DOI: http://doi.org/10.4038/sljastats.v23i2.8065

Sri Lankan Journal of Applied Statistics, Vol (23-2)

Statistical Approach of Identifying Crime Hotspots for GIS Mapping in Sri Lanka

M.A.P. Munasingha* and N.A.D.N. Napagoda

Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, 60200, Sri Lanka

*Corresponding author: poornimamunasingha@gmail.com

Received: 16^{th} March 2022/ Revised: 21^{st} August 2022/ Published: 25^{th} November 2022

ABSTRACT

Currently, crimes are well organized and difficult to investigate, and hence, the situation is more complex due to technological advancements. Therefore, there is a timely need to plan strategies to reduce crime using historical crime data. The objectives of the study are extraction of crime patterns using trend analysis, identifying the relationship between crime and geographic environment, and recognizing the district where crime is most prevalent. Five types of crimes that have a high impact on society from 2010 to 2019 in Sri Lanka were identified using the high mean cluster. There were House Breaking (HB), Hurt by Knife (HK), Robbery (RB), Rape (RP), and Cheating (CH). The result of Shapiro-Wilk test for normality wasn't normal therefore Kruskal-Wallis multiple comparisons were used. There was a significant difference between the mean of HB, HK, RB, RP, and CH. Using Spearman's rank correlation coefficient test, a very strong positive correlation was found between RP and HB (0.9472), RB and HB (0.9003) and HK and HB (0.9277). The type of crimes scattered throughout Sri Lanka was analyzed using Geographic Information System (GIS). Thematic maps were generated to identify hotspot areas. The risk of all five types of crime appeared to be high in Kurunegala and Gampaha districts. Prediction Accuracy Index (PAI) values were calculated to compare the predictive accuracy of crime types and the highest PAI value (3.98) was for RB crime. According to the findings of this study, a new security strategy can be developed to eradicate these trends from society.

Keywords: Crime, GIS, Hotspot, Mapping, Prediction Accuracy Index

1 Introduction

The wave of crime is a key social problem in Sri Lanka. Presently crimes are well organized and shows increasing trends with the rising population and rapid advancement of modern technologies than earlier. Crime is a social stimulus and our society suffers in many ways. Therefore, take necessary steps for the reduction of crime in the society is vital to maintain its well-being. Many types of crime are prevalent in society today. Kidnapping, Arson, Mischief over Rs. 5000, House Breaking, Grievous Hurt, Hurt by Knife, Homicide, Attempted Homicide, Rape, Riots, Robbery, Unnatural Offence/ Grave Sexual Abuse, Extortion, Cheating Crimes are targeted in this research. The increase in crime has severely affected the lives of the people. As a result, society is deprived of the opportunity to spend time freely. Therefore, it is very timely to formulate programs to reduce crimes. Hence, this research was focused to identify the pattern of crime scattered throughout Sri Lanka. Crime can be minimized by identifying high crime areas to facilitate providing greater security in those hotspot areas. This research is used annual total crime data ranges from 2010 to 2019 to generate hotspot maps, and test their accuracy for predicting where crimes will occur next. The Prediction Accuracy Index (PAI) was used to compare hotspot mapping accuracy based on crime types (housebreaking and theft, cheating, robbery, rape, and hurt by knife).

Crime statistics become a way to decrease future crimes. Our main goal was to identify the districts where crime is the most prevalent using Geographic Information System (GIS). GIS is one of the most effective quantitative data analysis methods currently used by security agencies to improve the quality of criminal investigations and it plays a very important role in crime mapping and analysis. This analysis can be used to determine where police officers should be deployed based on crime hotspot areas. While this research is not enough to predict exactly where crime will occur the next year, crime can be minimized by identifying hotspots and focusing more on those areas with minimal use of resources.

The rest of the paper is organized as follows: section 2 focuses on literature survey. The proposed methodology is presented in section 3. Section 4 contains details of the data analysis and the results of the analysis. Finally, Section 5 concludes with a discussion.

2 Literature Review

Criminal analysis has been found in several fields of academic literature to identify the relationship between crime and its prevalence. Researchers have developed various algorithms and classifications for this purpose, and the following is a summary of some of the research conducted on crime analysis around the world.

