Place and City: Toward Urban Intelligence
Integration of Local and Global Support Vector Machines to Improve Urban Growth Modelling

Babak Mirbagheri and Abbas Alimohammadi *

Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran 19967-15433, Iran; b_mirbagheri@email.kntu.ac.ir
* Correspondence: alimoh_abb@kntu.ac.ir; Tel.: +98-21-8887-7070

Received: 27 June 2018; Accepted: 20 August 2018; Published: 24 August 2018

Abstract: The use of local information for the classification and modelling of spatial variables has increased with the application of statistical and machine learning algorithms, such as support vector machines (SVMs). This study presents a new local SVM (LSVM) model that was developed to model the probability of urban development and simulate urban growth in a subregion in the southwestern suburb of the Tehran metropolitan area, Iran, for the periods of 1992–1996 and 1996–2002. Based on the focal training sample, the model was calibrated using the cross-validation method, and the optimal bandwidth was determined. The results were compared with those of a nonlinear global SVM (GSVM) model that was calibrated based on the ten-fold cross-validation method. This study then evaluated an integrated SVM model (LGSVM) obtained based on a weighted combination of the local and global urban development probabilities. A comparison of the probability maps showed a higher accuracy for the LGSVM than for either the LSVM or GSVM model. To assess the performance of the LSVM, GSVM and LGSVM models in the simulation of urban growth, probability maps were employed as the transition rules for urban cellular automata. The results show that a trade-off between local and global SVM models can enhance the performance of urban growth modelling.

Keywords: support vector machines; cellular automata; probability maps; urban development

1. Introduction

Rapid urban growth can have undesirable impacts on agricultural lands, natural landscapes and public open spaces [1]. In addition, urban planning, which influences the development of urban areas and the management of urban change [2], addresses various natural and socioeconomic factors [3–7]. Therefore, urban growth modelling can be an effective tool for investigating the results of different urban planning scenarios [8].

Over the past two decades, many studies have focused on the modelling and dynamic simulation of urban development with the help of the spatial data analysis functionality of Geographic Information Systems (GIS), as well as statistical and artificial intelligence methods and geosimulation models such as the cellular automata (CA) (e.g., [9–19]).

However, in most of these studies, the relationships between urban development and driving forces are considered stationary. Moreover, it has been shown that the assumption of spatial stationarity in analyses of spatial relationships is unrealistic, especially for processes of land use change and urban expansion [20,21].

Therefore, because of the ubiquity of spatial non-stationarity, it is inappropriate to assume constant values of model parameters over an entire study area [22]. In such cases, local models can represent spatial variations more adequately than can the global models [23].
Employing local models can improve the performance of dynamic simulations of land use change, and the definition of non-stationary transition rules in local CA models has led to improvements in urban growth predictions [24–28].

Local models were first applied in analyses of clusters [29] and hotspots [30]. Then, they were introduced into a new field of analysis encompassing the relationships between independent and explanatory variables based on geographically weighted regression (GWR) models [22,31]. Such models have been used to predict the values of spatial variables in various studies [32–35], and, in logistic form, these models been used for the production of urban growth probability maps [36–39].

Machine learning algorithms, such as support vector machine (SVM) algorithms, are not as highly interpretable as statistical models. However, researchers have employed these algorithms because they can readily adapt to complex data sets and have strong nonlinear modelling capabilities [40,41]. SVM has been applied to spatial data to produce probability maps for simulations of land use change (e.g., [42–44]).

In addition, the high level of efficiency of local learning algorithms, such as the k-nearest neighbours and radial basis function methods [45], provides a basis for the local use of other learning algorithms, such as SVMs. Therefore, the development of local SVM (LSVM) models and the associated comparisons with global models have been described in the literature.

For example, Gilardi and Bengio [40] demonstrated that local support vector regression displays a higher level of accuracy than does the global model. Considering the generalisation error, Ralaivola and d’Alché-Buc [46] presented a methodology for training SVMs known as incremental learning, which is based on the number of nearest neighbours. Yang et al. [47] proposed a weighted SVM model that was used to locally classify data based on the similarity between the training and testing data sets [48,49].

Ladicky and Torr [50] suggested a local linear SVM model that employs nonlinear manifold learning techniques. Moreover, Gu and Han [51] compared a local clustered SVM, a combination of the k-nearest neighbours and SVM algorithms and the K-means SVM algorithm with linear and nonlinear global SVM (GSVM) models. The results showed that the clustered SVM algorithm displayed higher efficiency than the other methods. In addition, Andris et al. [52] locally applied a linear SVM model to classify student enrolment in the United States.

