Weighted sum model for Spatial Analysis in Classification of Areas Prone to Diphtheria Tetanus

Anik Vega Vitianingsih[1], Achmad Choiron[2], Dwi Cahyono[3], Suyanto[4]
{vega@unitomo.ac.id[1], choiron@unitomo.ac.id[2], dwik@unitomo.ac.id[3]}

[1,2,3]Informatics Departments, Universitas Dr. Soetomo Surabaya, Indonesia

Abstract. The health sector in developing countries such as the State of Indonesia is still experiencing problems to find out the distribution of areas prone to tropical infectious diseases, including diphtheria and tetanus. The purpose of the discussion in this paper is to perform spatial analysis of geographical information system (GIS) fields through spatial data modeling for classification of diphtheria and tetanus-prone areas based on spatial parameters of datasets such as immunization status coverage, PD3I, epidemics, nutrition status, and population density. The Weighted sum model (WSM) method is used in the process of spatial analysis which is a multiple criteria decision making (MCDM). The classification results of diphtheria and tetanus-prone areas consist of good, fair, and poor categories. Preference values are generated using the Guttman method as a basis for spatial analysis in the next series of data. The test results can be used as recommendations in policymaking to anticipate the distribution of areas in the poor category.

Keywords: Spatial Analysis, Spatial Data Modeling, GIS, SAW, WPM, MCDM.

1. Introduction

Tetanus Neonaturium (TN) and Diphtheria are contagious tropical diseases which are very common in developing countries [1] which have tropical climates as well as the state of Indonesia, case studies in the discussion of this paper are based on case studies in East Java Province, Indonesia. Based on reports of diphtheria surveillance programs in East Java Regency including the category of re-emerging disease. There was an increase in cases of TN in 2012, 2013, 2014, namely 29 cases, 31 cases, 33 cases, then there was a decline in 2015 to 26 occurrences, 2016 found 19 cases, and 2017 again rose to 24 cases [2][3][4][5][6][7][8]. The case of Diphtheria in East Java is the largest contributor to cases in Indonesia, which amounted to 74% even in the world, wherein 2011, 2012, 2013, 2014, 2015, 2016 were 664 cases, 955 cases, 653 cases, 442 cases, 265 cases, 345 cases, and 489 cases, respectively [2][3][4][5][6][7][8]. The incidence of cases of the spread of tetanus-diphtheria (TD) in 38 districts/cities in East Java has always been a significant increase, even though there have been a few years of slight decline. Efforts made to suppress TD cases are to carry out basic immunization in infants with Diphtheria-Pertussis-Tetanus and Hepatitis B vaccine (DPTHB)[9].

One proven preventive measure is to provide complete immunization to infants. The vaccine aims to reduce morbidity, disability, and mortality due to immunization-preventable diseases (PD3I) [10]. Based on Minister of Health Regulation Number 1059 of 2004 concerning guidelines for administering immunization to the Minister of Health of the Republic of Indonesia, the risk of diphtheria and tetanus in children with incomplete

immunization status is 46.403 times greater than children with complete immunization status. International agreement on the achievement of immunization success must be achieved, one of which is national immunization coverage in 2011-2020 at least 90%, for immunization coverage in districts/cities at least 80%, this is based on the Minister of Health Republic of Indonesia regulation number 12 of 2017 concerning immunization.

Previous studies discussed mapping in spatial-temporal modeling to determine the coverage of diphtheria-pertussis-tetanus vaccine using the Bayesian model-based geostatistical estimates method, the results obtained showed DPT3 coverage of state administration [11]. The spatial regression model was used to map predictions of diphtheria-tetanus-pertussis and measles immunization coverage using regional demographic data and field surveys, the results of this study can be used as an indicator of the success of health mapping for regional development [12]. Spatial analysis to provide information on the implementation of immunization strategies in the global community [13]. Geospatial analysis using the spatial Bernoulli model is used to determine the distribution of pertussis outbreaks based on nonmedical vaccine exemptions on factors that influence the spread factor at the community level. However, the weakness of previous research has not resulted in multi-class classification for mapping vulnerable areas based on factor immunization status coverage, PD3I, epidemics, nutrition status, and population density. The selection of AI methods adjusts to the behavior of data to be made mathematical modeling in the process of spatial analysis because the data in the field are still analog maps that refer to one epidemics variable.

