

# Burglary Crime Susceptibility Assessment using Bivariate Statistics Approach of Information Value Model

S N Azmy, M A Asmadi, M Z A Rahman\*, S Amerudin, O Zainon

Faculty of Built Environment and Survey, Universiti Teknologi Malaysia, 81310  
UTM Johor Bahru, Johor

\*Email: mdzulkarnain@utm.my

**Abstract.** Geospatial technology advancement has boost the ability of crime assessment in terms of the accuracy of crime location and prediction. Aforetime, the crime assessment tend to focus on the development of sanction and law, as well as behaviour studies of why certain people are prone to be a victim of crime and why certain people are prone in committing crime, but none of them incorporating the idea of place of crime until 1971 (Jeffery, 1971). With technology advancement, the crime assessment of place has evolved from pin map to large scale digital mapping, effective inventory method, and adept crime analysis as well as crime prediction. The residential area of Damansara-Penchala, Kuala Lumpur and its vicinity are chosen as study area for its urban location and vastness of socioeconomic status. According to the data in Safe City Monitoring System (Sistem Pemantauan Bandar Selamat, SPBS), the monetary loss due to burglary crime activities in the study area for 2016 are sum up to RM 5,640,087 (RM 5.6 million) within 172 burglary incidence, with the mean loss of RM 32,791.00 with every offend of burglary. Apart from monetary loss, burglary also affecting the social values of the society and in terms of the perception of safe living. Instead of providing an analysis of area with high density of burglary, this paper embarks on finding the correlated social and environmental factor that leaning towards being the target of burglary crime. Utilizing the method of information value modelling, a bi-variate statistical method in the layout of raster data analysis, the vulnerability of each premise are calculated based on its association with the identified burglary indicators. The results finds that 17 significant indicators out of 18 indicators are identified as index contributing to burglary susceptibility. The burglary susceptibility mapping are acquired to contribute in predicting the premise's potential risk for the sake of future crime prevention.

Keywords: Burglary, Susceptibility, Information Value Modelling, Random Forest, GIS

## 1. Introduction

With the increasing numbers of crime committed yearly, the authority has conducted many efforts to deter crime. The most primitive effort to prevent the occurrence of crime includes the patrol by the neighbourhood watch on the places with high frequency of crime. As early as 1899 the basic idea of crime prevention through environmental design has been pitched by Enrico Ferri which commented on the character of spatial features which makes it prone to crime and outline several features that discourage the crime offence [1]. Later in 1971, the term of “environmental criminology” has been coined by Jeffery but has been ignored by the authority since the early studies of crime tend to focus three elements of crime that include the victim (what makes some people more susceptible to crime than



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

others), the law (how laws affect crime) and offenders (what makes some people commit crime) [2]. During this era, the theory for crime has been actively produced. The example of these theories includes Routine Activity Theory by Cohen and Felson [3] followed by Geometric Theory of Crime by Brantingham and Brantingham, 1981 [4], Rational Choice Theory by Clarke and Cornish, 1985 [5] and others. The underlying dynamics of location in crime occurrence has been proven by empirical researches conducted by Quetelet (1831) and Mayhew (1861) [6]. The criminology research involving the elements of geography in crime has been expanded with the improvement of data scale, from county to census tract to smaller area and today, to individual level of data. The improvement of technology indeed enhanced the human capability in collecting data in the more accurate manner.

Crime prevention initiative includes the Defensible Space concept by Oscar Newman in 1971, the Crime Prevention through Environment Design (CPTED) by Jeffery in 1971 and also the latest trend – the UN-Habitat Safe City Programme which has been adapted in local scale nationwide [7]. The Safe City programme has includes the element of Defensible Space and CPTED in terms of city design to provide a safe environment which may deter the crime occurrence. In addition, the Safe City is more comprehensive by encompassed the social and economic plan and design in order to create a safer living environment simultaneously increasing the life quality index. In Malaysia, the Safe City initiatives or also known as *Bandar Selamat* commence in 2010 as a part of National Key Results Area (NKRA) under the reducing crime policies. The implementation of Safe City in Malaysia are divided into two phase, which is the physical implementation of safety enhancing facilities and target hardening, meanwhile on the second phase, it is more to dispersal of awareness and education on safe city initiatives and the utilization of the developed aiding tools such as *Sistem Pemantauan Bandar Selamat* and *iSelamat*. The second phase on safe city implementation involve the government agencies and the community.