A review of the research on LSVM models shows that this type of model has only been applied for data classification. The objective of this study is to develop an LSVM model that can be used to estimate the probabilities of urban expansion and to present a novel approach that integrates LSVM and GSVM models into a new model (the LGSVM model) to provide inputs for a CA and perform urban growth modelling.

This paper is structured as follows. First, the study area and the data used are introduced. Then, the LSVM and GSVM models are briefly presented. The method of calibrating the local and global models and the determination of the optimal bandwidth for the LSVM model are then described. Next, the results of the integrated models are presented and discussed, followed by the conclusions of the study.

2. Study Area

Tehran has grown extensively since the 1940s due to public and private investment and an increase in the population. In 1938, Tehran’s population was twice that of Mashhad (the second largest city in Iran); by 1966, it was 6.6 times larger [53].

The population migration towards Tehran has been offset by factors such as high housing costs, but increased by factors such as the low cost of suburban properties [54]. These factors have led to the development of an urban agglomeration around Tehran [55]. High rates of development have occurred in some subregions in the suburbs of the Tehran metropolitan area.
For instance, one of these subregions, located in the southwestern Tehran suburbs, has developed around the Saeedi Highway. The city of Islamshahr has become the most important and largest urban centre in this subregion.

The study area investigated in this research covers the major parts of this urban region, which encompasses an area of 177 km$^2$ and extends from 35°27′24″ N to 35°36′05″ N and 51°04′47″ E to 51°18′12″ E. The study area includes cities and towns such as Islamshahr, Golestan, Nasimshahr, Salehieh, Nasirshahr, Vavan and Qaemieh (Figure 1).

3. Materials and Methods

3.1. Data and Variables

Land use/cover data from the study area were extracted from SPOT 2, 3 and 4 satellite images captured in 1992, 1996 and 2002, respectively. Table 1 presents the image specifications, and Figure 2 shows the SPOT satellite images.

Table 1. Specifications of the SPOT images.

<table>
<thead>
<tr>
<th>Satellite Image</th>
<th>Acquisition Date</th>
<th>Spatial Resolution (Meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPOT 2</td>
<td>25 June 1992</td>
<td>20</td>
</tr>
<tr>
<td>SPOT 3</td>
<td>19 June 1996</td>
<td>20</td>
</tr>
<tr>
<td>SPOT 4</td>
<td>13 July 2002</td>
<td>10</td>
</tr>
</tbody>
</table>

Because of the importance of data quality in local modelling, the land use maps were produced with maximum possible accuracy at the pixel level. Therefore, land use classes were manually extracted based on the visual interpretation of SPOT images. Then, the results were validated using high-resolution and historical images from Google Earth™ and 1:25,000 maps from the Iran National Cartographic Center (NCC). Finally, all suspicious cases were further investigated via field visits.
Additional road, watercourse and railway data were collected from 1:25,000 maps from the NCC. Figure 3 shows the land use map of the study area in 1992 and the urban areas that developed during the periods of 1992–1996 and 1996–2002.

Using these data sources, the distance to urban areas, indicator of bare land/agricultural land use, a combined variable of distance from industrial areas and the industrial land use density (known as the density-distance to industrial land use), and the distances to villages and roads were computed and considered predictive variables. The bare land/agricultural land use indicator is a binary variable that takes a value of 1 for bare land and a value of 0 for agricultural land use. The value of the density-distance to industrial land use increases at low distances and high densities and decreases at long distances and low densities. This combined density and distance metric is intended to highlight locations that are close to large industrial areas or sets of industrial areas from those that are near small industrial areas. The road, railroad and watercourse layers were converted to rasters and considered constraints to urban growth.
After standardisation to the 0 to 1 range using an ascending linear function (Figure 4), all the predictive variables were used to calibrate the SVM models. A binary variable for developed/undeveloped urban areas was also employed (developed = 1, undeveloped = 0).

In the present study, 2000 training samples for each of the developed (with a value of 1) and undeveloped (with a value of zero) areas were randomly selected for the calibration of the local and global models.

