

GIS Cellular Automata Using Artificial Neural Network for Land Use Change Simulation of Lagos, Nigeria

Onuwa Okwuashi¹, Mfon Isong¹, Etim Eyo², Aniekan Eyoh¹, Okey Nwanekezie³, Dupe Nihinlola Olayinka⁴, Daniel Okon Udoudo⁴ & Beulah Ofem⁵

¹ Department of Geoinformatics & Surveying, Faculty of Environmental Studies, University of Uyo, Nigeria

² School of Civil Engineering & Geosciences, Newcastle University, Newcastle upon Tyne, United Kingdom

³ Department of Estate Management, Faculty of Environmental Studies, University of Uyo, Uyo, Nigeria

⁴ Department of Surveying & Geoinformatics, Faculty of Engineering, University of Lagos, Lagos, Nigeria

⁵ Department of Urban and Regional Planning, Faculty of Environmental Studies, University of Uyo, Nigeria

Correspondence: Onuwa Okwuashi, Department of Geoinformatics & Surveying, Faculty of Environmental Studies, University of Uyo, Uyo, Nigeria. Tel: 234-813-044-4355. E-mail: onuwaokwuashi@yahoo.com

Received: February 27, 2012 Accepted: March 20, 2012 Online Published: May 18, 2012

doi:10.5539/jgg.v4n2p94

URL: <http://dx.doi.org/10.5539/jgg.v4n2p94>

Abstract

Tremendous land use change has occurred in Lagos in recent times. Modelling urban systems now extends beyond the use of geographic information systems models. This research therefore presents a loose coupling of geographic information systems and artificial neural network for simulating land use change in Lagos. The experiment is based on three land use epochs of Lagos: 1963-1978, 1978-1984, and 1984-2000. Twelve salient land use explanatory variables (distance to water, distance to residential structures, distance to industrial and commercial centres, distance to major roads, distance to railway, distance to Lagos Island, distance to international airport, distance to international seaport, distance to University of Lagos, distance to Lagos State University, income potential, and population potential) are used for the simulation. Using the Kappa statistic, the result of the simulation in terms of the order of best-fit of the reference data is: 1978-1984, 1984-2000, and 1963-1978. An evaluation of the simulation using the receiver operating characteristics corroborates the Kappa estimates. A non black-box experiment using a one-neuron neural network to assess the performance of the spatial independent variables used for the simulation indicates that for all three epochs distance to residential structures has the highest impact in the simulation while population potential has the lowest impact.

Keywords: land use change, GIS, artificial neural network

1. Introduction

Lagos (see Figure 1) is usually described as a laissez-faire urban society (Okwuashi, 2011); and the problem of urban sprawl in Lagos is generally viewed by most Nigerians as an intractable problem (Abiodun, 1974; Gandy, 2006).

Lagos has undergone rapid urban expansion in recent times (see Figure 2); therefore modelling an unregulated complex urban environment like Lagos may be unyielding without employing robust predictive tools that can realistically model their complexity, dynamism, and growth (Barredo et al., 2004). This research has adopted the Cellular Automata (CA) model due to their simplicity, dynamic properties, and inventive bottom-up approach (Clarke & Gaydos, 1998). Unlike Geographic Information Systems (GIS) based models, CA models are not subject to linear regression assumptions (Okwuashi, 2011). A major advantage of CA models is their compatibility with remote sensing and GIS (Torrens & O'Sullivan, 2001). Previous researches have shown that coupling GIS and CA models has helped improve dynamic spatial modelling (Park & Wagner, 1997). The objective of this research therefore is to explore the loose coupling of the GIS and Artificial Neural Network (ANN) based CA model for simulating land use change in Lagos, Nigeria.

