Available online at www.humg.edu.vn

ESASGD 2016

GIS-IDEAS (2016)

International Conference on GeoInformatics for Spatial-Infrastructure Development in Earth & Allied Sciences (GIS-IDEAS)

The Analysis of Spatial Distribution of Disabled People in Thailand using Geographically Weighted Regression Models

Amphawan KHAMCHIANGNGERN, Pathana RACHAVONG, Kampanart PIYATHAMRONGCHAI and Tanyalak SRIPHO

Faculty of Agriculture Natural Resources and Environment, Department of Natural Resources and Environment, 99 Moo 9 TambonTha Pho, AmphoeMuang, Phitsanulok, Thailand 65000,

Abstract

The study of the Analysis of Spatial Distribution of Disabled People in Thailand using Geographically Weighted Regression Models aims to evaluate spatial parameters that influence the spatial distribution patterns of disabled people in Thailand. Six factors have been used; numbers of road accidents occurred in 2015, urbanization rate, industrial development (numbers of industrial estate located in each province), poverty incidence, agricultural dependence using the ratio of irrigation area and percentage of natural hazard occurrence in each province.

The secondary data derived from the administrative authorities were used and were systematised into ordinal and/or interval scales. The Geographically Weighted Regression (GWR) was analysed for each pair of interested factors in order to find the highest R2 value. The selected R2 value was used to construct the GWR formula that can be represented an appropriate spatial distribution in the Geographic Information System. The research found that poverty is the factor that has the highest relationship with disability in Thailand followed by road accidents, industrial development and urbanization (R2 = 0.5090161, 0.509301, 0.504115 and 0.503677, respectively). Finally, the R2 of GWR formula was calculated at R2 = 0.500708. The urbanization and road accidents factors were shuffled to find the relationship of each factor represented by disability ratio. The study found that the GWR formula which consists of poverty, industrial development and road accidents has the higher R2 than the one that contains the poverty, industrial development, and urbanization parameters with the R2 = 0.650897 and 0.439129, respectively.

Finally, the study concluded that poverty, industrial development and road accidents factors are the most appropriate factors that can be used to describe the spatial distribution of disabled people in Thailand. Moreover, the relationship of spatial factors revealed the unique characteristics of each local setting in terms of Local R2, Local Predicted Values, Local Coefficient Intercepts, Local Coefficient Bi, and Local Residuals, which are represented in thematic maps to illustrate the geo-visualization of spatial distribution in the study area.

Keywords: Geographically Weighted Regression (GWR); GIS

Introduction

According to the National development plan for Disabilities of Thailand Act No. 4 (2012 to 2016), Thailand has set its strategy of action for people with disabilities addressing that people with disabilities must be able to live independently with everyone in the society with happiness. Disabilities should have an access to their rights of equality. The five strategies have been set which include the promotion of equality and equal rights without discrimination against disabled people and their carers, the creation of appropriate environment and information technology that people can access and use, creating empowerment for people with disabilities and those who care for the disabled, promoting and strengthening the capacity of disability organizations and networks, and the promotion of creative attitude towards disability and disabled people.

National Bureau of Statistics of Thailand reported that numbers of the disabled people who qualify to be promoted and developed are 1.406 million people, scattered in Central and Eastern regions 0.345 million, Northeast 0.544 million, Northern .325 million, Southern 0.156 million people, and could not identify around 0.036 million people. The study of the distribution of disability in the United States by McCoy, Devis and Hudson (1994) found that people with disabilities are distributed in relation to the topography, type of Industry, and economic development of the area. Rachavong, Fongmoon and Kamchiang-Ngern (2011) found that the distribution of the disabled in Phitsanulok province, Thailand is associated with urbanization, the distribution of poverty, and the prone to mudslides areas. The study of McCoy, Devis and Hudson (1994) showed only a number of disabled and summarized the relationship of the distribution of density but did not demonstrate the weight of each factor. Although the study of Rachavong, Fongmoon and Kamchiang-Ngern (2011), represented the weight of each factor but It is only an overview of the whole area, which is impossible to illustrate the weight variation related to the geography of the area. Moreover, the study area is the province that consists of a different set of factors that affected the weight of many factors not statistically significant.