Ratnayake (2015) focused on identifying the relationship between crime and geographic environment. In addition, finding out hot spots in a location

where different crimes keep occurring over an extended period of time and finding out the location which signals the police officers to be more attentive. The analysis was mainly carried out using Global Positioning System (GPS) and Geographic Information System (GIS) techniques. Crime records of arrests and crime events of the Mirihana Police Division in 2013 were used for this research. Results indicate commercial clusters and crimes had a positive relationship. Similarly, the demarcated low-income household and crimes also had a positive inter-relationship. This research was limited to identifying crime locations, time, and the factors that directly and indirectly affect crimes in the Mirihana Police Division of Sri Lanka. In addition, Wickrama, et al (2020) studied forecast homicides, rapes, and counterfeiting currency from 2013 to 2020 using Auto-Regressive Conditional Poisson (ACP) and Auto-Regressive Integrated Moving Average (ARIMA) models. Data were obtained from the Police Department of Sri Lanka and the Department of Census and Statistics. According to the results, the random K-closest neighborhood (RKNN) algorithm classified districts as safe and unsafe with a forecast accuracy of 84%. Moreover, it was found that the total migrant population and percentage of urban population was positively correlated with total crime. This research was limited to homicides, rapes and counterfeiting currency crimes in Sri Lanka.

Kedia (2016) utilized Geographic Information System (GIS) for crime mapping and its analysis for effective law enforcement and crime management. Heat map is created with the daily crime data of two month. The study proved that GIS gives us a better synoptic perspective to crime mapping, study, analysis, decision making, and thus prevents crime. This study was limited to explore the capability of GIS in crime mapping using the Faridabad District of Haryana in India. In addition, Chainey, et al (2008) introduced the Prediction Accuracy Index (PAI) as a measure to compare prediction abilities of crime mapping techniques. It is also identified if the most common hotspot mapping techniques used by practitioners differ in their ability to predict spatial patterns of crime. Daily crime data obtained by the Metropolitan Police Division in London were used from 2002 to 2003. This research examined whether hotspot maps for different types of crime differ in their predictive capabilities. But it appeared that considering the volume of input data and specifying the parameters carefully can be used to accurately predict spatial patterns of crime.

According to past studies, crimes in our country as well as in different countries have been investigated for various purposes. The main focus of our study was to identify the most common types of crime in Sri Lanka and to study their prevalence in districts. This investigation provided information on crime mitigation in Sri Lanka. The increase in crime has severely affected the lives of the people and deprives society of the opportunity to spend its free time. It is very timely to formulate programs to reduce crime. Crime can be minimized by identifying high crime areas and providing greater security for those areas. Data can be represented whether crime is rising or falling in tar-

geted areas. This can help criminal justice professionals understand whether their initiatives are successful. Also, map updating is a key resource for crime control.

3 Methodology

3.1 Data Sources

In this study, the required data were collected from the Department of Police and the Department of Census and Statistics, Sri Lanka (Sri Lanka Police, 2020). Annual total crime data ranges from 2010 to 2019 were used for further analysis. This dataset contained crime events collected district-wise with 14 different categories of crime. Missing values were imputed by a median of the respective variable.

3.2 Development of the design of the study

The crime rate varies across individual districts and could be more or less than the overall crime rate of Sri Lanka. Therefore, districts were ranked and categorized as safe and unsafe districts. If a crime rate of a district was below the overall crime rate, it was considered as a safe district and if the crime rate of a district was more than the overall crime rate, it was considered as an unsafe district (Wickrama et al., 2020). The Crime rate was calculated based on the population and land area of a district. Since the last census in Sri Lanka was conducted in 2012, 2012 population data was used to find the crime rate. Therefore, data from different crime types in 2012 were analyzed for each district.

$$Crime\ rate\ 10^5\ population = \frac{(Total\ crimes\ in\ a\ district)}{(Total\ population\ in\ that\ district)} \times 10^5 \tag{1}$$

and

$$Crime\ rate\ per\ 1km^2 = \frac{(Total\ crimes\ in\ a\ district)}{(Total\ area\ of\ a\ district\ in\ 1km^2)} \tag{2}$$

Using descriptive statistics, a description about the location distribution of criminal activities-including information was generated and five different types of crimes that were a higher impact on society were identified. They are House Break and Theft (HB), Cheating (CH), Robbery (RB), Hurt by knife (HK), and Rape (RP).

The Shapiro – Wilk Test (Shapiro & Wilk,1965) was a test of normality in frequentist statistics. This method was applied to check the normality in this

research.

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \tag{3}$$

Where,x_i is i[th] order statistic, $\bar{x}$ is sample mean, and a_i is the coefficients.

The Kruskal-Wall is test used ranks of ordinal data to perform an analysis of variance to determine whether types of crimes were similar to each other.

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{n} \frac{{R_j}^2}{n_j} - 3(N+1) \tag{4}$$

Where, R_j is sum of the ranks in the j[th] sample or groups,n_j is number of observations in the j[th] sample, k is number of samples, and N is number of observations in the combined sample.