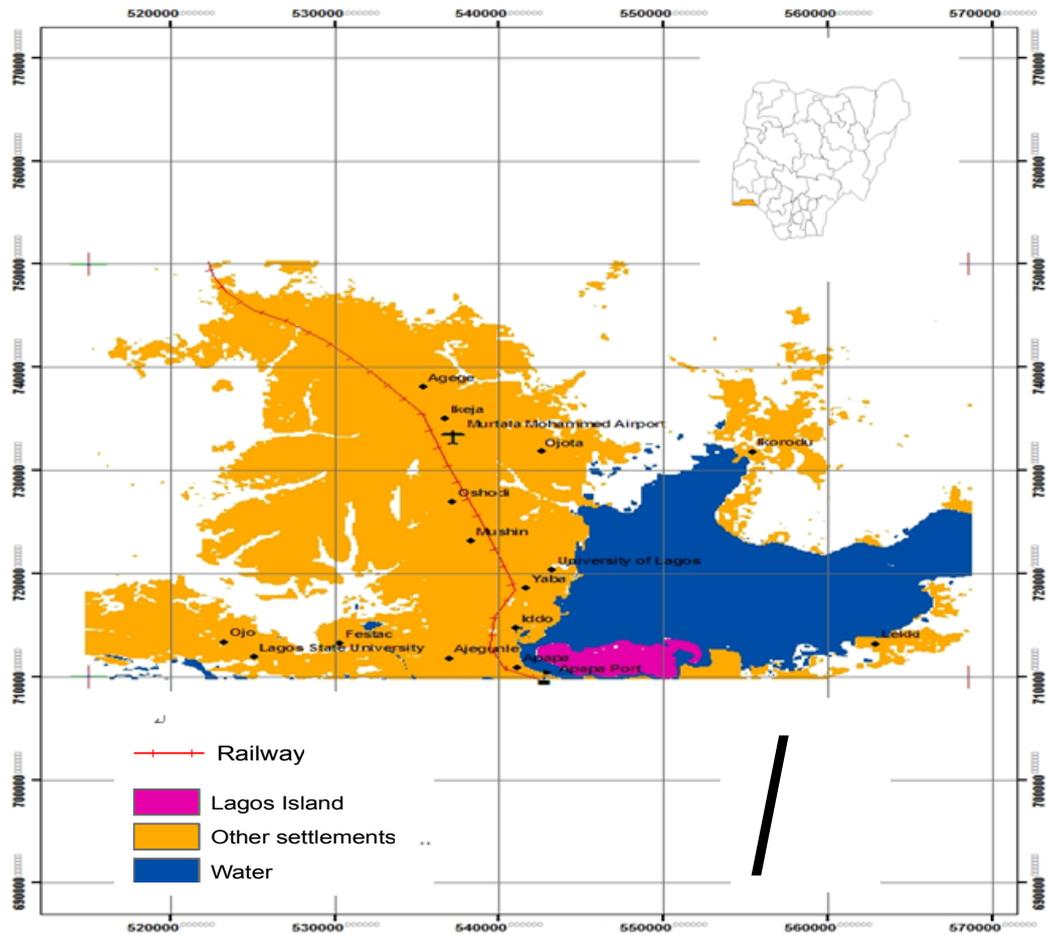


Figure 1. Lagos in relation to Nigeria

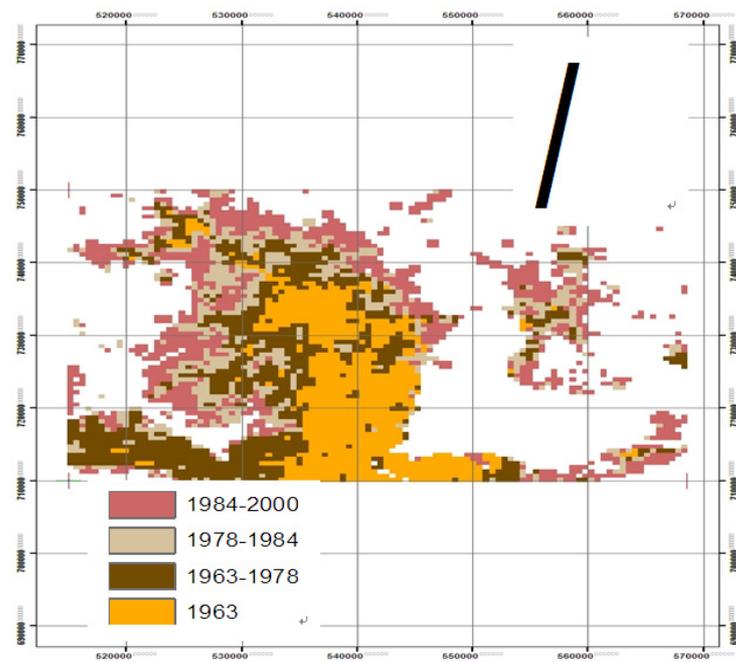


Figure 2. Land use of Lagos between 1963 and 2000

2. ANN Based CA Calibration

The ANN consists of three main components: the input layer, hidden layer, and the output layer (Figure 3). The hidden layer is the engine room of the neural network; it consists of n neurons ($n = 1, 2, 3, \dots$). The output layer consists of just a single neuron (Almeida et al., 2008). Basically a signal from neuron i of the first input layer of a cell x , at time t received by a neuron j of the hidden layer can be expressed as,

$$net_j(x,t) = \sum_j W_{i,j} S'_i(x,t) \tag{1}$$

where $S'_i(x,t)$ denotes the site attributes given by variable (neuron) i ; $W_{i,j}$ is the weight of the input from neuron i to neuron j ; $net_j(x,t)$ is the signal received for neuron j of cell x at time t .

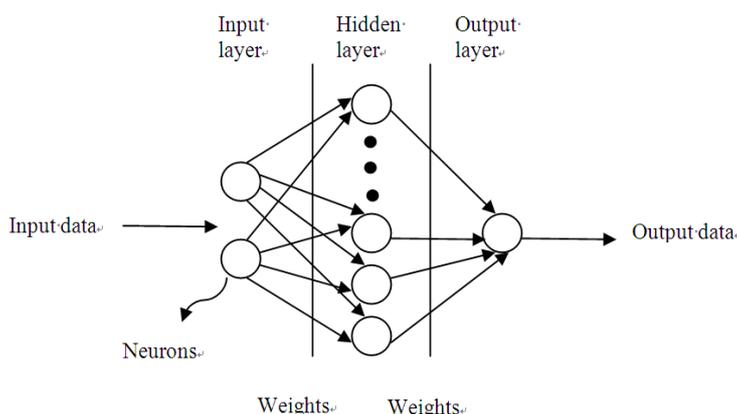


Figure 3. Generic example of an artificial neural network structure with a single output neuron

The activation of the hidden layer of the signal is,

$$\frac{1}{1 + e^{-net_j(x,t)}} \tag{2}$$