The present paper proposition is spatial analysis through modeling spatial data in the GIS field using the WSM method to describe data in spatial datasets that will be processed using artificial intelligence (AI) methods[14], process to obtain the results of the same decisions and policies in each interest through the MCDM method [15][16]. The WSM method is used to map the classification of TD disease-prone areas based on immunization status factors in complete and incomplete categories by finding the best results from normalization in the criteria for each alternative. The advantage of the WSM method is that it can determine the weight value of the parameters specified by each attribute, with a ranking process that will select the best alternative from each alternative, where the assessment will be more appropriate based on preference weighting, normalizing according to attribute values in determining disease spread mapping based on attribute values.

The results of the spatial analysis using the WPM method [17][18][19] will produce a preference value; then the preference value will be calculated by the Guttman method [20][21]. The process of data in this spatial analysis took a case study data from 2011-2016 in 38 districts/cities in East Java Province, Indonesia. The results obtained for the classification category of tetanus-prone areas are namely. If the condition of the area is good, then the value of A_i is greater 55.13, if the value of A_i is greater 46.66 and A_i is less than 55.13 then it is in the fair category, and if the value of A_i is less than 46.66 then it belongs to the category of poor territory. The classification in the area category for diphtheria is obtained, if the value of A_i is greater 26.6, then it belongs to the good category, if the value of A_i is greater 22.6 and A_i is less than 26.6 then it belongs to the region fair category, and if the value of A_i is less than 22.6 then belongs to the category of the poor region. The application of this preference value will be used as a basis for processing the next series of data.

The application of the information produced will significantly benefit the Health Office as policymakers to take steps to anticipate the area with the classification of the poor. East Java, which is a region with a tropical climate, is still quite a concern for the health world. This is because the high rate of outbreaks is one indicator of the success of preventive measures in the health sector in the field of epidemiological surveillance.

2. Spatial Datasets

The spatial analysis process requires parameters in the description of spatial datasets, this process will provide an overview relating to the needs of spatial data and attribute data data, giving range values for each spatial attribute datasets of immunization status coverage, PD3I, epidemics, nutrition status, and population density such as shown in Table 1[2][3][4][5][6][7].

The coverage of immunization status is one of the most effective health protection measures for children against several diseases that can be prevented by immunization[11]. PD3I is an infectious disease that can be prevented by giving complete vaccination so that the number of TD cases can be restricted and experience a decrease [11]. The epidemic is an extraordinary occurrence of the spread of disease cases between regions in a country. Nutrition status is a description of nutrition prone areas based on the number of low-income family heads. Population density is a parameter indicator for assessing population density in certain areas with the total population according to the sub-district.

**Table 1.** Spatial Datasets Multi-Criteria Parameter TD

Spatial Datasets	The Priority Value	Weight	Range Multi-Criteria Parameter	Level of importance	Descriptions
Coverage of Tetanus immunization status	1	1	TT1 > 44,4	4	The impact of tetanus transmission on pregnant women who not immunize or do not get TT immunization.
	1	1	TT2 > 44,4		
	1	1	TT3 > 44,4		
	1	1	TT4 > 44,4		
	1	1	TT5 > 44,4		
	1	1	TT2+ > 44,4		
	1	1	TT1 ≤ 44,4	8	Knowledge of pregnant women is more valid for TT immunization.
	1	1	TT2 ≤ 44,4		
	1	1	TT3 ≤ 44,4		
	1	1	TT4 ≤ 44,4		
	1	1	TT5 ≤ 44,4		
	1	1	TT2+ ≤ 44,4		
Coverage of Diphtheria immunization status	1	1	(DPT1-HB1) < 84,5	4	Children do not get immunization vaccines, so they are susceptible to DT disease.
			(DPT3-HB3) < 84,5	4	
	1	1	(DPT1-HB1) ≥ 84,5	8	Children get immunization vaccines, so they are more secure in the body.
			(DPT3-HB3) ≥ 84,5	8	
PD3I	2	0,80	PD3I case ≥12 months per year	4	If in one year or more there is an increase in the rate of transmission of TD, there will be more cases of PD3I which will influence the results of the classification of TD (poor) disease-prone areas.

