An Enhanced Support Vector Machine Model for Classification of Transit-Oriented Development

Rupali Khare
IIT (ISM): Indian Institute of Technology

Vasanta Govind kumar Villuri (vgkvilluri@iitism.ac.in)
IIT (ISM): Indian Institute of Technology

Devarshi Chaurasia
School of Planning and Architecture

Research Article

Keywords: Machine learning, Urban planning, Bhopal, Transportation, Geographical Information System (GIS), Multicriteria analysis

Posted Date: April 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-385513/v1

An enhanced Support Vector Machine model for classification of Transit-oriented development

Rupali Khare
Ph.D. Scholar, Indian Institute of Technology (Indian School of Mines), Dhanbad, India 826004,
rupali.16dr000204@me.ism.ac.in

Vasanta Govind Kumar Villuri
Faculty of Geomatics, Indian Institute of Technology (Indian School of Mines), Dhanbad, India Tel.:0326-2235445; fax: + (0326)2296628
Corresponding Author: vgkvilluri@iitism.ac.in

Devarshi Chaurasia
Faculty, School of Planning and Architecture, Bhopal, India
dchaurasia@spabhopal.ac.in

Highlights

- A new model using the support vector machine (SVM) is proposed to classify an area into transit-oriented development (TOD) and non-TOD.
- In the model, bus rapid transit (BRT) station and land use parameters under TOD mode are considered.
- Output of Multicriteria analysis was considered for the SVM model.
- The SVM approach was found suitable for the data samples used in the study.

Abstract

Transit-oriented development (TOD) can invigorate sustainable development by conveying a more coordinated transit and surrounding land use. Given the dynamic nature of urban community development, urban planners find it hard to precisely respond to questions such as where TOD planning around the transit hubs can succeed in the city. The present study proposes a framework utilizing a support vector machine (SVM) to enhance the TODness prediction of an area to address this issue. An SVM model has successfully applied to 16 bus rapid transit station areas in Bhopal city, India, using the tenfold cross-validation resampling methods and thirteen predictor variables. The models performance was in good agreement with 93.75% precision, utilizing the sigmoid kernel function and the regularization parameter esteem equivalent to 4. This methodology could be used at any scale, and the outcomes could offer recommendations for more accurate urban planning, fortifying the relationship between TOD and spatial association. The study provides the basis for predicting better future TODness classification, which will help the urban planner for sustainable urban planning and policy making.

Keywords- Machine learning, Urban planning, Bhopal, Transportation, Geographical Information System (GIS), Multicriteria analysis

1. Introduction

Transit-Oriented Development (TOD) is an urban development approach that keeps land use and transportation system under one umbrella (Khare et al. 2020). TOD is a powerful tool to help shape and assess urban development. It refers to the moderate to high-density, mixed-use land designed to maximize access by public transit and non-motorized vehicle, with other features to encourage transit ridership (Ewing & Cervero 2001; Khare et al. 2020b). Automobile dependency, urban sprawl, congestion, long trip length, poor health, pollution concerns are among the main drivers of sustainable urban development (Bibri 2018). TOD is a solution to the unsustainable, automobile-dependent, and poor urban sprawl that has characterized cities' growth worldwide. The work of Calthorpe (Calthrope 1993) discovered the concept of TOD. Introduce the designs for the redevelopment of a region. With the perspective of new urbanism, we can adapt TOD in multiple ways. Urban regenerated areas in the US (Arlington County, New Jersey, Portland, San Francisco bay area, Virginia, etc.) encompassed TOD principles in their planning to ameliorate mixed use of land, pedestrian facility, affordable housing, and promoting public health (Jamme et al. 2019). Whereas in European countries (Netherland, Paris, etc.) TOD mainly implements for curbing the urban design to make it suitable for transportation modes, mixed land uses, and public space within a range (Singh et al. 2017; Thomas et al. 2018). However, in South American cities, Curitiba is one of the oldest and most successful TOD examples and delivered high-density urban development around transit stations and controlled rapid growth. In Mexico City, the TOD policy promotes public transit (Cervero R. 1998). In some large Asian cities (Ahemdabad, Beijing, Delhi, Hong Kong, Singapore, Tokyo, etc.), TOD practice successfully reduced automobile dependency, urban sprawl, affordable housing, and pedestrian-friendly (Loo et al. 2010; Singh et al. 2014). Bus-based TOD tactic in Ahmedabad (India) curb sprawl by promoting compact high-density development structure near public transit (Joshi et al. 2017). Singapore is well known for its affordable housing, liveable high density, pedestrian-friendly ring pattern decentralized strategy. In Australia (Brisbane, Melbourne, Perth, Sydney, etc.), cities use TOD to solve lower density urban sprawl, affordable housing, connection to the town centers, and public transport-focused development (Newman 2005). Based on these studies, TOD could be considered as a useful approach for the development of cities. From a new urbanism point of view, there are multiple definitions of TOD. TOD is a pedestrian-friendly, compact, mixed-use development surrounding a transit station (Calthorpe 1993; Newman 2006). TOD is acquainted with assistance both the public transport and to utilize the openings offered by such a system (Belzer and Autler 2002). A couple of assessments have prescribed that organized land-use plans around public transit stations can provoke an increase in ridership (Li et al. 2019). TOD is not just any improvement close to transit; it improves territory capability so that individuals can walk, cycle, and use public transport. Likewise, it diminishes private vehicle usage by providing a mix of residential, jobs, recreational decisions, offers opportunities for the individual and private sector, and gives personal satisfaction to its inhabitants (Dittmar & Ohland 2004; Loo et al. 2010). A TOD zone could encompass a span of 400 or 800 meters from a transit station (Bhatt et al. 2012). Evaluation of the existing situation permits policymakers to perceive its outcome in the proposed policy (Joumard & Nicolas 2010; Pupphachai & Zuidema 2017) and thus working efficiently in the direction of changed land use. Sustainability concerning urbanization essentially relies upon the encompassing land use and transportation design and how productively pertinent assets are assigned (Ma et al. 2018). TOD is

additionally associated with the assessment of the presentation of sustainability. So as opposed to exclusively consider transit system and land use, TOD incorporates both viewpoints.

It is undeniable that modeling for the prediction of TODness requiring a deep understanding of the factor affecting its performance. Selection of the parameters according to the geographical context to find the relationship between predictor variables and TODness is an essential step for evaluation. Over the last few years, various techniques based on multicriteria analysis and data envelopment analysis have been developed to measure the TODness of an area. (Singh et al. 2014; 2017) proposed a GIS-based study to measure TODness through TOD index at a regional and local level. And some studies about the transit planning and selection of indicators (Cervero 1996; 2005; Cervero & Kockelman 1997; Curtis & Perkins 2006; Renne 2009; Curtis 2012, Khare et al. 2020a). Xie & Ding (2013) studied the relationship between rail transit and land use using the data envelopment analysis model. (Khare et al. 2020b) evaluated the coordinated relationship between land use and BRT station by using the data envelopment analysis model under TOD mode. For measuring TODness, these studies are a good example. But no attempt was done for predicting the TODness of an area. The present work is motivated by the above literature but differs significantly from the modeling approach. It is needful to classify the area into TOD and non TOD for evaluation and planning purposes. Due to significant growth in computational power in the previous two decades, machine learning (ML) strategies have pulled into consideration modelers and researchers to provide savvy solutions in different urban study fields. Through the inventive ML strategies alongside the incorporation of current procedures, researchers are driving close to more exact and capable models (Mosavi et al. 2018; Papacharalampous et al. 2019; Singh et al. 2021). ML is an inventive data-driven method, a zone of AI to get reliable results fixated on past insights and utilizations it to foresee the examples of future conduct. The ML techniques applied to urban planning studies include artificial neural network systems (ANN), support vector machines (SVM), and deep learning. Currently, SVM is the most extensive among all ML methods and has attracted many researchers' attention in Geospatial analysis. Its consolidation with GIS for Geospatial investigation has improved the capacity to produce a precise model yield in a spatial domain. In most Geospatial studies, SVM has been applied to satellite data classification and land use/land cover analysis. SVM also avoids over-fitting and ensures good generalization performance. Therefore SVM can be beneficial for modeling TODness. It has been applying widely in many studies of the evaluation area. (Lu & Liu 2018) proposed a model for sustainable development of urban transportation and the ecological environment in 22 cities in China using the SVM and Delphi method. The system was reasonable and, in real-time, evaluate the status of the cities. (Jiang et al. 2016) presented a study for the classification of landslide stability based on SVM. (Karimi et al. 2019) predict the urban expansion of Guilford County, NC, using the SVM model. However, to the best of our knowledge, SVM has not been applied to evaluate land use and transportation sustainability. The present paper aims to describe the TOD statues characteristics through SVM and provide useful reference and guidance for enhancing urban function. As discussed earlier, the present study's novelty lies in the development of a framework utilizing a support vector machine (SVM) to enhance the TODness prediction of an area. The model performance and prediction accuracy are evaluated by evaluation matrices, particularly for the TODness change case study.