**Figure 4.** Standardised factors used to predict land use changes: (a) distance to urban built-up areas, (b) density-distance to industrial land use, (c) distance to villages, (d) distance to roads, and (e) bare land/agricultural land use indicator.
3.2. Methodology

3.2.1. Research Outline

Using the training data, the cross-validation method, and the developed/undeveloped binary variable as the response variable, a local linear SVM model was calibrated based on data from the period of 1992–1996. The optimal bandwidth considering the number of nearest neighbours was extracted, and the probability map of urban land use development was determined. The accuracy of the resulting probability map was compared with that obtained using the nonlinear GSVM model, which was calibrated using the ten-fold cross-validation method. The final model was ultimately obtained as a weighted combination of the local and global probabilities. The accuracy of the resulting probability map was then compared with the accuracies obtained using the local and global SVM models in the periods of 1992–1996 (calibration phase) and 1996-2002 (validation phase). Finally, the probability maps were used to establish the transition rules of cellular automata for the simulation of urban development (hereafter deemed the Urban-CA). Then, the performances of the GSVM, LSVM and LGSVM models for urban growth simulation were evaluated for the periods of 1992–1996 and 1996–2002.

3.2.2. Global Support Vector Machines

In a classification problem, the goal is to separate two classes from one another using a function with a small generalised error [56]. Unlike neural networks, SVMs attempt to both separate two classes from each other and to maximise the margin between them [57]. SVMs search for the hyperplane that simultaneously maximises the margin and reduces the classification error [58]. However, this process consequently reduces the VC dimension [59] and the generalised errors [60]. Thus, this approach leads to the construction of a hyperplane with a soft margin. The trade-offs between the margin and the classification error are controlled by a constant $C$, where $0 < C < \infty$. Finally, the linear classification problem is transformed into the following optimisation problem:

$$
\text{max. } \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \\
\text{subject to: } \sum_{i=1}^{n} \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C
$$

(1)

where $x$ is the vector in the input space, $y$ is the class value and $\alpha$ is the Lagrange multiplier.

If it is not possible to classify a given data set using a linear function, the classification can be performed using nonlinear transfer functions within high-dimension space. As a result, data from space $X$ are mapped to space $H$ or the feature space to solve the nonlinear classification problem. The purpose of this process is to convert a nonlinear space into a linear space in $H$ [61]. Computations in the feature space can be a computationally intensive because such spaces may have many dimensions; thus, kernel functions, which are known as kernel tricks [62], are applied herein. Given the possibility of the internal multiplication of $(x_i^T x_j)$ in the feature space, the optimisation problem is converted to the following form using a kernel function $K(x_i, x_j)$.

$$
\text{max } \left( \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right) \\
\text{subject to: } \sum_{i=1}^{n} \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C ,
$$

(2)

Although there are various types of kernel functions, the radial basis function (RBF) is one of the most widely used [63]. This function is defined as follows:

$$
K(x_i, x_j) = e^{-\frac{(x_i - x_j)^2}{2\sigma^2}}
$$

(3)
where $\sigma$ is the width of the kernel.

Although the SVM model offers the abovementioned advantages, it is strict and unable to provide probabilistic predictions. Therefore, Platt [64] proposed a method for obtaining the optimal values of $A$ and $B$ using the following sigmoid probability function:

$$P(y = 1|f(x)) = \frac{1}{1 + \exp(Af(x) + B)} \quad (4)$$

where $P$ is the probability of variable $y$ equalling 1 and $f(x)$ is the decision function. The resulting probability is highly efficient for modelling land use changes, especially urban land development, for which developed/undeveloped binary variables are used.

The k-fold cross-validation method is commonly used to calibrate nonlinear GSVM models, prevent overtraining and accelerate calibration. The calibration process is implemented using $n$ training samples processed $k$ times. In each repetition, the data are randomly divided into $k$ categories, such that $k - 1$ categories are classified as training data and the remaining categories are classified as test data. Ten-fold cross-validation is typically used to calibrate SVM models [65,66].

3.2.3. Local Support Vector Machines

The LSVM approach is mainly based on using a limited quantity of training data, instead of all available data, for the calibration of SVM models. In this approach, multiple SVM models cover the entire data space of the problem, in contrast to GSVM models.

The training samples in an LSVM can be used in the calibration process with varying levels of importance [47]. In geographical problems, this importance can be defined by weights that are assigned to training samples based on their spatial similarity to the focal location of calibration. As a result, these types of SVM problems can be defined as follows [49]:

$$\begin{align*}
\text{max.} & \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \\
\text{subject to :} & \quad 0 \leq \alpha_i \leq W_i C, \quad \sum_{i=1}^{n} \alpha_i y_i = 0
\end{align*} \quad (5)$$

where $W_i$ is the weight assigned to training sample $i$.

