

Spatial Relationship of Drug Smuggling in Northern Thailand Using GIS-based Knowledge Discovery

Sumethat Niamkao and Ornprapa Robert*

Department of Environmental Science, Faculty of Science, Silpakorn University, Nakornpathom 73000, Thailand

ARTICLE INFO

Received: 30 Aug 2019
Received in revised: 6 Mar 2020
Accepted: 23 Mar 2020
Published online: 20 Apr 2020
DOI: 10.32526/ennrj.18.3.2020.26

Keywords:

Drug smuggling/ GIS knowledge discovery/ Spatial relationship/ Decision tree/ Prediction

* Corresponding author:

E-mail: robert_o@silpakorn.edu

ABSTRACT

The number of drug users has been growing, likely caused by oppressive social conditions. The drug situation in Thailand has changed so that it is no longer a production source. However, Thailand is one of the transit sites for narcotics smuggling. Drug smuggling occurs most recurrently along the border of Northern Thailand by topographic roads. Chiang Mai and Chiang Rai Provinces have been shown to have the highest statistics in terms of drug trafficking. In this investigation, eight districts adjacent to neighboring countries were chosen as the areas of study. These areas are Mae Chan, Mae Fa Luang, and Mae Sai located in Chiang Rai Province, as well as Fang, Chiang Dao, Mae Ai, Chai Prakan, and Wiang Haeng situated in Chiang Mai Province. This research studied the spatial relationship of factors related to narcotic smuggling using a data mining-based decision tree technique. The geographic locations of drug trafficking arrests were transferred into a data-mining process in order to assess the spatial relationships among types of exhibited drugs, season, land use, distance from checkpoint and smuggling routes. Drug smuggling risk areas were further predicted using decision tree modeling. The results revealed that the geographic locations of drug trafficking arrests in Mae Chan, Mae Sai, Mae Ai, and Fang Districts were related to the season factor. The distance from checkpoint showed a spatial relationship with drug smuggling arrests in the Chai Prakan District. Narcotic trafficking arrests in Mae Fa Luang were mostly related to land use and type of drug exhibited. Geo-locations of drug smuggling illustrated an independent relationship with smuggling routes. The results retrieved from the prediction-based decision tree method indicated that Chai Prakan, Mae Chan, Mae Sai, Mae Fa Luang, Mae Ai, Fang, Wiang Haeng and Chiang Dao Districts were high-risk drug smuggling areas. The precision value of prediction was found to be 0.652. These results could support spatial decision making for national drug smuggling monitoring and surveillance.

1. INTRODUCTION

According to the United Nations Office on Drug and Crime (UNODC, 2018), there were 275 million drug users worldwide in 2016, a number expected to reach 346 million drug users in 2019. It was observed that the demand of drug users was enhanced, which affected the increasing amount of drug suppliers. Narcotics production also showed an impact on the environment. The cultivation of substances used in production causes deforestation as well as ecological changes (UNODC, 1995). The study of Medel et al. (2015) examined marijuana and

opium smuggling in Mexico. Physical, socio-demographic, and drug violence factors were considered and included in smuggling cost estimation using the cost surface principle to predict the trafficking routes for marijuana and opium. Robert et al. (2018) applied the Analytic Hierarchy Process to determine the importance of drug smuggling factors including physical, narcotic and social demographic aspects prior to performing potential surface analysis. The results showed spatial risk areas for drug smuggling in northern Thailand. Ours and Williams (2012) studied the physical and health impact of

cannabis use. They explained that cannabis use reduced the mental and physical well-being of men and women. The number of drug addicts in Thailand has increased, with approximately two million users discovered (ONCB, 2017). The ONCB (2017) also reported that drug smuggling was found mainly along the border of northern Thailand, originating from the 'Golden Triangle', which includes Mae Chan, Mae Fa Luang, and Mae Sai Districts in Chiang Rai Province, as well as Fang, Chiang Dao, Mae Ai, Chai Prakan, and Wiang Haeng Districts in Chiang Mai Province. In this research, the study areas covered these districts.