In the weighing of the factors to the incident illustration, the statistical method was used to analyze the use of the least squares regression. Results are expressed as the weight of the perspective of the study area. Fortheringham, Brundon and Charton (2002) mentioned that this analysis provides only one statistics value that uniforms the characteristics of the area of interest. Although it can be displayed on a map, it does not imply the spatial characteristics. Moreover the program is not suitable for analysis with GIS. They proposed analytical method that can express the weight of each factor, which varies to the different characteristics of the area. This method is called "Geographically Weighted Regression Analysis". This method can be provided the statistical values that satisfactorily present the details of spatial differences of interested area, which makes it possible to analyse the spatial statistics in the software packages and show the results on map conveniently.

With the benefits of the spatial analysis of geographically weighted regression analysis that is used to present the difference of local characteristics and can be shared information with a geographic information system in various forms of map, this research tends to analyze the influence of the spatial distribution patterns of people with disabilities in Thailand.

Main Concept and Methodology

GWR is an outgrowth of ordinary least squares regression (OLS). It has a level of modeling sophistication by allowing the relationships between the independent and dependent variables to vary by locality. GWR is useful as an exploratory procedure. It allows visualization of stimulus-response relationships and how that relationship varies in space. It also accounts for spatial autocorrelation of variables. Briefly, GWR constructs a separate OLS equation for every location in the dataset, which incorporates the dependent and explanatory variables of locations falling within the bandwidth of each target location. Bandwidth can be manually entered by the user or it can be determined by the statistical software.

GWR was originally developed for the analysis of spatial point data and allows for the interpolation of values that are not included in the data set. It is applied under the assumption that the strength and direction of the relationship between a dependent variable and its predictors may be modified by contextual factors. GWR has high utility in epidemiology, particularly for infectious disease research and evaluations of health policies or programs. Limitations of GWR include problems of multicollinearity and the approaches to calculating goodness of fit statistics. We have included two articles that specifically address these concerns.

How GWR works:
1. OLS models are run to determine the global regression coefficients (β) for the independent variables:

$$yi = \beta 0 + \beta 1x1i + \beta 2x2i + \ldots + \beta nxni + Ɛi$$

with the estimator:

$$\beta' = (XT\ X)\text{-}1\ XT\ Y$$

2. Once the independent variables that you wish to retain in the model are identified, and there is a theoretical basis for thinking that the relationships may differ by space, GWR may be an appropriate next step. The regression models that underlie GWR:

$$yi = \beta 0 + \beta 1x1i + \beta 2x2i + \ldots + \beta nxni + Ɛi$$

With the estimator:

$$\beta'(i) = (XTW(i)\ X)\text{-}1XTW(i)Y$$

Where W(i) is a matrix of weights specific to location i such that observations nearer to i are given greater weight than observations further away.

Data used in the study represent the characteristics of seven variables which are the number of disabled people, poverty, urbanization, agricultural promotion area, industrial promotion zone, disaster encounter, and traffic accidents in the area. The data in this study were collected by coordinating with the responsible organization in order to obtain information directly and some of the data were queried from the relevant authorities. They are arranged into the system that has been set up using a proportion/ratio per unit area to make it ready to perform with GIS software both for the descriptive statistics and analysis of geographically weighted regression.

Table 1 The Six Factors of PWD Distribution

Object ID	VARNAME	VARIABLE					
		URB	IND	POV	IRR	DIS	ACC
1	Bandwidth	225186.03	179617.45	170359.83	213066.67	211026.55	200994.34
2	Residual Square	7.735019	7.728204	6.387197	8.614766	8.207947	7.647373
3	EffectiveNumber	15.218435	16.563335	21.543258	14.673532	17.139015	17.312264
4	Sigma	0.353835	0.357593	0.339374	0.371780	0.370293	0.357943
5	AICs	71.993653	75.246155	72.970507	79.032619	80.490826	76.156173
6	R2	0.503677	0.504115	0.590161	0.447228	0.473332	0.509301
7	R2Adjusted	0.389454	0.376417	0.438342	0.325958	0.331338	0.375197
8	Depend. Field 0	PPWD	PPWD	PPWD	PPWD	PPWD	PPWD
9	Explan. Field 1	URB	IND	POV	IRR	DIS	ACC

Note: URB: Urbanization; IND: Industrial Development; POV: Poverty Rate;
IRR: Irrigation Zone; DIS: Disasters; and ACC: Traffic Accidents

The attribute of the shapefile polygon layers consist of the information of Province Code, Shape Length, Shape Area and province name and were arranged for the retrieval in conjunction with geographic information. Geographically weighted regression analysis is used to analyze the data in order to determine a wide range of kernel using AICc method to acquire the local parameters R2, Local R2, Local Coefficient, Local Intercept Coefficient and Local Residuals using GIS are detailed below.