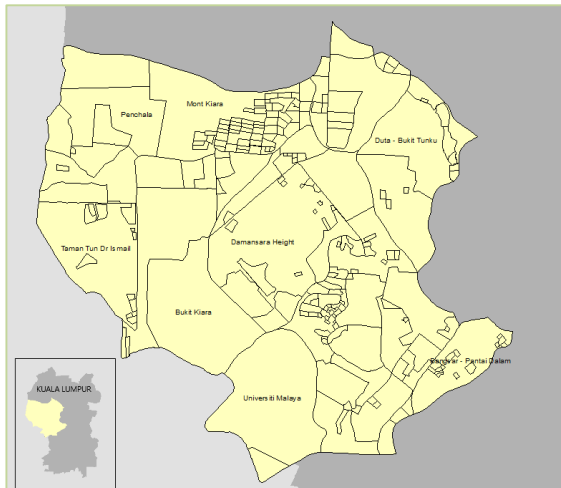
Even though there is a lot of studies has been conducted for crime analysis, it is find that there are loophole in the susceptibility assessment of burglary crime in local scale. This paper addressed the assessment of burglary crime in terms of susceptibility of each building in study area by applying the concept of Information Value Model in finding the indirect prediction model of burglary crime. The contribution of this paper is in terms of obtaining the multiple causal factor of burglary crimes in urban areas in Malaysia, along with its corresponding weights and the application of bivariate spatial statistics method of Information Value Modelling into acquiring the susceptibility of burglary crime of places.

## 2. Study Area

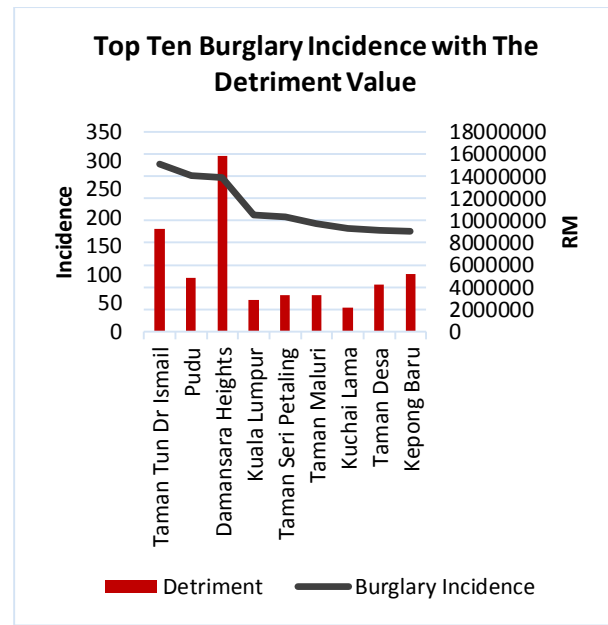
The study area of Damansara-Penchala is an area located at the side of the area of Kuala Lumpur City with the total area of 45.17782 km<sup>2</sup>. There are 226 residential areas of various typology varies from traditional Malay settlement, to land-based and high-rise planned residential development bounded inside this region. Damansara-Penchala is an official strategic zoning demarcated by Kuala Lumpur City Hall (Dewan Bandaraya Kuala Lumpur) for strategic planning and urban development under the Kuala Lumpur Structure Plan 2020. **Figure 1** shows the study area and its location at the edge of Kuala Lumpur District.

This area are chosen as the study area due to the highest association of burglary incidence and detriment value. Figure 3 shows the top ten list of number of burglary incidence with the detriment value per residential area in Kuala Lumpur. From the chart, two highest committed offences and loss values are from the residential area in the region of Damansara-Penchala. The detriment value of RM 15,796,027.00 for 270 burglary incidence for residential area of Damansara Height is quite high with the mean of RM 58,503.80 for each offend. Apart from monetary loss, burglary also affecting the social values of the society and in terms of the perception of safe living. The variety of social and demography make up of Damansara-Penchala zone is also another factor of site selection. This area inhibit by 41% of Malay, 23% Chinese, 13% Indian and 18% of foreigner (non-malaysian). In terms of housing typography, this area comprises of various house type which reflects the socioeconomic gaps. Affluence

and inequality is one of the reason that attracting the offending of burglary [8]–[10]. Another factor of consideration in site selection is due to its geographical placement between the urban areas of Kuala Lumpur and the highly populated district of Petaling Jaya which causing high commute and mobility which affects the risk of burglary offences, in line with research findings of [6], [10], [11] which concern of cognitive space awareness and familiarity which heightens the opportunity to become a burglary target.



**Figure 1.** The study area of Damansara-Penchala and the placement of study area inside of Kuala Lumpur.



**Figure 2.** The top ten burglary incidence and corresponding detriment value according to residential area in Kuala Lumpur.

### 3. Methodology

The methodology to achieve the objectives of this paper started off with the data collection of burglary incidence data and the burglary causal factors (the indicators) from the data custodians in various department of government agencies. The causal factor maps are compiled from the literature reading of the previous conducted studies. The causal factors or the parameters will undergo the process of defining the rank of importance through the Random Forest (RF) Algorithm. The results of RF were used in the sensitivity analysis to optimize the susceptibility model in later stage. Meanwhile, both burglary incidence data and all causal factor maps were assigned as data input to model burglary vulnerability using Information Value method. Information Value Model is the bivariate statistical analysis between the area of previous burglary incidence with the identified indicator (causal factor) which in form of demography, social economic and the sites of crime generator. The model are validated using Area under Curve (AUC) of Receiver Operating Characteristics (ROC). The model are optimized by eliminating the least importance causal factor. The optimized model achieve when the value of AUC is no longer increasing.