The probability can be expressed as,

$$P(x,t) = \sum_j W_j \frac{1}{1 + e^{-net_j(x,t)}} \tag{3}$$

where, P is the development probability (see Li & Yeh, 2002).

By introducing the Moore neighbourhood function $\Omega_{3 \times 3}$ (Wu, 2002), a coefficient Q , constraints contributions $cons_{ij}$, and a stochastic function $1 + (-\ln \gamma)^\alpha$ (γ is a uniform random variable; while α controls the magnitude of the perturbation) (White & Engelen, 1993), equation 3 can be revised to derive the final development probability (Okwuashi, 2011),

$$P'_{ij} = Q * \left(\sum_{i=1} W_j \frac{1}{1 + e^{-net_j(x,t)}} \right) * (1 + (-\ln \gamma)^\alpha) * \Omega_{3 \times 3}^{t-1} * \prod_{i=1}^m cons_{ij} \tag{4}$$

A threshold probability value (ψ) is set as a benchmark for determining undeveloped cells that are eligible to transit to developed cells:

$$\begin{cases} P'_{ij} \geq \psi & \text{developed} \\ \text{Otherwise} & \text{undeveloped} \end{cases} \tag{5}$$

Q can be used to regulate the value of P'_{ij} with respect to ψ ; in order to either decrease or increase the number of iterations required for the simulation.