As a result, data of higher importance will have a larger negative influence on the model in the case of an incorrect prediction. Thus, in this study, the weighted LSVM models are calibrated and validated considering the spatial nature of the urban growth phenomenon and the associated similarity to the logic behind the development of the local models (e.g., GWR). The bi-square weighting function, which is frequently used to assign weights to training points in GWR models [38], is employed in this study, as shown in Equation (6):

$$w_i = \begin{cases} 
1 - \left( \frac{d_i}{b} \right)^2 & \text{if } d_i < b \\
0 & \text{otherwise}
\end{cases} \quad (6)$$

where $d_i$ is the distance of training sample $i$ from the calibration focal point and $b$ is the bandwidth.

In all classification techniques, it can be assumed that the decision boundary is nearly linear within a small area, such that the data can be linearly classified [50]. In contrast, the determination of the type of kernel function, particularly the calibration of the associated parameters, represents a challenge in the application of nonlinear SVM models [67], especially an LSVM model, which requires multiple calibrations. Therefore, in the present study, the LSVM model is linearly implemented in accordance with the optimisation equations.
In this research, based on the bandwidth used, a local linear SVM is separately calibrated for the neighbourhood of each training sample, and this focal training sample does not participate in the calibration process (i.e., in finding the optimum value of C) (Figure 5).

The calibration aims to achieve the lowest possible error for each training sample, and the process is repeated for all the sample points. One advantage of this method is that it reduces the calibration time required for the local model. Because k-fold cross-validation is not applied for each LSVM model, this method can facilitate the computation of the cross-validation error, which is an appropriate criterion for determining the optimal bandwidth.

Determining the Optimal Bandwidth

After calibration of the LSVMs with a specified bandwidth, the optimal bandwidth can be determined based on the minimum value of the cross-validation error. The cross-validation method based on the focal training sample has an advantage in that by reducing the bandwidth, it prevents the residual sum of squares (RSS) from tending to zero. This method is used to determine the bandwidth of local regression [68]. The total error is obtained using Equation (7) through cross-validation based on the focal training samples:

$$CV_b = \sum_{i=1}^{n} (y_i - p_i^{\neq x_i}(b))^2$$

where $CV_b$ refers to the RSS obtained through cross-validation at bandwidth $b$, $y_i$ is the value of $y$ (0 or 1) at sample point $i$, and $p_i^{\neq x_i}(b)$ is to the estimated posterior probability of $y_i = 1$ for focal training sample $i$ at bandwidth $b$, as estimated without using the predictive data for sample $i$. In this study, an adaptive bandwidth based on the number of nearest neighbours for the focal training location is used due to the non-uniform distribution of the sample points within the study area [38].

Estimation

After determining the optimal bandwidth, the probability of urban development was estimated for all locations (prediction points) in the study area for the periods of 1992–1996 and 1996–2002. In addition to being used as an input for simulations of urban development with the CA, the output map can be used to validate the local and global models and determine their generalisation capabilities.
In the probability estimation, each prediction point is empirically near several training samples and, hence, several LSVM models; thus, the probabilities can be calculated by the LSVMs in a specific neighbourhood. The final probability for the prediction point is then estimated as the weighted average of the probability values produced by the neighbouring LSVMs (Equation (8)):

\[
p_{i}^{\text{prediction}} = \frac{\sum_{j=1}^{n} W_{ij} P_{j}(x_i)}{\sum_{j=1}^{n} W_{ij}}, \quad W_{ij} = \frac{1}{d_{ij}^2}
\]

where \(p_{i}^{\text{prediction}}\) is the final estimated probability of prediction point \(i\), \(n\) is the number of neighbouring samples, \(W_{ij}\) is the weight assigned to the calculated probability of the LSVM calibrated at training point \(j\) (\(P_{j}\)) based on the squared inverse distances between the prediction and training points, \(d_{ij}\) is the distance between prediction point \(i\) and training point \(j\), and \(x_i\) is the vector of predictive variables at point \(i\) (Figure 6).

**Figure 6.** Calculation of the urban development probability at prediction point \(i\) using the neighbour-calibrated LSVM models.