#### 3.1 Description of data

The main data of the study are the burglary incidence data dated from 2011 to 2016 which obtained from Sistem Pemantauan Bandar Selamat (SPBS). The burglary data is in form of location point of incidence individually and accompanied by the attributive information of the incidence including address, offence date and time, as well as the monetary loss value. This data were the “template” of unsupervised classification in identifying the consistency of each burglary incidence with the correlated burglary indicators.

As for the indicators to burglary crime in the study area, 18 indicators has been identified as significant to contributing to burglary using the analysis of Ordinary Lease Square as first layer of filtering. These 18 indicators also supported by previous literature which elaborated as physical factor of the premise [12] the social factor [13], surveillance [14] and crime generators area [6]. These 18 indicators are derived from various sources which narrowed down by:

- Indicators of physical factor of premised derived from building 2013 data from Dewan Bandaraya Kuala Lumpur and verification on Google Maps.
- Indicators of social and demography data from census data (2010) from Jabatan Perangkaan Malaysia.
- Indicators for security and surveillance element are derived from visual observation on Google Map and Google StreetView.
- Indicators of crime generator areas are derived spatially using spatial operations of buffering, overlay and queries.

The indicators and the sub indicators are as listed in **Table 1**:

<b>Indicator</b>	<b>Data / Sub Indicator</b>
<b>Active site</b>	Burglary Incident
<b>Wealth</b>	Area Type (Private land, Urban / Old / Usual development) Type of House (Single dwelling, Luxury housing, Middle cost strata, Middle cost terrace, Low cost housing)
<b>Accessibility</b>	Level of Floor (Strata, Non strata) Route Design (Cul-de-sac, Linear, Curvilinear) Point of Entrance (not more than two, three and above)
<b>Security</b>	Access Permit (Gated / Guarded / Not Gated)
<b>Surveillance</b>	Mixed Function a) with business and/or leisure activities b) with other activities c) mixed residential area d) area completely residential (buildings for apartments) e) area completely residential (detached houses) f) area typically monofunctional Proximity to Police Station (100, 200, 300 – 1000 m, more than 1000m)
<b>Social</b>	Race Domination (Malay-dominated, Chinese-dominated, Indian-dominated, Integrated) Percentage of Chinese Percentage of Malay Percentage of Indian Percentage of Foreigner
<b>Crime Generator</b>	Proximity to Low Cost Housing (100, 200, 300 – 1000 m, more than 1000m) Proximity to Shopping Mall

(100, 200, 300 – 1000 m, more than 1000m)
Proximity to Night Club
(100, 200, 300 – 1000 m, more than 1000m)
Proximity to Traditional Settlement
(100, 200, 300 – 1000 m, more than 1000m)

### 3.2 Indicator Prioritization Based on Random Forest Algorithm

Random Forest is non parametric modelling whereby multiple decision trees classifier are ensemble to yield the feature importance by accumulating the misclassification (error) of each trees. Ensemble referring to a process of combining several weak learners (trees) to form a strong decision trees. In simple word, it goes by majority inside the ensemble of multiple decision trees. According to [15], Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The internal processing inside the Random Forest algorithm monitor the error, strength and correlation between predictors during the splitting which yield the **variable importance** and Mean Decrease Gini Index [15].

In foundation, a tree stratifies the data – based on the if-yes rules that dividing the dataset into a non-overlapping regions [16]. There are two type of random forest splitting – regression splitting for continuous data, meanwhile classification splitting for categorical data. This study applying the classification splitting which utilized the Gini Index in splitting decision based on node purity. Higher purity yields lesser the uncertainties in decision making.

In order to optimize the burglary susceptibility model which consist of eighteen (18) geospatial indicators, the level of importance of each indicator towards the burglary crime based on the incidence data is required. This were achieved using Random Forest (RF) processing. The feature importance is very critical in order to find the combination of parameter for that is the best in defining the scenario of burglary in our study area. The indicators are eliminated one by one starting from the less important indicator, until the model stops showing the increment of accuracy. The optimization of indicators are required at the stage of obtaining the weighted sum of burglary susceptibility of each building.

The first step in running the RF algorithm is the data preparation. In order to “teach” the RF model to learn the condition, equal number of data with presence of burglary of burglary and the absence of burglary are provided as the input data. With that, the classification can be achieved efficiently without biased. As it known, RF is the ensemble of decision trees which works with probability and permutation, thus the process of RF need to be run for several sets, in this case, the lowest error term are accepted as final indicator rank prioritization.

The feature importance is critical to find a combination of parameter to define the best scenario of burglary in the study area. The indicators were eliminated one by one starting from the less important indicator, until the model stops showing the increment of accuracy. The optimization of indicators are required at the stage of weighted sum of Information Value.