Knowledge Discovery from Database (KDD) is a multi-step process involving data warehousing, pattern searching or data mining, knowledge evaluation, and refinement with repetition after modification in order to extract useful information or knowledge retrieved from a collection of data or a set of factors (Fayyad et al., 1996; Harvey and Jiawei, 2009). KDD was used to study the relationship between suicide and temperature change in Mexico (Fernández-Arteaga et al., 2016). Warehousing of suicide using hanging means from 2005 to 2012 was prepared in order to investigate association rules-based data mining. The results illustrated that the highest number of suicides among Mexican men was found mainly on dry days from 30 °C to 40 °C. KDD was also applied to study airline passenger management in Taiwan (Wong and Chung, 2007). In this paper, we applied Geographic Information System (GIS) KDD to discover the spatial relationship of factors related to drug smuggling in northern Thailand. This approach made it different from the study of Fernández-Arteaga et al. (2016). GIS KDD could provide data concerning the geo-specific relationship in knowledge or a set of factors. Moreover, the importance of drug smuggling factors was excluded, which was different from the study of Robert et al. (2018). To the extent of our knowledge, GIS KDD has not previously been applied in drug trafficking investigation.

As mentioned earlier, data mining is one of the GIS KDD multi-step processes. The data mining technique is a process used to analyze key data, identify data relationships, and eventually predict the probability of an expected incident. Data mining is divided into two methods, including unsupervised and supervised learning techniques (Hand et al., 2001). The unsupervised technique comprises clustering and association rule. Thomas et al. (2018) applied

clustering means to monitor the chemical process in the Eastman chemical industry. Clustering means could be also applied in environmental study, as seen in the study of Ge et al. (2019), who used clustering in an unsupervised learning technique to extract soil structure information. Calçada et al. (2019) studied the relationship of parameters in fertilizer using the association rule to discover the appropriate parameters of decomposition. Xu et al. (2018) employed the association rule to investigate the relationship for factors affecting traffic congestion. On the other hand, the supervised learning technique in data mining is composed of two methods, which are regression and classification. Data mining-based regression is often applied in environmental studies. Oliveira et al. (2017) predicted sea mussel populations using the data mining-based regression method. Weather conditions, phytotoxin episodes, stock-biomass indicators per species, tourism levels, and surveyed mussel populations were considered in multiple linear regressions. The second method for the supervised learning technique is classification. Kesavaraj and Sukumaran (2013) explained that the supervised learning-based classification technique is a method used to categorize data (a set of factors) for discovering data relationships and prediction. The supervised learning-based classification technique is comprised of Bayesian network, decision tree, K-nearest neighbor (KNN), and neural network algorithms. These classification means are referred to as the prediction technique. Nguyen et al. (2017) used the KNN data mining technique to predict the volume and water retention in soil. Peter et al. (2018) applied the supervised learning-based classification technique in an epidemiology study. Bayesian network was applied to predict the risk areas for malaria in northern Thailand. Zhao et al. (2016) used the decision tree method to explain the relationship in the change of underground water level, rainfall and reservoir level as well as predict landslides in Shaxi city, China. In this study, the aim was to discover the relationship of factors related to drug smuggling in northern Thailand using the data mining-based decision tree method, as well as to predict the geographical risk areas for drug smuggling using the classification-based decision tree technique.

2. METHODOLOGY

The methodology of GIS KDD for investigating the spatial relationship of drug smuggling factors and predicting drug smuggling risk

areas comprises three sections, which are (2.1) the collection and manipulation of drug trafficking arrest data, (2.2) finding the spatial relationship for factors related to drug smuggling-based data mining, and

(2.3) the prediction of drug smuggling risk areas. The conceptual framework of the study is illustrated in Figure 1 and explained below.

