model's ability to represent built-up land expansion in Magelang City and its peri-urban areas. Accuracy testing can be conducted in two ways, namely the accuracy of the quantity of change and the accuracy of the location of change. The accuracy test was conducted using two inputs, consist of the results of the simulation of built-up land expansion Magelang City in 2025 from the four models and the existing built-up land map in Magelang City in 2025 obtained from the results of multispectral classification of Landsat 9 OLI-2 pan-sharpened imagery.

The accuracy test for the quantity of change can be performed using overall accuracy. Overall accuracy is calculated by comparing the percentage of correctly classified pixels with the total number of pixels in the image[59]. This calculation is conducted by comparing the simulation prediction results with the existing built-up land map in 2025. The result will be a percentage that shows the extent to which the simulation results can represent the development of built-up land in Magelang City. Meanwhile, the location accuracy test is carried out by calculating the figure of merit (FOM). FOM shows the percentage of overlap between the predicted change area and the existing change area. FOM ranges from 0 to 100%, where the higher the FOM, the better the model's performance in representing the location of built-up land development [60]. FOM was selected as the main evaluation metric because it measures the agreement between observed and simulated land-use changes by focusing specifically on areas where changes occur. Unlike overall accuracy, which can be influenced by large areas of unchanged land, FOM evaluates the model's ability to correctly predict spatial changes. Therefore, it provides a more appropriate measure for assessing the performance of land-use change simulations. FOM is calculated using three components, namely *hits, misses, and false alarms* using Equation 8:

$$FOM = \frac{Hits}{Hits + Misses + False\ alarms}$$

Equation 8

Where:

Hits = number of pixels predicted to change and actually change
Misses = number of pixels predicted not to change but actually change
False alarms = number of pixels predicted to change but do not actually change

3. Result and Discussion

3.1 LULC Classification

The 2013, 2019, and 2025 LULC classification maps were classified using the random forest method with pan-sharpened Landsat imagery. LULC in 2013, 2019, and 2025 changed over time. The LULC classification maps are presented in Figure 3. All LULC classes except water body experienced changes in area over time. The change in area for each LULC class is presented in Figure 4. The largest increase in area was in residential buldings, with an increase of 526.77 ha from 2013 to 2019 and a further increase of 242.96 ha by 2025. Nonresidential buildings increased by 30.17 ha from 2013 to 2019 and then by 9.56 ha by 2025. Meanwhile, agriculture land increased by 322.58 ha from 2013 to 2019 and then increased by 435.92 ha by 2025. Agroforestry land decreased by 1,218.04 ha from 2013 to 2019 and then increased by 1,273.64 ha by 2025. The LULC that decreased was shrubland, which increased by 373.57 ha from 2013 to 2019 and then decreased by 1,159.76 ha by 2025. The largest decrease in LULC was in bare land, which decreased by 33.89 ha from 2013 to 2019 and will decrease by 802.42 ha by 2025. The LULC map shows that residential and non-residential buildings are concentrated in the central part, which is the Magelang City area. Meanwhile, agriculture land, agroforestry land, and shrubland are widely spread across the peri-urban areas surrounding Magelang City. Agriculture land is mostly found in the western part, which is part of the Bandongan District. Agroforestry land is mostly found in the south-eastern part, which is part of the Candimulyo District. This shows that the Magelang City area is mostly dominated by built-up land, while the surrounding peri-urban areas are still dominated by non-built-up land. The peri-urban area, which is still dominated by non-built-up land, allows for the expansion of built-up land from Magelang City to the surrounding areas.

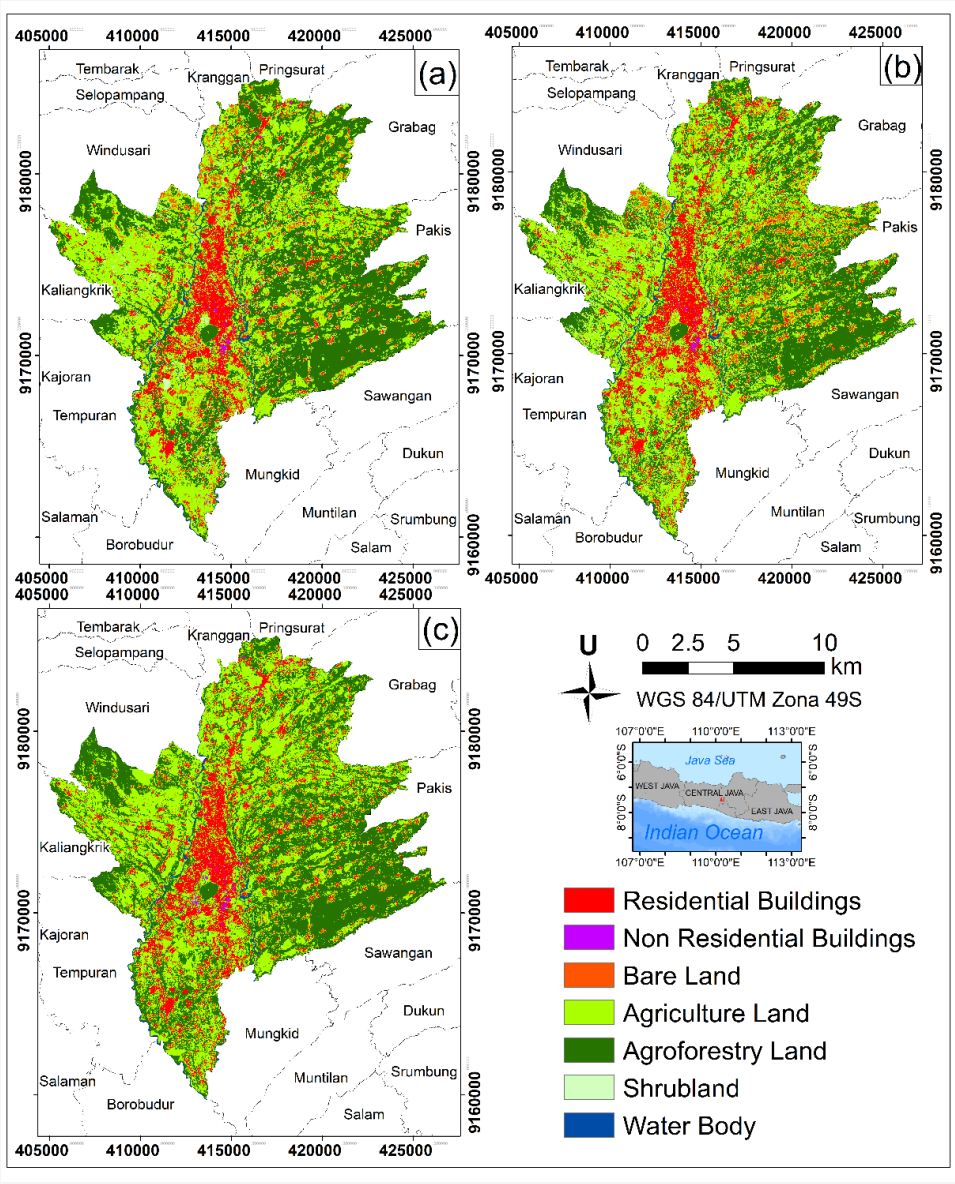

Figure 3: Land Use Land Cover (LULC) Map of: (a) 2013, (b) 2019, (c) 2025

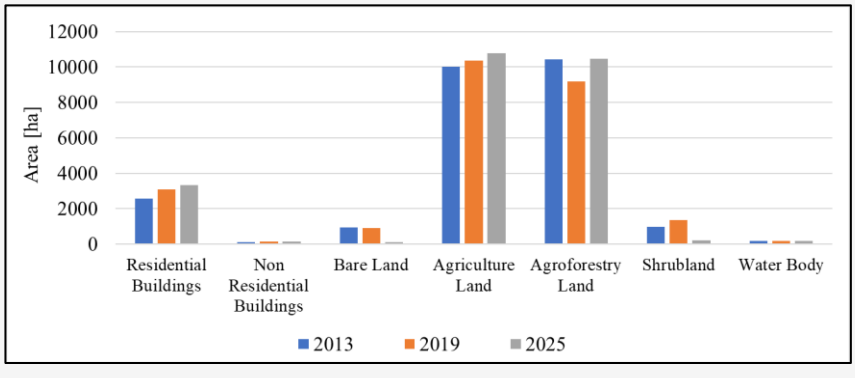

Figure 4: LULC area between 2013, 2019, and 2025

Gains, losses, and net changes in each LULC class are presented in Table 4. Residential buildings had the largest net change from 2013 to 2019 at 2.09%, indicating that residential buildings increased significantly in area and contributed significantly to LULC changes. Then, from 2019 to 2025, residential buildings had a positive net change of 0.96%, indicating that residential buildings in Magelang City and its peri-urban areas increased over time. This shows that Magelang City and its surrounding areas experienced a large percentage increase in residential buildings over time. Furthermore, non-residential buildings also had a positive net change of 0.12% from 2013 to 2019 and 0.04% from 2019 to 2025. This indicates that non-residential buildings also increased, but their contribution was not as significant as residential buildings. Bare land had a net change of -0.13% from 2013 to 2019 and -3.18% from 2019 to 2025. This shows that the area of bare land is decreasing over time. Agriculture land increased in area with a net change of 1.28% from 2013 to 2019 then increased from 2019 to 2025 with a net change of 1.73%. This indicates that the area of agriculture land increased. Agroforestry land decreased in area from 2013 to 2019 with a net change of -4.83% but then increased from 2019 to 2025 with a net change of 5.05%. This shows that the area of agroforestry land is increasing over time. Shrubland increased in area with a net change of 1.48% from 2013 to 2019 but then decreased with a net change of -4.60% from 2019 to 2025. This shows that LULC changes in Magelang City and its peri-urban areas are dynamic. Meanwhile, water body did not experience any gain or loss, so the net change was 0.