Spatial Datasets	The Priority Value	Weight	Range Multi-Criteria Parameter	Level of importance	Descriptions
			PD3I case <12 months per year	8	If in one year or less there are no cases of increasing rates of TD transmission, then there will be fewer or no cases of PD3I that will provide results of regional classification of categories not prone to TD (good).
Epidemics	3	0,60	Epidemic case ≥12 months per year	4	If in one year or more cases of outbreaks of TD disease occurs, there are still many who have not received incomplete diphtheria (poor) immunization coverage.
			Epidemic case < 12 months per year	8	If there is no outbreak case in one year or less, then it is classified as getting complete diphtheria immunization coverage (good).
Nutrition status (standard deviation/sd)	4	0,40	$sd \geq 2$	8	More nutrition, classified as a nutrient with a higher volume value
			$sd < 2$ && $sd \geq -2$	6	Good nutrition, classified as the maximum value.
			$sd < -2$ && $sd \geq -3$	4	Poor nutrition, classified as a lack of nutritional value.
			$sd < -3$	2	Poor nutrition, classified as a bad life and dense occupancy.
Population density	5	0,20	PD < 500	1	If the value of population volume is getting denser, then the spread of TD disease is increasing so that the weaker value of the expected cure rate.
			500 – 1.249	2	
			1.250 – 2.499	3	
			2.500 – 3.999	4	
			4.000 – 5.999	5	
			6.000 – 7.499	6	
			7.500 – 8.499	7	
			PD > 8.500	8	

3. Method

3.1 Spatial Analysis

Spatial analysis using the WSM method which is part of MCDM is a model by determining the priority value weights in each spatial datasets attribute as seen in Table 1, the weight will be multiplied by the data on each attribute to obtain an alternative multi-criteria decision value [18][19][22].
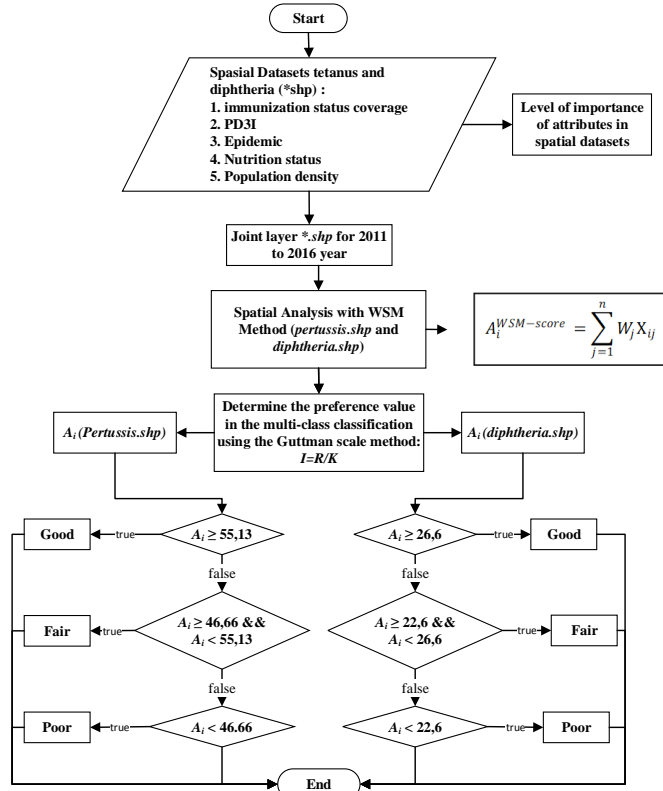
Fig. 1. Spatial Analysis of Prone Areas in Pertussis and Diphtheria Disease Using the WSM Method on MCDM