1. The six variables were analyzed with analysis by Geographically weighted regression analysis by exploiting the coordinates U and V into the equation. The key is to get the parameters R-Square, Residual Square, AICc and all six variables.

2. The calculation of the weighted spatial function is performed by GWR analysis of each variable, one at a time to get the important parameters of R-Square, Residual Square and AICc.

3. Analysis to select an appropriate model to perform four steps analysis to create a GWR model that can predict the proportion of disabled people appropriately by calculating based on the R-Square with the highest values and Residual Square with lowest values as follows:

• Step 1: Choose a variable with the highest R-Square GWR analysis of these variables.

• Step 2: Compare the values of R-Square equation in Step 1 to R-Square in step 1.

• Step 3: Test to switch variable in the equations to find with the highest R-Square and lowest values of Residual Square.

• Step 4: Select the equation that contains highest R-Square and lowest values of Residual Square to explain the distribution of people with disabilities in each province of the country.

Area of Study

Thailand is a Southeast Asian country. The country is located on the southeastern part of Burma, sharing its borders with the Andaman Sea and the Gulf of Thailand. The total area of Thailand occupies an area of 514,000 square kilometers. The countries sharing border with Thailand are Myanmar, Cambodia, Laos and Malaysia. Thailand shares a total of 4863 kilometers with its neighboring countries.