Dunn's Multiple Comparison Test was used to pinpoint which specific means were significant from the others.

$$Z = \frac{\left|\frac{R_i}{n_i} - \frac{R_j}{n_j}\right|}{\sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \tag{5}$$

Where, R_i and R_j is rank sum of the groups being compared, n is the total sample size, and n_i and n_j is the size of the groups being compared.

The Spearman rank coefficient was calculated to determine how well two variables for individual data points can predict each other. Test statistic (ρ) is from -1 to 1, with -1 being a perfect negative correlation and 1 a perfect positive correlation.

$$\rho = 1 - \frac{6 \sum {d_i}^2}{n(n^2 - 1)} \tag{6}$$

Where, ρ is spearman's rank correlation coefficient, $d_i$is difference between the two ranks of each observation, and n is number of observations.

Trend analysis was used to fit a general trend model to detect crime data and to provide forecasts. The trend equation was used to calculate the forecast for specific time values.

$$y_t = \beta_0 + \beta_1 t + e_t \tag{7}$$

Where, y_t is the value off the time series in period t and e_t is the error term in the time period t. The intercept β_0 and the slope β_1 are unknown constants, and they are both called regression coefficients.

Quantum Geographic Information System (QGIS) is a free, open-source Geographic Information System (GIS) application allowing users to visualize, manage, edit, and analyze data. Hotspots were determined by selecting

the uppermost thematic class calculated based on the five classes (The five classes are shown in Figure 8,9,10,11,12,13) and the default values of the five classes were automatically generated by applying the equal count (Quantile) thematic range method in QGIS. Districts with values in only the top thematic class were then selected, with these areas determined as the hotspots. This approach is used as the number of classes falls within the upper and lower settings specified and the quantile method is chosen. It distributes the data in an approximately equal balance between the classes, resulting in a visually balanced map pattern (Chainey et al., 2008).

Prediction Accuracy Index (PAI) (Chainey et al., 2008) was utilized to consider the hit rate against the areas where crimes were predicted to occur with respect to the size of the study area.

$$PAI = \frac{Hit\ Rate}{Area\ percentage} \tag{8}$$

4 Results and Discussion

Table 1 shows the ranks of districts based on total crimes per 100,000 population (i.e., population criteria) and per 1 square Kilometer (i.e., area criteria) basis. Total crimes of each district in 2012 were used for this analysis.

Colombo and Gampaha have the highest crime rates based on both population and area criteria and have been ranked in first and second positions respectively, whereas Nuwara Eliya district records the lowest based on the population criteria (per 100,000 people). Based on the area criteria, the Mullativu district records the lowest.

Table 2: Classification of districts as safe and unsafe

Safe district	Polonnaruwa, Rathnapura, Kalutara, Galle, Mannar, Kandy, Matara, Kurunegala, Monaragala, Batticaloa, Matale, Puttalam, Trincomalee, Jaffna, Ampara, Badulla, Mullaitivu,Nuwara Eliya
Unsafe district	Colombo, Gampaha, Kegalle, Kilinochchi, Anuradhapura, Vavuniya, Hambantota

Table 2 describes the status of districts as safe or unsafe based on the country's total crime rate based on population criteria in which safe districts have their crime rate below the overall total crime rate and unsafe districts have their crime rate higher than the overall total crime rate. The overall crime rate per 100,000 population in 2012 is 205.8. According to the classification, there are 18 safe and 7 unsafe districts in Sri Lanka.

Table 1: Ranks of districts based on total crimes

	Population criteria ($Per10^5people$**)**		**Area criteria (**$Per1km^2$**)**	
Rank	**District**	**Rate**	**District**	**Rate**
1	Colombo	321.9396	Colombo	10.70529
2	Gampaha	301.1064	Gampaha	5.003605
3	Kegalle	253.8518	Kalutara	1.478098
4	Kilinochchi	253.7221	Kegalle	1.260484
5	Anuradhapura	236.4698	Galle	1.237893
6	Vavuniya	210.9055	Kandy	1.184536
7	Hambantota	207.8669	Matara	1.052221
8	Polonnaruwa	199.9567	Jaffna	0.742439
9	Rathnapura	199.4472	Rathnapura	0.662595
10	Kalutara	193.2979	Kurunegala	0.543397
11	Galle	192.3196	Hambantota	0.477961
12	Mannar	189.8162	Puttalam	0.388997
13	Kandy	167.0809	Matale	0.388861
14	Matara	165.8379	Nuwara Eliya	0.383113
15	Kurunegala	161.6964	Badulla	0.369451
16	Monaragala	161.62	Batticaloa	0.297477
17	Batticaloa	161.233	Anuradhapura	0.283466
18	Matale	159.9485	Polonnaruwa	0.246584
19	Puttalam	156.7427	Kilinochchi	0.225176
20	Trincomalee	155.4509	Trincomalee	0.216355
21	Jaffna	130.3346	Ampara	0.19094
22	Ampara	129.8117	Vavuniya	0.184545
23	Badulla	129.6288	Monaragala	0.129278
24	Mullaitivu	110.5835	Mannar	0.094689
25	Nuwara Eliya	93.72664	Mullaitivu	0.038976