The flow of spatial analysis of prone areas in pertussis and diphtheria disease using the WSM method on MCDM in Figure.1, with the following steps:
1. Make a matrix of the weighted data, which consists of rows and columns, where the rows consist of attributes in the spatial dataset, and columns consist of sub-district name attributes.
2. Determining the value of the variable W_j which contains the weight of each criterion, for pertussis shown in equation (1), and for diphtheria in equation (2).

$$W_{j(Pertussis)=}[1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0,80 \quad 0,60 \quad 0,40 \quad 0,20] \qquad (1)$$
$$W_{j(Diphtheria)=}[1 \quad 1 \quad 0,80 \quad 0,60 \quad 0,40 \quad 0,20] \qquad (2)$$

3. Calculate the priority value for each spatial datasets attribute using equation (3) [19], where the variable n is the number of criteria used, W_j is the weight of each criterion in

spatial datasets, and the variable X_{ij} is the value on the X matrix. The value of A_i for pertussis, and diphtheria by equation (4), and equation (5), respectively.

$$A_i^{WSM-score} = \sum_{j=1}^{n} W_j X_{ij} \tag{3}$$

$$A_{i\,(Pertussis)} = (1 * TT1_i) + (1 * TT2_i) + (1 * TT3_i) + (1 * TT4_i) + (1 * TT5_i) + (1 * TT2+_i) + (0,80 * PD3I_i) + (0.60 * Epidemic_i) + (0.40 * Nutrition\_status_i) + (0.20 * Population\_Density_i) \tag{4}$$

$$A_{i\,(Diphtheria)} = (1 * DPT1\_HB1_i) + (1 * DPT3\_HB3_i) + (0,80 * PD3I_i) + (0.60 * Epidemic_i) + (0.40 * Nutrition\_status_i) + (0.20 * Population\_Density_i) \tag{5}$$

4. Calculate preference values to rank alternative values A_1 to A_i.

3.2 The Gutmann Scale Method

Calculates the value of the preference interval to determine the rank of each criterion in the multi-class classification using the Guttman scale method. The Guttman scale method is used to make measurements by looking at the scale value on the multivariate frequency distribution [23]. The reason for using this method is because the data in the field have not been found in quantitative data types [21] with definite value (ambiguous) for the range used for multi-class classification in diphtheria-susceptible pertussis disease areas in the good, fair, and poor categories as shown in Table 2.

This method will also provide measurements of classification values that are not necessarily true values by human perception, where the assessment is still not objective [20][24]. The Guttman scale method based on equation (5), where the value is taken from the score calculated by A_i in equation (5) for pertussis and (6) for diphtheria. R variable is the range of maximal A_i value reduced by A_i minimum value, K is the number of multi-class classification.

$$I = \frac{R}{K} \tag{6}$$

Multi-class classification for PD with results in formulas (7), and (8), where if the value of A_i gets smaller then the area belongs to the classification category of areas prone to diphtheria (poor) pertussis, and if the value of A_i is greater then it includes areas that are not vulnerable.

Table 2. Pertussis and Diphtheria scale assesment

Pertussis scale assessment	Diphtheria scale assessment
R = $A_{i\,(Pertussis)}Max - A_{i\,(Pertussis)}Min =$ 63,6 – 38,2 = 25,4	R = $A_{i\,(Pertussis)}Max - A_{i\,(Pertussis)}Min =$ 30,6 – 18,6 = 12
K = 3 ; where, $I = \frac{25,4}{3} = 8,47$	K = 3 ; where, $I = \frac{12}{3} = 4$
$assesment = A_{i\,(Pertussis)}Max - I \qquad = 63,6 - 8,47 = 55,13$	$assesment = A_{i\,(Pertussis)}Max - I =$ 30,6 – 4 = 26,6