3. Application

3.1 Data

The land use data of Lagos consist of remotely sensed Landsat Thematic Mapper images, acquired in 1978, 1984, and 2000 respectively; and an analogue base map acquired in 1963. The Landsat images were classified with the k -means algorithm using the MATLAB software. The analogue map was processed with ArcGIS. The analogue and remote sensing data were geo-referenced to ensure both data were in the same coordinate system. Twelve salient land use independent variables were used for the experiment. They were grouped into two categories: (i) proximity variables: distance to water, distance to residential structures, distance to industrial and commercial centres, distance to major roads, distance to railway, distance to Lagos Island, distance to international airport (1984-2000 only), distance to international seaport, distance to University of Lagos, distance to Lagos State University (1984-2000 only); and (ii) weighted variables: income potential and population potential. The proximity variables were extracted in ArcGIS while the weighted variables were extracted in MATLAB.

3.2 Modelling

The ANN experiment was implemented in MATLAB with the method of backpropagation (Rumelhart et al., 1986) using a two-layer feed-forward neural network. A backpropagation network has an input layer, an output layer, and one or two hidden layers; however, there is no limit to the number of hidden layers (Anderson, 1995; Chauvin & Rumelhart, 1995). It is very important to first choose the *random seed number* and the required *number of neurons* in the hidden layer. A neural network is initialised with initial weights; hence different results are always obtained every time the ANN model is run. To ensure the results remain the same at every run of the neural network the random seed number must be set. The random seed number is an arbitrary constant chosen by trial-and-error. After the random seed number is set, the ANN neurons then remain the only parameter that can be adjusted to vary the simulation results of the ANN. The training of the neural network is simply the adjustment of the number of neurons in the hidden layer in order to minimise the training error. The training error is the discrepancy between the predicted and the actual value. The adjustment of the number of neurons is sustained until a training error that falls below a pre-determined threshold is found (Wang, 1994).

The experiment was based on three epochs of urban change: 1963-1978, 1978-1984, and 1984-2000. Based on the stratified random sampling, one thousand training points were selected for each of periods 1963-1978, 1978-1984, and 1984-2000. The modelling was based on the k -fold cross-validation procedure (where $k=10$). Since there is no precise technique for selecting an optimal number of neuron required for training the network (or optimising the number of neurons in the hidden layer), 14 designated number of neurons: 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 were used for optimising the number of neurons in the hidden layer. The dependent variables were represented as developed = +1, and undeveloped = -1. The 'random seed number' was set at 3558583436. Results of the computed Kappa coefficient for each designated neuron for $k=1,2,\dots,10$ are depicted in Figures 4a, 4b, and 4c; while the overall mean Kappa coefficient for each designated neuron is given in Figure 4d.

From Figure 4d for period 1963-1978, 3 neurons yielded the highest overall mean kappa coefficient while 50 neurons yielded the lowest. For 1978-1984, 2 neurons yielded the highest overall mean kappa coefficient while 50 neurons yielded the lowest. For 1984-2000, 1 neuron yielded the highest overall mean kappa coefficient while 50 neurons yielded the lowest.

Since the appropriate number of neurons for the modelling has been selected, equation 4 was used to simulate the maps for 1978, 1984, and 2000. The modelling is stochastic; therefore two hundred designated iterations thresholds were run to determine where the most accurate maps for the three periods will lie by running each iteration threshold ten times and comparing the simulated maps with the actual maps (Figure 5a). For 1978-1984 and 1984-2000, the highest mean kappa coefficients were yielded at the 130th iterations, while for 1963-1978, the highest mean kappa coefficient was yielded at the 120th iterations (Figure 5a). A modelling performance assessment was done using the Receiver Operating Characteristics (ROC) plot (Figure 5b).