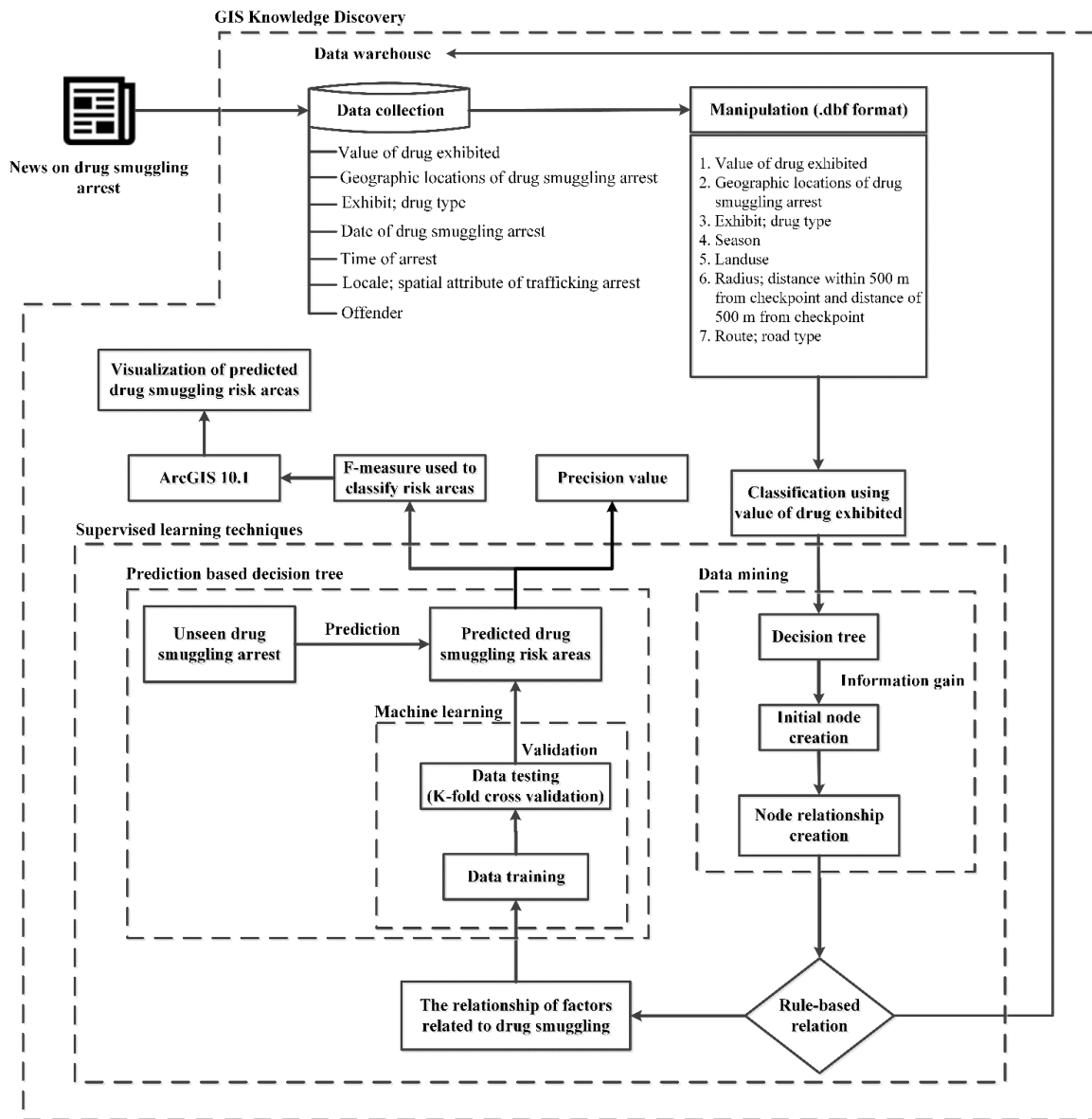


Figure 1. The conceptual framework of the study

2.1 Collection and manipulation of drug trafficking arrest data

Table 1 displays an example of the narcotic trafficking arrest data transformed into a database file (.dbf format) containing X and Y, composed of geographic locations for drug smuggling arrest based datum WGS 1984 Universal Transverse Mercator, date of drug smuggling arrest, time of arrest, Locale; spatial attribute of trafficking arrest, offender, exhibit; drug type and value of drug exhibited in United States Dollar (USD). The collected data shown in Table 1

was manipulated and transformed into seven factors prior to transfer into the data mining process. This data manipulation can be elaborated as follows: (1) Value of drug exhibited factor was categorized into five ranges including 0.033 USD to 33,000 USD, 33,001 USD to 330,000 USD, 330,001 USD to 3,300,000 USD, 3,300,001 USD to 33,000,000 USD and above 33,000,000 USD; (2) Location factor (L1-L5) referred to the spatial sites for drug smuggling arrests, which were divided by five district boundaries including Mae Chan and Mae Sai District (L1),