| Pertussis scale assessment		Diphtheria scale assessment	
$\begin{cases} \text{Good,} & \text{if } A_i \geq 55{,}13 \\ \text{Fair, } if\ A_i \geq 46{,}66\ and\ UA_i < 55{,}13 \\ \text{Poor,} & \text{if } A_i < 46{,}66 \end{cases}$	(7)	$\begin{cases} \text{Good,} & \text{if } A_i \geq 26{,}6 \\ \text{Fair, } if\ A_i \geq 22{,}6\ and\ UA_i < 26{,}6 \\ \text{Poor,} & \text{if } A_i < 22{,}6 \end{cases}$	(8)

4. Results and discussion

Trial data obtained from the East Java Education Office in Indonesia in 2011-2016, there are 38 Districts / Cities with 657 Subdistricts. The results of the classification of pertussis diphtheria tropical disease-prone areas for the good, fair and poor categories are shown in Table 3. These results indicate that the occurrence of diphtheria pertussis disease cases that increase and decrease every year such as the incidence of cases described in the introduction chapter shows the same results with the classification values in Table 3.

Table 3. The classification results of areas prone to diphtheria pertussis tropical disease

	Pertussis tropical disease					
Class	Sub-District					
	2011	2012	2013	2014	2015	2016
Good	551	601	587	576	546	492
Fair	106	59	75	79	108	163
Poor	6	3	1	8	9	8
	Diphtheria Tropical Disease					
Good	566	595	513	77	507	558
Fair	50	27	44	485	12	18
Poor	47	41	106	101	144	87

The results of the classification of pertussis disease-prone areas in Klabang-Bondowoso Subdistrict are based on Figure 1 with values of DPT1, DPT3, TT3, TT4, TT5, TT2 +, PD3I, epidemic, nutrition status, and population density in Figure 2. 4) so that the class results are obtained according to equation (7).
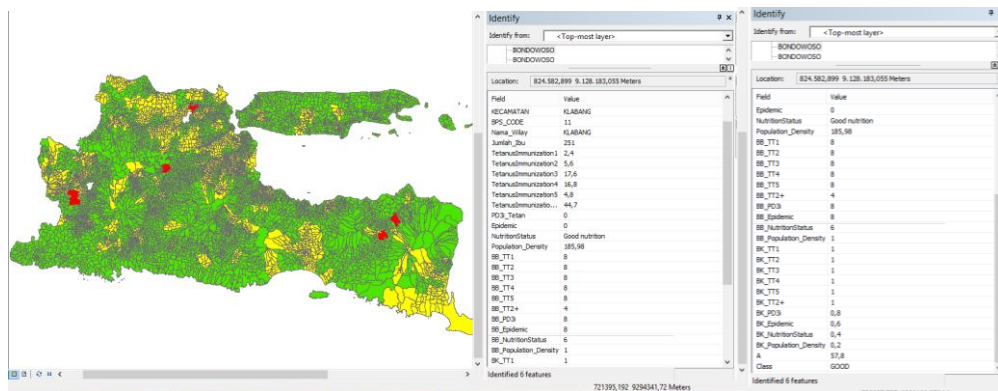
Fig. 2. Spatial Analysis of Prone Areas in Pertussis Disease Using the WSM Method on MCDM