The accuracy of the LULC classification results was tested against the actual LULC classes using a confusion matrix. A total of 120 samples were used for the accuracy test. The confusion matrix shows the overall accuracy, user accuracy, and producer accuracy values presented in Table 5. Overall accuracy increased from year to year, from 90.83% in 2013 to 91.67% in 2019 and 93.33% in 2025. This also occurred in the kappa value, which increased from 0.88 to 0.89 and then further to 0.91. This increase could have occurred because the quality of Landsat images improved from year to year so that LULC classification could be carried out with higher accuracy. Overall accuracy in all three years showed fairly good results, as it was above 90%. LULC classification results can be used for further analysis if they have a minimum accuracy of 85% [61]. The LULC classification results in this study can be used for further analysis because they already meet this requirement.

Table 4: Gain, loss, and net change for each LULC classification

LULC	2013-2019			2019-2025		
	Gain (%)	Loss (%)	Net Change (%)	Gain (%)	Loss (%)	Net Change (%)
Residential Buildings	2.17	-0.8	2.09	0.99	-0.03	0.96
Non Residential Buildings	0.13	-0.01	0.12	0.04	0	0.04
Bare Land	1.20	-1.33	-0.13	0.14	-3.32	-3.18
Agriculture Land	7.63	-6.35	1.28	5.08	-3.35	1.73
Agroforestry Land	3.00	-7.83	-4.83	5.55	-0.5	5.05
Shrubland	2.21	-0.73	1.48	0.26	-4.86	-4.60
Water Body	0	0	0	0	0	0

Table 5: User accuracy, producer accuracy, overall accuracy, and Kappa value for each LULC classification

LULC	2013		2019		2025	
	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)
Residential Buildings	100.00	87.50	100.00	88.89	100.00	100.00
Non-Residential Buildings	80.00	100.00	100.00	100.00	100.00	100.00
Bare Land	71.43	83.33	80.00	100.00	80.00	80.00
Agriculture Land	82.50	91.67	90.48	86.36	93.02	88.89
Agroforestry Land	97.62	89.13	90.24	94.87	92.50	94.87
Shrubland	100.00	100.00	83.33	100.00	80.00	100.00
Water Body	100.00	100.00	100.00	100.00	100.00	100.00
Overall Accuracy (%)		90.83		91.67		93.33
Kappa Values (%)		0.88		0.89		0.91

Based on the accuracy evaluation results, each LULC class showed generally high classification performance throughout the observation years. The residential buildings class had consistently high UA and PA, even reaching 100% in 2019 and 2025, which demonstrated the model's excellent ability to recognise residential buildings. The non-residential buildings class also showed perfect accuracy in 2019 and 2025, indicating no pixel recognition or omission errors. In the bare land class, the 2013 accuracy (UA 71.43%, PA 83.33%) showed that there were still commission and omission errors, although performance improved in subsequent years. The agriculture land class maintained high accuracy with UA and PA above 80%, reflecting stable classification in agricultural areas. The agroforestry land class had the highest UA and PA throughout the year (>90%), indicating excellent class separation. For shrubland, accuracy decreased in 2019 (UA 83.33%), but PA remained at 100% in 2025, indicating that no shrubland pixels were missed but that pixels from other classes were misclassified as shrubland. Meanwhile, water body consistently had 100% UA and PA throughout the year, indicating that water body are the most easily recognisable class. Overall, this pattern shows that most classes have excellent accuracy, with minor variations in certain classes that are more prone to classification errors. The LULC classification results were generalised into two classes, consisting of built-up land and non-built-up land. Residential and non-residential buildings were classified as built-up land, while other LULC was classified as non-built-up land. The maps of built-up land in 2013, 2019 and 2025 are presented in Figure 5. Built-up land increased in area from year to year, growing from 2,674.35 ha in 2013 to 3,231.29 ha in 2019.

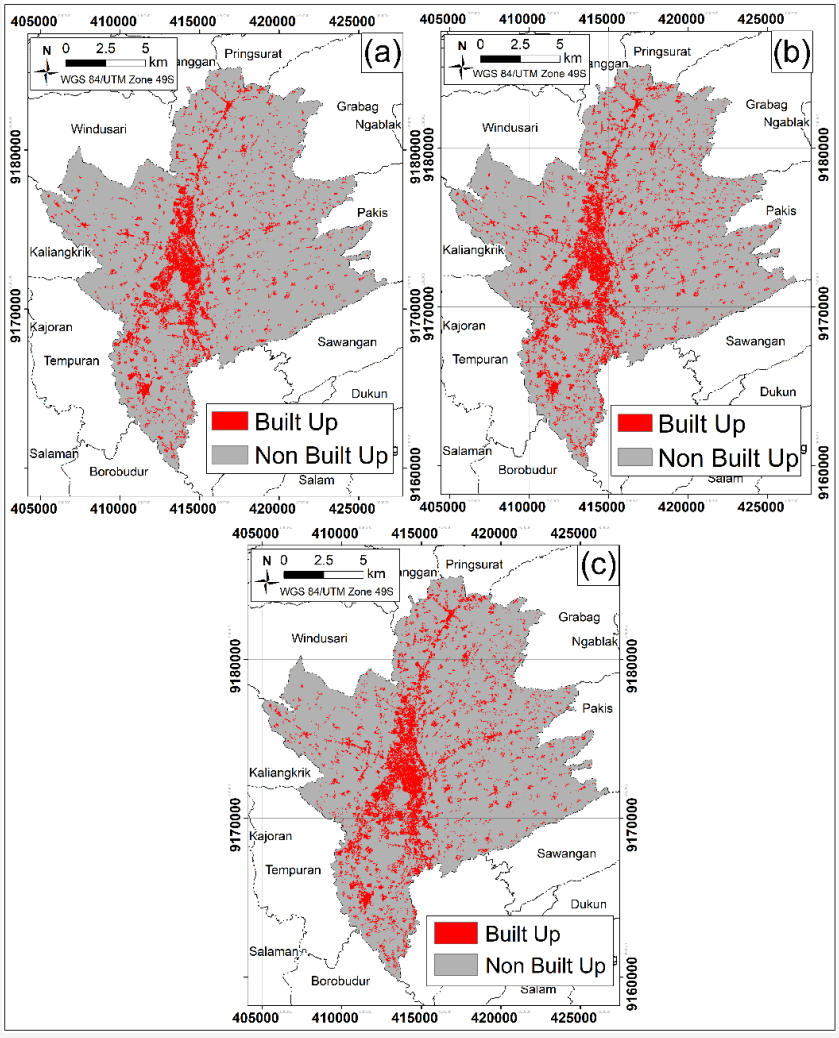

Figure 5: Built-Up and Non Built-Up land map of (a) 2013, (b) 2019, and (c) 2025

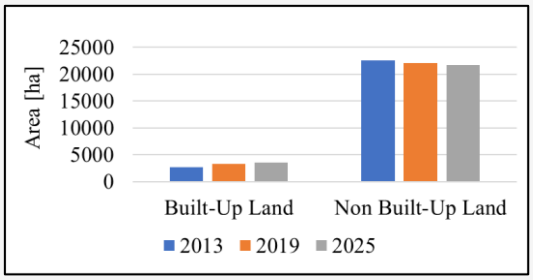

Figure 6: Built-Up and Non Built-Up area between 2013, 2019, and 2025

Then, in 2025, it will increase to 3,483.81 ha. Meanwhile, the area of non built-up land has decreased over time, from 22,555.2 ha in 2013 to 21,998.27 ha in 2019. Then, in 2025, it decreased to 21,745.76 ha. The graph showing changes in built-up and non-built-up land areas is presented in Figure 6. The reduction in non-built-up land area was due to the conversion of LULC from non-built-up land to built-up land. This shows that in Magelang City and its peri-urban areas, there has been an increase in the area of built-up land over time, indicating the expansion of built-up land.