### 3.2.4. Integration of the Local and Global Models

To determine whether the integration of the local and global models (LGSVM model) leads to increased accuracy, the probability of urban development (\(P_{\text{LGSVM}}\)) is calculated based on the weighted average of the local (\(P_{\text{LSVM}}\)) and global (\(P_{\text{GSVM}}\)) model probabilities:

\[
P_{\text{LGSVM}} = W_{\text{Local}} \times P_{\text{LSVM}} + W_{\text{Global}} \times P_{\text{GSVM}}
\]

where \(W_{\text{Local}}\) and \(W_{\text{Global}}\) are the weights of the local and global models, respectively, and \(W_{\text{Local}} + W_{\text{Global}} = 1\). Then, when \(W_{\text{Local}} = 0\) or \(W_{\text{Global}} = 0\), the model is converted to a GSVM or LSVM, respectively.
3.2.5. Urban Cellular Automata

The efficiency of calculating the probabilities based on the local, global and integrated models was evaluated, and the probabilities were used as transition rules to simulate urban development using the Urban-CA. Transition rules are considered the most important components of CAs [69–71].

In the CA model developed in this research, the final urban development probability of cell $c$, i.e., $P_c$, in each iteration is calculated using Equation (10) [27,72]:

$$P_c = P_{c}^{LGSVM} \times \Omega_c \times \gamma \times S_c$$

(10)

where $\Omega_c$ is the number of developed cells in the extended Moore’s neighbourhood [73] of cell $c$ divided by 24; $\gamma$ is the stochastic disturbance term, in which $\alpha$ controls the dispersion of urban development; and $r$ is a random value between 0 and 1 (Equation (11)). The formula for $\gamma$ is given as follows [74].

$$\gamma = 1 + (-\ln(r))^\alpha$$

(11)

The value of $S_c$ for cell $c$ is 0 when the cell is considered a constraint and 1 otherwise. After optimising the $\alpha$ value for the global, local and integrated models, the performance of the Urban-CA model was evaluated for the periods of 1992–1996 and 1996–2002. In the following sections of this paper, the CA model developed based on the integrated probabilities is referred to as the LGSVM-CA, whereas the CA models developed based on the local and global probabilities are referred to as the LSVM-CA and GSVM-CA, respectively.

3.2.6. Validation

Area Under the ROC curve

The area under the receiver operating characteristic (ROC) curve is a suitable criterion for validating the probability of an event estimated by a model via a comparison with its real value of either one or zero [75]. Higher values of the area under the ROC curve (AUC) represent stronger relationships between the real values and estimated probabilities [76]. A non-parametric method [77] is employed here to assess the standard errors of the AUCs and the significance of the differences between them.

Kappa Coefficient

In this study, the kappa coefficient was applied to compare the simulated maps generated by the CA models based on different transition rules. An advantage of the kappa coefficient is that it accounts for correct predictions by random chance [43].

4. Results

Figure 7 illustrates the calibration results obtained with the LSVM model using the cross-validation method and based on the focal training sample. The X-axis represents the bandwidth considering the number of nearest neighbours, and the Y-axis shows the cross-validation error. Since the optimal bandwidth is based on the minimum value of the cross-validation error, the optimal bandwidth for the LSVM model is equivalent to 40 nearest neighbours.

The average RSS values obtained from the calibration of the GSVM model using the ten-fold cross-validation method with $C$ and $\sigma$ near the optimal values ($C = 1$ and $\sigma = -4$) are shown in Table 2. Additionally, Figure 8 shows the probability maps obtained from the calibrated LSVM and GSVM models.
probability maps obtained from the global, local and integrated models. High AUC values indicate the GSVM, LSVM and LGSVM models are low, the non-parametric statistical test [77] shows that the high reliabilities for the probability maps. However, although the differences in AUC values among ISPRS Int. J. Geo-Inf. 2018

disturbance term is 0.01 for both the LSVM-CA and LGSVM-CA and 0.3 for the GSVM-CA. These results indicate that the accuracy of the LSVM-CA is greater than that of the GSVM-CA and that the integration of the local and global models increases the accuracy, where

weight. The kappa value of the optimum LGSVM-CA model increases by 18.1% and 5.3% relative to the GSVM, LSVM and LGSVM models at the 0.001 significance level (Table 4; period: 1992–1996).

In this study, the kappa coefficient was applied to compare the simulated maps generated by the CA models based on different transition rules. An advantage of the kappa coefficient is that it accounts for correct predictions by random chance [43].

Figure 7 illustrates the calibration results obtained with the LSVM model using the cross-validation method and based on the focal training sample. The optimum bandwidth of the LSVM model resulting from the cross-validation error is shown in Table 2.