Chiang Dao and Wiang Haeng District (L2), Chai Prakan District (L3), Mae Ai and Fang District (L4), and Mae Fa Luang District (L5); (3) Exhibit factor referred to different smuggled drug types, which comprised Amphetamine or Ya-Ba (D1), Crystalline methamphetamine or ICE (D2), *Papaver somniferum* L. (D3), *Mitragyna speciosa* K. (D4), Heroine (D5), Substances (D6), and Exhibit more than one type (D7); (4) Season factor was categorized into three seasons including summer (Feb-May), rainy (Jun-Oct) and winter (Nov-Jan); (5) Land use data was retrieved from the Thailand Land Department and

classified into 5 categories including agriculture (A), residence (R), forest (F), water (W) and miscellaneous (M); (6) Radius factor means the distance within 500 meters from checkpoint (T) and the distance more than 500 meters from checkpoint (F), where drug smuggling occurred, and (7) Route factor referred to the road type where drug smuggling occurred. Route characteristics were divided into main or highway (M), sub-route (S) and topographic types (T). The seven factors mentioned above were transferred into the data mining process to discover the spatial relationships among them.

Table 1. Collected data for narcotics trafficking arrests

Value of drug smuggling exhibited	X	Y	Exhibit	Date of arrest	Time	Locale	Offender
500,000	510238.00	2186339.00	20,000 tablets of Amphetamine	02052011	12.00	Lychee garden ¹	Name A Age 36
52,008	514246.00	2172072.00	1,970 tablets of Amphetamine	25082011	13.00	Phahong checkpoint ²	Name B Age 23
37,880	592503.00	2253528.00	3.5 kg of Heroine	20022011	13.00	Sub-route ³	Name C Age 30

¹Located in Mae Ngon sub-district, Fang District, Chiang Mai; ²Located in Sridongyen sub-district, Chai Prakan District, Chiang Mai; ³Located in Pong Pha sub-district, Mae Sai District, Chiang Rai

2.2 Finding the spatial relationship in factors related to drug smuggling-based data mining

Data mining using decision tree-based classification is one of the supervised learning techniques. It is implemented to find out the spatial relationship in factors related to drug smuggling. Data mining using decision tree-based classification comprises two steps, which are node creation and ruling spatial relationship of mined data or factors transferred from step 2.1. In the node creation process, factors are transformed into nodes prior to investigation of the spatial relationship. Data training with information gain is employed to specify the initial node describing the most important factors and the non-initial nodes of the study. In the second step of data mining, the designated initial node was investigated for spatial relationship associated with the remaining nodes or factors. In this research, Rapid miner data mining software (Kotu and Deshpande, 2015) was applied. Drug trafficking data was discrete. Hence, the C4.5 principle was used in this investigation. The results of the spatial relationship in factors related to drug smuggling are illustrated as a hierarchy or decision tree structure (Figure 3).

2.3 Prediction of drug smuggling risk areas

The spatial relationship of factors related to drug smuggling retrieved from a previous step was transferred into prediction modeling based on the decision tree approach to reveal the geographic risk areas for drug smuggling. The prediction approach comprised three steps including (1) machine learning for training data related to spatial relationship in drug smuggling, (2) data testing-based K-fold cross validation, and (3) the prediction of drug smuggling risk areas. Firstly, machine learning was applied to investigate the accuracy of spatial relationship for factors related to drug smuggling. Secondly, the data related to spatial relationship of drug smuggling was divided into four categories using K-fold cross validation. One quarter of the data was tested with the remaining data to study the accuracy of data training. Thirdly, unseen drug smuggling arrest data in 2018 was used to validate the predicted drug smuggling risk areas. The prediction value explains the accuracy of predicted drug smuggling areas, which is between 0.0 and 1.0. A precision value close to 1.0 shows higher accuracy. GIS classification technique in ArcGIS 10.1 software was applied to prepare the drug smuggling

risk map. Drug smuggling risk was classified into 4 classes using the F-measure value retrieved from validation analysis as a classifier (Figure 4).

3. RESULTS AND DISCUSSION

3.1 Drug smuggling data warehouse

A drug smuggling data warehouse was prepared by capturing news of drug smuggling arrests in northern Thailand. There were 564 news articles collected related to drug smuggling arrests from 2011 to 2017. These articles were manipulated and transformed into GIS format, as seen in Figure 2. It shows the geographic locations of drug smuggling arrests in northern Thailand overlaid with land use, route, checkpoints, and distance from checkpoints layers. The study area covers eight districts of Chiang Mai and Chiang Rai Provinces. The population of

these two provinces is approximately 505,403 people (The Bureau of Registration Administration, 2019). The majority of the local population is related to agriculture. This area is a central point connecting to neighboring countries, and the pathway for the Belt and Road Initiative. Hence, this area attracts international investors. The spatial attribute of drug smuggling arrests in the study area was also included in order to investigate the geo-spatial relationship of factors related to narcotics smuggling. Prior to delivering these factors into the Rapid miner data mining software, determining the factors or nodes was carried out. There were seven nodes, as explained in section 2.1. The factor of drug value exhibited was designated to use in the data classification process prior to implementing data mining, as shown in the results in section 3.2.