Table 3: Crime prevalence according to districts

District	Mean	Variance
Anuradhapura	110.4857	153.9806
Ampara	45.30714	57.65176
Badulla	58.9	80.38955
Batticaloa	35.58571	49.58663
Colombo	447.2071	800.2358
Galle	110.1571	158.8142
Gampaha	359.9929	706.5201
Jaffna	48.45	65.82332
Kalutara	131.0929	231.1631
Kandy	124.928	188.8028
Kegalle	107.5214	181.3989
Kilinochchi	14.41429	17.39511
Kurunegala	156.9643	223.1185
Mannar	10.67857	14.782
Matale	40.17857	48.62398
Matara	69.35	106.412
Monaragala	39.52143	56.7219
Mulllativu	7.564286	10.55933
Nuwara Eliya	37.02143	48.39057
Polonnaruwa	44.53571	62.00666
Puttlam	78.16429	110.1159
Ratnapura	130.4071	205.452
Hambanthota	65.87143	44.19259
Vavuniya	20.37143	28.93962

Table 3 illustrates the mean and variance between districts in Sri Lanka from 2010 to 2019. Colombo is recorded as the highest number of crimes which is 447.2071 while counts of Gampaha account for 359.9929 and Mannar and Mullativu districts have the lowest number of crimes.

Table 4 illustrates the mean and variance between different crime types. This shows that the majority of crimes committed in Sri Lanka from 2010 to 2019 are related to crimes in which House Break and Theft and Cheating.

According to that result, five crime types that were a higher impact on society were identified and that five crimes are used to further analysis. They are House Break and Theft (HB), Cheating (CH), Robbery (RB), Hurt by knife (HK), and Rape (RP). First, the normality of the dataset was tested using the Shapiro-Wilk normality test. The result of the normal test confirms that the data was not normal, so the non-parametric Kruskal-Wallis rank sum test was used for the dataset. According to the test results, Chi-square value is 269.722 and asymmetric p-value is 0.000. It can be concluded that there is very strong evidence to suggest that there is a difference between at least one pair of groups.

Table 4: Crime prevalence according to crime types

Crime type	Mean	Variance
Kidnapping	40.016	31.10392
Arson	20.656	15.91473
Mischief over Rs5000/=	32.588	34.98276
House Breaking & Theft	527.552	703.3154
Grievous Hurt	52.644	41.26744
Hurt by Knife	99.224	88.44583
Homicide/Abet To commit suicide	22.408	17.42873
Attempted Homicide	7.408	5.32041
Rape/Incest	73.716	45.15261
Riot	1.108	4.675873
Robbery	181.344	284.0546
Unnatural Offence /Grave Sexual Abuse	24.688	19.44859
Extortion	5.356	8.779155
Cheating/Misappropriation C.B. Trust over Rs.100,000/=	213.96	335.8693

Dunn's Multiple Comparison Test that was a post hoc non-parametric test was carried out on each pair of groups.

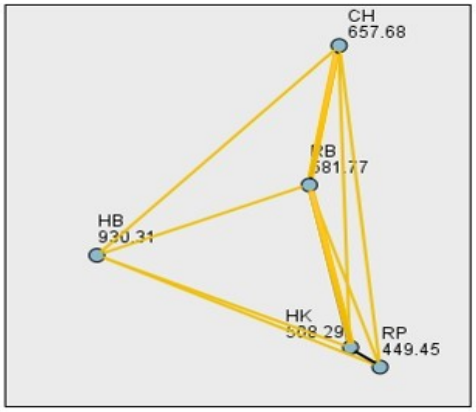

Fig. 1: Pairwise Comparisons of crime types

Significant differences are highlighted using an orange line to connect the two different groups in the diagram which shows the mean rank for each group. According to the results, only one pair is not a significant difference that is HK and RP. Other all pairs are significant differences.

HB - House Break and Theft; CH - Cheating; RB - Robbery; HK - Hurt by knife; RP - Rape.