Table 2. Average RSS values of the GSVM calculated by the ten-fold cross-validation method with \( \sigma \) and \( C \) near the optimal values (\( C = 1, \sigma = -4 \)).

<table>
<thead>
<tr>
<th>( C )</th>
<th>( \sigma )</th>
<th>−6</th>
<th>−5</th>
<th>−4</th>
<th>−3</th>
<th>−2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.125</td>
<td>21.88</td>
<td>22.10</td>
<td>22.01</td>
<td>23.19</td>
<td>25.90</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>21.89</td>
<td>18.97</td>
<td>18.58</td>
<td>21.22</td>
<td>24.95</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>22.42</td>
<td>15.74</td>
<td>15.93</td>
<td>19.36</td>
<td>23.47</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>27.95</td>
<td>15.23</td>
<td>14.78</td>
<td>18.19</td>
<td>22.12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>61.18</td>
<td>16.52</td>
<td>14.86</td>
<td>17.20</td>
<td>20.70</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>98.36</td>
<td>18.20</td>
<td>15.49</td>
<td>16.37</td>
<td>19.58</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>71.07</td>
<td>25.06</td>
<td>15.92</td>
<td>16.29</td>
<td>19.17</td>
<td></td>
</tr>
</tbody>
</table>

Figure 7. Optimum bandwidth of the LSVM model resulting from the cross-validation error.

Table 3 presents the relationship between \( W_{\text{Local}} \) and the AUCs calculated from the ROCs of probability maps obtained from the global, local and integrated models. High AUC values indicate high reliabilities for the probability maps. However, although the differences in AUC values among the GSVM, LSVM and LGSVM models are low, the non-parametric statistical test [77] shows that the

Figure 8. Map of urban development probabilities based on the (a) LSVM model and (b) GSVM model.

Table 3 presents the relationship between \( W_{\text{Local}} \) and the AUCs calculated from the ROCs of probability maps obtained from the global, local and integrated models. High AUC values indicate high reliabilities for the probability maps. However, although the differences in AUC values among the GSVM, LSVM and LGSVM models are low, the non-parametric statistical test [77] shows that the...
AUC of the optimum LGSVM model (i.e., $W_{\text{Local}} = 0.6$) is significantly higher than the AUCs of the GSVM and LSVM models at the 0.001 significance level (Table 4; period: 1992–1996).

### Table 3. Relationship between $W_{\text{Local}}$, the AUC and the kappa coefficient.

<table>
<thead>
<tr>
<th>GSVM ($W_{\text{Local}} = 0$)</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.979</td>
<td>0.987</td>
<td>0.988</td>
<td>0.989</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.989</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.489</td>
<td>0.578</td>
<td>0.619</td>
<td>0.642</td>
<td>0.660</td>
<td>0.667</td>
<td>0.670</td>
<td>0.667</td>
<td>0.660</td>
</tr>
</tbody>
</table>

The mean values of kappa obtained from 10 simulations of urban development using the Urban-CA for the period of 1992–1996 are highlighted in Table 3. The optimum value of $\alpha$ in the stochastic disturbance term is 0.01 for both the LSVM-CA and LGSVM-CA and 0.3 for the GSVM-CA. These results indicate that the accuracy of the LSVM-CA is greater than that of the GSVM-CA and that the integration of the local and global models increases the accuracy, where $W_{\text{Local}} = 0.6$ is the optimal weight. The kappa value of the optimum LGSVM-CA model increases by 18.1% and 5.3% relative to that of the GSVM-CA and LSVM-CA, respectively.


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in AUCs</td>
<td>0.0104</td>
<td>0.00403</td>
</tr>
<tr>
<td>Standard error (DeLong et al. 1988)</td>
<td>0.000238</td>
<td>0.000483</td>
</tr>
<tr>
<td>z statistic</td>
<td>43.513</td>
<td>8.33</td>
</tr>
<tr>
<td>Significance level</td>
<td>p-value &lt; 0.001</td>
<td>p-value &lt; 0.001</td>
</tr>
</tbody>
</table>

Figure 9a shows the probabilities obtained from the LGSVM model, and Figure 9b illustrates the prediction of urban development from 1992–1996 using the LGSVM-CA. The accuracies of the urban development probabilities and the CAs for the validation period of 1996–2002 are presented in Table 5. These predictions are based on the optimised parameters using the calibration period of 1992–1996.