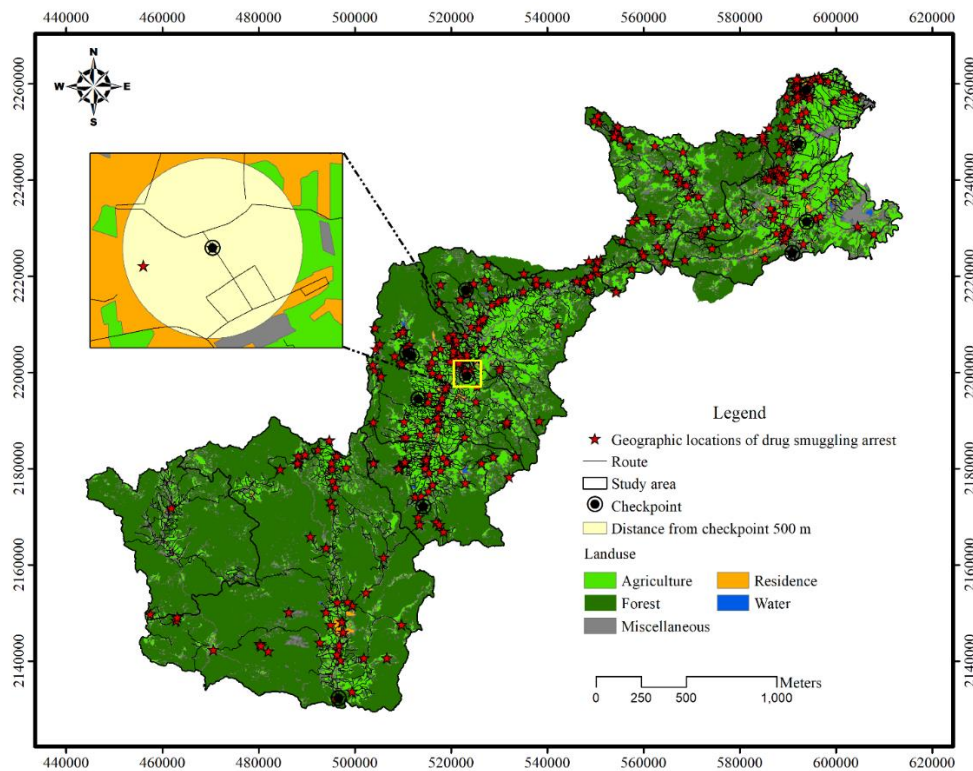


Figure 2. Geographic locations of drug smuggling arrests in northern Thailand

3.2 Spatial relationship of drug smuggling factors

From node creation using data training method, the information gain for factors including geographic location of drug smuggling arrest, type of exhibited drug, land use, season, distance from checkpoint and smuggling route was studied. The information gain for the factors was found to be 0.518, 0.348, 0.194, 0.179, 0.135, and 0.060, respectively. The most influential factor to be specified as an initial node was

the geographic locations (L1-L5) of drug smuggling arrests, as shown in Figure 3. It can be further explained that the narcotics smuggling incidents found in Mae Chan and Mae Sai Districts (L1) and Mae Ai and Fang Districts (L4) were related to seasons. The peak season for illegal drug trafficking-based value of drug exhibited (in 330,001-33,000,000 USD range) was discovered in the rainy season. The amount of trafficked drugs per time in the rainy