Table 5: Spearman's rank correlation test results

	HB	RB	RP	CH
RB	0.947196	-	-	-
RP	0.8868798	0.8565825	-	-
CH	0.7759841	0.769895	0.7271467	-
HK	0.927738	0.9002913	0.8118393	0.717504

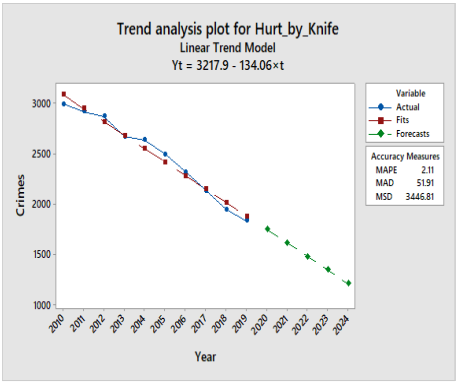

Fig. 2: Trend analysis plot for HK

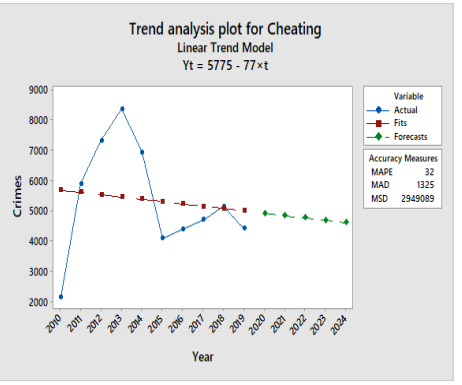

Fig. 3: Trend analysis plot for CH

Results show that there is a very strong positive correlation between RB-HB and HK-HB. All crime types have shown a positive high correlation for the occurrence of crimes in consecutive years from 2010 to 2019.

The criminal databases of HB, HK and RB are showing a declining trend and it can be concluded that there is a tendency for those crimes to decrease in the future. However, the crime data point of RP and CH fluctuates around a straight line (Fig. 3 & Fig. 5), and since they do not represent a clear trend, it is not possible to determine whether crime is low or high in the future.

Table 6 shows the number of crimes for coming years using linear trend model equations. Crime types with a clear trend are used for forecasting. This shows a decrease in the number of crimes in all crime types over the next few years.

A heat map of total crime based on area criteria is indicated in Fig. 7. It shows that crimes are more prevalent in the Western Province of Sri Lanka. Furthermore, Kegalle, Galle, Kandy, Matara, and Jaffna districts have a significant number of total crimes per area. Moreover, most of the crimes can

Table 6: Forecasting for the near future

Year	HB	HK	RB
2020	6107.40	1743.27	2115.87
2021	4819.87	1609.21	1676.28
2022	3532.35	1475.15	1236.69
2023	2244.82	1341.08	797.10
2024	957.29	1207.02	357.52