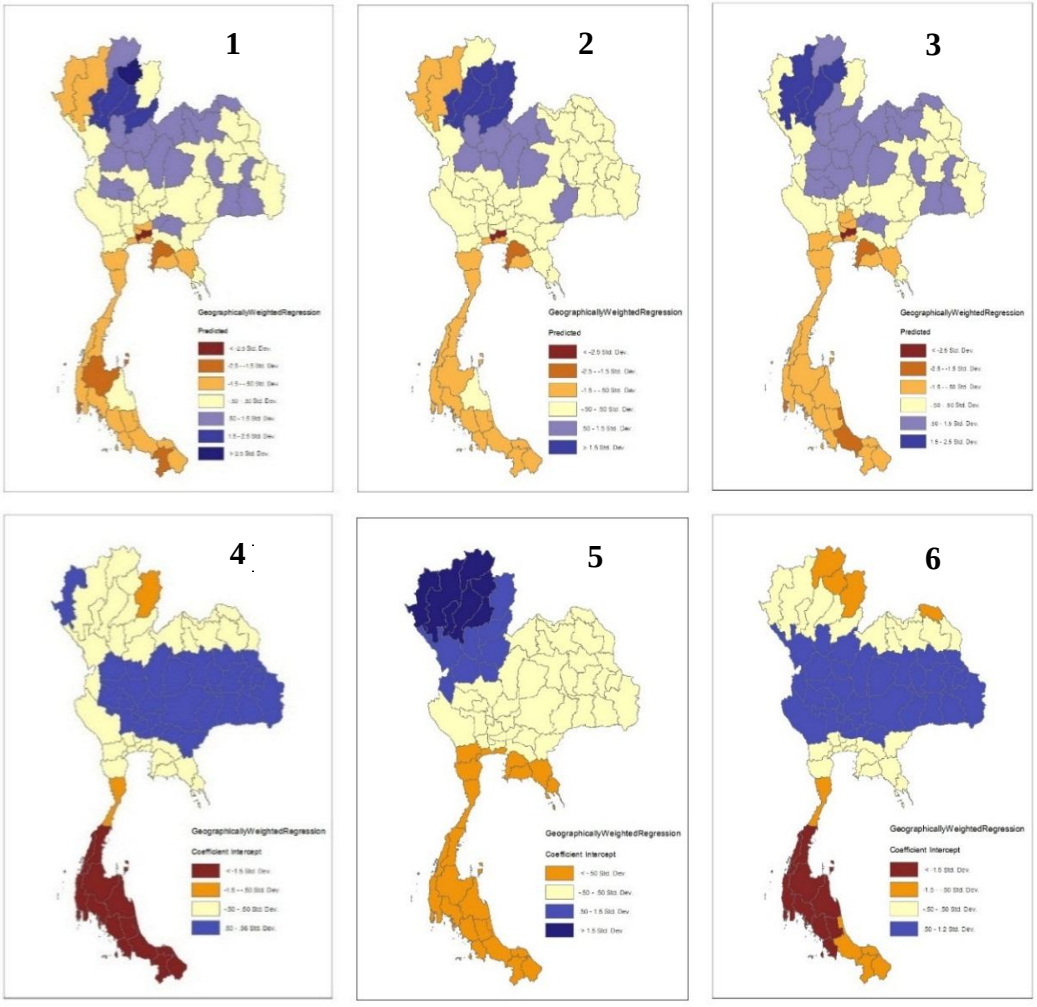
Results

The regression analysis of the variables that influences the distribution of people with disabilities using GWR has R2 value of 0.527493 including six variables that were statistically significant. It can be used to predict the proportion of disabled people in each province. However, in order to get the best regression result, the model was built to find alternative equation. Three models were constructed. GWR_01 (R2 = 0.500708) consists of four independent variables; poverty, industry development, road accident, and urbanization. GWR_02 (R2 = 0.650897) has three independent variables; poverty, industry development, and road accidents. GWR_03 (R2 = 0.439129) uses three variables; poverty, industry development, and urbanization. The study found that GWR_02 model could be able to predict the proportion of people with disabilities the best with the highest R2.

Table 2 GWR Three Scenarios of PWD Distribution

Object ID	VARNAME	VARIABLE		
		GWR_01	GWR_02	GWR_03
1	Bandwidth	340575.65	200390.93	319736.86
2	Residual Square	7.781297	5.440655	8.740985
3	Effective Number	18.461643	27.01085	16.19312
4	Sigma	0.364591	0.329904	0.379144
5	AICs	81.977307	80.200197	85.025795
6	R2	0.500708	0.650897	0.439129
7	R2Adjusted	0.351772	0.469248	0.298991
8	Dependent Field 0	PPWD	PPWD	PPWD
9	Explanatory Field 1	POV	POV	POV
10	Explanatory Field 2	IND	IND	IND
11	Explanatory Field 3	ACC	ACC	URB
12	Explanatory Field 4	URB		

R-Square value of GWR_02 model is higher than those of GWR_01 and GWR_03 models and also provides to the lower Residual Square (5.440655), which is less than the Residual Square of GWR_01 (7.781297) and GWR_03 (8.750985). The predicted distribution of the disabled people using geographically weighted regression analysis found that. Poverty, Industry development, and road accident variables are the most suitable variables.



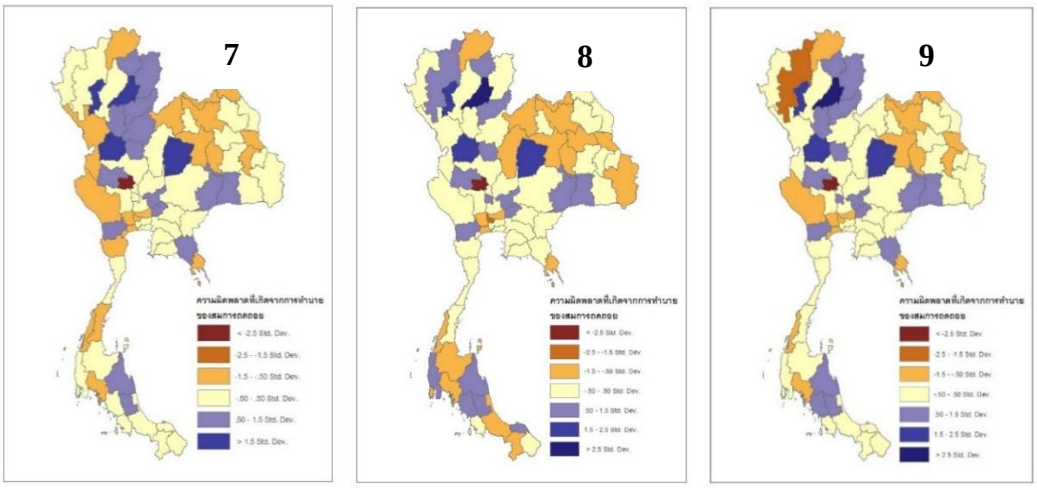


Fig 1. Results from Geographically Weighted Regression Analysis

1. GWR_01 Predicted
2. GWR_02 Predicted
3. GWR_03 Predicted
4. GWR_01 Coefficient Intercept
5. GWR_02 Coefficient Intercept
6. GWR_03 Coefficient Intercept
7. GWR_01 Residual
8. GWR_02 Residual
9. GWR_03 Residual