Based on flow in Figure 1, the classification of diphtheria-prone areas was tested in Dasuk District, Sumenep Regency with a value of DPT1-HB1, DPT3-HB3, PD3I, epidemic, nutrition status, and population density in Figure 2, then the value was included with the formula (5) to obtain the classification results according to (8).
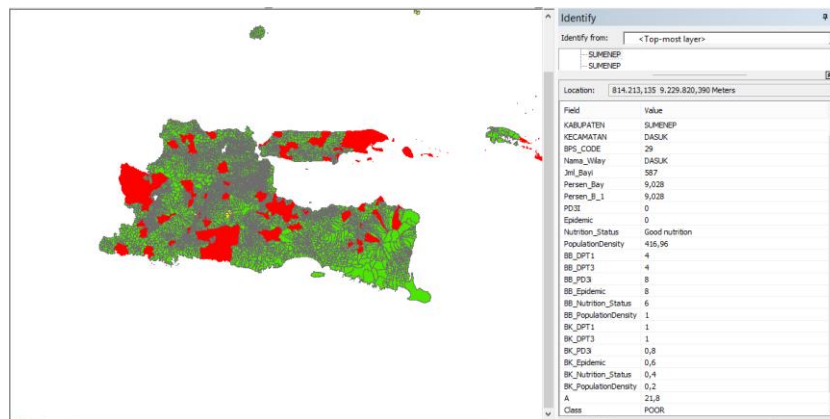
Fig. 3. Spatial Analysis of Prone Areas in Diphtheria Disease Using the WSM Method on MCDM

5. Conclusion

The discussion in this paper uses the WSM method on MCDM for the spatial GIS analysis process; the resulting method gets a significant trial value on the results report in the field with the resulting classification value. The description of the value of spatial datasets by determining the priority value, weight, range of multi-criteria criteria, and level of importance can provide a solution to the determination of MCDM to describe the sensitivity level of the parameters used.

The results of the preventive interval value with the Guttman scale method can be used as a reference for further research on quantitative series data types by utilizing the results of multi-class classification results that have been produced in formulas (7) and (8). Further research can be done by hybridizing the method to determine the level of sensitivity between the data used with the calculation of the resulting method.

Areas with the poor category will be a reference by the Government as Policy Makers to make preventive steps as part of disaster mitigation programs in handling the spread of diseases in a region [25]. It can also provide an understanding of the public to be more concerned about immunizing children or pregnant women as anticipation for increasing cases of diphtheria pertussis disease. The role of the community is very important for the implementation of free infectious disease programs in the community by utilizing GIS technology, such as GIS technology which can be used to find out the distribution of geospatial databases of health facilities [26], and strengthening the health sector [27].

Acknowledgments

References

[1] S. Kalra and P. Gupta, "50 Years Ago in THE JOURNAL OF PEDIATRICS: Epidemiology of Diphtheria—Role of Cutaneous Infection", *Journal of Pediatrics*, Vol. 209, No. August 2017, p.51, 2019.

[2] Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur Th 2011*. Dinas Kesehatan Provinsi Jawa Timur, 2012.

[3] Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur Tahun 2012*. Dinas Kesehatan Provinsi Jawa Timur, 2013.

[4] D. K. P. J. Timur, *Profil Kesehatan Provinsi Jawa Timur 2013*. Dinas Kesehatan Provinsi Jawa Timur, 2014.

[5] Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur 2014*. Dinas Kesehatan Provinsi Jawa Timur, 2015.

[6] Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur 2015*. Dinas Kesehatan Provinsi Jawa Timur, 2015.

[7] Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur Tahun 2016*. Dinas Kesehatan Provinsi Jawa Timur, 2017.

[8] Dinkes Jatim, *Profil Kesehatan Propinsi Jawa Timur 2017*. Dinas Kesehatan Provinsi Jawa Timur, 2018.

[9] J. A. Horney, "History of Disaster Epidemiology" , *Disaster Epidemiology*, pp.1–10, 2018.

[10] P. Pezzotti, S. Bellino, F. Prestinaci, S. Iacchini, P. Stefanelli, and G. Rezza, "Reply to Letters to the Editor: Bellavite P. Factors That Influenced the Historical Trends of Tetanus and Diphtheria. Donzelli A, Duca P. More than 70,000 Deaths Prevented by Vaccination against Three Diseases in about 75 years? The Estimation Seems Exaggerated" , *Vaccine*, Vol. 36, No. 37, pp.5508–5509, 2018.

[11] J. F. Mosser *et al.*, "Mapping Diphtheria-Pertussis-Tetanus Vaccine Coverage in Africa, 2000–2016: A Spatial and Temporal Modelling Study" , *The Lancet*, Vol. 393, No. 10183, pp.1843–1855, 2019.