However, the rate of increase in built-up land declined significantly from the period 2013 to 2019 to the period 2019 to 2025. From 2013 to 2019, there was an increase in built-up land area of 556.94 ha, while from 2019 to 2025 there was only an increase of 252.52 ha. This shows that during the period from 2013 to 2019, the annual growth rate of built-up land was 3.47%. However, in the period from 2019 to 2025, the annual growth rate decreased to 1.3%. This phenomenon may be due to the limited availability of land over time. If more lands are developed into built-up areas, the availability of potential land for development with flat slopes and strategic locations will decrease. In addition, the COVID-19 pandemic from 2020 to 2022 was also a key factor in this slowdown. The COVID-19 pandemic caused economic uncertainty, a decline in construction activity, and delays in housing projects. This certainly had a significant impact on the drastic slowdown in the addition of built-up land. In addition, the increase in built-up land occurred more in peri-urban areas than in Magelang City. From 2013 to 2025, there was an increase in built-up land area of 710.62 ha in peri-urban areas. Meanwhile, in Magelang City, there was only an increase in built-up land area of 98.84 ha. This is due to the limited availability of land near the city centre, resulting in more built-up land development occurring in the surrounding peri-urban areas. The peri-urban area that experienced the most expansion of built-up land was Mertoyudan Subdistrict, located in the southern part. This may be influenced by Mertoyudan District's position as a residential centre in Magelang Regency and its conditions that are conducive to development. Mertoyudan District has a low slope and elevation, is close to the main road, and is close to the city centre and public facilities. This will encourage development in Mertoyudan District to be faster than other peri-urban areas around Magelang City.

The land-use transition analysis from non built-up land to built-up land is presented in Table 6. During the period 2013–2019, agriculture land represented the largest source of newly developed built-up areas, suggesting that urban growth largely occurred through the transformation of productive land surrounding existing settlements. In contrast, during the period 2019–2025, the conversion of bare land and agroforestry land became more dominant, while the contribution of agriculture land decreased. This shift indicates a change in the pattern of urban expansion, where development increasingly occurred on previously unused or less intensively managed land rather than on agriculture areas. Overall, the results indicate that the availability of convertible land types plays an important role in shaping the pattern of built-up expansion in the study area.

3.2 Driving Factors Map

This study utilised 10 driving factors of built-up land expansion, including distance to roads, distance to rivers, distance to educational facilities, distance to health facilities, distance to industrial areas, distance to commercial facilities, distance to existing built-up land, population density, elevation, and slope. Maps of all driving factors are presented in Figure 7. Driving factors in the form of distance to objects were calculated using Euclidean distance in metres. Meanwhile, the population density map was created using polygon to raster. The Cramer's V values for each driving factor are presented in Table 7.

Table 6: Land Use transition matrix

Land Use Type	Converted to Built-up (2013–2019) (ha)	Converted to Built-up (2019–2025) (ha)
Agriculture Land	367.04	52.29
Bare Land	96.21	110.61
Agroforestry Land	101.19	89.24
Shrubland	58.64	33.21
Water Body	0.00	0.00

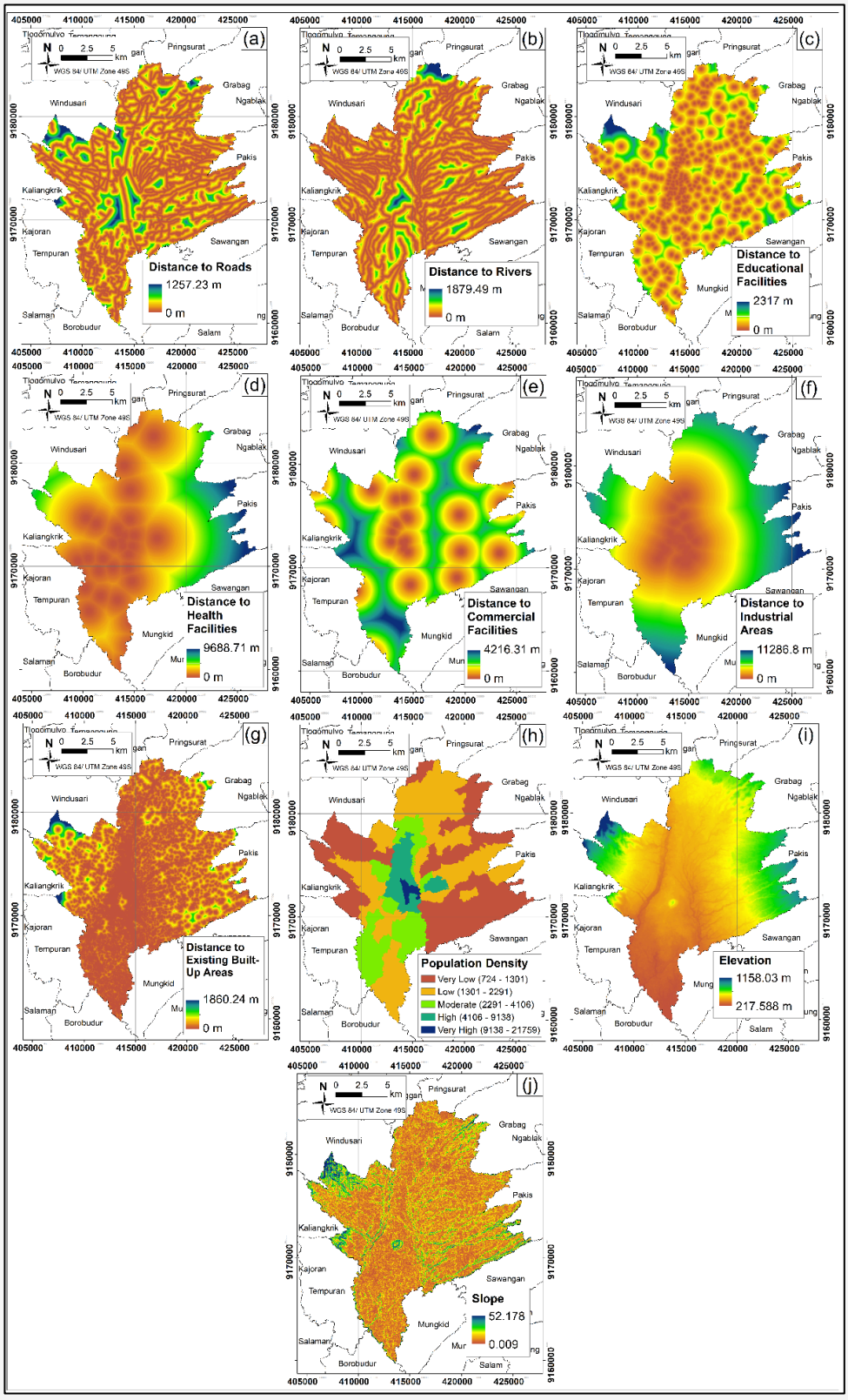

Figure 7: Driving factors map, (a) Distance to roads, (b) Distance to rivers, (c) Distance to educational facilities, (d) Distance to health Facilities, (e) Distance to commercial facilities, (f) Distance to industrial areas, (g) Distance to existing built-up areas, (h) Population density, (i) Elevation, and (j) Slope

Table 7: Cramer's V values for each driving factors

Driving Factor	Cramer's V
Distance to Road	0.195
Distance to River	0.151
Distance to Educational Facilities	0.157
Distance to Health Facilities	0.176
Distance to Commercial Facilities	0.048
Distance to Industrial Areas	0.156
Distance to Existing Built Up Areas	0.418
Population Density	0.198
Elevation	0.151
Slope	0.151

Table 8: Multicollinearity test results for each driving factors

Driving Factor	Tolerance	VIF
Distance to Road	0.857	1.167
Distance to River	0.915	1.093
Distance to Educational Facilities	0.805	1.243
Distance to Health Facilities	0.273	3.661
Distance to Industrial Areas	0.424	2.360
Distance to Existing Built Up Areas	0.546	1.833
Population Density	0.697	1.434
Elevation	0.317	3.156
Slope	0.546	1.327

The driving factors were tested using Cramer's V to determine the influence of each factor on LULC change from built-up to non-built-up land. The driving factor with the highest Cramer's V value is distance to existing built-up land, with a value of 0.4184. A driving factor can be categorised as good if it has a value greater than 0.4. This shows that distance to existing built-up land has a significant influence on LULC change. This is because new built-up land tends to be built near existing built-up land [62]. Another factor that also has a significant influence on the development of built-up land is population density, with a Cramer's V value of 0.1982. This is because population growth tends to drive the development of built-up land as a result of increasing housing needs Areas with high population densities have greater potential to experience LULC changes from non-built-up to built-up land [63].