### Table 5. The AUC and kappa values of the probability maps and CA models for the period of 1996–2002.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>Model</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSVM</td>
<td>0.896</td>
<td>GSVM-CA</td>
<td>0.483</td>
</tr>
<tr>
<td>LSVM</td>
<td>0.939</td>
<td>LSVM-CA</td>
<td>0.493</td>
</tr>
<tr>
<td>LGSVM</td>
<td>0.943</td>
<td>LGSVM-CA</td>
<td>0.512</td>
</tr>
</tbody>
</table>

Figure 9. (a) LGSVM probability map and (b) urban growth prediction simulated by the LGSVM-CA.
The results show that the AUC of the LSVM probability map is higher than that of the GSVM probability map and that the AUC of the LGSVM probability map is higher than the AUCs of the other two maps, indicating higher accuracy. The differences in AUC values between the LGSVM and the GSVM and LSVM models are significant at the 0.001 significance level (Table 4, period: 1996–2002).

In addition, similar to the results for the calibration period, the accuracy of the urban development prediction simulated by the LGSVM-CA is higher than the accuracies obtained with the global and local models. The kappa value of predicted urban growth using the LGSVM-CA is 1.9% to 2.9% higher than the values obtained using the LSVM-CA and GSVM-CA.

Table 5. The AUC and kappa values of the probability maps and CA models for the period of 1996–2002.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>Model</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSVM</td>
<td>0.896</td>
<td>GSVM-CA</td>
<td>0.483</td>
</tr>
<tr>
<td>LSVM</td>
<td>0.939</td>
<td>LSVM-CA</td>
<td>0.493</td>
</tr>
<tr>
<td>LGSVM</td>
<td>0.943</td>
<td>LGSVM-CA</td>
<td>0.512</td>
</tr>
</tbody>
</table>

5. Discussion

The local and global probabilities indicate that the deficiencies of one model can generally be offset by the advantages of the other model.

For example, Figure 10a shows the map of differences between the local and global probabilities, i.e., $P_{Local} - P_{Global}$, around the cities of Golestan and Salehieh for the period of 1992–1996. Positive values indicate locations at which the local probability is greater than the global probability, and negative values indicate the opposite condition. Figure 10b shows the land use in this area. At location 1, the high negative values indicate that the global probabilities are much higher than the local values despite the absence of a developed urban area at this location from 1992–1996. Thus, the LSVM model can increase the accuracy of urban prediction at this location. As illustrated in Figure 10b, the land use at location 2 is agricultural, with no developed areas. Moreover, at this location, the probability of urban development derived from the GSVM model is low, whereas the LSVM model incorrectly estimates high probabilities. Notably, the influence of agricultural land use on urban development was considered in the GSVM model but not in the LSVM model.

Figure 10. (a) Difference between the local and global probabilities and (b) the land use map.
Therefore, both the LSVM and GSVM models may yield low or high predictions at different locations. As a result, the integration of the LSVM and GSVM models can lead to higher accuracies than can either model alone.

In addition, a high correlation can be observed between the probability of urban development estimated by the LSVM and GSVM models and the density of urban development. To verify this finding, the kernel density function (KDF) [78] was executed in GIS. First, developed urban cells were converted to points. Then, the KDF was implemented for these points. As shown in Figure 11, in the KDF process, the number of points within the specified radius R is determined, and the points closer to the central point with respect to distance (r) will receive higher weights.

Figure 11. Schematic illustration of how the kernel density function works.

Figure 12 shows the urban development density maps produced by KDE for the periods of 1992–1996 and 1996–2002. Additionally, Table 6 presents the correlation between the density of urban development and calculated probabilities obtained by the SVM models from 1992–1996 and 1996–2002. As expected, the correlation values in the first period are relatively high and suggest that when the density of urban development is high, the probabilities estimated by the SVM models are also high.

However, Table 6 shows that the accuracy of the urban growth prediction decreases in the 1996–2002 period compared to that in the 1992–1996 period. This result is likely associated with the relationship between the density and probability. As Table 6 shows, the correlation between the density of urban development in this period and the estimated probability obtained by the GSVM, LSVM and LGSVM decreases by 14.5%, 12.5% and 11.9%, respectively. Thus, the location of urban development in the new period is not necessarily in the vicinity of the recently developed areas in the previous period, and differences exist in the density of urban development at nearby locations in the two periods.

In addition, the results of the correlation tests between probability maps of the two periods show that there is a relationship between probabilities of the calibration phase and validation phase. The correlation values are 0.718 and 0.698 for LSVM and GSVM models respectively.