[12] C. E. Utazi *et al.*, "A Spatial Regression Model for the Disaggregation of Areal Unit Based Data to High-Resolution Grids with Application to Vaccination Coverage Mapping" , *Statistical Methods in Medical Research*, 2018.

[13] M. E. Maravi, L. E. Snyder, L. D. McEwen, K. DeYoung, and A. J. Davidson, "Using Spatial Analysis to Inform Community Immunization Strategies" , *Biomedical Informatics Insights*, Vol. 9, p.117822261770062, 2017.

[14] M. Irfan, A. Koj, M. Sedighi, and H. Thomas, "Design and Development of a Generic Spatial Decision Support System, Based on Artificial Intelligence and Multicriteria Decision Analysis" , *GeoResJ*, Vol. 14, pp.47–58, 2017.

[15] M. Bystrzanowska and M. Tobiszewski, "How Can Analysts Use Multicriteria Decision Analysis?" , *TrAC - Trends in Analytical Chemistry*, Vol. 105, pp.98–105, 2018.

[16] A. Boggia, G. Massei, E. Pace, L. Rocchi, L. Paolotti, and M. Attard, "Spatial Multicriteria Analysis for Sustainability Assessment: A New Model for Decision

Making" , *Land Use Policy*, Vol. 71, No. November 2017, pp.281–292, 2018.

[17] E. Triantaphyllou and S. H. Mann, "An Examination of the Effectiveness of Multi-Dimensional Decision-Making Methods: A Decision-Making Paradox" , *Decision Support Systems*, Vol. 5, No. 3, pp.303–312, 1989.

[18] V. Maliene, R. Dixon-Gough, and N. Malys, "Dispersion of Relative Importance Values Contributes to the Ranking Uncertainty: Sensitivity Analysis of Multiple Criteria Decision-Making Methods" , *Applied Soft Computing Journal*, Vol. 67, pp.286–298, 2018.

[19] E. Triantaphyllou, *Multi-Criteria Decision Making Methods: A Comparative Study*. Springer, Boston, MA, 2000.

[20] R. E. Tractenberg, F. Yumoto, P. S. Aisen, J. A. Kaye, and R. J. Mislevy, "Using the Guttman Scale to Define and Estimate Measurement Error in Items over Time: The Case of Cognitive Decline and the Meaning of 'Points Lost'" , *PLoS ONE*, Vol. 7, No. 2, 2012.

[21] L. Guttman, "A Basis for Scaling Qualitative Data" , *American Sociological Review*, Vol. 9, No. 2, p.139, 1944.

[22] E. Mulliner, N. Malys, and V. Maliene, "Comparative Analysis of MCDM Methods for the Assessment of Sustainable Housing Affordability" , *Omega (United Kingdom)*, Vol. 59, pp.146–156, 2016.

[23] S. Psychology, G. Thomson, and W. Ledermann, *The Determinacy of Factor Score Matrices With Implications for Five Other Basic Problems of Common-Factor Theory*, Vol. VIII, No. 2. 1955.

[24] A. Stegeman, "A New Method for Simultaneous Estimation of the Factor Model Parameters, Factor Scores, and Unique Parts" , *Computational Statistics and Data Analysis*, Vol. 99, pp.189–203, 2016.

[25] P. J. Christopher and W. A. Lopez, "Hydatid Disease Notifications in New South Wales." , *Medical Journal of Australia*, Vol. 1, No. 2, pp.54–56, 1970.

[26] Q. Shafi, J. Qadir, and F. Jalal, "Using Geographic Information System (GIS) to Develop Health Information System (HIS) for Srinagar City, Jammu and Kashmir" , *International Journal of Advanced Remote Sensing and GIS*, Vol. 7, No. 1, pp.2589–2602, 2018.

[27] E. Asuo-mante, "A Pilot Trial of Applying Geographic Information System Technology to Health System Strengthening in The ..." , *Journal of Medical Informatics and Decision Making*, Vol. 1, No. 1, pp.1–9, 2017.