The data used for random forest modelling comprises of both sample of sites with and without the incidence of burglary as pattern comparison. The levels of importance were calculated based on the total value of Gini Impurity, (GI). GI is a metric used in decision tree determination of splitting in terms of variable and the value of threshold. GI measures how often a randomly chosen record from the dataset used to train the model will be labelled incorrectly. In short, the highest value of node impurity indicate the unique quality that probably determined the phenomena, while the lowest node of impurity indicate that the indicator is less significant in determining the scenario of burglary. The results of feature of importance are shown in **Table 2**.

**Table 2.** The results of feature of importance.

Indicator	Importance
Distance to Mall	100.000

Distance to Low Cost Housing	78.124
Distance to Police Station	69.168
Distance to Night Club	55.907
Chinese Percentage	53.287
Mixed Function	45.086
Area Type	43.812
Immigrant Percentage	41.181
Distance to Kampung	40.788
Type Of Building	36.018
Indian Percentage	29.968
Street Design	26.645
Security	26.489
Malay Percentage	23.076
Level Of Floor	11.294
Point Entrance	9.845
Education	3.630
Race Domination	0.000

### 3.3 Burglary Susceptibility Modelling

Information Value Modelling (IVM) is a geospatial integrated bi-variate statistical method to analyse the probability of a site to experience landslide based on its morphology, geological, physical and the element at risk on certain site, pioneered by [17]. According to [18], IVM is a statistical analysis method that was developed from information theory. [18] Uses IVM to characterize the possibility of landslide occurrences based on the predisposing factors. Considering burglary as one of the risk of modern living, this approach has been adapt to assess the susceptibility of residential premise towards the crime of burglary. The information value  $i(x_i, H)$  of each indicator,  $x_i (i = 1, 2, \dots, n)$  contributing to burglary susceptibility can be expressed as

$$I(x, H) = \ln \left( \frac{densclass}{densmap} \right) \quad (1)$$

$$I(x, H) = \ln((Ni/N)/(Si/S)) \quad (2)$$

Where *densclass* is a cross value of burglary densities is for each class and *densmap* is overall burglary densities for entire map. Meanwhile, *H* represent the likelihood of burglary occurrence, *S* is total number of pixels in certain parameter class, *N* is the total area (sum) of burglary pixel in study area, *Si* is the number of pixels with the presence of predisposing indicators,  $x_i$  and  $N_i$  is the total area of burglary with the presence of predisposing indicator.

Meanwhile, the procedure flow of the Information Value started with splitting the incidence of burglary data into two parts of training and prediction. The process of Information Value calculation take place in raster form with pixel size of 1m x 1m, whereby the operation will occur on overlapping raster grid of the 18 underlying indicators with the active sites of burglary incidence. The model are optimized by eliminating the less importance indicator as produced in random forest operation which conducted earlier.

### 3.4 Model Evaluation using Receiver Operating Characteristics

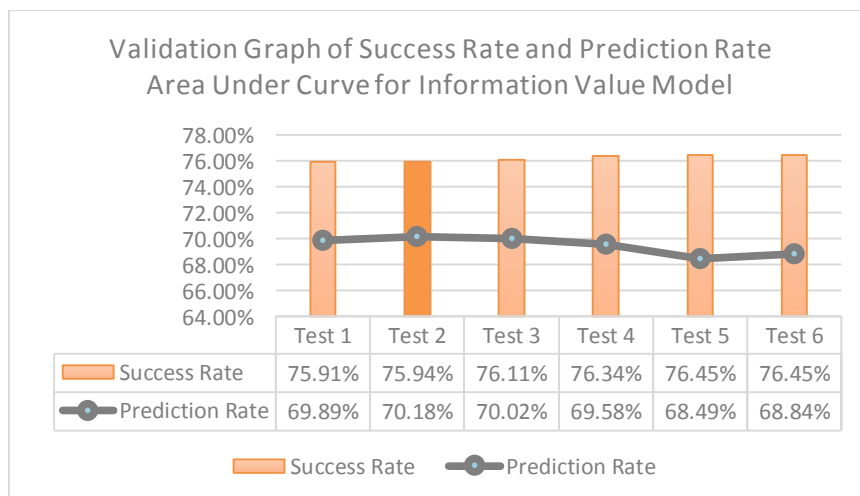
In susceptibility modelling, there are two major types of evaluation model. The first method is the evaluation of model fitness, and the second method is evaluation of model prediction performance. Receiver Operating Characteristics (ROC) is one of the example of the latter method.

ROC works in evaluating a model in terms of measuring how accurate does the models predicting the susceptibility of the unknown incidence location, which the test data is not inclusive in the stage of model training and development. ROC - AUC is the term that has been used interchangeably to describe this method of model evaluation. AUC referring to Area under Curve of True Positive Rate (TPR) which represent the percentage of accuracy of the model and indicator in classifying the susceptibility of the predicted incidence. In this case, the ROC-AUC has been used to evaluate the model and its reliability to represent the susceptibility of burglary of each building in the study area. In optimizing the model, the least importance indicator are eliminated one by one according to the lowest rank as obtained from the random forest algorithm procedure as mentioned in **section 3.2**. This loop of elimination continues until the AUC is no longer increasing. The highest value of AUC is considered as the optimized model in classifying the susceptibility of study area.