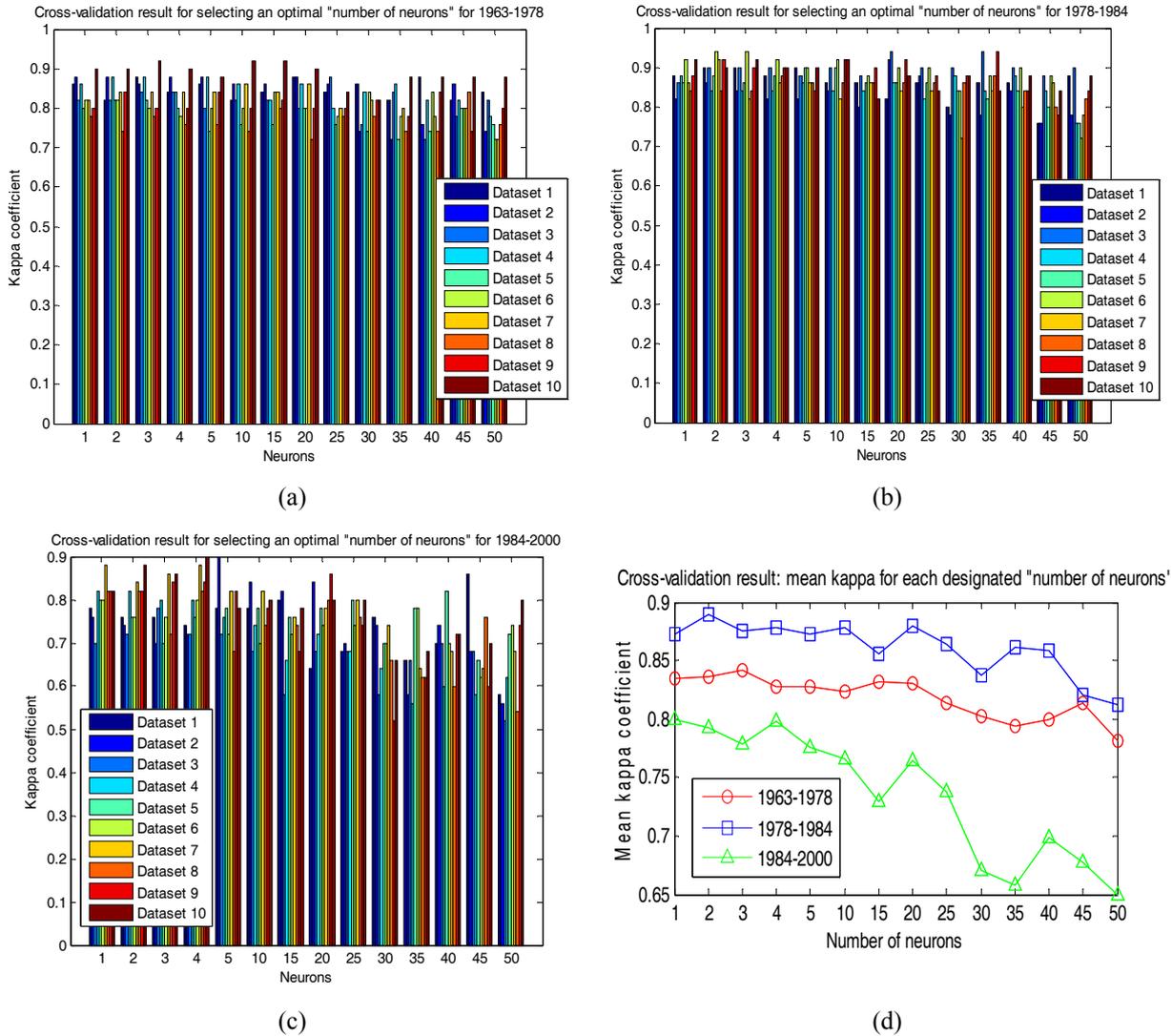


Figure 4. (a), (b), & (c) Cross-validation result: Mean Kappa coefficients for 1963-1978, 1978-1984, and 1984-2000 respectively; (d) Overall cross-validation result for appropriate choice of ‘number of neurons’ in the hidden layer