2. Support Vector Machine

The SVM is a supervised ML approach and was proposed in the 1960s by Vapnik and Lerner (1963). Gives high performance in solving classification as well as logistic regression problems in many areas like remote sensing image classification, prediction, bio-informatics field, image processing, pattern recognition, etc. in comparison with logistic regression depends on the known model to predict incidence by fitting the data to a logistic curve (Yu et al. 2010). Using the training set of a different class, SVM finds an optimal separating decision surface between classes trying to get the maximum margin between classes closet training sample (fig.1). It is robust, especially for small training sets (Huang 2002; Foody & Mathur 2004) because the SVM approach is model-free and data-driven (Yu et al. 2010). It has crucial classification power, especially where the variable number is more and the training size is small. SVM decision function relies upon the small subset of the training data points, called support vectors (SV) (Wang et al. 2007). The principal working principle of SVM is to create (n-1) dimensional separating decision surfaces to classify the different classes in an n-dimensional space (Yu et al. 2010). Sample data points are an n-dimensional vector. In the best case, two classes are discriminated by a line (linear hyperplane), dividing a plane into two parts where each class lay on either side. Assume a training dataset having 'n' data points that can be separated into two classes shown as -

$$T = ((x_1, c_1), (x_2, c_2) \ldots\ldots\ldots (x_n, c_n)) \varepsilon X \times C, X \varepsilon R_n, C \varepsilon (1,-1)$$

C_1 is either 1 (yes), or -1 (no) shows the class to which the dataset lies. SVM eventually distinguish the training data through the dividing hyperplane is-

$$w.\ x + b = 0 \qquad (1)$$

Where w = weight vector, x = number of properties (input vector), and b = bias (allow us to increase the margin). As we are keen on the SV and the parallel hyperplane closest to these SV in one or the other class. Equations describing the parallel hyperplane.

$$w.x + b = 1 \qquad (2)$$
$$w.x + b = -1$$

For the linear separable data points, we can choose these decision surfaces. The maximum distance between the two hyper-planes can be denoted by $2 / \|w\|^2$. So we can maximize the margin by minimizing the 'w' (Karimi et al. 2019). To eliminate data points, make sure that-

$$C_i\ (w.\ x_i + b) \geq 1 \qquad 1 \leq i \leq n$$

In nonlinear SVM, firstly, convert the input data using a nonlinear mapping $\Phi(x)$ into a high dimensional space (Rienow & Goetzke 2015). The decision surface defined is-

$$w.\ \Phi(x) + b = 0 \qquad (3)$$

Where w and b are variables, that weight to be determined. SVM constructs an optimal separating hyperplane (OSH) by solving the quadratic optimization problem.

$$\text{Minimize} \quad 1 / 2\ \|w\|^2$$
$$\text{s.t.} \quad C_i\ (w.\ x_i + b) \geq 1 - \xi_i, \qquad (4)$$
$$\xi_i \geq 0\ ,\ i = 1,2,3,\ldots\ldots n$$

Where 'ξ' is the slack variable used to control the error caused by misclassification of points, which triggers the problem in finding the OSH (Su et al. 2007), 'c' is the regularization parameter. To solve difficult problems (not separable or appear too costly to separate because of overlapping of data points), directly use the Langrange multipliers by transferring into a dual space from primal space (Wang et al. 2005). Nevertheless, in the real-world, some data points are not separable by a linear line. Then SVM solves this problem using kernel function, inner products in the dual problem is replaced by a suitable kernel trick $k(x_i, x_j)$. Resulted from two classes could be separable in high dimensional space (Nanda et al. 2018) .

Thus the hyperplane equation can be written as-

$$k(x_i, x_j) = \Phi(x_i)\,\Phi(x_j) \qquad (5)$$

$$F(x_i) = a_n c_n\, k(x_i, x_j) + b$$

Where x_n is SV data, α_n is the Langrange multiplier, and C_n is the label of membership class with n = 1,2,3….N (Nanda et al. 2018). Here, studied the comparison of using the four kernel function at the SVM algorithm, i.e., linear kernel, RBF kernel, sigmoid kernel, a polynomial kernel (table 7). Each kernel function has explicit boundaries that should be improved to get the best performance.

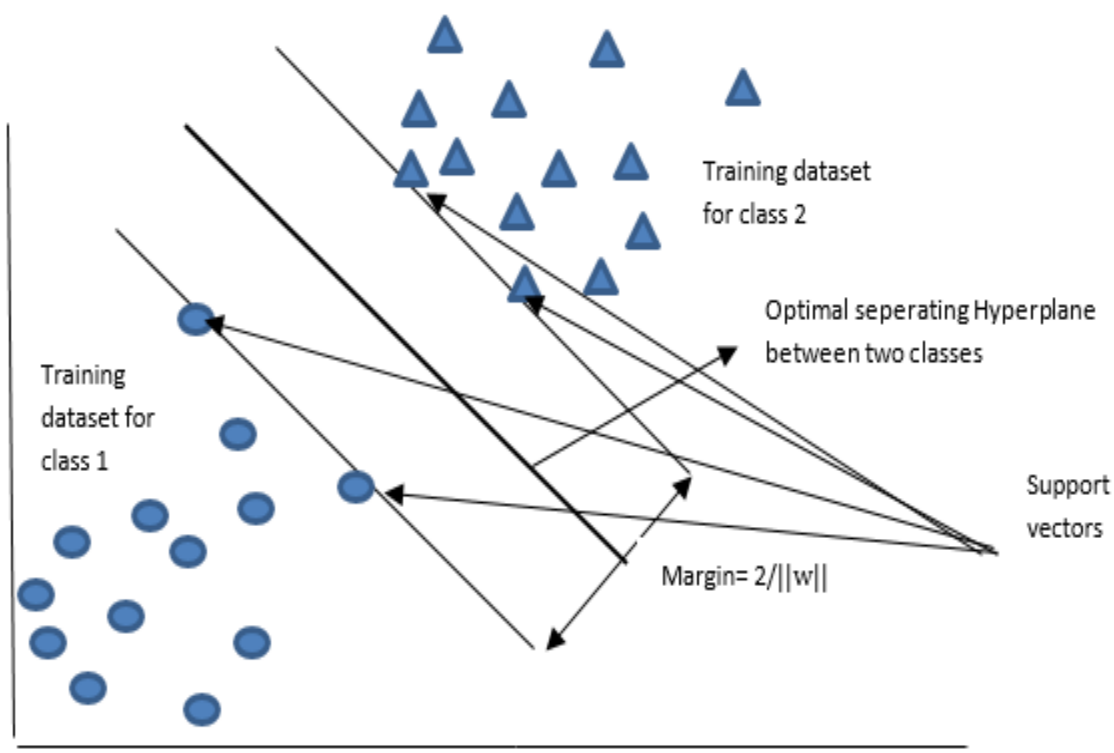

Figure 1- Optimal separating hyperplane schematic diagram

3. Study area and data used

3.1 Case study area

In the present study, the developed SVM model is applied to classify the TOD region in Bhopal city, India. Bhopal is the capital, and the administrative center of Madhya Pradesh, has coordinates 23°25'99" N and 77°41'26" E and an area of 285.9 km^2 and 2,254,000 inhabitants (Deshmukh et al. 2016). The Bhopal city as a case for studying TOD is applicable to numerous causes it has a well-established BRTS system characterized by a network structure and high transit ridership. Presently development of MRTS and corridor level TOD implementation is going on in the city, proposed to assume a transformative role in the cities growth. City buses are a fundamental part of their transportation about 44% of the work trips are made by public transport. In India, most TOD examinations have been coordinated only for metro cities as these urban communities need a political and administrative point of view. Relatively few examinations were coordinated for small and medium cities because of their slightest need in planning sectors while planning such cities. The study was performed on 16 BRT station area (table1 and fig.2) (Khare et al. 2020b). Sixteen stations of different land use (commercial, residential, industrial, semi-industrial) were selected from the city.

Table 1- Description of the study area and their land uses.

TOD area	Description
Ashima Mall	The area along with the Hoshangabad road commercial and residential
Aura Mall	Residential area with mixed land use (E-8 Extention)
Beema Kunj	Residential area with mixed land use on Kolar area
Bhopal Railway Station	Old Bhopal area densely populated and commercial area
Board Office	Near to M.P. Nagar area (Newly developed CBD of Bhopal) mainly office cum commercial area
C-21 Mall	The area along with the Hoshangabad road commercial and residential
Gandhi Nagar	Residential area with neighborhood commercial area
Habibganj	Mixed land-use area
Halalpur	The area near busiest road junction Lalghati commercial area which is periphery of the city
Karound Square	The area near busy road junction residential dominated with low-income group periphery of the city
Koh-E-Fiza	Residential high-income group area
New Ashok Garden	Residential low-medium income group area
New Market	CBD (Central Business District) main commercial area of Bhopal city
People Mall	The periphery of Bhopal main city mixed land use with residences
Piplani	Residential area near to BHEL industrial area
Sai Board	Residential area with neighborhood commercial area