**4. Results and Discussion**

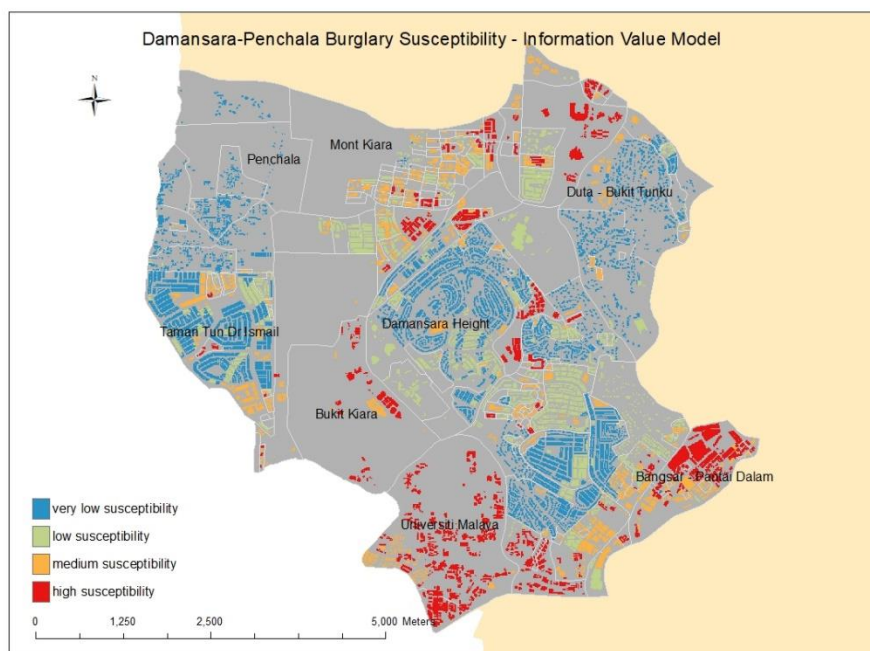
Information Value Method has been used in order to identify the level of burglary susceptibility of each building in the study area based on the attributes value of each indicator corresponding to the distribution of buildings with burglary incidence (active sites). The process involving the bivariate relationship establishment of burglary data, where the unit is the building footprint itself, with the underlying factor (the indicators) that possibly contributing to the susceptibility towards burglary.

The data were divided into two portions, the training dataset and the test dataset for prediction purpose. The data was divided on ratio of 60:40 for training dataset and test dataset. On the training section, firstly, the active site for burglary incidence are created and will be used as the basis of correlation to generate weight map with other indicators. In this phase, the weight map of each indicator has been obtained individually. This process followed by separating “no data” area inside the bounding box of raster data. Following, the weighted sum are calculated from the layers of indicators and active site to define the overall area’s susceptibility. Success rate are calculated using Area under Curve (AUC) of Receiver Operating Characteristics (ROC). For the purpose of obtaining the prediction rate, the test data were processed to become an active map in a separate geodatabase. The same weight raster obtained in training data part were used to calculate the prediction rate. The model were optimized by scenario testing of eliminating the least important indicator one by one based on the rank of importance results obtained from the Random Forest Modelling as shown in correlated **Table 3**. The AUC in form of Success Rate and Prediction Rate of multiple scenario mapping were compared and the best model were chose as charted in **Figure 3** below.



**Figure 3.** The graph of Success Rate and Prediction Rate of multiple indicator mapping.

From the results in **Figure 4**, it is found that the prediction rate peaked at 70.18% by which when 17 indicators are used as best combination of parameter in Hazard Model training and prediction. Meanwhile the mapping with other scenarios does increase the model success rate but not the prediction rate. Only one (1) indicator has been found to have less significant in predicting the burglary crime susceptibility, which is the Race Domination. The results of susceptibility modelling using Information Value method has found that 24.66% of the area in Damansara Penchala has high susceptibility towards burglary meanwhile 18.89% has medium susceptibility, 21.12% low susceptibility and the rest of 35.34% of the balance area has very low susceptibility.



**Figure 4:** The susceptibility map of Information Value Modelling

The map for susceptibility are differentiated in category of very low susceptibility, low susceptibility, medium susceptibility and high susceptibility are shown in **Figure 4**, meanwhile the weight value for each of 17 significant indicator are tabulated in **Table 3**. The weight of each indicator towards burglary susceptibility are differs in location and among the premise's attributes. The value of susceptibility were obtained with the combination of unique weight values of all 17 indicators varies to location depending on the social setup, the building characteristics, accessibility and surveillance feature for burglary offending.