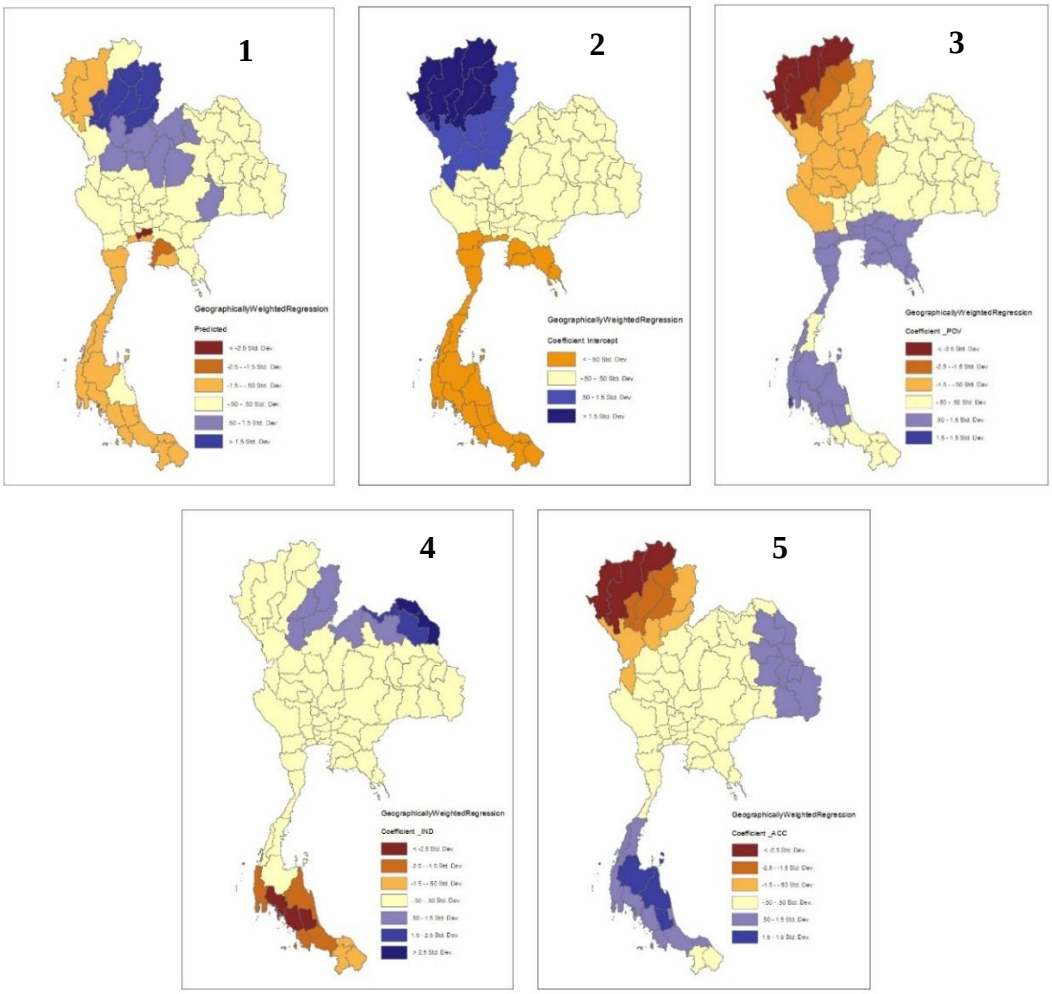


Fig 2 Results from Geographically Weighted Regression Analysis of GWR_02 model

1. GWR Predicted
2. Coefficient Intercept
3. GWR Coefficient of poverty
4. GWR Coefficient of industrial development
5. GWR Coefficient of road accident

Conclusions and Discussions

Geographically weighted regression analysis provides multiple values of the statistics which shows the details of spatial differences properly. This method can be used to display information into the various types of map. Brundon, Fortheringham and Charton (1998) have written a specialized program in order to find an exception of the local event in any local. Brunsdon, Fortherington, and Charlton (1996) suggested that the nature of the model tends to modify and manage over the area to reflect the actual structure of the data package.