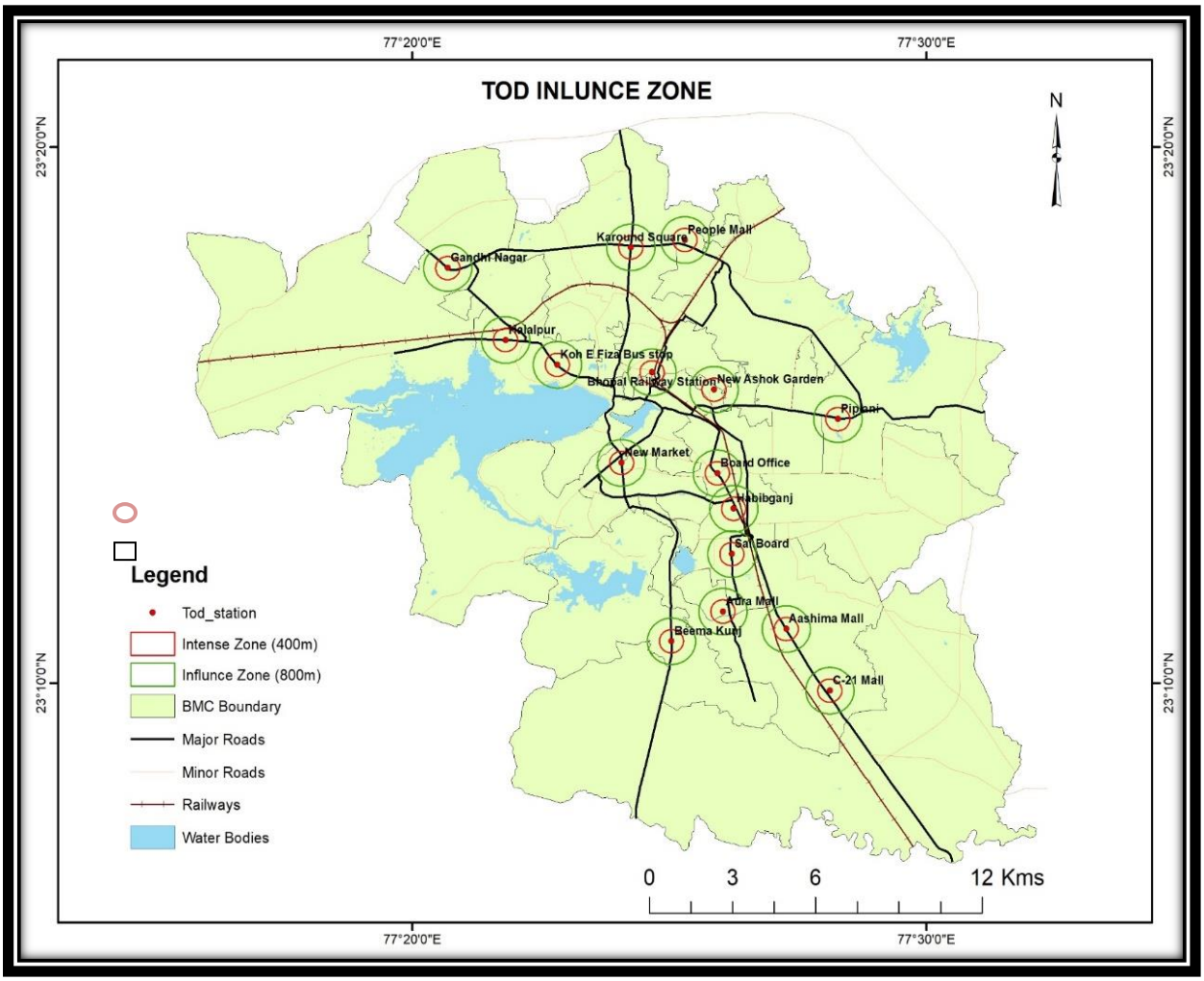

Figure 2- Location of the BRTS and its surroundings

3.2 Data used (Selection of Criteria and Performance measure)

According to the summary of the most widely used factors in the literature (Renne 2005; 2009; Performance based transit-oriented development typology guidebook 2010; Singh et al. 2017; Vilela et al. 2019), 13 predictor variables has been selected based on the different attributes of the study area (table 2). Selected predictor variables include the principle of 3Ds (Density, Diversity, and Design) (Cervero and Kockelman 1997; Joshi et al. 2017). In a built environment, the importance of the high population density is to maximize transit systems utilization. Land use diversity is a significant factor affecting non-motorized and public transport-based trips, especially for work purposes, and creates a balanced passenger flow consistently of the day. Mixed-use is commonly observed in urban area countries like India to accelerate pedestrian users. The urban area design is iconic in TOD since walkability and cyclability play a significant role in TOD success, and the built environment, which is pedestrian-friendly, will

emphasize individuals for foot or cycle-based trips (Schlossberg and Brown 2004). Knowledge of the travel behavior of the transit-supportive user is necessary.

Table 2- List of the performance measures used in the model.

Performance measure		Description
Population Density		The population density computed as the number of persons per hectare.
Employment Density		The number of job per hectare
Diversity		Different lands use in a given area and the degree to which they are present in that area
Mixed-use of land		How another land use is mix with residential use
Walkable catchment area		Walkable catchment area calculated as the ratio of the area accessed by the 5-minute walking distance to the actual area within 5-minute walking distance.
No. of High-frequency transit service available		Availability of transportation
Vehicle kilometer traveled (VKT) per capita		Kilometer traveled by a resident of TOD area
Travel behavior (Number of trips per day %)	Motorized vehicle	Gives an idea of how people use transport and what people do, to realize desired public transit ridership numbers
	Public transport,	
	Walking or cycle	
Walkable path		The pedestrian accessible road length
Intersection density		Presence of crossing points per km^2 in a region
Economic development		Several employment opportunities in the area

4. Methodology

The steps followed in the study were outlined in Figure 3 to explain the workflow of the research. The framework majorly consists of three parts: (a) Data preparation, (b) Model implementation (c) Model evaluation. The selection

of the data (performance measures) was discussed in the data preparation section. In the model implementation, regulation of the SVMs model parameters is explained to determine the most effective setup of SVM based TODness model. Moreover, in the model evaluation, define for efficient prediction, define the optimization of SVM parameters.

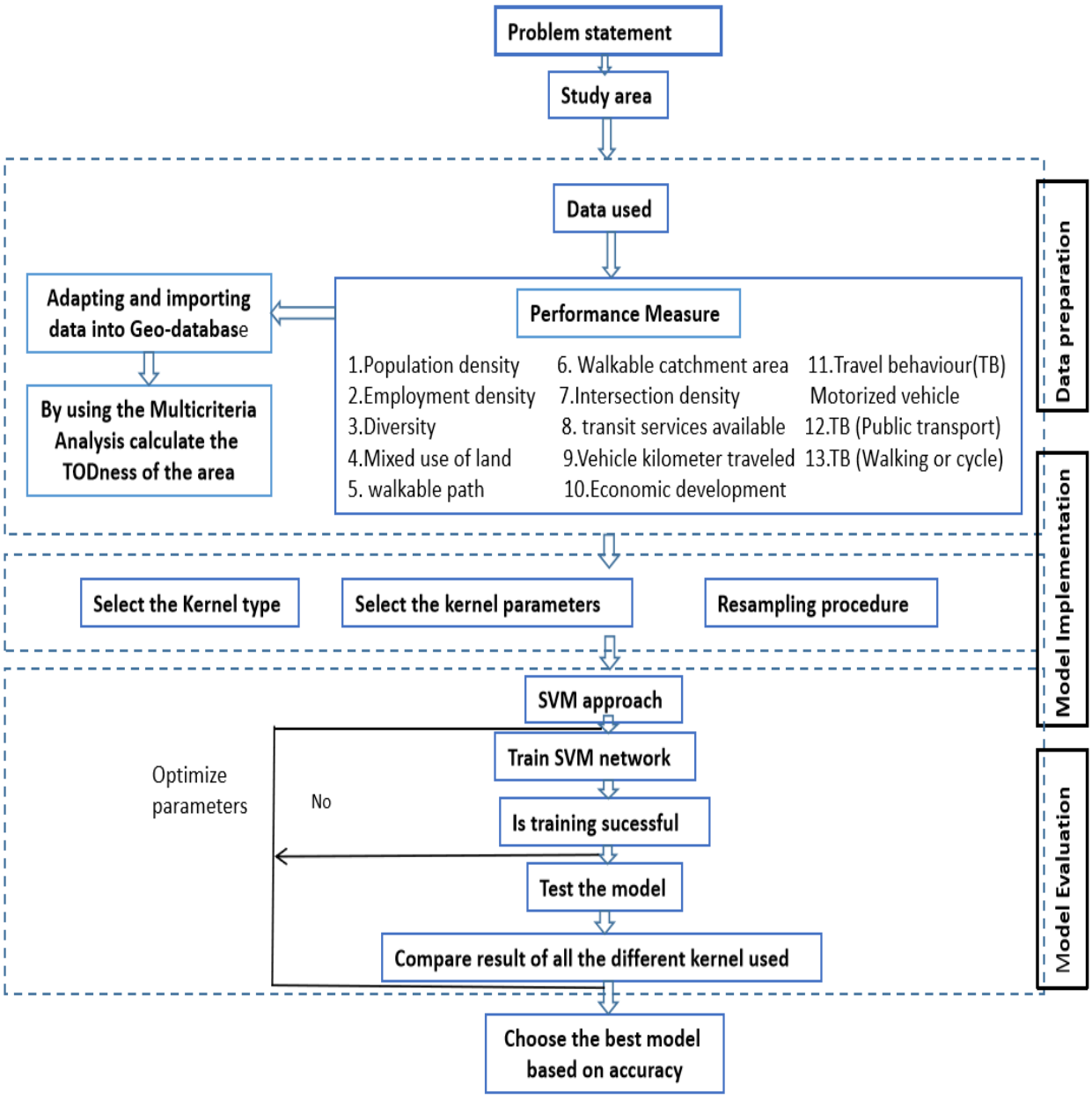

Figure 3- Flow chart of the applied methodology

4.1 Data Preparation

For effectual planning, it is essential to know about the region's existing circumstances and sort out which variable is possible to collect and set up standard data (Lagarias 2015). For land use land cover analysis study includes the satellite data collected from the National Land Cover Database (USGS) of spatial resolution 30m for the past 4 decades, and Google Earth image with the accuracy of 91.3%, 89.92%, 92.06%, and 87.83% respectively. The vector data of built-

up areas, water bodies, road networks, land use types are extracted from the LULC map. Registered vehicle data, census data, and field survey data are used for defining the pattern of Bhopal. The BMC boundary, number of houses, population data are taken from the Bhopal Municipal Corporation, and the vehicle's registered number is taken from the road transport office. GIS software is used for the spatial analysis of the data. The study needed location-dependent information to limit the area boundary up to 800m around the transit station. A TOD zone could envelop a radius of 400m or 800m from a transit station (Bhatt et al. 2012). To train and assess the SVM based TODness prediction model, the readied data layers are utilized as a predictor variable at time t, including the two classes TOD and non-TOD zone at time t, and a TODness prediction model of the two classes of TOD and non-TOD of 16 BRTS is used by using the tenfold cross-validation approach.