**Table 3.** The weight value of each properties of indicator obtained from IVM.

No	Indicator	Sub-Indicator	Weight Value
1	Area Type	New Development	0.2541
		Usual Development	0.2259
		Old Development	0.1240
		Construction Site	0
		Non Residential	-0.4784
		Private Land	-0.8832
		Private Lot	-0.9532



<b>2</b>	Chinese Percentage	<b>55% (max)</b>	<b>2.1504</b>
		27%	0
		36% (min)	-2.5529
<b>3</b>	Distance to Police Station	<b>100</b>	<b>0.8066</b>
		200	0.5750
		300	0.4533
		400	-0.3766
		500	0.7083
		600	-0.2321
		700	-0.0992
		800	-0.5607
		900	-0.5276
		1000	-0.4616
		>1000	0.0148
		<b>4</b>	Distance to Kampung
200	-0.4855		
300	-1.0813		
<b>400</b>	<b>0.5073</b>		
500	-0.0905		
600	-0.0876		
700	-0.9510		
800	-0.7582		
900	0.3547		
1000	-0.2959		
>1000	0.0610		
<b>5</b>	Distance to Low Cost Housing		
		200	-0.1137
		300	0.4398
		400	-0.0548
		500	-0.6914
		600	-0.7287
		700	-0.7544
		<b>800</b>	<b>0.9226</b>
		900	-0.8317
		1000	0.1021
		>1000	0.0035
		<b>6</b>	Distance to Shopping Mall
200	0.1926		
300	0.4211		
400	0.2432		
500	-0.2669		
600	0.2842		
700	0.2231		
800	-0.2865		
900	0.4460		
1000	-0.5158		
>1000	-0.2017		
<b>7</b>	Distance to Night Club		

		200	0.1976
		300	0.9068
		400	0.7818
		500	-0.0924
		600	-0.4955
		700	-0.1943
		800	-0.0511
		900	-0.0530
		1000	-0.9926
		>1000	-0.0659
<b>8</b>	Education	50% – 79% population with tertiary education	-1.3759
		20% – 49% population with tertiary education	-0.4534
		<b>Less than 20% population with tertiary education</b>	<b>0.3643</b>
<b>9</b>	Immigrant Percentage	<b>16 %</b>	<b>1.8309</b>
		30%	0
		2%	-1.8216
<b>10</b>	Indian Percentage	20%	-2.0093
		39%	0
		<b>40%</b>	<b>1.8979</b>
<b>11</b>	Level of Floor	<b>Strata</b>	<b>0.6228</b>
		Non Strata	-0.8055
<b>12</b>	Malay Percentage	<b>43%</b>	<b>0.9338 – 2.1535</b>
		19%	-0.4259 – 0.1319
		12%	-2.316 - -1.2416
<b>13</b>	Mixed Function	area completely residential (terrace houses)	-1.1351
		area completely residential (detached houses)	-0.850700
		mixed residential area	-0.823200
		residential area with other activities	-0.775800
		area completely residential (apartments buildings)	0.239900
		<b>residential area with business and leisure</b>	<b>0.534100</b>
		Non Residential	0.559300
<b>14</b>	Point of Entrance	<b>Not more than 2 entrance</b>	<b>0.3078</b>
		Three and above	-0.5754
<b>15</b>	Race Domination	Chinese	1.4876
		Indian	0.911400
		Malay	0.288800
		Integrated	-0.155000
<b>16</b>	Street Design	Cul-de-sac	-1.069000
		Grid	-0.244800
		<b>Curvilinear</b>	<b>0.195800</b>
<b>17</b>	Security	<b>Guarded</b>	<b>0.704900</b>

	Gated and Guarded	0.367500
	Not Gated and Not Guarded	0.181000
<b>18</b> Type of Building	Gated	-0.955500
	Middle Cost Terrace	-1.595500
	Single Dwelling	-0.946500
	Istana Negara	-0.849900
	Utility	-0.482700
	Construction Site	0
	Low Cost Housing	0.111500
	Middle Cost Strata	0.491400
	Non Residential	0.493000
	<b>Luxury Housing</b>	<b>0.528200</b>
	Institution	0.616400
	Shophouses	0.636000
Commercial	0.6684	

Weight of Area Type indicator is highest at the area of New Development. One reason that can relate to this results is the factor of occupancy. New Development usually has low occupancy hence limited surveillance and historical track record from patrolling police. New development usually have a less density of resident at the beginning causing less eyes on the road and reduced occupancy [14], [19], [20].

In terms of racial make-up, the susceptibility weight is found to be high when the Chinese Percentage is low to medium from 0% to 28%, Indian Percentage at 0% to 13.6%, Immigrant Percentage at 14% to 28% meanwhile Malay Percentage at 38% to 74%. As for education, the area with lowest percentage of tertiary education has the highest weight on susceptibility while area with 50% to 79% of the tertiary education has the lowest weight value. This is contradicting with our null hypothesis that positively relate the affluence and vulnerability as target with education level.

Mixed Function referring to the building and the surrounding's function which contribute in terms of surveillance to prevent crime. The highest weight for mixed function is Non Residential which includes commercial area, institution and other non-housing area. In terms of housing category, the residential area with business and leisure activities shows the second highest weight contributing to burglary susceptibility. Meanwhile, residential area that completely residential (terrace house) has the lowest correlation.