season was generally higher than the quantity of smuggled drugs per time that occurred in the summer and winter. This was to minimize the risk of damage to drugs while smuggling. Drug trafficking incidents in Chiang Dao and Wiang Haeng Districts (L2) displayed the highest value of drug exhibited (More than 33,000,000 USD) and no spatial relationship with other factors. The drug trafficking information in Chiang Dao and Wiang Haeng District was unclassified during relationship creation, showing less of a relationship in drug trafficking information. This was because the entropy of drug trafficking information in Chiang Dao and Wiang Haeng Districts was too high. Regarding the factor of radius from checkpoint, the results illustrated that the distance within 500 meters from checkpoint (T) and the distance more than 500 meters from checkpoint (F) explained the spatial relationship in Chai Prakan District (L3). The comparison between the drug value exhibited within 500 meters from a checkpoint and the distance of 500 meters from a checkpoint described that the value of drugs exhibited within the distance 500 meters from a checkpoint was less than the drugs exhibited from the distance more than 500 meters from a checkpoint. Smugglers aimed to be less suspicious while passing a checkpoint. Narcotics

trafficking incidents in Mae Fa Luang District (L5) were related spatially to land use and drug type. The value of drugs exhibited in 0.033-330,000 USD range was shown in the residential area of Mae Fa Luang District. In the agricultural area of this district, the higher value of drugs exhibited was displayed as 3,300,001-33,000,000 USD and the most prevalent drug type for those arrested was heroine. In accordance with the news for illegal drug trafficking arrests in Mae Fah Luang District, the percentage of amphetamine and heroin smuggling in agricultural area was found at 55 and 31, respectively. While heroin trafficking illustrated a lower percentage of arrest information than amphetamine smuggling, the cost of heroin production and exhibited values were higher than for amphetamine. Additionally, the factor of smuggling route was shown to have an independent relationship with other factors. According to the study of [Medel et al. \(2015\)](#), the transportation type could reduce the cost of drug trafficking. In this investigation, it was found that drug smuggling incidents occurred at sub-routes and topographic routes at percentages of 85 and 15, respectively. This was because the drug trafficking data displayed in public news was related to an intercepting and blockade search strategy.

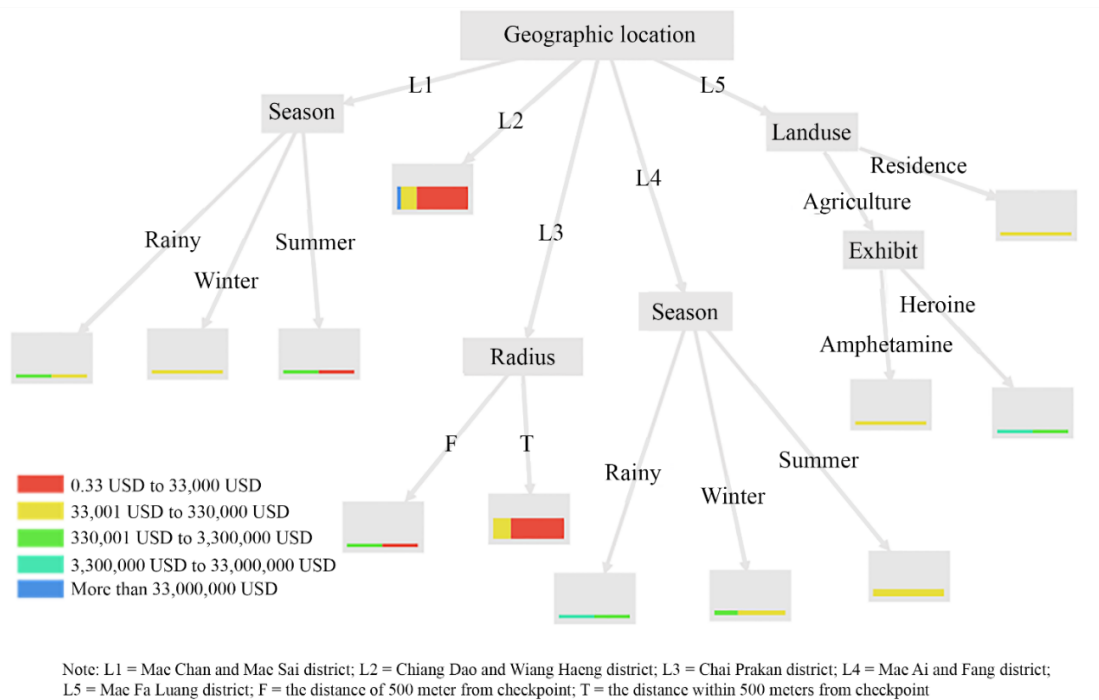


Figure 3. The decision tree structure of spatial relationship for drug smuggling factors