Multicriteria Analysis

GIS-based Multicriteria analysis is a method that combines and transforms spatial and non-spatial data input into a resultant decision (output) (Malczewski 1999). The procedure includes geographical data and the stackholders' preferences and combines them into uni-dimensional values of alternative decisions. Decision-making for analyzing different objectives can be achieved by MCE and can be used to observe different alternatives, resulting from selecting a suitable alternative for the function (Eastman 1999). After selecting criteria, two important steps of the MCE method are standardization and weight determination for each performance measure. Standardization of indicators required because all indicators are different in units. To make comparisons possible, set the suitability values of the factors to a common. In the present study, the maximum method is used for standardization and the rank-sum method to determine each indicator's weight after ranking derived from stockholders. Because not all indicator has equal importance in the calculation of TOD, depending on the perception of experts involved in planning professionals rank the indicators according to their importance in TOD calculation. The ranks of each criterion and indicator are presented in table 2. Weights can reflect the involvement of indicators in planning for TOD (table 3). Therefore, weight assigns to the indicators before the TOD index calculation. The rank-sum method is used to convert the weight from the ranking offered by the stockholders. As given below, in Eq. (6).

$$W_k = x + 1 - k / \sum x_i (x + 1 - i) \qquad (6)$$

Where, W_k = normalized weight for the criterion having rank 'k,' x = total number of the criterion in the study, i = index of summation that takes the value from one to x (Malczewski 1999; Singh et al. 2014)

Table 3. Based on the ranking of stakeholder, the weights assigned to the criteria

CRITERIA	INDICATOR	RANK	WEIGHTS
Density		***2***	***0.18***
	Population density	1	0.67
	Employment density	2	0.33
Diversity		**1**	**0.23**
	Diversity	1	1
Mixed-use		**1**	**0.23**
	Mixed-ness index	1	1
Walkable Catchment		**3**	**0.13**
	Pedshed	1	1

Travel Behavior			***5***	***0.04***
	Vehicle KM traveled per household		3	0.30
	% of trips per day		2	0.31
		Motorized vehicle	3	0.22
		Public transport	1	0.45
		Walking	2	0.33
	% No. of high-frequency transport service available		1	0.39
Street Block Pattern			**4**	**0.06**
	Walkable path		1	0.67
	Intersection density		2	0.33
Economic Development			**3**	**0.13**
	Economic %		1	1

4.2 Model Implementation

The SVM based TODness prediction model is advanced, by utilizing R. All the data layers of Bhopal is converted into CSV files to make the data compatible with R. In the model, 13 predictor variable for 16 BRT stations are considered (fig. 4). A tenfold cross resampling method is applied to evaluate the model. For limited data set the evaluation of the machine leaning method is done by using the k-fold cross-validation resampling procedure. It is a trendy technique since it is easy to comprehend and less idealistic evaluations of the model expertise than different strategies, for example, a straightforward train/test split. It shuffles the dataset randomly and splits the dataset into equal size k groups. The one group acts as a validation set, and the method applies on the remaining k-1 groups.

Data layers are standardized (by taking the average) because the scale value of input features significantly influences SVM optimal hyperplane determination. Therefore the standardization of the value of the indicator makes the equal importance of every indicators in the SVM model. The SVM model configuration includes a regularization parameter 'c', kernel function, and kernel function parameters that affect the model performance. The regularization parameter c is used to avoid the misclassification of each training. A larger ‘c’ value chooses a smaller margin hyperplane resultant, getting all the training points classified correctly. However, a larger 'c' value increased the complexity of the model and was influenced by the local support-vector. In contrast, a smaller 'c' value reduces the model complexity by producing smoother surfaces and a simpler model. Therefore, to achieve the model best generalization performance, an optimal value of c should be identified. In this study, the 'c' value of (2^{-5}, ….2^{5}) are examined to attain the SVM models foremost performance. Moreover, the learning of hyperplane also has a major impact on the performance of the SVM model.

Selection of Kernel Function

In SVM, there are the linear kernel, polynomial kernel, Sigmoid kernel, Gaussian radial basis function (RBF) kernel, and so on. The analogous kernel function can likewise be developed for particular issues. At the point when we are utilizing SVM to solve practical issues, at that point choosing kernel function is the key factor.

In the present study, linear kernel, polynomial kernel, RBF kernel, and Sigmoid kernel functions are tested, each kernel function has some parameters (cost (c), gamma (γ), degree (d), coefficient (r))for achieving the good result we

have to optimize them (table 4). For tuning the parameters, we used the grid search method. In this, we set the upper bound and lower bound value (c (2^{-5},…2^{5}), γ (2^{-3}…2^{4}), d (0,1,2,3), r (2^{-3},…2^{3}) in the region of the grid to find the finest value of parameters. In the RBF kernel, it is not needful to set the parameters. The no. of variable γ increased; thus, no. of support vector is decreased. In the polynomial kernel, the kernel parameter's increased value gives better rationalization, but after a threshold value, the models' performance is reduced due to over-fitting (Nanda et al. 2018). Then Sigmoid kernel function is selected to consign the non-linearity issue in the TODness prediction model based on the SVM.

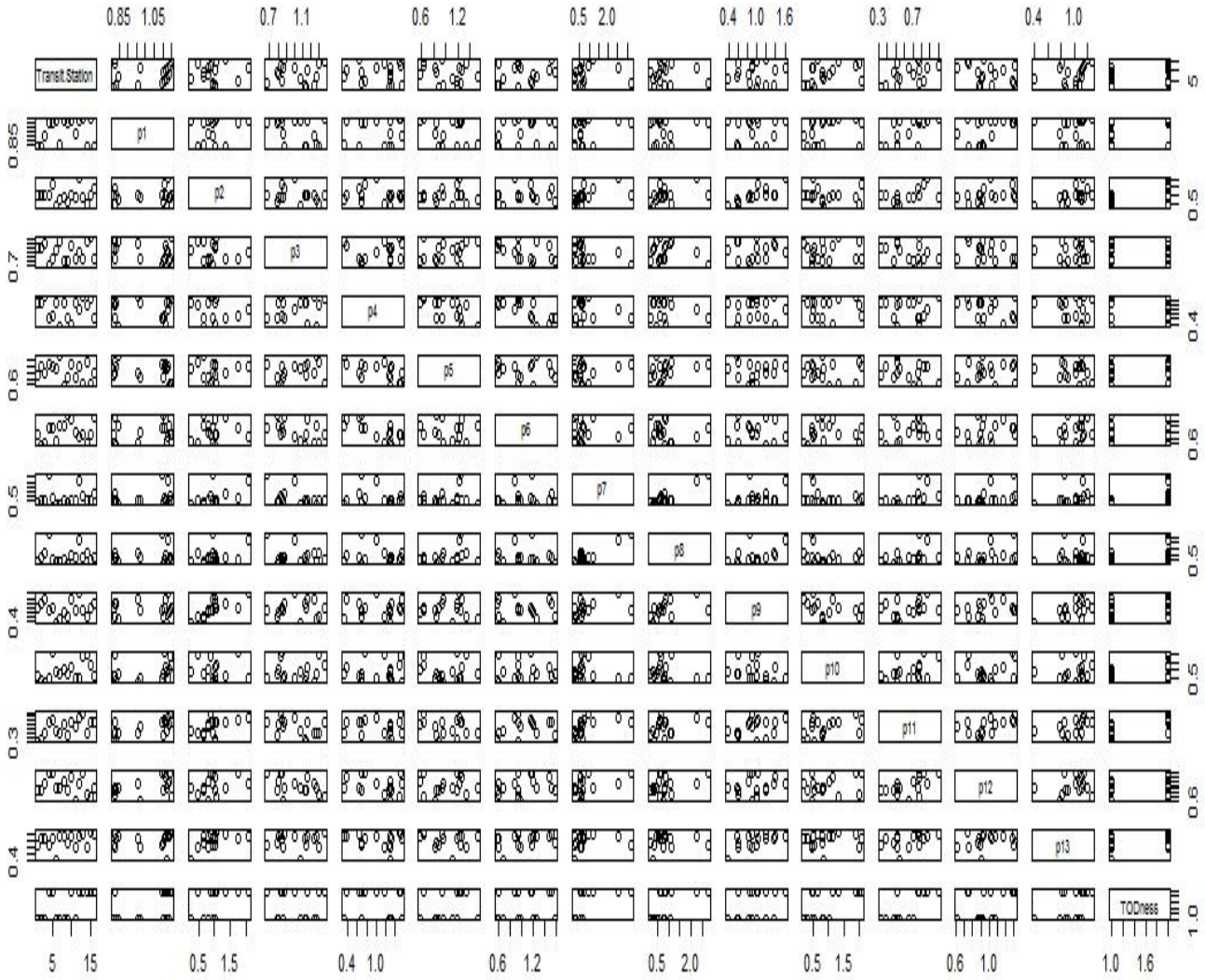

Figure 4 - Data used in the study

Table 4- Type of kernel function used in the study (Nanda et al.,2018).

Type of Kernel	Full name	Function	Parameters
Linear	Linear Kernel	$K(x, x_j) = \Phi(x).\ \Phi(x_j)$	none
Sigmoid	Sigmoid Kernel	$K(x,x_j) = \tanh(ax^T x_j + c)$	cost, gamma, coefficient
PK	Polynomial kernel	$K(x, x_j) = (1 + x.\ x_j)^q$	cost, gamma, degree, coefficient
RBF	Radial basis function kernel	$K(x, x_j) = \exp(-\lambda \lVert x - x_j \rVert^2)$	cost, gamma

4.3 Model Evaluation

By using various accuracy, the SVM model is evaluated. Total classification accuracy depends on the training data accuracy set how well we train the model. Furthermore, it measures the robustness of the model. Tenfold cross-validation was carried out for data set training, divide the data set into ten equivalent subsets. The process was repeated ten times, and each subset was allowed to act once as a data set. Test the model to validate the capability of the model. Calculation of various parameters such as accuracy, sensitivity, specificity, and precision is calculated for testing the performance. The confusion matrix presented in table 5 is based on the presence and absence of TOD.