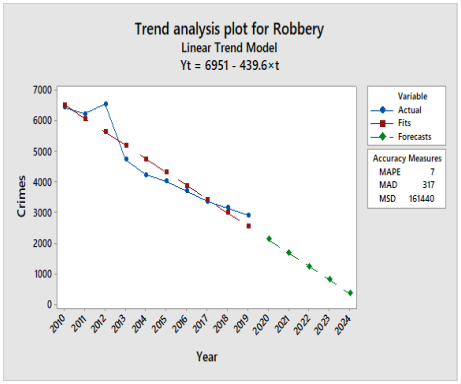

Fig. 4: Trend analysis plot for RB

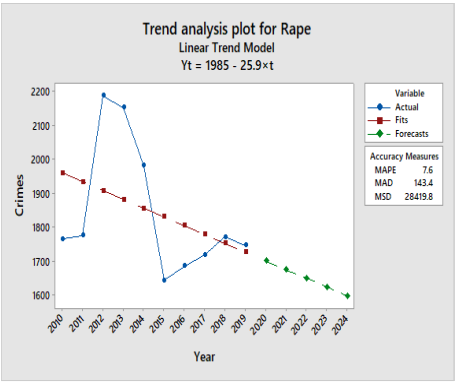

Fig. 5: Trend analysis plot for RP

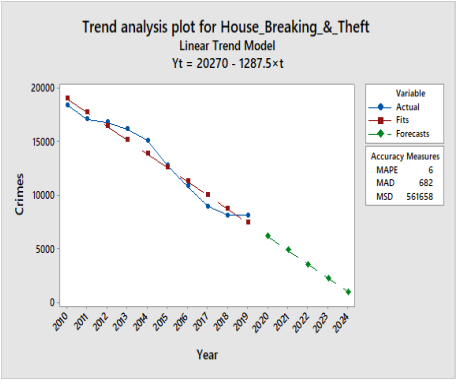

Fig. 6: Trend analysis plot for HB

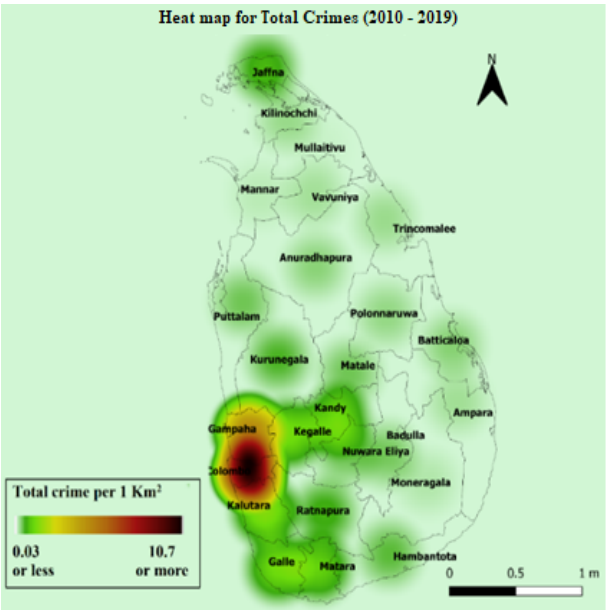

Fig. 7: Heat map for Total Crimes

be observed in Colombo and Gampaha districts and spread over to the down-south in decreasing magnitude.

The following figures show the relationship between crimes and the geographic environment. Hot areas are represented in dark red color.