Driving factors can be used in modelling if they have a value greater than 0.15. All driving factors except distance to commercial facilities have values greater than 0.15 and can therefore be used for further analysis. This indicates that most of the driving factors used in this study have an influence on built-up land expansion in Magelang City and its peri-urban areas. However, there is one factor, distance to commercial facilities, which has a value of less than 0.15, which has a value of 0.0475. This indicates that this factor cannot be used for further analysis and must therefore be eliminated. The small Cramer's V value indicates that distance to commercial facilities does not have a significant influence on LULC changes from non-built-up land to built-up land. These results indicate that there are only 9 driving factors that will be used in modelling.

Before being included in the modelling process, the driving factors were tested for multicollinearity to avoid highly correlated variables that could affect the stability and reliability of the model. Multicollinearity was evaluated using the Variance Inflation Factor (VIF) and tolerance values. The results of the multicollinearity test in Table 8 show that all driving factors have VIF values below 5 and tolerance values above 0.2. This indicates that there is no significant multicollinearity among the driving factors used in the modelling. In other words, the selected variables do not have strong linear relationships with one another and therefore can independently contribute to explaining the spatial variation of built-up land expansion. Ensuring the absence of multicollinearity improves the robustness of the modelling process and allows each driving factor to provide meaningful information in predicting the transition potential to built-up areas.

3.3 Transition Potential Map

Transition potential maps were created using logistic regression, SVM, MLP, and RF models in R Studio. Each model analysed the relationship between driving factors and LULC change in different ways, resulting in different outputs. The transition potential map shows the probability of a pixel changing from non-built-up land to built-up land. The value ranges from 0 to 1, where the closer the value is to 1, the

greater the likelihood that the pixel will change to built-up land. The transition potential maps created using the four models are presented in Figure 8. The results show that the transition potential maps created using each model have different pixel value ranges but remain within the range of 0 to 1. The largest range is owned by the random forest model with values between 0 and 0.996. Meanwhile, the smallest range is possessed by the MLP model with a value range between 0.12459 and 0.735342.

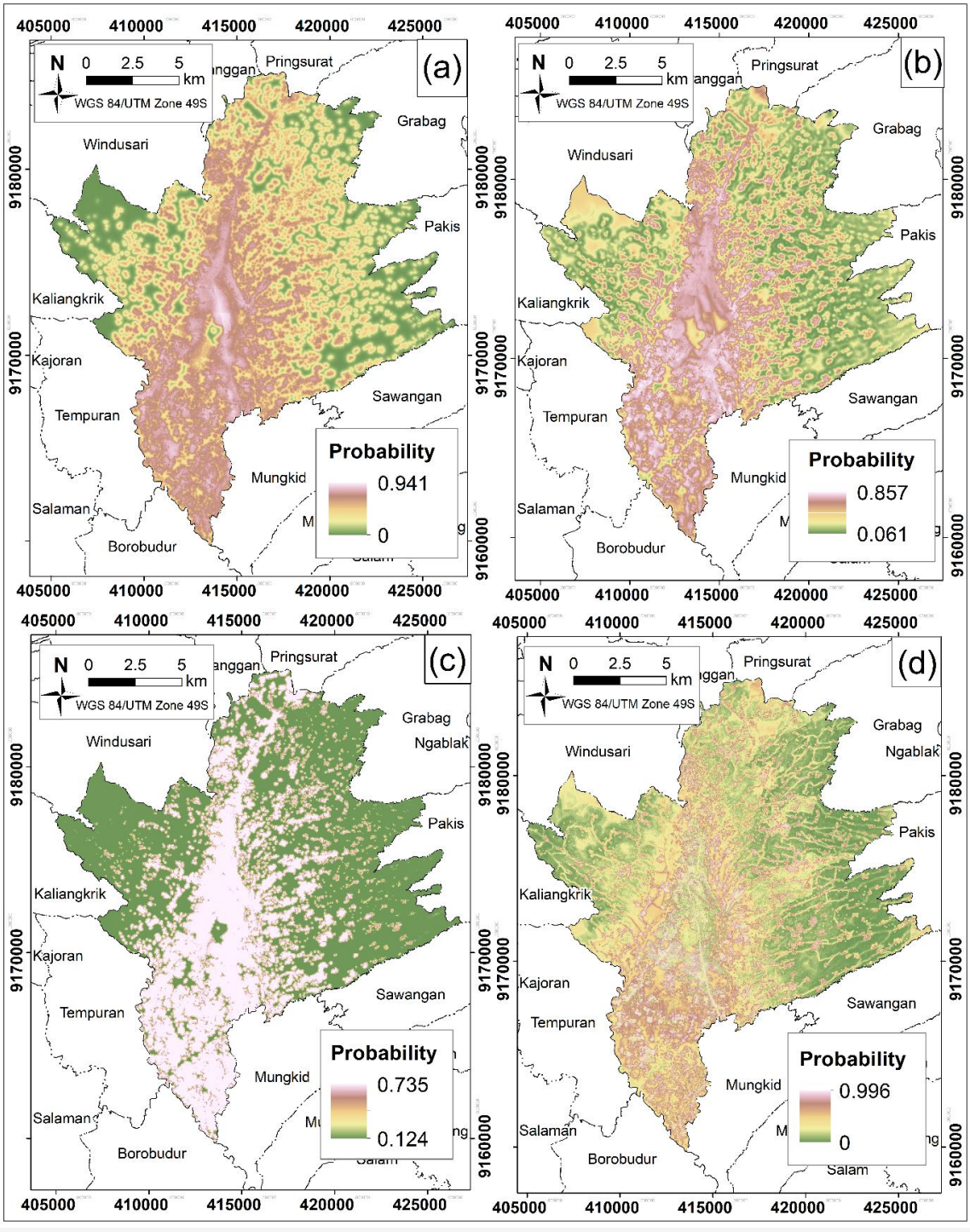

Figure 8: Transition potential map, (a) Logistic regression, (b) SVM, (c) MLP, and (d) RF

The differences in the transition potential maps across the four models are also reflected in the distribution of pixel values. In the logistic regression map, pixels with high probability values are predominantly concentrated in the city centre, particularly along the main road. In contrast, pixels with low probabilities are dispersed across the Bandongan in the western part and Candimulyo subdistricts in the south-eastern part, where agroforestry is the dominant land use. In the SVM transition potential map, high-probability pixels are more widely scattered throughout Magelang City in the centre part and Mertoyudan subdistrict in the southern part, while low-probability pixels are fewer and distributed across Bandongan in the western part, Tegalrejo in the eastern part, and Candimulyo in the south-eastern part. Similarly, in the MLP transition potential map, high-probability pixels are extensively distributed in Magelang City in the centre part, Mertoyudan in the southern part, and Secang Districts in the northern part. Conversely, Bandongan, Tegalrejo, and Candimulyo Districts are largely dominated by low-probability pixels, with values approaching zero. In the RF transition potential map, high-probability pixels are less numerous compared to the other models and are primarily located near the main roads. Meanwhile, low-probability pixels are widely distributed across areas characterized by agroforestry land use, particularly in Bandongan, Tegalrejo, and Candimulyo subdistricts.

In all four maps, it can be observed that pixels with high probability values tend to be located in the Magelang City and Mertoyudan District areas. This may be influenced by several factors, including their proximity to the city centre, existing built-up areas, main roads, industrial areas, and public facilities such as educational and health facilities. In addition, areas in both regions also tend to have low elevation and slope. Built-up land tends to be developed on land with low elevation and slope because it has better carrying capacity for development. Land with high elevation and slope has higher surface runoff, which increases the risk of flooding and landslides, making it unsuitable for development [64].

3.4 Simulated Built-Up Map

Transition potential maps created using logistic regression, SVM, MLP, and RF models are integrated with transition probability matrices created with Markov Chain to predict the development of built-up land in Magelang City and its peri-urban areas using cellular automata. The map of the simulated development of built-up land is presented in Figure 9. The map of the simulated development of built-up land shows different results for the four models. The differences are shown in the distribution of built-up land pixels between models. The logistic regression and MLP models show simulation results indicating that the development of built-up land is more concentrated in the Magelang City area near the city centre. This is evidenced by highly dense built-up land pixels within the city. Meanwhile, SVM and random forest show simulation results of built-up land development that is more spread out across the peri urban areas surrounding Magelang City.