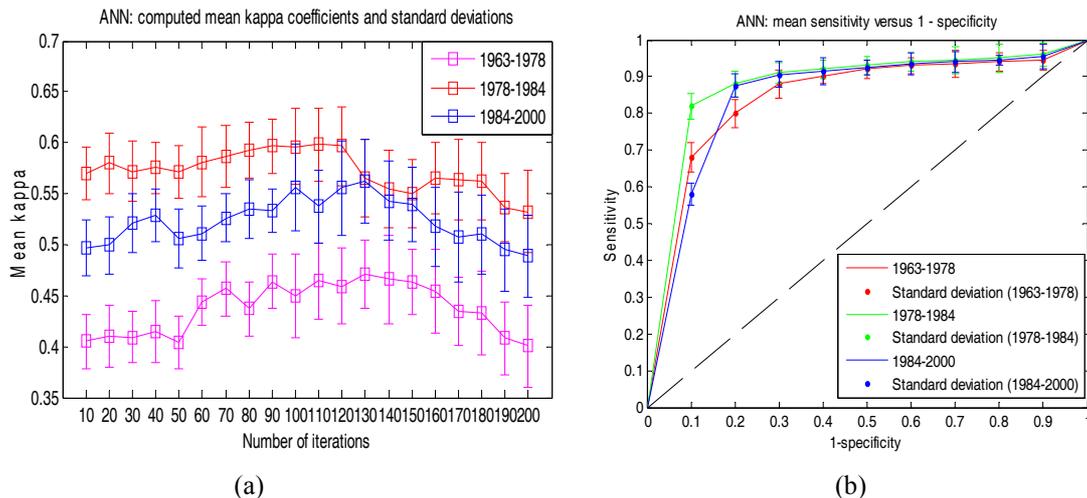


Figure 5. (a) Overall mean Kappa and standard deviations for 200 designated iteration thresholds; (b) Plotted

mean sensitivity versus 1-specificity and standard deviations

The ROC is a plot of sensitivity against 1-specificity. Figure 5b presents plots computed from 10 ROC curves, sampled at fixed 1-specificity points: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 (see Fawcett, 2004). The computed Area Under Curve (AUC) resulting from the ROC plots and their respective standard deviations were 0.7574 ± 0.0304 , 0.7850 ± 0.0308 , and 0.7613 ± 0.0288 , for 1963-1978, 1978-1984, and 1984-2000 respectively. The higher the AUC value the more accurate the simulation. Therefore 1978-1984 was the most accurate simulation while 1963-1978 was the least accurate simulation. The simulated maps in 1978, 1984, and 2000 are given in Figure 6.