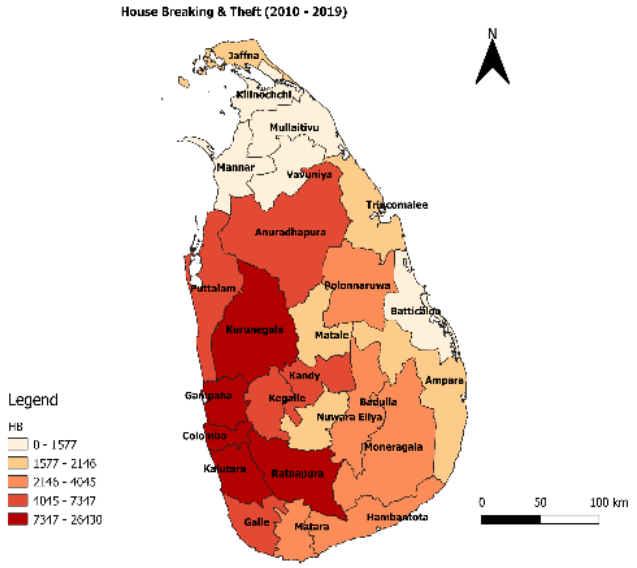

Fig. 9: Hotspot Areas of HB Crime

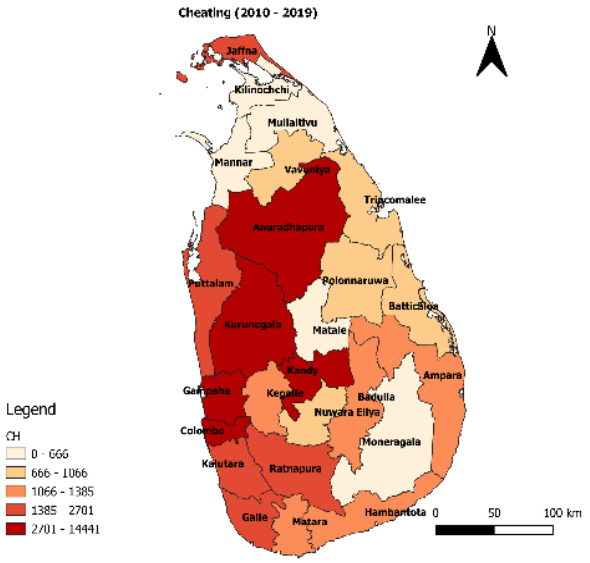

Fig. 8: Hotspot Areas of CH Crime

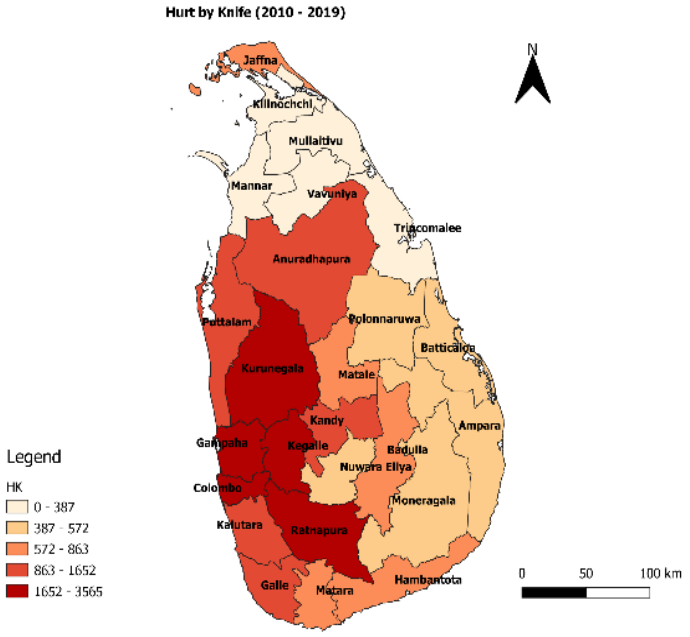

Fig. 10: Hotspot Areas of HK Crime

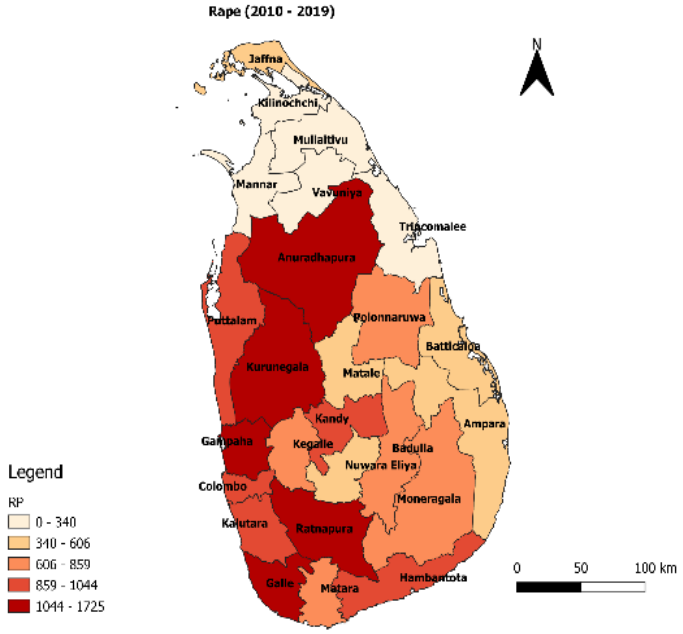

Fig. 11: Hotspot Areas of RP Crime

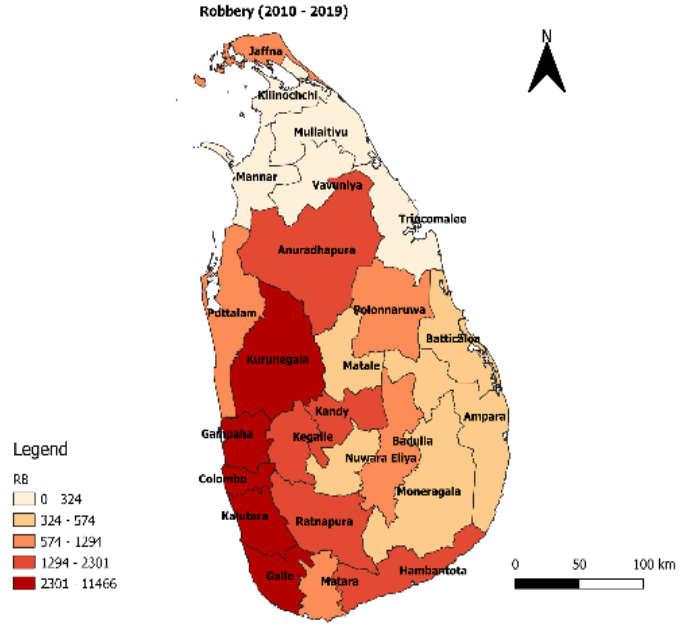

Fig. 12: Hotspot Areas of RB Crime

Table 7: Comparison between hotspot maps

Crime Type	PAI	Crimes committed (2010-2019)	Number of crimes in hotspots	Percentage of crimes in hotspots (Hit rate)
CH	2.45	53490	31864	59%
HB	3.12	131888	75224	57%
HK	2.61	24806	11931	48%
RB	3.98	45336	28447	63%
RP	1.35	18429	6931	37%

According to the results, the risk of all five types of crime appears to be high in Kurunegala and Gampaha districts. The Colombo district appears to be at risk for four crime types: CH, HB, HK, and RB. Kilinochchi, Mullaitivu, Mannar, Vavuniya, and Trincomalee are the districts with the lowest incidence of RP, RB, and HK. Kilinochchi, Mullaitivu, Mannar, Matale, and Monaragala are the districts with the lowest incidence of CH. Mullaitivu, Mannar, Vavuniya, and Batticaloa are the districts with the lowest incidence of HB.