Table 5- TOD- non-TOD area confusion matrix

Predicted/ observed	TOD	Non -TOD
TOD	True positive	False negative
Non- TOD	False positive	True negative

4.4 Statistics for performance calculation

Based on the predicted outcomes and actual outcomes, we calculated the overall accuracy, precision, sensitivity, and specificity using the following formula (Yu et al. 2010; Karimi et al. 2019).

$$\text{Overall accuracy} = \frac{Tp+Tn}{Tp+Tn+Fp+Fn}$$

$$\text{Precision} = \frac{Tp}{Tp+Fp}$$

$$\text{Sensitivity} = \frac{Tp}{Tp+Fn}$$

$$\text{Specificity} = \frac{Tn}{Tn+Fp}$$

Where, TP = True positive

TN = True negative

FP = False positive

FN = False negative

5. Results

5.1 MCA Analysis

The TOD index value for all selected 16 BRT station areas in Bhopal city is shown in table 6 and figure 5, for which, on a scale of 1, the maximum index value is 0.74 for New Market and the minimum index value is 0.39 for Ashima mall BRT station. Since Newmarket is the CBD (Central Business District) main commercial area of Bhopal city, it was expected that they would have high TOD scores. Similarly, the Ashima mall, C-21 mall station is a newly developed commercial area with less population density, and others like them were expected to score very low. Thus the predictive performance of the TOD index is well demonstrated.

Table 6- Performance score of selected BRT stations by using the MCE method.

Transit stations	**TODness Score Value**	**Transit stations**	**TODness Score Value**
Ashima Mall	0.39	Halalpur	0.45
Aura Mall	0.44	Karound Square	0.62
Beema Kunj	0.46	Koh-E-Fiza	0.48
Bhopal Railway Station	0.63	NewAshok Garden	0.70
Board Office	0.71	New Market	0.74
C-21 Mall	0.40	People Mall	0.62
Gandhi Nagar	0.48	Piplani	0.65
Habibganj	0.56	Sai Board	0.60

For validating the results, a sensitivity analysis was performed. In GIS-based sensitivity analysis, we measured the variation in the output defined by the variation in inputs. We changed the weights to a certain degree for analyzing their effects on the final TOD index. For analyzing the TOD index variations due to minor changes in weights, whether significant or not, Sensitivity analysis needs to be performed. The study did the sensitivity analysis on the criteria level for cunning TOD index robustness. Performed the analysis at the criteria level because the performance measure weights are not as much the criteria weight. The analysis demonstrates that the 10% changes in the criteria weight would not influence the TOD index. The impact just affected the slight difference in ranking of the station with little contrast of the TOD index. Areas having a score value greater than 0.60 means TOD is present (fig. 4) otherwise need to improve TODs level of services (Singh et al. 2017). The output TODness value from Multicriteria analysis was used in the present SVM model and other performance measures data.

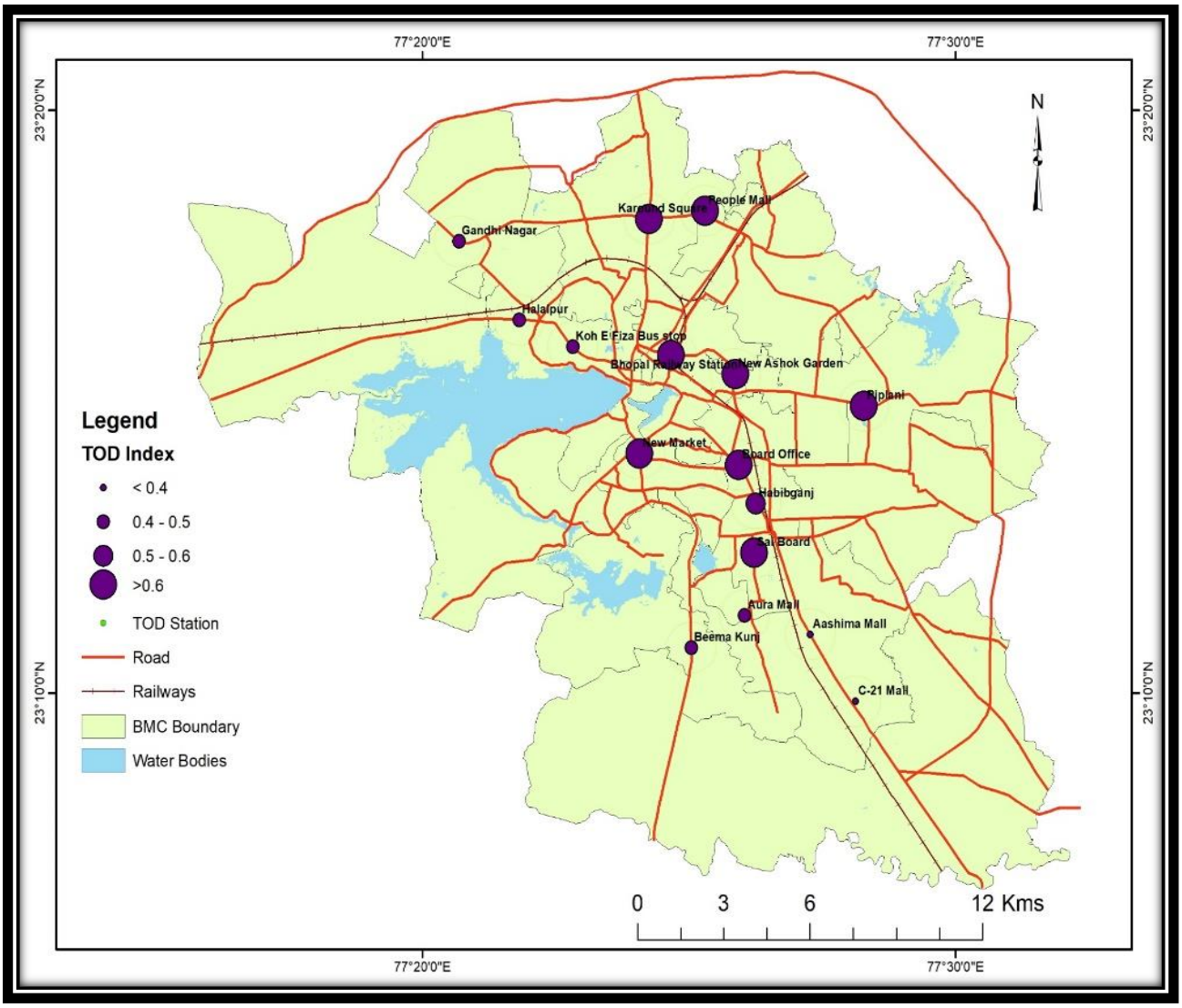

Figure 5- TOD Index along selected TOD station area

5.2 SVM

Through SVM implementation in R, a prediction is made on the data collected from 16 BRTS to predict the TODness. Choosing the most informative predictor variables by removing the variables with little importance for the model upgrades the model's precision and lessens the complexity (Karimi et al. 2019) and the necessary training and testing span. To do so, predictor variables are ranked using the perception of urban decision-makers using the rank sum method. As presented in table3, The integration of multicriteria decision-making with GIS has considerably advanced the conventional decision-making approaches (Jelokhani-niaraki and Malczewski 2015). Present SVM model performs for classifying the BRT station areas in two classes based on earlier mentioned multicriteria analysis data. The dataset target consists of binary data, i.e., TOD and non-TOD areas across the selected BRT stations. The SVM training algorithm was executed dependent on the SMO technique. Linear, polynomial, RBF, and the sigmoid kernel were utilized, and the generalization error of the model was assessed utilizing the tenfold cross-validation method. The SVM model with the sigmoid kernel function performed best in prediction compared to linear, polynomial, and

RBF kernel function, as shown in table 7. Sigmoid kernel function with the value of regularization parameter (c), epsilon, gamma, and coefficient value equal to 4, 0, 1, and 1 simultaneously gives the best result, for enhancing the accuracy of prediction; the model is tested for a different combination of 'c' value and kernel function. Statistics parameters for performance evaluation of models such as sensitivity, accuracy, precision, specificity are presented in table 8. The overall accuracy achieved of the model is 93.75%. The SVM approach appears to perform well. SVM model with sigmoid kernel function is equal to a two-layer, perceptron neural network. It was pretty prominent for SVM because its roots are from neural network theory.

Table 7 - Comparison of statistical parameters used in the study with the different kernel functions.

Kernel Function	Accuracy	Specificity	Sensitivity	Precision
Linear	0.75	0.75	0.6	0.75
Polynomial	0.6875	1	0.61	1
Sigmoid	0.9375	1	0.88	1
RBF	0.66	1	0.62	1

Table 8 - Statistics parameters used in the study for performance evaluation of the final model.

Statistic parameters	Accuracy	Precision	Sensitivity	Specificity
Value %	93.75	100	88	100

In the investigation, a confusion matrix was utilized to gauge the SVM model's presentation regarding classification. In the confusion matrix (table 9), the quantities of correct and incorrect expectations are summed up by each class. The confusion matrix's diagonal elements show that each class's true positive value (TODness) implies the model effectively predicted the target value, and the column addresses the model's predicted value, although the row represents the actual value. For example, 1 and 7 values in a column of TOD show model wrongly predicted 1 BRT station non-TOD as TOD area (false positive) and correctly predicted 7 BRT station TOD area as TOD area (True positive).