Even though the point of entrance (door) is considered as the weakest part of the building for its tendency towards breaking and entry, it is found that premises with more than 3 entrance does not has the higher weight towards the susceptibility, but the premise with not more than 2 entrance, does. It is the same with strata building which has higher weight rate compared to non-strata building that supposedly has higher accessibility than the former.

Crime generator susceptibility weight has been recognized via the proximity with four type of POIs – Kampung or private land of traditional housing, Low Cost Housing, Night Club and Shopping Mall. It is found that the highest weight of susceptibility for proximity to Kampung is at the distance of 400 meter, 100 meter for Night Club and Shopping Mall, meanwhile 800 meter for proximity to Low Cost Housing.

Accessibility in terms of street design shown that restricted street such as Cul-de-sac has lowest weight for burglary susceptibility meanwhile open route such as curvilinear has highest weight. Grid street design has the medium association to the susceptibility.

In terms of security, it is found that buildings that equipped with guard features only has higher association of weight with burglary, meanwhile premise with gated features has the lowest association. Meanwhile gated and guarded community has the second highest weight on burglary incidence and not

gated and not guarded premise has a fairly association. From this results, it can be assumed that affluence (of gated and guarded community) and vacancy (of the guarded building) is still a strong factor in target selection.

As for the Type of Building indicator, the weight is high on the commercial area, followed by shophouses and Institution, which is all the non-residential building. For housing type, Luxury Housing has the highest weight, followed by Middle Cost Strata and Low Cost Housing. Middle Cost Terrace has the lowest weight value towards burglary susceptibility, followed by Single Dwelling. In this view, the affluence factor is still prominent in defining the susceptibility towards burglary risk uniform with the theory of rational choice perspective [23]–[26].

In terms of area percentage, it is found that all six (6) scenario of parameter combination yields almost the same range of percentage of very low, low, medium and high susceptibility prediction. Very low susceptibility area is around 32% to 34% of total area, meanwhile low susceptibility area ranging from 20% to 24% of the area, medium susceptibility area ranging around 18% to 20% of the study area and last but not least, high susceptibility area ranging around 23% to 25% of the area.

## 5. Conclusion

The availability of spatial data has enabled a more efficient approach to model crime susceptibility which heightens the ability of predicting the trend of crime in a large scale spatial layout. An efficient geospatial tools allow the pre-requisite and planning of urban composition in terms of social and demography in curbing and naturally controlling the occurrence of crime in conjunction with the intention of safe city implementation. This susceptibility maps has outline multiple scenario of modelling to outline the interaction of each indicator in effectively predicting the burglary crime. The outcome of this research has contributed in listing the predictor of burglary crime in local context as well as providing the spatial information of the location of high susceptibility buildings and attributes in the area of Damansara-Penchala. With the deliverables of the susceptibility map, several crime prevention measures can be taken by the local authority on the high weighted areas in order to prevent the burglary crime from occurring.

There are 17 indicators that are usable in predicting the vulnerability of a premise towards the crime of burglary. The indicators are Area Type, Chinese Percentage, Distance to Police Station, Distance to Kampung, Distance to Low Cost Housing, Distance to Shopping Mall, Distance to Night Club, Education, Immigrant Percentage, Indian Percentage, Level of Floor, Malay Percentage, Mixed Function, Point of Entrance, Street Design, Security and Type of Building. Meanwhile, the Race Domination indicator does not significantly affect the vulnerability prediction accuracy. In terms of type of building, it is confirmed that the luxury housing of condominiums has higher susceptibility than the others, correlates with the perception of affluence among burglary offender.

In developing this model, the challenges is more on the data gathering and data processing of different data sources. This model can be improved by incorporating more specific indicator to crime such as rate of unemployment, the income and tax value in the scale of house individuals, as well as some information on the burglary offender living in the vicinity of Damansara-Penchala. This model can be used by the local authority to provide a better crime prevention method in the identified high risk area to reduce the crime and elevate the safety perception among the resident. In depth, this model also can be referred by the urban planners and the local authority to enhance the residential area safety by design.

## Acknowledgement

The authors would like to express their appreciation for the support of the sponsors from Ministry of Higher Education and Universiti Teknologi Malaysia with Skim Latihan Akademik Muda and SLAB SLAI. The authors also like to convey appreciation to Malaysian Royal Police for their cooperation in providing data and information on burglary offences.