Each model also produced different simulation results regarding the increase in built-up land area. The logistic regression model indicated that by 2025, Magelang City and its surrounding areas would have a built-up land area of 3,412.78 hectares. Meanwhile, the SVM model indicated a larger built-up land area of 3,439.67 hectares. Meanwhile, the MLP model also indicates a larger built-up area compared to the previous two models, covering 3,483.25 hectares. Furthermore, the random forest model indicates the largest built-up area, covering 3,587.69 hectares. The differences in area produced by each model are influenced by how each model constructs the transition potential map. Logistic regression produces the smallest predicted area because it is linear and conservative, thus only assigning high probability values to areas that very clearly meet the criteria. This results in a more limited number of eligible cells, leading to the smallest area. Meanwhile, SVM can capture non-linear relationships between driving factors and changes in land use using an RBF kernel, resulting in a greater number of pixels with high probability values compared to logistic regression. Furthermore, MLP has the ability to model complex relationships between driving variables and changes in land use, resulting in a more flexible and extensive distribution of pixel probability values. Meanwhile, RF works by combining numerous decision trees, thereby tending to assign high potential values to a greater number of pixels, so that the resulting area is maximised.

The map of simulated built-up land development was then tested against the existing built-up land map for 2025 to obtain the best model with the highest accuracy. The accuracy test was carried out using two methods, consisting of overall accuracy to measure the accuracy of quantitative changes and figure of merit (FOM) to measure the accuracy of location changes. FOM is calculated using three aspects, namely hits, misses, and false alarms. The distribution map of hits, misses, and false alarms in the four models is presented in Figure 10. The results of the accuracy and FOM tests are shown in Table 9. The logistic regression model has an overall accuracy of 98.3% and an FOM of 9.7%.

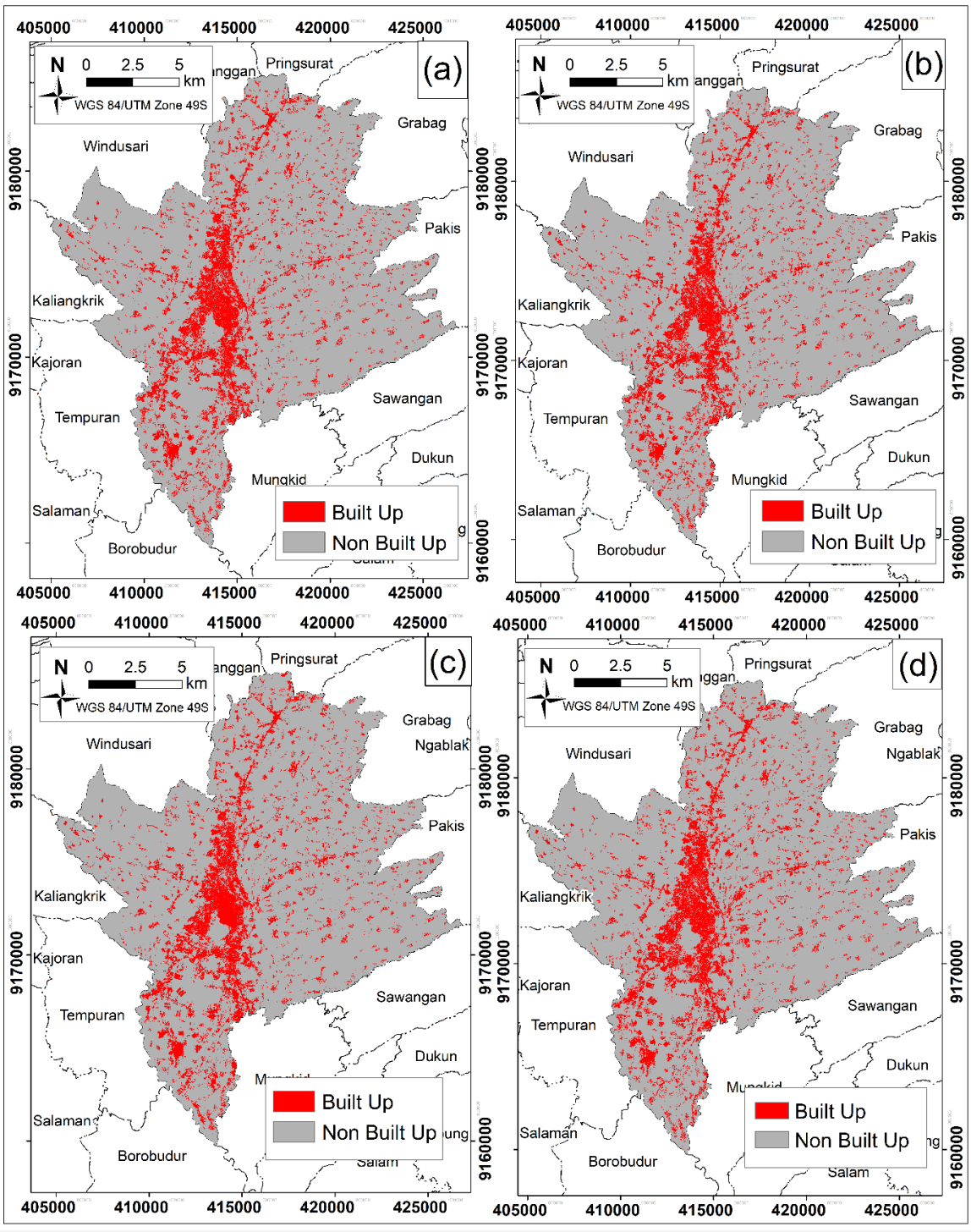

Figure 9: Simulated built-up map, (a) Logistic regression, (b) SVM, (c) MLP, and (d) RF

The SVM model has an overall accuracy of 98.4% and an FOM of 15.3%. The MLP model has an overall accuracy of 98% and an FOM of 7.4%. The RF model has an overall accuracy of 97.4% and an FOM of 2.9%. These results indicate that the SVM model has the highest accuracy because it has a higher overall accuracy and FOM than the other three models. Meanwhile, the model with the lowest accuracy is RF because it has a lower overall accuracy and FOM than the other three models.

Table 9: Overall accuracy and FOM of logistic regression, SVM, MLP, and RF models

Model	LR (%)	SVM (%)	MLP (%)	RF (%)
Overall Accuracy	98.3	98.4	98.0	97.4
Figure of Merit	9.7	15.3	7.4	2.9