HB - House Break and Theft; CH - Cheating; RB - Robbery; HK - Hurt by knife; RP - Rape.

Robbery (RB) has the highest Prediction Accuracy Index (PAI) value. There is a 63% probability of robbery occurring in hotspot areas. Robbery crime hotspots map is consistently better in their ability to predict spatial patterns of Robbery crime than any of the other crime types in their ability to predict crimes of their respective type. Rape has the lowest PAI value. There is a 37% probability of rape occurring in hotspot areas. Hit rates can be used to measure the probability of crime using resources in hotspot areas.

According to Fig. 13, Kurunegala, Colombo, Kalutara, Gampaha, and Ratnapura are the hotspot areas of total crimes from 2010 to 2019. Kilinochchi, Mullaitivu, Mannar, Vavuniya, and Trincomalee are the districts with the lowest incidence of total crimes.

5 Conclusion

Crimes have been disturbing threats to all the Sri Lankans all over the country. All five crime types have shown a positive high correlation with each other for the occurrence of crimes and there was a significant difference between the mean of all other pairs without RP-HK. Data on crime in HB, HK, and RB had a downward linear trend pattern. The other two types of crime trends had fluctuated. The risk of all five types of crime appeared to be high in Kurunegala and Gampaha districts. The Colombo district appeared to be at risk for four crime types: CH, HB, HK, and RB. Considering total crimes in Sri Lanka, the area defined as 'hot' covered 18.3% of the total area and had a 52.7% proba-

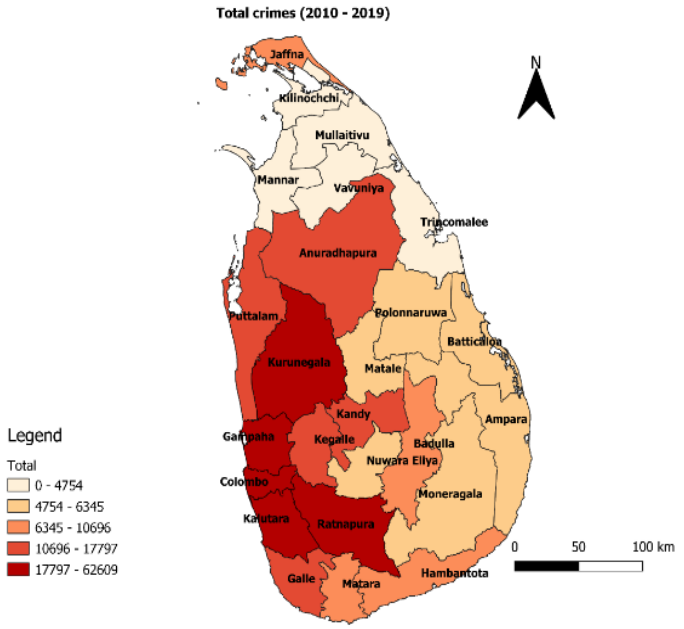

Fig. 13: Hotspot Areas of Total crimes

bility of committing the crime in hotspot areas. The robbery hotspot map was consistently better at predicting where Robbery crime will occur in the future, than hotspot maps of other crime types. Using our results, a decision-making process that considers where the best to target enforcement of crime and how-to prevention can be started. Paying more attention to hotspots areas has the potential to reduce crime next years.

References

Department of Police (n.d.) *Sri Lanka Police, Crime Statistics.*. Available at: https://www.police.lk/index.php/item/138-crime-statistics (Accessed: December, 2020).

Ratnayake R.M.K. (2015) Distribution pattern of urban crimes in Sri Lanka; with special reference to Mirihana Police Division, *International Journal of Multidisciplinary Research and Development* , 2(8):441-453.

Wickrama, C.B., Nawarathna, R.D., Nawarathna, L.S.(2020) Forecasting homicides, rapes and counterfeiting currency: A case study in Sri Lanka, *Biom Biostat Int J* , 9(6): 209-215.

Kedia P. (2016) Crime mapping and analysis using GIS,*International Institute of Information Technology*, 1(1):1-15

Chainey, S., Tompson, L., Uhlig, S.(2008) The utility of hotspot mapping for predicting spatial patterns of crime,*Security journal* , 21(1): 4-28.

Monmonier M. (1996) Temporal generalization for dynamic maps, *Cartography and Geographic Information Systems* , 23(2):96-98.

Shapiro S.S. and Wilk M.B. (1965) An analysis of variance test for normality (complete samples), *Biometrika* , 52(3/4):591-611.