Table 9 - Confusion Matrix

Predicted/ Observed	Non-TOD	TOD
Non-TOD	8	1
TOD	0	7

6. Conclusion

The present study constructs a TODness prediction model to evaluate the transportation systems sustainability status and land-use system under TOD mode. The study incorporated the output of multicriteria analysis and the performance measures for classifying the TODness of the area by the SVM approach. Total thirteen performance measures namely population density, employment density, diversity, mixed-use of land, walkable catchment area, number of high-

frequency transit service available, vehicle kilometer traveled (VKT) per capita, travel behavior (Number of trips per day %) motorized vehicle, public transport, walking or cycle, walkable path, intersection density, and economic development are applied on 16 BRTS area in Bhopal. Sensitivity analysis was performed to distinguish the overall significance of TOD performance measures. Model performance was evaluated in the present study with the assistance of the confusion matrix. Besides accuracy, any remaining assessment measurements like precision, sensitivity, and specificity were also determined from the confusion matrix for dealing with lopsidedness and dis-proportionality in datasets.

Understanding the significance of these performance measures in TOD is very important for classifying the BRTS area. As per our concerns, this is the primary investigation wherein the SVM method was effectively used to predict the TODness of a region. Without any assumption about the data, such as dependency, the SVM distribution gives an efficient solution to prediction problems. Many SVM models are evaluated. The modeling process includes selecting the best sampling method, predictor variables, investigation of SVM model parameter, and evaluation metrics development. The finding reveals that different values of c, kernel function, and kernel parameters affect the model performance. Therefore these parameters are explicitly controlled for each case to amplify the productivity of the model. The assessment results demonstrate that the SVM model with the Sigmoid kernel function is the most efficient model for TODness prediction of an area with an accuracy of 93.75%. The model can be used to assess the impact of land-use patterns, transportation, and travel behavior. Therefore the model will help the urban decision-maker in sustainable urban planning policies. According to the study area and data availability, investigating more predictor parameters, developing a multiclass SVM model is recommended for forthcoming studies to enrich TODness modeling accuracy.

Declaration

Data availability- The dataset used or analyzed during the current study are available from the corresponding author on reasonable request. Partial Data used in the study are included in the published article (DOI- https://doi.org/10.1016/j.asej.2020.08.012).

Animal Research- Not applicable

Consent to participate- Informed consent was obtained from all the individuals who participated in the author's questionnaire survey.

Consent to publish- I, undersigned, give my consent for the publication of opinion of the BRT station area residents details, survey done by the author within the text to be published in the journal.

Plant reproducibility- Not applicable

Clinical trial registration- Not applicable

Author Contribution-

Rupali Khare- Conception and design, acquisition of data, analysis, interpretation, and writing the manuscript.

Vasanta Govind Kumar Villuri- Interpretation of data, Drafting and revisiting the article, funding

Devarshi Chaurasia: Analysed the data, conceptualization, reviewing, and editing.

All authors read and approved the final manuscript.

Conflicts of interest/Competing interests- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Funding- Not applicable

Reference

Belzer D & Autler G (2002) Countering Sprawl with Transit-Oriented Development. Issues in Science and Technology 19, no. 1.

Bhatt C, Paradkar P, and Fliert N van de (2012) Station area planning A Guide to Planning and Implementing Transit Oriented Development in Indian Cities. Institute for Transportation and Development Policy. India.

Bibri SE (2018) A foundational framework for smart, sustainable city development:Theoretical, disciplinary, and discursive dimensions and their synergies. Sustain Cities Soc, 38, 758–794. https://doi.org/10.1016/j.scs.2017.12.032

Calthorpe P (1993) The next American metropolis: Ecology, community, and the American dream. Princeton architectural press.

Census of India MADHYA PRADESH (2011), SERIES-24. http://www.censusindia.gov.in/2011census/dchb/2327_PART_B_DCHB_BHOPAL.pdf

Cervero R (1996) Mixed land-uses and commuting: Evidence from the American housing survey. Transp. Res. A, 361–377. https://doi.org/10.1016/0965-8564 (95)00033-X

Cervero R (1998) The transit metropolis: A global enquiry, edition illustrated, island press, ISBN no. 1559635916, 9781559635912.

Cervero R (2005) Accessible Cities and Regions : A Framework for Sustainable Transport and Urbanism in the 21st Century Paradigm Shift : The Accessibility-Based Approach to Planning. Transport

Cervero R, and Kockelman K (1997) Travel demand and the 3Ds: Density, diversity, and design. Transport Res D-Tr E, 2(3), 199–219. https://doi.org/10.1016/S1361-9209 (97)00009-6

Curtis C (2012) Delivering the “D” in transit-oriented development: Examining the town planning challenges.

Journal of Transport and Land Use, 5(3), 83–99. https://doi.org/10.5198/jtlu.v5i3.292

Curtis C & Perkins T (2006) Travel Behaviour : A review of recent literature. Urban, (3), 1–77.

Deshmukh H, Khare R, Pandey K, Kumar D (2016) REIS: A spatial decision system for land valuation. JoG. Apr;10(1).

Dittmar H, and Ohland G (2004) The new transit town. Best Practices in Transit-Oriented Development. TRIS.

Eastman JR (1999) Multi-criteria evaluation and GIS. Geographical Information Systems, 493–502.

Ewing R, and Cervero R (2001) Travel and the Built Environment: A Synthesis. Transp. Res. Rec.

https://doi.org/10.3141/1780-10

Foody GM & Mathur A (2004) A Relative Evaluation of Multiclass Image Classification by Support Vector Machines, 42(6), 1335–1343.

Huang C (2002) An assessment of support vector machines for land cover classification. Int. J. Remote Sens., 23(4), 725–749.

Jamme HT, Rodriguez J, Bahl D & Banerjee T (2019) A Twenty-Five-Year Biography of the TOD Concept: From Design to Policy, Planning, and Implementation. JPER, 39(4), 409–428. https://doi.org/10.1177/0739456X19882073

Jelokhani-niaraki M & Malczewski J (2015) A group multicriteria spatial decision support system for the parking site selection problem : A case study. Land Use Policy, 42, 492–508.

https://doi.org/10.1016/j.landusepol.2014.09.003

Jiang T, Lei P, Qin Q, Jiang T, Lei P, & Qin Q (2016) An Application of SVM-Based Classification in Landslide Stability An application of SVM-based Classification in Landslide Stability. Intelligent Automation and Computation, 8587(February), 0–5. https://doi.org/10.1080/10798587.2015.1095480

Joshi R, Joseph Y, Chandran VM & Darji V (2017) Transit-Oriented Development: Lessons from Indian Experiences, (January), 1–31. http://cept.ac.in/UserFiles/File/CUE/Working Papers/Revised New/36CUEWP-

36_TOD Lessons from Indian Experiences.pdf

Joumard R & Nicolas J (2010) Transport project assessment methodology within the framework of sustainable development. Ecol, 10, 136–142. https://doi.org/10.1016/j.ecolind.2009.04.002

Karimi Firoozeh, Sultana Selima, Babakan Ali Shirzadi, Shan Suthaharan (2019) An enhanced support vector machine model for urban expansion prediction. Comput Environ Urban Syst .https://doi.org/10.1016/j.compenvurbsys.2019.01.001

Khare R, Villuri VGK, Chaurasia D, Kumari S (2020a) Evaluating Transit-Oriented Development Using a Sustainability Framework: A Case Study of Bhopal. In: Ghosh J., da Silva I. (eds) Applications of Geomatics in Civil Engineering. Lecture Notes in Civil Engineering, vol 33. Springer, Singapore https://doi.org/10.1007/978-981-13-7067-0_26

Khare R, Villuri VGK and Chaurasia D (2020b) Urban sustainability assessment: The evaluation of coordinated relationship between BRTS and land use in transit-oriented development mode using DEA model, Ain Shams Engineering Journal, https://doi.org/10.1016/j.asej.2020.08.012

Lagarias A (2015) Exploring land-use policy scenarios with the use of a cellular automata-based model : urban sprawl containment and sustainable development in Thessaloniki. Geocarto Int, 6049(April 2016). https://doi.org/10.1080/10106049.2015.1013066

Li Z, Han Z, Xin J, Luo X, Su S & Weng M (2019) Transit-oriented development among metro station areas in Shanghai , China : Variations , typology , optimization, and implications for land use planning. Land Use Policy, 82(129), 269–282. https://doi.org/10.1016/j.landusepol.2018.12.003

Loo BPY, Chen C & Chan ETH (2010) Rail-based transit-oriented development: Lessons from New York City and Hong Kong. Landsc Urban Plan, 97(3), 202–212. doi:10.1016/j.landurbplan.2010.06.002

Lu S & Liu Y (2018) An evaluation system for the sustainable development of urban transportation and ecological environment based on SVM. J INTELL FUZZY SYST, 34, 831–838. https://doi.org/10.3233/JIFS-169376

Malczewski J (1999) GIS and multicriteria decision analysis. John Wiley & Sons.

Ma X, Chen X, Li X, Ding C, & Wang Y (2018) Sustainable station-level planning: An integrated transport and land use design model for transit-oriented development. J. Clean. Prod, 170, 1052–1063.

https://doi.org/10.1016/j.jclepro.2017.09.182

Mosavi A, Ozturk P, and Chau KW (2018) Flood prediction using machine learning models: Literature review. Water, 10(11), p.1536. https://doi.org/10.3390/w10111536

Nanda Achirul , Muhammad, Boro Seminar, Kudang; Nandika, Dodi; Maddu, Akhiruddin (2018) A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection. Information 9, no. 1: 5. https://doi.org/10.3390/info9010005

Newman P (2005) Transit oriented development: an Australian overview.