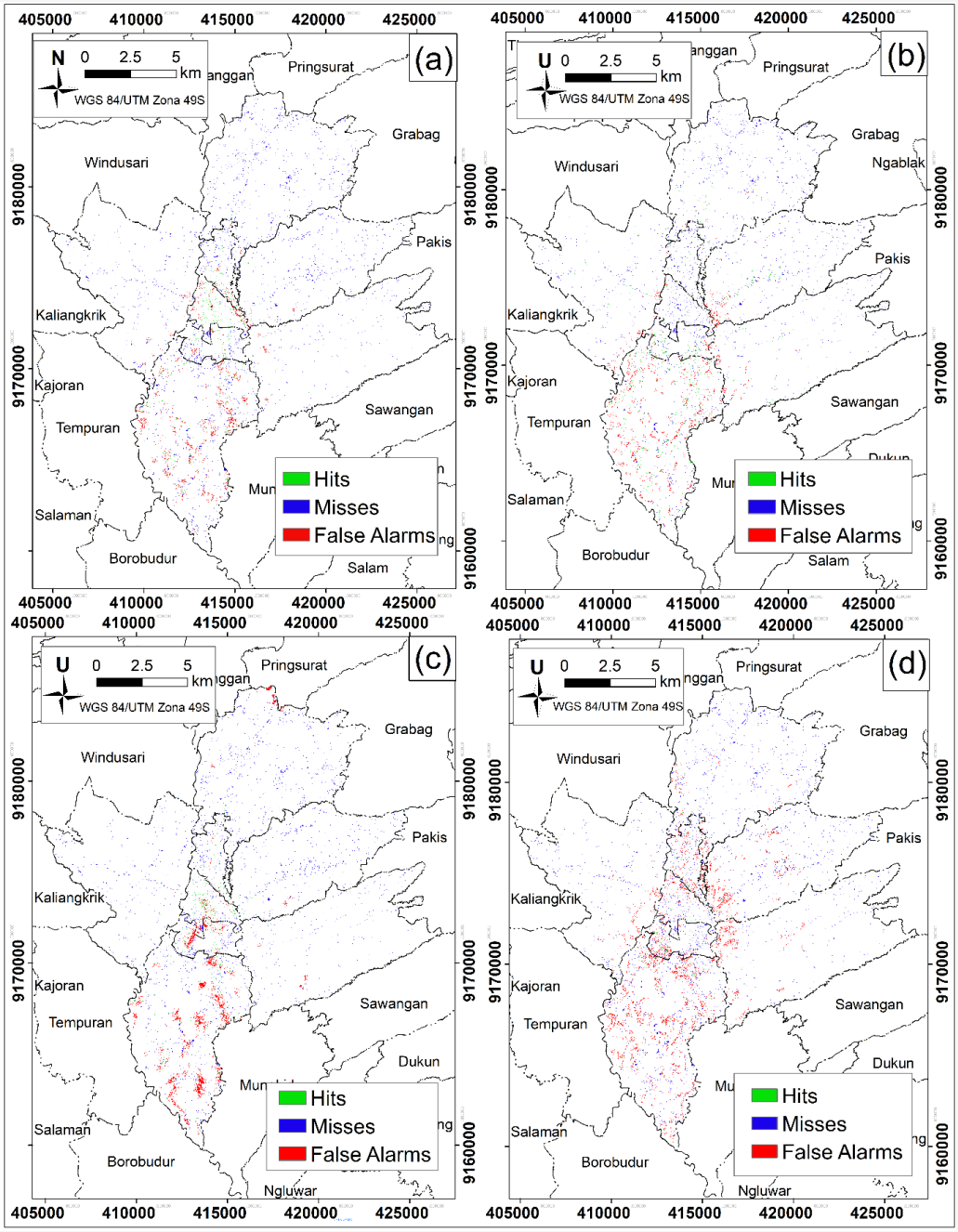

Figure 10: Distribution map of hits, misses, and false alarms:
(a) Logistic regression, (b) SVM, (c) MLP, (d) RF

These findings underscore the importance of evaluating both the quantitative and spatial accuracy of land use prediction models. While overall accuracy reflects the model's ability to reproduce the correct amount of built-up land, the Figure of Merit (FOM) provides a more stringent assessment by measuring the overlap between predicted and actual locations of change. A higher FOM indicates that the model not only predicts the correct quantity of change but also allocates it to the correct spatial locations, which is critical for urban planning and land management applications. The superior performance of the SVM model suggests that it is more effective in capturing the spatial dynamics of urban expansion in Magelang City and its surrounding peri-urban areas. This may be attributed to SVM's ability to handle complex and non-linear relationships between LULC transitions and driving factors, particularly when the RBF kernel is used. Therefore, the SVM model is recommended as the most suitable approach for simulating future built-up land development in this Magelang City and its peri-urban areas, particularly when spatial precision is essential for policy making, infrastructure planning, and environmental impact assessments.

4. Conclusion

This study shows that LULC changes occurred in Magelang City and its peri-urban areas between 2013 and 2025. Residential buildings, non-residential buildings, agriculture land, and agroforestry land increased in area over time. Meanwhile, bare land and shrubland decreased in area over time. The results also show that there was an expansion of built-up land in Magelang City and its peri-urban areas from 2013 to 2025, from 2,674.35 ha to 3,483.81 ha. The annual growth rate of built-up land decreased from the 2013 to 2019 period to the 2019 to 2025 period, which may be due to the limited availability of land. In addition, built-up land expansion occurred more in peri-urban areas than in Magelang City.

This study predicts built-up expansion using CA-Markov integrated with four models, namely logistic regression, SVM, MLP, and RF. The driving factors of built-up land expansion are determined using Cramer's V analysis, and it is found that the factor that has a significant influence is the distance to existing built-up areas. Meanwhile, the factor with the smallest influence is the distance to commercial facilities, and this factor was not included in the modelling because it had a value of <0.15. The built-up expansion prediction shows that logistic regression, SVM, MLP, and RF models can be used to construct transition potential maps as input for CA-Markov and produce relatively good accuracy. The model with the highest accuracy is SVM with an overall accuracy of 98.4% and FOM of 15.3%. These results indicate that the SVM model is the most suitable model to use in predicting built-up land expansion in Magelang City and its peri-urban areas because it has the highest accuracy.

5. Limitation

This study still has several limitations that can be improved in future research. One limitation of this study is that the modelling framework focuses primarily on predicting the expansion of built-up areas. Consequently, transitions among other land-use types, such as agriculture, forest, or bare land, were not explicitly simulated. While this approach allows the study to focus on urban expansion dynamics, it may not fully capture the complexity of overall land-use change processes. Another limitation is that the model does not incorporate socio-economic factors that may influence built-up land expansion, such as land prices, distance to workplaces, regional income distribution, and land ownership status. Future studies could address these limitations by incorporating multi-class land-use transitions and integrating socio-economic variables into the modelling framework to provide a more comprehensive representation of land-use change dynamics.

Acknowledgements

The author gratefully acknowledges the Faculty of Geography, Universitas Gadjah Mada, for academic support and facilities provided during the research process. The author also extends sincere appreciation to the Department of Public Works and Spatial Planning (DPUPR) of Magelang City and Magelang Regency for providing spatial data and supporting information essential to this study.

References

[1] Hafid, S., (2021). Simulasi Spasial Berbasis Citra Landsat, Cellular Automata dan Regresi Logistik Biner Untuk Prediksi Perubahan Penggunaan Lahan Pertanian di Kabupaten Karawang [Spatial Simulation Based on Landsat Imagery, Cellular Automata, and Binary Logistic Regression for Predicting Agricultural Land Use Change in Karawang Regency]. Master's Thesis, Remote Sensing, Universitas Gadjah Mada. Available: https://etd.repository.ugm.ac.id/.

[2] Rahmadewi, R. and Kurniati, E., (2025). Dampak Alih Fungsi Lahan Terhadap Pembangunan Daerah: Studi Kasus Di Kabupaten Kendal [The Impact of Land

Conversion on Regional Development: A Case Study in Kendal Regency]. *Jurnal Ilmu Ekonomi*, Vol. 4(1); 298–322. https://doi.org/10.59827/JIE.V4I1.225.

[3] Prabandari, A. A., Dwi, M. and Manessa, M., (2024). Analisis Perkembangan Lahan Terbangun Berdasarkan Metode Supervised Classification Menggunakan Google Earth Engine (Studi Kasus: Desa Ciputi, Kecamatan Pacet, Kabupaten Cianjur) [Analysis of Built-Up Land Development Based on the Supervised Classification Method Using Google Earth Engine (Case Study: Ciputi Village, Pacet District, Cianjur Regency)]. *Jurnal Tanah dan Sumberdaya Lahan*, Vol. 11(2); 403–412. https://doi.org/10.21776/UB.JTSL.2024.011.2.11.

[4] Yuliastuti, N., Fatchurochman, A. and Sudarto, J., (2012). Pengaruh Perkembangan Lahan Terbangun Terhadap Kualitas Lingkungan Permukiman (Studi Kasus: Kawasan Pendidikan Kelurahan Tembalang) [The Influence of Built-Up Land Development on Settlement Environmental Quality (Case Study: Educational Area of Tembalang Subdistrict)]. *Jurnal Presipitasi: Media Komunikasi dan Pengembangan Teknik Lingkungan*, Vol. 9(1); 10–16. https://doi.org/10.14710/presipitasi.v9i1.10-16.

[5] Beckers, V., Poelmans, L., Van Rompaey, A. and Dendoncker, N., (2020). The Impact of Urbanization on Agricultural Dynamics: A Case Study in Belgium. *Journal of Land Use Science*, Vol. 15(5); 626–643. https://doi.org/10.1080/1747423X.2020.1769211.

[6] Perroux, F., (1950). Economic Space: Theory and Applications. *Quarterly Journal of Economics*, Vol. 64(1); 89–104. https://doi.org/10.2307/1881960.

[7] Orbawati, E. B. and Muhammad, I., (2024). Challenges and Opportunities: Assessing Smart City Readiness in Magelang City for Sustainable Change. *Sosiohumaniora*, Vol. 26(3); 422–429. https://doi.org/10.24198/SOSIOHUMANIORA.V26I3.56120.

[8] Bere, R. C., Precup, I. B. and Silvestru, C. I., (2015). On Growth Poles from EU Countries in the Framework of Europe 2020. *Procedia Economics and Finance*, Vol. 23; 920–925. https://doi.org/10.1016/S2212-5671(15)00519-5.

[9] Suprayogi, R., (2021). Kesesuaian Perubahan Penggunaan Lahan Dengan Rencana Tata Ruang Di Kawasan Peri-Urban [Suitability of Land Use Change with Spatial Planning in Peri-Urban Areas]. *Jurnal Kajian Ruang*, Vol. 1(2); 238–254. https://doi.org/10.30659/jkr.v1i2.20031.