## References

- [1] G. Nicotri, “Enrico Ferri & Criminal Sociology,” *J. Am. Inst. Crim. Law Criminol.*, vol. 20, no. 2, p. 179, 1929, doi: 10.2307/1134783.
- [2] C. R. Jeffery, *Crime prevention through environmental design*. Florida: Sage Publication, 1971.
- [3] L. E. Cohen and M. Felson, “Social Change and Crime Rate Trends: A Routine Activity Approach,” *Am. Sociol. Rev.*, vol. 44, no. 4, p. 588, 1979, doi: 10.2307/2094589.
- [4] J. E. Eck and D. L. Weisburd, “Crime Places in Crime Theory,” *Am. J. Sociol.*, vol. 95, no. 6, p. 1605, 1995, doi: 10.1086/229477.
- [5] R. Paternoster, C. M. Jaynes, and T. Wilson, “Rational Choice Theory and Interest in the ‘Fortune of Others,’” *J. Res. Crime Delinq.*, vol. 54, no. 6, pp. 847–868, 2017, doi: 10.1177/0022427817707240.
- [6] P. L. Brantingham and P. J. Brantingham, “Criminality of place Crime generators and crime attractors,” 1995.
- [7] UN-Habitat, “Safer Cities Programme.” [Online]. Available: <https://unhabitat.org/urban-initiatives/initiatives-programmes/safer-cities/>.
- [8] W. Bernasco and P. Nieuwebeerta, “How Do Residential Burglars Select Target Areas?: A New Approach to the Analysis of Criminal Location Choice,” *Br. J. Criminol.*, vol. 44, pp. 296–315, 2005, doi: 10.1093/bjc/azh070.
- [9] W. H. Chiu and P. Madden, “Burglary and income inequality,” vol. 69, no. August 1997, pp. 123–141, 1998.
- [10] D. W. Sorensen, “The Nature and Prevention of Residential Burglary: A Review of the International Literature with an Eye Toward Prevention in Denmark,” *Int. Lit. with an Eye Toward. ...*, 2003.
- [11] W. Bernasco and T. Kooistra, “Effects of residential history on commercial robbers’ crime location choices,” *Eur. J. Criminol.*, vol. 7, no. 4, pp. 251–265, 2010, doi: 10.1177/1477370810363372.
- [12] G. Agarbati, “Burglary Risk Assessment of Buildings : a semi-quantitative method,” Università Politecnica delle Marche, 2015.
- [13] N. Malleon and M. Birkin, “Estimating Individual Behaviour from Massive Social Data for An Urban Agent-Based Model,” pp. 1–6, 2012.
- [14] K. A. Ham-Rowbottom, R. Gifford, and K. T. Shaw, “Defensible space Theory and the Police: Assessing the Vulnerability of residences to Burglary,” *J. Environ. Psychol.*, vol. 19, no. 2, pp. 117–129, 1999, doi: 10.1006/jevp.1998.0108.
- [15] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [16] M. Saraswat, “Practical Tutorial on Random Forest and Parameter Tuning in R,” 2019. [Online]. Available: <https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/tutorial-random-forest-parameter-tuning-r/tutorial/>.
- [17] C. J. van Westen, E. Castellanos, and S. L. Kuriakose, “Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview,” *Eng. Geol.*, vol. 102, no. 3–4, pp. 112–131, 2008, doi: 10.1016/j.enggeo.2008.03.010.
- [18] Q. Ba, Y. Chen, S. Deng, Q. Wu, J. Yang, and J. Zhang, “An improved information value model based on gray clustering for landslide susceptibility mapping,” *ISPRS Int. J. Geo-Information*, vol. 6, no. 1, 2017, doi: 10.3390/ijgi6010018.
- [19] Y. Chun and J. Lee, “Traces of Occupancy and Its Effect upon Burglar’s Residential Target Selection,” *Soc. Sci.*, vol. 2, no. 3, p. 135, 2013, doi: 10.11648/j.ss.20130203.15.
- [20] B. Snook, M. K. Dhama, and J. M. Kavanagh, “Simply Criminal : Predicting Burglars ’ Occupancy Decisions with a Simple Heuristic,” 2010, doi: 10.1007/s10979-010-9238-0.
- [21] A. A. Bujang, H. Abu Zarin, and N. Jumadi, “The Relationship Between Demographic Factors

- and Housing Affordability,” *Malaysian J. Real Estate*, vol. 5, no. 1, pp. 49–58, 2010.
- [22] G. D. Breetzke and E. G. Cohn, “Burglary in Gated Communities: An Empirical Analysis Using Routine Activities Theory,” *Int. Crim. Justice Rev.*, vol. 23, no. 1, pp. 56–74, 2013, doi: 10.1177/1057567713476887.
- [23] B. Chastain, F. Qiu, and A. R. Piquero, “Crime Theory Evaluation Using Simulation Models of Residential Burglary,” *Am. J. Crim. Justice*, pp. 814–833, 2016, doi: 10.1007/s12103-016-9336-8.
- [24] I. Hearnden and C. Magill, “Decision-making by house burglars: offenders’ perspectives,” London, 2004.
- [25] C. Nee and A. Meenaghan, “Expert Decision-Making in Burglars,” *Br. J. Criminol.*, vol. 46, pp. 935–949, 2006.
- [26] A. Sidebottom, “Repeat burglary victimization in Malawi and the influence of housing type and area-level affluence,” *Secur. J.*, vol. 25, no. 3, pp. 265–281, Jul. 2012, doi: 10.1057/sj.2011.22.