3.3 Predicted drug smuggling risk areas

The results of the narcotics smuggling prediction-based decision tree approach are as follows. The machine learning technique presented the accuracy of spatial relationship for factors related to drug smuggling at a percentage of 63.5. K-fold cross validation was applied to test the data used in the machine learning process and explained the validation accuracy at a percentage of 57.1. Increasing the amount of data transferred into the prediction approach is possible to enhance the validation accuracy. The predicted geographic risk of drug smuggling was classified into high, medium, low and very low risk level, as illustrated in Figure 4. The

drug smuggling risk area for each level was Chai Prakan (F-measure at 0.676), Mae Chan, Mae Sai (F-measure at 0.412), Mae Fa Luang (F-measure at 0.222), Mae Ai, Fang, Chiang Dao, and Wiang Haeng District (F-measure at 0.000). The precision value of prediction was discovered at 0.652, describing moderate accuracy for predicted drug smuggling areas. In order to improve the accuracy, other factors should be considered including drug production site, transit site, and ambush and blockade location. These factors are not publicly available due to data confidentiality issues; hence, it is a limitation of this study.

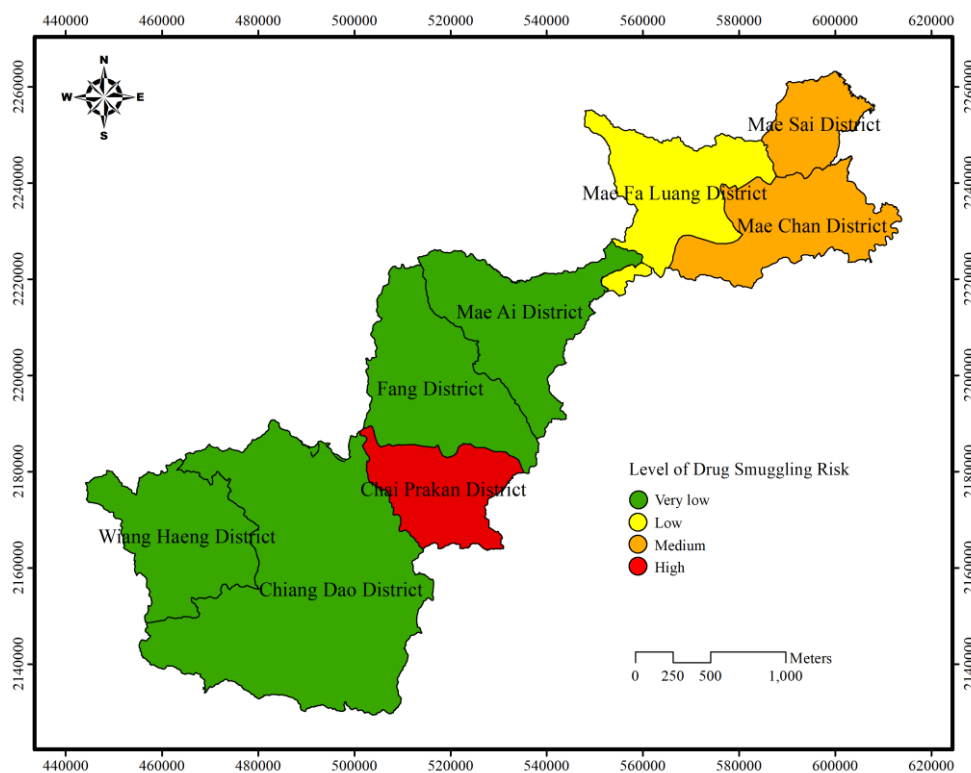


Figure 4. Predicted Drug Smuggling Risk Areas

4. CONCLUSION

To the best of our knowledge, we demonstrated that the GIS KDD method was able to search for patterns and extract key information related to drug smuggling. Data mining-based decision tree made it possible to identify spatial relationships for narcotic smuggling in northern Thailand. The results of spatial relationship for narcotic smuggling are illustrated in graphic format. This approach is less complex and easier to manage since the amount of data is limited. Drug smuggling risk areas were predicted by using a supervised learning-based decision tree classification

technique. The precision value for predicted drug risk areas was discovered at 0.652, which was moderately acceptable. The precision of prediction relies on the entropy of data as well as the amount of input data or factors.

ACKNOWLEDGEMENTS

This research was partially supported by the Department of Environmental Science, Faculty of Science, Silpakorn University, Thailand and the geodatabase project for military and defense. The authors would like to thank the Office of the Narcotics Control

Board for its assistance and comments, which greatly improved the manuscript.