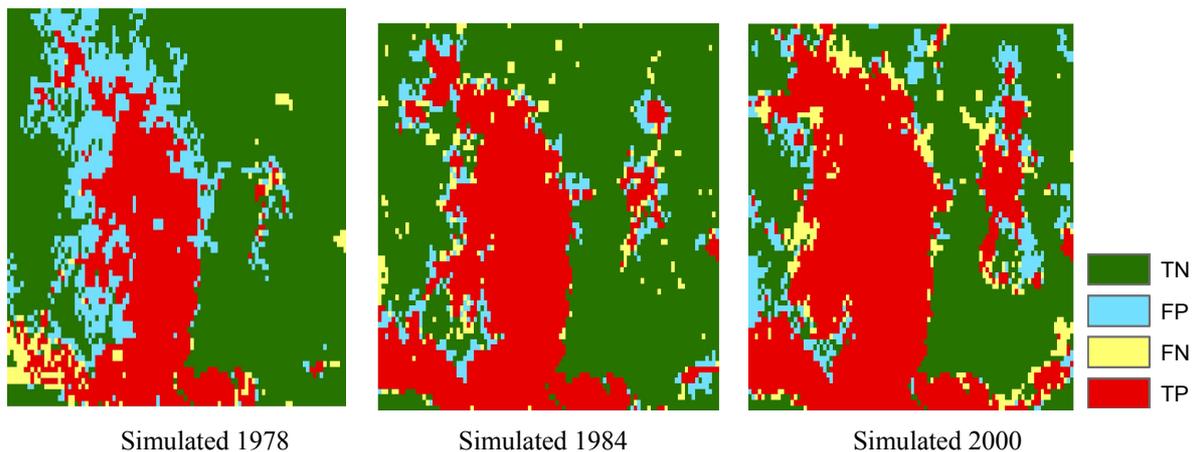


Figure 6. The simulated maps (TN=true negative; FP=false positive; FN=false negative; and TP=true positive)

Tables 1-3 were derived from the simulated maps given in Figure 6. The computed Kappa coefficients for 1963-1978, 1978-1984, and 1984-2000 resulting from the confusion matrices in Tables 1-3 were 0.5900, 0.7825, and 0.7161 respectively. ANNs are insintrically called black-box models because information about the independent variables is obscured by the model. It is important to assess the impact of each independent variable in the model. A one-neuron neural network furnishes information about the actual weights of the independent variables in the model. A one-neuron ANN is not a black-box because its input weights can be equated with the logistic regression coefficients. The result of the experiment using a one-neuron neural network is presented in Table 4. From Table 4, a variable with the highest magnitude had the most impact in the model; while a variable with the lowest magnitude had the least impact in the model.

Table 1. Confusion matrix for period 1963-1978

	Reference data 1978	
	Developed	Undeveloped
Predicted data 1978		
Developed	1609	1092
Undeveloped	183	4116

Table 2. Confusion matrix for period 1978-1984

	Reference data 1984	
	Developed	Undeveloped
Predicted data 1984		
Developed	2232	434
Undeveloped	276	4058

Table 3. Confusion matrix for period 1984-2000

	Reference data 2000	
	Developed	Undeveloped
Predicted data 2000		
Developed	2996	558
Undeveloped	436	3010

Table 4. Input weights from a one neuron neural network

Variables	1963-1978	1978-1984	1984-2000
	Input weights	Input weights	Input weights
Distance to water	-0.9023	-0.4545	0.13579
Distance to residential	-10.5806	-35.6294	20.3025
Distance to industrial and commercial	-2.5625	-0.8232	3.2947
Distance to major roads	-4.0632	-0.6593	1.04639
Distance to railway	-4.8828	-0.2645	0.6693
Distance to Lagos Island	-0.5946	-0.0060	0.1587
Distance to international airport	Not applicable	Not applicable	0.0745
Distance to international seaport	-0.2302	-0.1045	0.1877
Distance to University of Lagos	-0.4228	0.0577	0.0308
Distance to Lagos State University	Not applicable	Not applicable	0.5797
Income potential	0.2682	0.2062	-0.0791
Population potential	0.1141	0.0034	-0.0206
Input weight bias	-21.8324	-36.3103	25.0515
Model bias	0.0396	0.0765	0.1394