Newman P, Kenworthy J (2006) Urban Design to Reduce Automobile Dependence", Opolis: An International Journal of Suburban and Metropolitan Studies, Vol. 2: Article 3. http://repositories.cdlib.org/cssd/opolis/vol2/iss1/art3

Performance-Based Transit-Oriented Development Typology Guidebook (2010) CTOD

Papacharalampous, G., Tyralis, H. & Koutsoyiannis, D (2019) Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. Stoch Environ Res Risk Assess 33, 481–514. https://doi.org/10.1007/s00477-018-1638-6

Pupphachai U & Zuidema C (2017) Sustainability indicators : A tool to generate learning and adaptation in sustainable urban development. Ecol, 72, 784–793. https://doi.org/10.1016/j.ecolind.2016.09.016

Renne J (2005) Transit-Oriented Development in Western Australia: Attitudes, Obstacles, an opportunities. Development: Making It Happen, Perth. https://trid.trb.org/view.aspx?id=783726

Renne JL (2009) Evaluating Transit-Oriented Development Using a Sustainability Framework: Lessons from Perth's Network City. In Tsenkova s (Ed) Planning Sustainable Communities: Diversity of Approaches and Implementation Challenges, University of Calgary, 115–148. Retrieved from http://www.vtpi.org/renne_tod.pdf

Rienow A & Goetzke R (2015) Supporting SLEUTH – Enhancing a cellular automaton with support vector machines for urban growth modeling. Comput Environ Urban Syst, 49, 66–81. doi:10.1016/j.compenvurbsys.2014.05.001

Schlossberg M, and Brown N (2004) Comparing Transit-Oriented Development Sites by Walkability performance measures. Transp. Res. Rec. 34–42. https://doi.org/10.3141/1887-05

Singh RK, Soni A, Kumar S, Pasupuleti S, Villuri VGK (2021) Zonation of flood prone areas by an integrated framework of a hydrodynamic model and ANN. Water Supply, 21 (1): 80-97. https://doi.org/10.2166/ws.2020.252

Singh YJ, Fard P, Zuidgeest M, Brussel M, van Maarseveen M (2014) Measuring transit oriented development: a spatial multi criteria assessment approach for the City Region Arnhem and Nijmegen J. Transp. Geogr. https://doi.org/10.1016/j.jtrangeo.2014.01.014

Singh YJ, Lukman A, Flacke J, Zuidgeest M, Van Maarseveen MF (2017) Measuring TOD around transit nodes-Towards TOD policy. Transp Policy. 5696-111. https://doi.org/10.1016/j.tranpol.2017.03.013

Su L, Chopping MJ, Rango A, Martonchik JV & Peters DPC (2007) Support vector machines for recognition of semi-arid vegetation types using MISR multi-angle imagery, 107, 299–311. https://doi.org/10.1016/j.rse.2006.05.023

Thomas R, Pojani D, Lenferink S, Bertolini L, Stead D & van der Krabben E (2018) Is transit-oriented development (TOD) an internationally transferable policy concept? Regional Studies, 52(9), 1201–1213. doi:10.1080/00343404.2018.1428740

Vapnik VN (1982)Estimation of Dependence Based on Empirical Data. Springer-Verlag, Berlin.

Vilela APL, Reboita MS, Silva LF, Gerasimova MK & Sant Anna DO (2019) Sustainability neighbourhoos in Brazil: a comparasion of concepts and applications. Environ Dev Sustain. Doi:10.1007/s10668-019-00439-9

Wang, J., Neskovic, P., & Cooper, L. N. (2005). Training Data Selection for Support Vector Machines. Lecture Notes in Computer Science, (3600), 554–564.

Wang J, Neskovic P, Cooper LN (2007) Selecting Data for Fast Support Vector Machines Training. In: Chen K., Wang L. (eds) Trends in Neural Computation. Studies in Computational Intelligence, vol 35. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-36122-0_3

Xie B & Ding C (2013) An evaluation of the coordinated relationship between urban rail transit and land-use under

TOD mode. J. Transp. Syst. Eng. Inf. Technol, 13(2), 9–13. https://doi.org/10.1016/S1570-6672(13)60101-4

Yu W, Liu T, Valdez R, Gwinn M, & Khoury MJ (2010) Application of support vector machine modeling for prediction of common diseases : the case of diabetes and pre-diabetes. BMC Med. Inform. Decis. Mak, 10(16), 1–7. https://doi.org/10.1186/1472-6947-10-16