Geographically Weighted Regression analysis is an attempt to detect this variability by revising the multiple regression analysis model in order to reveal the relationship of different spatial points and display the uniqueness of spatial relationship shown on the map.

The research showed that the analysis of regression equations for describing the distribution of the disabled people in Thailand is the equation of GWR_02 model: PWD = f (POV, IND, ACC). The influence of poverty on the proportion of people with disabilities when considered in conjunction with the development of industrial and road accidents has the highest rate in Phuket followed by all of the Eastern region area and most provinces of Central and Southern regions. When consider the details of the local situation, the study found that Mae Hong Son, Tak, Kalasin, Buri Ram, Yala, Pattani and Narathiwat have relatively high poverty rate.

When consider with the influence of road accident and poverty, the study found that Nakorn Phanom and Bung Karn have relatively high proportion of people with disabilities. Whereas when consider with the influence of industrial development and poverty, the study found that SuratThani, Pattalung, and Nakornsri Thammarat have relatively high proportion of people with disabilities. There are examples of the study of the distribution of disability in the United States. McCoy, Devis and Hudson (1994) found that disabled people scattered in various areas associated with geographic features, the density of industry sector, and economic conditions of the area. The research of Rachavong, Fongmoon and Kamchiang-Ngern (2011) found that the distribution of the disabled in Phitsanulok is associated with urban area, the distribution of poverty, and the mudslide risk area.

Predicted results from the analysis of a regression model to predict the proportion of disabled body showed that areas with disabled density is Lamphun, Lampang, Phayao, Phrae, Nan and Uttaradit, while the real distribution of the disabled are more common in Phrae, Phayao, Lamphun, Uttaradit, Kamphaeng Phet, Chaiyaphum, Nan, Lampang, Sukhothai, Phitsanulok, Phichit, Uthai Thani. Buri Ram, Nakhon Nayok and Saraburi According to residual value of the regression model, although the regression would be appropriate for the reasons mentioned but when it is used to predict the actual situation, mistakes can be happened. It can be seen that the Kamphaeng Phet, Lamphun, Phrae, and Chaiyaphum their faulty prediction are higher than other provinces.

Determining the default coefficient Intercept value in the exponential distribution with disabilities determinants consists of poverty, industry development, and road accident, the study found that individual province has different results. The provinces that have have default value are including Mae Hong Son, Chiang Rai, Phayao, Lampang and Phrae. The group of provinces that has low default value are including all areas of the Southern region, six provinces in the Central region; Samut Sakhon, Samut Songkhram, Ratchaburi, Phetchaburi and Prachuap Khiri Khan, and four eastern provinces which are Chon Buri, Rayong, Chanthaburi and Trat. Moderate default value include all areas of the Northeast, most areas of the Central region, and four provinces of eastern and northern province which are Prachin Buri, Sa Kaeo, and Phetchabun.

References

Brunsdon, Chris; Fortherington, A. Steward; and Charlton, Martin E. (1998). "Geographically Weighted Regression-Modelling Spatial Nonstationarity." THE STATISTICIAL. 47: pp. 431-443.

Fotheringham, A. Stewart; Brunsdon, Chris; and Charlton, Martin. (2002). GEOGRAPHICALLY WEIGHTED REGRESSION: THE ANALYSIS OF SPATIALLY VARYING RELATIONSHIPS. Chichester: John Wiley and Sons.

McCoy, John L.; Devis, Miles and Hudson, Russell E. (1994) "Geographic Patterns Disabilities in the United States." SOCIAL SECURITY BULLETIN. Vol.57 No.1 Spring.

Mennis J. (2006). Mapping the Results of Geographically Weighted Regression. The Cartographic Journal 43(2): 171-179.

Rachavong, Fongmoon and Kamchiangngern (2011). "Spatial Distribution of Rehabilitation Rights of People with Disabilities: A Case Study of Local Administrative Authorities in Phitsanulok, Thailand" The Institute of Health Promotion for People with Disability.