In this situation, because of the temporal non-stationarity that exists for urban development at some locations, we can expect a decrease in the performance of urban growth simulations.


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{\text{GSVM}} )</td>
<td>0.712(^\uparrow)</td>
<td>0.567(^\downarrow)</td>
</tr>
<tr>
<td>( P_{\text{LSVM}} )</td>
<td>0.787(^\uparrow)</td>
<td>0.662(^\downarrow)</td>
</tr>
<tr>
<td>( P_{\text{LGSVM}} )</td>
<td>0.792(^\uparrow)</td>
<td>0.673(^\downarrow)</td>
</tr>
</tbody>
</table>

Figure 13 shows the difference between the density of urban development in the periods of 1996–2002 and 1992–1996 (i.e., \( \text{KDE}_{1996–2002} - \text{KDE}_{1992–1996} \)) near the cities of Golestan and Salehieh. High positive and negative values reflect the temporal non-stationarity of urban development, which may be a reason for the decrease in the prediction accuracy.

Figure 13. Differences in the density of urban development in the periods of 1996–2002 and 1992–1996 near the cities of Golestan and Salehieh.
The results of this study indicate that like similar studies, the definitions of non-stationary transition rules can lead to improvements in simulation performance. Some improvements can be made by implementing a zoning approach [24,26]; however, in this study, a local SVM model was developed to estimate the probability of urban development over a continuous surface to avoid discretisation issues. In addition, the improvement in simulation performance attained through integrating the local and global SVMs is another unique aspect of this study.

In this research, the number of explanatory variables used in urban development modelling was limited to only five based on the availability of local data. Generally, as additional local and historical socioeconomic data, such as population or land price data, become available, the results of urban growth modelling will be more similar to reality.

6. Conclusions

In this research, a local SVM model was designed to generate probability maps of urban development in a subregion southwest of the metropolitan area of Tehran. The model was calibrated using local training data, and the optimum bandwidth was determined by the cross-validation method. The output is considered the basis for determining the optimum bandwidth.

For comparison, a nonlinear GSVM model was also calibrated by determining the optimal values of the parameters $C$ and $\sigma$ using the ten-fold cross-validation method. Closer investigation of the behaviours of the global and local models showed that both models have advantages and limitations in estimating the development probability. Therefore, an integrated model developed from a linear combination of the local and global SVM probabilities was also evaluated.

A comparison of the urban development probability maps produced from the local and global SVM models and the integrated models based on the AUCs for the periods of 1992–1996 and 1996–2002 revealed a higher accuracy for the LSVM model than the GSVM. Additionally, the integrated model exhibited a higher accuracy than of either of the other two models, and the LGSVM-CA outperformed the other CAs developed from local or global SVMs in the prediction of urban development for the periods of 1992–1996 and 1996–2002. The results showed that considering the temporal stationarity of urban development based on the location and area can improve simulations of urban growth.

The important contributions of this research are the development of a local SVM model for calculating the urban development probability and the integration of this model with a global SVM model to improve the accuracy of urban growth prediction. The results of the present study suggest that the integration of the LSVM and GSVM models can improve the predictive accuracy compared to that obtained using either the LSVM or GSVM model alone.

Our future work will focus on comparing the efficiencies of the LSVM and LGSVM models with the efficiency of the well-known logistic form of the GWR model in simulating land use changes and in urban development prediction.

Author Contributions: Investigation: Babak Mirbagheri; Methodology: Babak Mirbagheri and Abbas Alimohammadi; Supervision: Abbas Alimohammadi; Writing (original draft): Babak Mirbagheri; Writing (review & editing): Abbas Alimohammadi

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References
19. Yao, Y.; Li, J.; Zhang, X.; Duan, P.; Li, S.; Xu, Q. Investigation on the expansion of urban construction land use based on the CART-CA model. *ISPRS Int. J. Geo-Inf.* 2017, 6, 149. [CrossRef]  
47. Yang, X.; Song, Q.; Wang, Y. A weighted support vector machine for data classification. *Int. J. Pattern Recogn.* 2007, 21, 961–976. [CrossRef]


52. Andris, C.; Cowen, D.; Wittenbach, J. Support vector machine for spatial variation. Trans. GIS 2013, 17, 41–61. [CrossRef]


55. Ghamami, M. Urban Agglomeration of Tehran and Around Cities; Center of Studies and Researchers of Iran’s Architecture and Urban Planning; Ministry of Roads and Urban Development: Tehran, Iran, 2000.