[10] Fahyudi, D. I., Christiawan, P. I. and Sarmita, I. M., (2020). Perkembangan Luas Permukiman Dan Penggunaan Lahan Pada Daerah Peri-Urban Kota Singaraja Tahun 2010, 2015 Dan 2020 [Development of Settlement Area and Land Use in the Peri-Urban Area of Singaraja City in 2010, 2015, and 2020]. *Jurnal Pendidikan Geografi Undiksha*, Vol. 8(3); 140–147. https://doi.org/10.23887/JJPG.V8I3.29541.

[11] Mardiansjah, F. H., Sugiri, A. and Ma'rif, S., (2021). Examining Small-Town Growth and Expansion in Peri-Urban Areas of Small Cities: Evidence from Peripheries of Three Small Cities in Central Java. *Journal of Regional and City Planning*, Vol. 32(3); 216–232. https://doi.org/10.5614/jpwk.2021.32.3.2.

[12] Destianto, R. and Pigawati, B., (2014). Analisis Keterkaitan Perubahan Lahan Pertanian Terhadap Ketahanan Pangan Kabupaten Magelang Berbasis Model Spatio Temporal SIG [Analysis of the Relationship Between Agricultural Land Change and Food Security in Magelang Regency Based on a Spatio-Temporal GIS Model]. *Geoplanning: Journal of Geomatics and Planning*, Vol. 1(1); 21–32. https://doi.org/10.14710/GEOPLANNING.1.1.21-32.

[13] Hafiz, A. A., Usman, F., Rohman, A. R., Hidayat, T. and Zakiyah, D. M., (2025). Spatiotemporal Dynamics of Urban Sprawl Types in the Peri-Urban Area of Malang Municipality, Indonesia. *Urban Science*, Vol. 9(8). https://doi.org/10.3390/URBANSCI9080313.

[14] Saraswati, Z. F., Purqon, A., Malik, I. B. I., Awfa, D., Awan, F. N., Risky, M., Permata, M. V., Paramitha, M., Menanza, I. W. and Putra, N., (2022). Model of Land Cover Change Caused by Toll Road Access Using Cellular Automata in Sumatra Island. *Jurnal Arsitektur*, Vol. 12(1); 51–66. https://doi.org/10.36448/ja.v12i1.2323.

[15] Suryaningsih, W., Siregar, H. and Mulatsih, S., (2024). Pengaruh Pembangunan Jalan Tol terhadap Perubahan Tata Guna Lahan dan Luas Lahan Sawah di Kabupaten Brebes [The Effect of Toll Road Development on Land Use Change and Rice Field Area in Brebes Regency]. *TATALOKA*, Vol. 26(3); 154–164. https://doi.org/10.14710/TATALOKA.26.3.154-164.

[16] Noraini, A. and Mabrur, A. Y., (2020). Perbandingan Visualisasi Hasil Deteksi Area Terbangun Berdasarkan Metode Maximum Likelihood Classification (MLC) dan Normalized Difference Built-Up Index (NDBI) [Visualization Comparison of Built-Up Area Detection Results Based on Maximum Likelihood Classification (MLC) and Normalized Difference Built-Up Index (NDBI) Methods]. *Buletin LOUPE*, Vol. 16(1).

[17] Liu, H., Homma, R., Liu, Q. and Fang, C., (2021). Multi-Scenario Prediction of Intra-Urban Land Use Change Using a Cellular Automata-Random Forest Model. *ISPRS International Journal of Geo-Information*, Vol. 10(8). https://doi.org/10.3390/IJGI10080503.

[18] Zahrotunisa, S. and Wicaksono, P., (2017). Prediksi Spasial Perkembangan Lahan Terbangun Melalui Pemanfaatan Citra Landsat Multitemporal di Kota Bogor [Spatial Prediction of Built-Up Land Development Through the Utilization of Multitemporal Landsat Imagery in Bogor City]. *Jurnal Online Informatika*, Vol. 2(1); 30–35. https://doi.org/10.15575/JOIN.V2I1.88.

[19] Rokni, K., (2023). Investigating the Impact of Pan Sharpening on the Accuracy of Land Cover Mapping in Landsat OLI Imagery. *Geodesy and Cartography*, Vol. 49(1); 12–18. https://doi.org/10.3846/GAC.2023.15308.

[20] Lestari, W. and Pratomoatmojo, N. A., (2019). Pemodelan Spasial Prediksi Perkembangan Kawasan Permukiman Berbasis Cellular Automata dengan Pendekatan Kependudukan di Surabaya Timur [Spatial Modeling for Predicting Settlement Area Development Based on Cellular Automata with a Population Approach in East Surabaya]. *Jurnal Teknik ITS*, Vol. 8(2); 507–808. https://doi.org/10.12962/J23373539.V8I2.48665.

[21] Rusdin, M., Putra, R. and Rudiarto, I., (2018). Simulasi Perubahan Penggunaan Lahan Dengan Konsep Cellular Automata Di Kota Mataram [Simulation of Land Use Change Using the Cellular Automata Concept in Mataram City]. *Jurnal Pengembangan Kota*, Vol. 6(2); 174–185. https://doi.org/10.14710/jpk.6.2.174-185.

[22] Ouma, Y. O., Nkwae, B., Odirile, P., Parida, B. and Qi, J., (2024). Land-Use Change Prediction in Dam Catchment Using Logistic Regression-CA, ANN-CA and Random Forest Regression and Implications for Sustainable Land–Water Nexus. *Sustainability*, Vol. 16(4). https://doi.org/10.3390/SU16041699.

[23] Parasdyo, M. M. and Susilo, B., (2016). Komparasi Akurasi Model Cellular Automata untuk Simulasi Perkembangan Lahan Terbangun dari Berbagai Variasi Matriks Probabilitas Transisi: Kasus Bagian Timur Kota Yogyakarta [Comparison of Cellular Automata Model Accuracy for Simulating Built-Up Land Development from Various Transition Probability Matrix Variations: Case of Eastern Yogyakarta City]. *Jurnal Bumi Indonesia*, Vol. 5(4).

[24] Sankarrao, L., Ghose, D. K. and Rathinsamy, M., (2021). Predicting Land-Use Change: Intercomparison of Different Hybrid Machine Learning Models. *Environmental Modelling and Software*, Vol. 145. https://doi.org/10.1016/J.ENVSOFT.2021.105207.

[25] Kamusoko, C. and Gamba, J., (2015). Simulating Urban Growth Using a Random Forest-Cellular Automata (RF-CA) Model. *ISPRS International Journal of Geo-Information*, Vol. 4(2); 447–470. https://doi.org/10.3390/IJGI4020447.

[26] BPS, (2025). Kota Magelang Dalam Angka 2025 [Magelang City in Figures 2025]. Badan Pusat Statistik, Magelang City.

[27] BPS, (2025). Kabupaten Magelang Dalam Angka 2025 [Magelang Regency in Figures 2025]. Badan Pusat Statistik, Magelang Regency.

[28] Narotama, N., Sunoko, K. and Pramesti, L., (2021). Penerapan Prinsip Transit-Oriented Development Pada Perencanaan Stasiun Kereta Api Di Kota Magelang [Implementation of Transit-Oriented Development Principles in Railway Station Planning in Magelang City]. *Senthong*, Vol. 4(1); 33–42. Available: https://jurnal.ft.uns.ac.id/index.php/senthong/article/view/1223.

[29] Magelang Regency Government, (2024). Peraturan Daerah (Perda) Kabupaten Magelang Nomor 7 Tahun 2024 tentang Rencana Tata Ruang Wilayah Kabupaten Magelang Tahun 2024–2044 [Regional Regulation of Magelang Regency Number 7 of 2024 concerning the Spatial Plan of Magelang Regency 2024–2044]. Magelang Regency Government, Magelang Regency.

[30] Sholikhah, H. F., (2021). Fenomena Pembayaran Pajak Bumi dan Bangunan (PBB) Di Kecamatan Bandongan Kabupaten Magelang [The Phenomenon of Land and Building Tax (PBB) Payments in Bandongan District, Magelang Regency]. *The Journalish: Social and Government*, Vol. 2(2); 30–35. https://doi.org/10.55314/TSG.V2I2.141.

[31] BPS, (2024). Kecamatan Secang Dalam Angka 2024 [Secang District in Figures 2024]. Badan Pusat Statistik, Magelang Regency.

[32] BPS, (2024). Kecamatan Tegalrejo Dalam Angka 2024 [Tegalrejo District in Figures 2024]. Badan Pusat Statistik, Magelang Regency.