4. Conclusion

The computed Kappa coefficients of the simulated maps indicated that the order of best fit of the reference data for the three periods was: 1978-1984, 1984-2000, and 1963-1978. The calculated AUC estimates implied that the order of best fit of the observed data was: periods 1978-1984, 1984-2000, and 1963-1978; which corroborated the results from the Kappa estimates. According to Landis and Koch (1977), the computed Kappa coefficients of the simulated maps can be appraised based on the interpretation given in Table 5. The computed Kappa coefficients for periods 1963-1978, 1978-1984, 1984-2000 were 0.5900, 0.7825, and 0.7161 respectively; therefore using Table 5, the 1963-1978 simulated map had a moderate agreement with the reference data; while the 1978-1983 and 1984-2000 simulated maps had a substantial agreement with the reference data. The computed weights of the independent/explanatory variables from the one-neuron ANN indicated that distance to residential had the highest impact in the model while population potential had the lowest impact in the model for all the three periods. The one-neuron neural network is therefore not a black-box, since it can furnish information regarding the impact of each independent variable in the model. The result of this experiment will be useful to urban planners for understanding the spatial dynamics responsible for rapid urbanisation in Lagos.

Table 5. Interpretation of Kappa statistic

Kappa	Interpretation
< 0	No agreement
0.0 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 1.00	Almost perfect agreement

References

- Abiodun, J. O. (1974). Urban growth and problems in metropolitan Lagos. *Urban Studies*, 11(33), 341-347. <http://dx.doi.org/10.1080/00420987420080601>
- Almeida, C. M., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S. (2008). Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science*, 22(9), 943-963. <http://dx.doi.org/10.1080/13658810701731168>
- Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.
- Barredo, J. I., Demicheli, L., Lavelle, C., Kasanko, M., & McCormick, N. (2004). Modelling future urban scenarios in developing countries: An application case study in Lagos, Nigeria. *Environment and Planning B: Planning and Design*, 32, 65-84. <http://dx.doi.org/10.1068/b29103>
- Chauvin, Y., & Rumelhart, D. E. (Eds.) (1995). *Backpropagation: Theory, architectures and applications*. Hillsdale, NJ: Erlbaum.
- Clarke, K. C., & Gaydos, L. (1998). Long term urban growth prediction using a cellular automaton model and GIS. *International Journal of Geographical Information Science*, 12(7), 699-714. <http://dx.doi.org/10.1080/136588198241617>
- Fawcett, T. (2004). *ROC graphs: Notes and practical considerations for data mining researchers* (Technical report HPL-2003-4). Palo Alto, CA: HP Laboratories.
- Gandy, M. (2006). Planning, anti-planning and the infrastructure crisis facing metropolitan Lagos. *Urban Studies*, 43(2), 371-396. <http://dx.doi.org/10.1080/00420980500406751>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174. <http://dx.doi.org/10.2307/2529310>
- Li, X., & Yeh, A. G. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 16(4), 323-343. <http://dx.doi.org/10.1080/13658810210137004>
- Okwuashi, O. (2011). *Application of geographic information systems cellular automata based models to land use change modelling of Lagos, Nigeria* (Unpublished doctoral dissertation). Victoria University of Wellington, Wellington, New Zealand.
- Park, S., & Wagner, D. F. (1997). Incorporating cellular automata simulators as analytical engines in GIS. *Transactions in GIS*, 2(3), 213-231. <http://dx.doi.org/10.1111/j.1467-9671.1997.tb00012.x>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition* (pp. 318-362). Cambridge, MA: MIT Press.
- Torrens, P. M., & O'Sullivan, D. (2001). Cellular automata and urban simulation: Where do we go from here? *Environment and Planning B*, 28, 163-168. <http://dx.doi.org/10.1068/b2802ed>
- Wang, F. (1994). The use of artificial neural networks in a geographical information systems for agricultural land suitability assessment. *Environment and Planning A*, 26, 265-284. <http://dx.doi.org/10.1068/a260265>
- White, R., & Engelen, G. (1993). Fractal urban land use patterns: A cellular automata approach. *Environment and Planning A*, 25, 1175-1199. <http://dx.doi.org/10.1068/a251175>
- Wu, F. (2002). Calibration of stochastic cellular automata: The application to rural urban land conversions. *International Journal of Geographical Information Science*, 16(8), 795-818. <http://dx.doi.org/10.1080/13658810210157769>