REFERENCES

- Calçada DB, Rezende SO, Teodoro MS. Analysis of green manure decomposition parameters in northeast Brazil using association rule networks. *Computers and Electronics in Agriculture* 2019;159:34-41.
- Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery: An overview. In: Fayyad U, Piatetsky-Shapiro G, Smyth P, Ulthurusamy R, editors. *Advances in Knowledge Discovery and Data Mining*. Cambridge, United States: MIT Press; 1996. p. 1-34.
- Fernández-Arteaga V, Tovilla-Zárate CA, Fresán A, González-Castro TB, Juárez-Rojop IE, López-Narváez L. Association between completed suicide and environmental temperature in a Mexican population, using the Knowledge Discovery in Database approach. *Computer Methods and Programs in Biomedicine* 2016;135:219-24.
- Ge Q, Xiong F, Xie L, Chen J, Yu M. Dynamic interaction of soil-structure cluster. *Soil Dynamics and Earthquake Engineering* 2019;123:16-30.
- Hand D, Mannila H, Smyth P. *Principles of Data Mining*. Cambridge, United States: MIT Press; 2001.
- Harvey JM, Jiawei H. *Geographic Data Mining and Knowledge Discovery*. Boca Raton, United States: Taylor and Francis Group; 2009.
- Kesavaraj G, Sukumaran S. A study on classification techniques in data mining. *Proceedings of the Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*; 2013 Jul 4-6; Tiruchengode: India; 2013.
- Kotu V, Deshpande B. *Predictive Analytics and Data Mining: Concepts and Practice with Rapidminer*. Waltham, United States: Elsevier; 2015.
- Medel M, Lu Y, Chow E. Mexico's drug networks: Modeling the smuggling routes towards the United States. *Applied Geography* 2015;60:240-7.
- Nguyen PM, Haghverdi A, Pue J, Botula YD, Le KV, Waegeman W. Comparison of statistical regression and data-mining techniques in estimating soil water retention of tropical delta soils. *Biosystems Engineering* 2017;153:12-27.
- Office of the Narcotics Control Board. Drug situation in the northern region January 2017 [Internet]. 2017 [cited 2018 Aug 6]. Available from: <https://www.oncb.go.th>.
- Oliveira MM, Camanho AS, Walden JB, Miguéis VL, Ferreira NB, Gaspar MB. Forecasting bivalve landings with multiple regression and data mining techniques: The case of the Portuguese artisanal dredge fleet. *Marine Policy* 2017;84:110-8.
- Ours JV, Williams J. The effects of cannabis use on physical and mental health. *Journal of Health Economics* 2012;31:564-77.
- Peter H, Hasan I, Kasantikul R, Lawpoolsri S, Sa-angchai P, Kaewkungwal J. Spatiotemporal bayesian networks for malaria prediction. *Journal of Artificial Intelligence in Medicine* 2018;84:127-38.
- Robert O, Kumsap C, Janpengpen A. Simulation of counter drugs operations based on geospatial technology for use in a military training simulator. *International Journal of Simulation and Process Modeling* 2018;13:402-15.
- The Bureau of Registration Administration. Official statistics registration systems [Internet]. 2019 [cited 2020 Feb 26]. Available from: <http://stat.bora.dopa.go.th>.
- Thomas MC, Zhu W, Romagnoli JA. Data mining and clustering in chemical process databases for monitoring and knowledge discovery. *Journal of Process Control* 2018;67:160-75.
- United Nations Office on Drugs and Crime. World drug report 2018 [Internet]. 2018 [cited 2019 Jun 16]. Available from: <https://www.unodc.org/wdr2018>.
- United Nations Office on Drugs and Crime. The social impact of drug abuse [Internet]. 1995 [cited 2019 Oct 13]. Available from: https://www.unodc.org/pdf/technical_series_1995-03-01_1.pdf.
- Wong JY, Chung PH. Managing valuable Taiwanese airline passengers using knowledge discovery in database techniques. *Journal of Air Transport Management* 2007;13:362-70.
- Xu C, Bao J, Wang C, Liu P. Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. *Journal of Safety Research* 2018;67:65-75.
- Zhao Y, Li Y, Zhang L, Wang Q. Groundwater level prediction of landslide based on classification and regression tree. *Geodesy and Geodynamics* 2016;7:348-55.