[33] BPS, (2024). Kecamatan Candimulyo Dalam Angka 2024 [Candimulyo District in Figures 2024]. Badan Pusat Statistik, Magelang Regency.

[34] Maurer, T., (2013). How to Pan-Sharpen Images Using the Gram-Schmidt Pan-Sharpen Method: A Recipe. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 40(1W1); 239–244. https://doi.org/10.5194/ISPRSARCHIVES-XL-1-W1-239-2013.

[35] Pálsson, F., Sveinsson, J. R., Benediktsson, J. A. and Aanæs, H., (2012). Classification of Pansharpened Urban Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 5(1); 281–297. https://doi.org/10.1109/JSTARS.2011.2176467.

[36] Marlina, D., (2022). Klasifikasi Tutupan Lahan pada Citra Sentinel-2 Kabupaten Kuningan dengan NDVI dan Algoritme Random Forest [Land Cover Classification on Sentinel-2 Imagery of Kuningan Regency Using NDVI and Random Forest Algorithm]. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, Vol. 7(1). https://doi.org/10.30998/STRING.V7I1.12948.

[37] Sugiyono, (2017). Metode Penelitian Kuantitatif, Kualitatif, dan R&D [Quantitative, Qualitative, and R&D Research Methods]. Alfabeta, Bandung.

[38] Congalton, R. G., (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, Vol. 37(1); 35–46. https://doi.org/10.1016/0034-4257(91)90048-B.

[39] Megahed, Y., Cabral, P., Silva, J. and Caetano, M., (2015). Land Cover Mapping Analysis and Urban Growth Modelling Using Remote Sensing Techniques in Greater Cairo Region Egypt. *ISPRS International Journal of Geo-Information*, Vol. 4(3); 1750–1769. https://doi.org/10.3390/IJGI4031750.

[40] Kim, J. H., (2019). Introduction to Multicollinearity and Misleading Statistical Results. *Korean Journal of Anesthesiology*, Vol. 72(6); 558–569. https://doi.org/10.4097/kja.19087.

[41] Gaur, S., Mittal, A., Bandyopadhyay, A., Holman, I. and Singh, R., (2020). Spatio-Temporal Analysis of Land Use and Land Cover Change: A Systematic Model Inter-Comparison Driven by Integrated Modelling Techniques. *International Journal of Remote Sensing*, Vol. 41(23); 9229–9255. https://doi.org/10.1080/01431161.2020.1815890.

[42] Ae, P. H., (2015). An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. *Journal of Korean Academy of Nursing*, Vol. 43(2); 154–164. https://doi.org/10.4040/JKAN.2013.43.2.154.

[43] Thammaboribal, P., and Tripathi, N. (2024). Predicting Land Use and Land Cover Changes in Pathumthani, Thailand: A Comprehensive Analysis from 2013 to 2023 Using Landsat Satellite Imagery and CA-ANN Algorithm, with Projections for 2028 and 2038. *International Journal of Geoinformatics*, Vol. 20(5), 13–27. https://doi.org/10.52939/ijg.v20i5.3225.

[44] Aji, A., Husna, V., and Purnama, S. (2024). Multi-Temporal Data for Land Use Change Analysis Using a Machine Learning Approach (Google Earth Engine). *International Journal of Geoinformatics*, Vol. 20(4), 19–28. https://doi. org/10.52939/ijg.v20i4.3145.

[45] Fremmuzar, P. and Baita, A., (2023). Uji Kernel SVM dalam Analisis Sentimen Terhadap Layanan Telkomsel di Media Sosial Twitter [SVM Kernel Testing in Sentiment Analysis of Telkomsel Services on Twitter Social Media]. *Komputika: Jurnal Sistem Komputer*, Vol. 12(2); 57–66. https://doi.org/10.34010/KOMPUTIKA.V12I2.9460.

[46] Lin, H. T., Lin, C. J. and Weng, R. C., (2007). A Note on Platt's Probabilistic Outputs for Support Vector Machines. *Machine Learning*, Vol. 68(3); 267–276. https://doi.org/10.1007/S10994-007-5018-6.

[47] Chan, K. Y., Abu-Salih, B., Qaddoura, R., Al-Zoubi, A. M., Palade, V., Pham, D.-S., Del Ser, J. and Muhammad, K., (2023). Deep Neural Networks in the Cloud: Review, Applications, Challenges and Research Directions. *Neurocomputing*, Vol. 545. https://doi.org/10.1016/j.neucom.2023.126327.

[48] Khalil, M., Alsmadi, S., Bin Omar, K. and Noah, S. A., (2009). Back Propagation Algorithm: The Best Algorithm among the Multi-Layer Perceptron Algorithm. *IJCSNS International Journal of Computer Science and Network Security*, Vol. 9(4).

[49] Jagtap, A. D. and Karniadakis, G. E., (2023). How Important Are Activation Functions in Regression and Classification? A Survey, Performance Comparison, and Future Directions. *Journal of Machine Learning for Modeling and Computing*, Vol. 4(1); 21–75.

[50] Breiman, L., (2001). Random Forests. *Machine Learning*, Vol. 45(1); 5–32. https://doi.org/10.1023/A:1010933404324.

[51] Belgiu, M. and Drăgu, L., (2016). Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 114; 24–31. https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011.

[52] Pholgerddee, P., Pattanakiat, S., Nakmuenwai, P., Sattraburut, T., and Phutthai, T. (2026). Relative Importance of Environmental and Spatial Predictors for Slipper Orchid (Paphiopedilum spp.) Distribution in Thailand Using Random Forest . *International Journal of Geoinformatics*, Vol. 22(4), 160–176. https://doi.org/10.52939/ijg.v22i4.4945.

[53] Eastman, R., (2006). *IDRISI Andes Guide to GIS and Image Processing*. Clark University, Worcester.

[54] Behera, M. D., Borate, S. N., Panda, S. N., Behera, P. R. and Roy, P. S., (2012). Modelling and Analyzing the Watershed Dynamics Using Cellular Automata (CA)–Markov Model: A Geo-Information Based Approach. *Journal of Earth System Science*, Vol. 121(4); 1011–1024. https://doi.org/10.1007/S12040-012-0207-5.

[55] Kumar, S., Radhakrishnan, N. and Mathew, S., (2014). Land Use Change Modelling Using a Markov Model and Remote Sensing. *Geomatics, Natural Hazards and Risk*, Vol. 5(2); 145–156. https://doi.org/10.1080/19475705.2013.795502.

[56] Aldileemi, H., Zhran, M., and El-Mewafi, M. (2023). Geospatial Monitoring and Prediction of Land Use/Land Cover (LULC) Dynamics Based on the CA-Markov Simulation Model in Ajdabiya, Libya. *International Journal of Geoinformatics*, Vol. 19(12), 15–29. https://doi.org/10.52939/ijg.v19i12.2973.

[57] Batty, M., (2007). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. MIT Press, Cambridge.

[58] Shi, W., Goodchild, M., Batty, M., Kwan, M. P. and Zhang, A., (2021). *Urban Informatics*. Springer, Singapore.

[59] Yuhendra and Yulianti, E., (2019). Multi-Temporal Sentinel-2 Images for Classification Accuracy. *Journal of Computer Science*, Vol. 15(2); 258–268. https://doi.org/10.3844/JCSSP.2019.258.268.

[60] Pontius, R. G. Jr., Boersma, W., Castella, J. C., Clarke, K., Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C. D., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T. N., Veldkamp, A. T. and Verburg, P. H., (2008). Comparing the Input, Output, and Validation Maps for Several Models of Land Change. *Annals of Regional Science*, Vol. 42(1); 11–37. https://doi.org/10.1007/S00168-007-0138-2.

[61] Anderson, J. R., (1971). Land Use Classification Schemes Used in Selected Recent Geographic Applications of Remote Sensing. *Photogrammetric Engineering*, Vol. 37(4); 379–387.

[62] Kim, M., Kim, D., Jin, D. and Kim, G., (2023). Application of Explainable Artificial Intelligence (XAI) in Urban Growth Modeling: A Case Study of Seoul Metropolitan Area, Korea. *Land*, Vol. 12(2). https://doi.org/10.3390/LAND12020420.

[63] Angel, S., (2023). Urban Expansion: Theory, Evidence and Practice. *Buildings and Cities*, Vol. 4(1); 124–138. https://doi.org/10.5334/BC.348.

[64] Benchelha, N., Bezza, M., Belbounaguia, N., Benchelha, S., and Benchelha, M., (2022). Modeling Dynamic Urban Growth Using Cellular Automata and Geospatial Technique: Case of Casablanca in Morocco. *International Journal of Geoinformatics*, Vol. 18(5); 27–40. https://doi.org/10.52939/ijg.v18i5.2369.