Figures

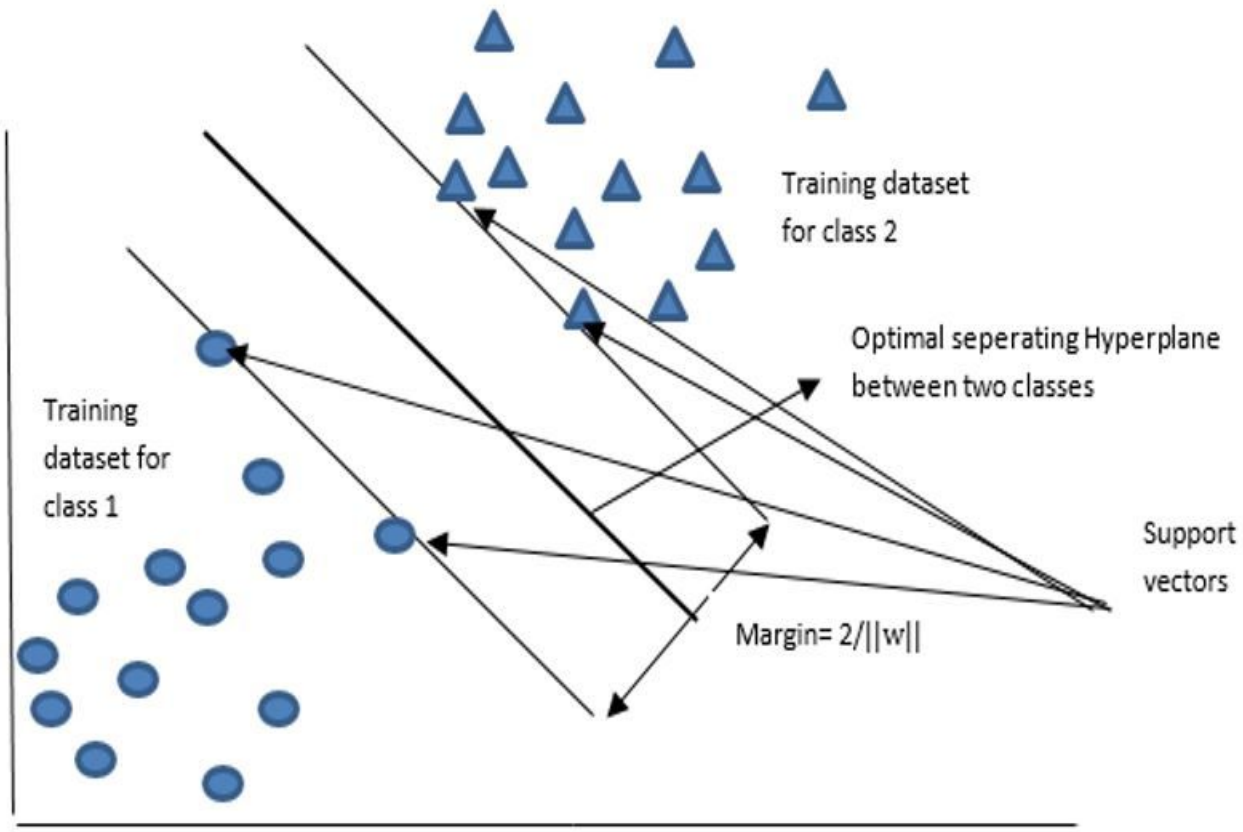

Figure 1

Optimal separating hyperplane schematic diagram

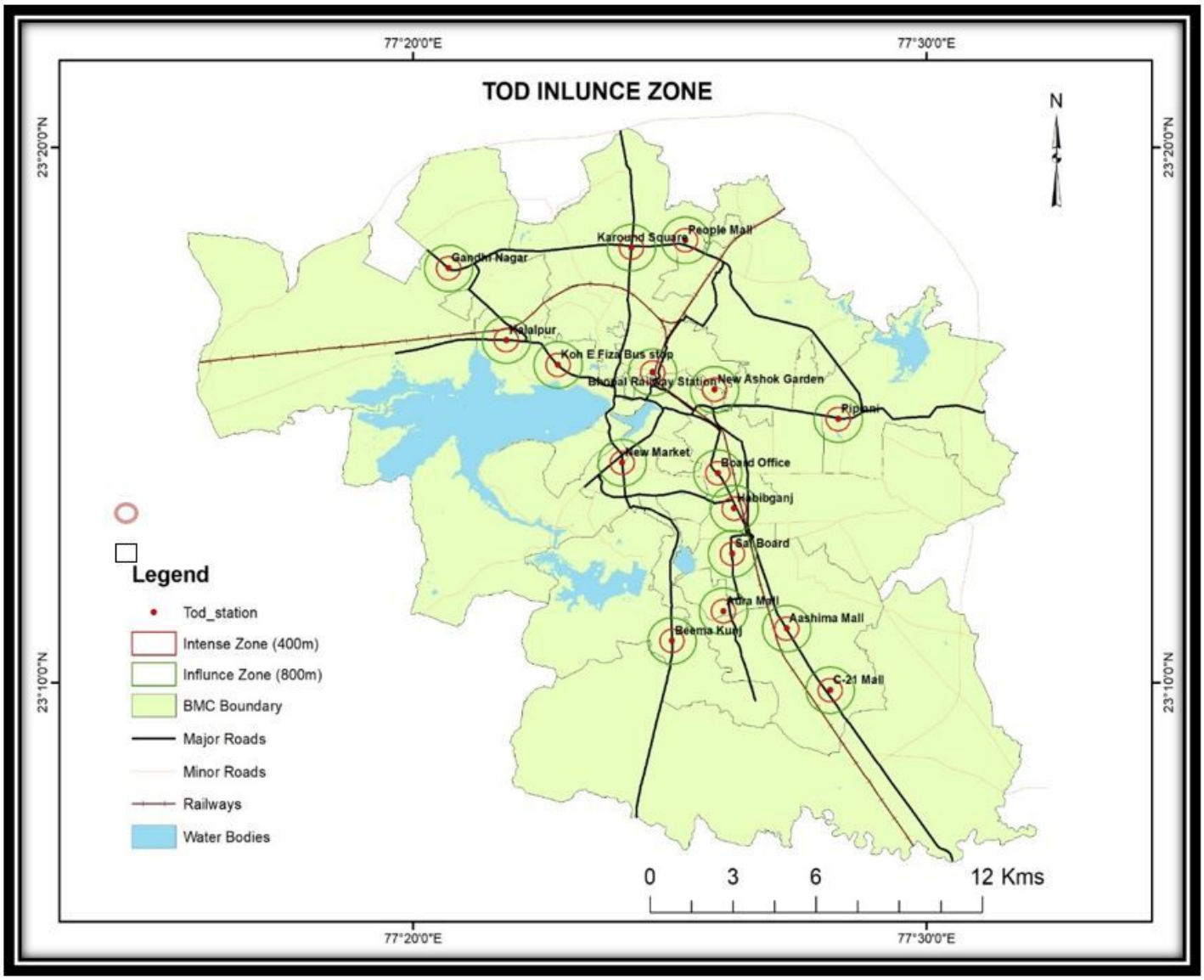

Figure 2

Location of the BRTS and its surroundings Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

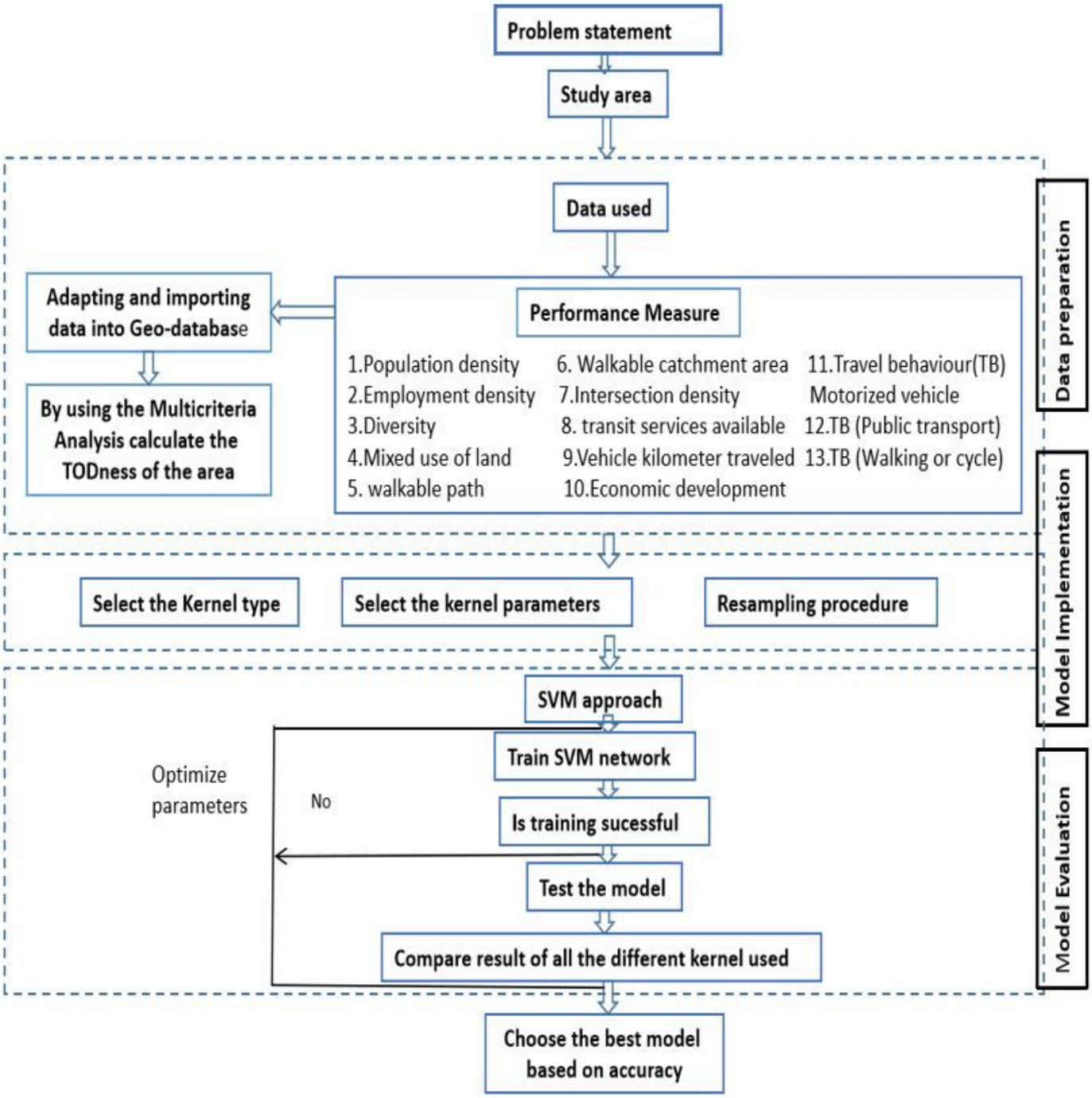

Figure 3

Flow chart of the applied methodology

Figure 4

Data used in the study

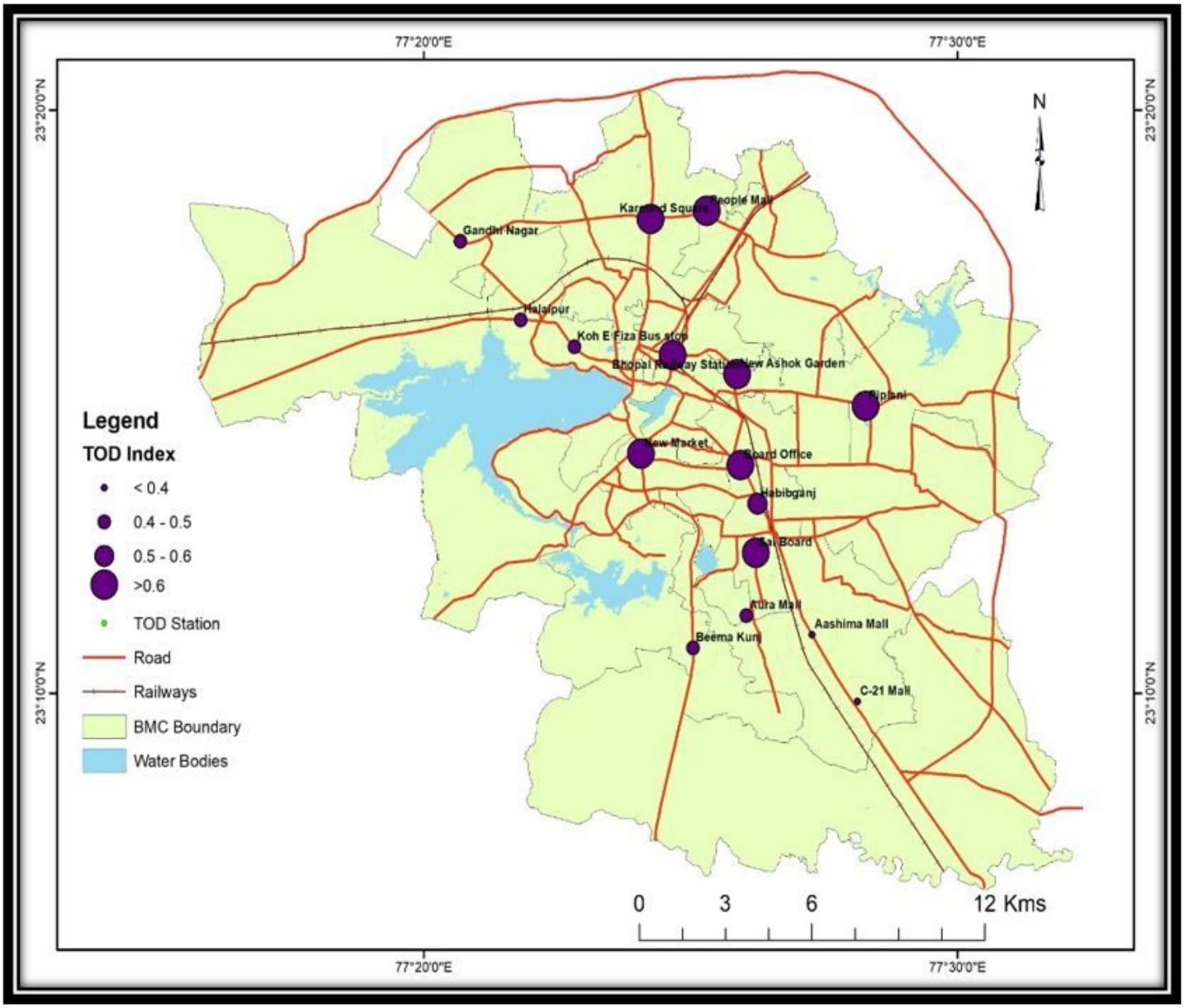

Figure 5

TOD Index along selected TOD station area Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.