

Identification of Reasons for Culpable Homicides and Attempted Murders: A Case Study for the Kelaniya Police Division

H. L. A. Weerakoon* and N.V. Chandrasekara

Department of Statistics & Computer Science, University of Kelaniya

*Corresponding author: lakniweerakoon@gmail.com
<https://orcid.org/0009-0004-4880-8137>

Received: 20th May 2023/ Revised: 21th June 2023/ Published: 30th June 2023

©IAppstat-SL2023

ABSTRACT

Culpable Homicides and attempted murders are ultimate crimes that could create ripple effects on a society which could go far beyond the original loss of human life. Owing to the unpredictable nature of such crimes that require complex investigations the objective of this study was to come up with an appropriate model to identify the reason for a culpable homicide or an attempted murder using a statistical approach. This study use data collected from 12 Police stations in Kelaniya Police Division relating to the incidents happened between 2010 and 2020. The Pearson Chi-square test was used in identifying the influential explanatory variables. Out of the 18 variables, 8 predictors including Weapon used, Relationship, Location, Civil Status of the perpetrator were statistically associated with the identified reasons at 5 % level of significance. Multinomial logistic regression followed by four data mining models including classification tree, support vector machine (SVM), k-nearest neighbour (KNN), and probabilistic neural network (PNN) were employed initially with a training and testing set which was randomly selected in the ratio 90:10. The 4 data mining models were then fitted separately by using the bagging technique. The accuracies were compared using the confusion matrixes and rates of misclassifications of the critical classes. Out of the fitted models, the highest accuracy of 93.75 % was shown by the PNN model with a spread of 0.6. The identified model can be used as a decision support tool by crime investigators and relevant authorities for wise decision making.

Keywords: Crime, Bagging, Data Mining, Multinomial Logistic Regression, Pearson Chi-square

1 Introduction

A crime is an illegal conduct for which a person can be prosecuted under a country's laws. According to the United Nations Office of Drugs and Crime's (UNODC) International Classification of Crime for Statistical Purposes (ICCS), "crime" is defined as a criminal transgression of the constraints on human behaviour (ICCS,2015). Every year more than 400,000 people die from homicides and in some countries, it is one of the leading causes of death. The penal code of Sri Lanka under sections 293 and 294 describes homicides as a killing of a human being by another human being which could be either lawful or unlawful. Lawful homicides where the accused had a valid reason to commit the crime may include justifiable and excusable homicides. Whereas unlawful homicides are inflicted upon a person to cause death or serious injury. Unlawful homicides include murder, and culpable homicides not amounting to murder. Culpable homicide can be considered as the genus and murder would be species. Further, attempted murder is considered a crime when a person acts purposefully, willfully, or carelessly with severe contempt for human life. Section 300 and 301 of the penal code of Sri Lanka states the terms and conditions along with the punishments for attempted murder and there is a tight line between murders and attempted murder.

The punishments meted out to criminals under Sri Lanka's Penal Code heavily rely on proving "intent" to commit the crime. If illegal conduct (culpable homicide or attempted murder) was committed unintentionally, the sentence is often reduced to 10 to 20 years in jail (Code, 1991). However, if the crime is pre-planned and the investigating officers have sufficient evidence to confirm the criminal's culpability, such as in the case of a homicide (Penal Code, 1991), the perpetrator is sentenced to death or life imprisonment. Hence it is evident that whether it is an intentional murder or an attempted murder, intention plays a crucial role in every crime.

Countries worldwide have created various classifications for different criminal offences from different perspectives. The ICCS has classified criminal offences into homogeneous categories aggregated at four different hierarchical levels for statistical feasibility. In Sri Lanka, criminal offences and homicides are classified based on the Penal Code of Sri Lanka. Besides these categorizations, murders and attempted murders have had major detrimental consequences on the lives of surviving family members, particularly children. Anxiety, post-traumatic stress disorder, aggressiveness, and despair are just a few of the significant psychological consequences that survivors face. The complex mix of factors that cause homicides include sexual motives, family disputes, and organized gang-related violence. Poverty, economic inequality, and the use of weapons and drugs act as catalysts for homicide rates in a country.

Several studies from Sri Lanka and other countries have conducted research in the domain of interest, as revealed by a comprehensive literature review. Two of these studies were directly relevant to this research study, one

by (Jayasundara, 2021) which identified factors influencing individuals to become murderers in Sri Lanka using a random sample of 63 offenders from Welikada prison, and the other by (Jayathunga, 2011) which analyzed homicides in the Rathnapura police division from a sociological aspect and included a sample of 20 cases from police reports between 2001 and 2002. Based on the review of those studies, it was observed that there is potential for further expansion of previous studies by utilizing statistical and data mining modelling techniques to gain a deeper understanding of the interconnection between the factors that influence such criminal acts and the underlying reasons for them.

The main objective of this study is to develop a more accurate and predictive model for identifying the reason behind culpable homicides or attempted murders, which will serve as a decision-support tool for relevant authorities. A decision support tool usually refers to a model, a system or a software application that can incorporate statistical methods, and algorithms to assist individuals or organizations to enhance decision making process by generating insights. The significance of this model lies in its ability to provide insights into the potential motive behind a crime by utilizing available data pertaining to the victim, perpetrator, and the event, thus aiding in pre-assessing the cause of the crime. The subsidiary objective is to identify significant factors that can differentiate reasons for culpable homicides and attempted murders specifically in the Kelaniya Police Division, where the division holds Kelaniya City, one of the top 5 violent cities in Sri Lanka (Towards Data Science, 2020). The available summary statistics on homicide cases in Sri Lanka are not sufficient for obtaining statistically meaningful insights, and therefore, a retrospective study using case-by-case data collection and statistical analysis is needed.

Although a number of descriptive or sociological studies have been published in Sri Lanka most of them concentrate on the counts and percentages of demographic features of the perpetrators and victims and mostly aiming the bigger picture of crime. Little or no effort has been made to analyse homicides and attempted murders in a statistical sense that could be used as a decision support tool for the relevant authorities to predict the reason for a particular homicide or an attempted murder. Moreover, the study uses statistical as well as data science approaches to develop the models which makes the study significant in multiple ways. Therefore, this research is considered to be a beneficial investigation for a nation as it directly impacts both social and economic factors.

The rest of the paper will discuss the methodology, data collection, analysis techniques, and the findings of the study on identifying the significant risk factors and modelling the reasons for homicides and attempted murders.

2 Methodology

Considering the significant impact of predictor variables on categorical response variables, Multinomial Logistic Regression, classification tree, support vector machine (SVM), k-nearest neighbour (KNN), and probabilistic

neural network (PNN) were used to model the motives for considered crimes.

Pearson Chi-Square Test of Independence

The Chi-Square test of independence is used to determine if there is a significant relationship between two categorical variables. The frequency of each category for one categorical variable is compared across the categories of the second categorical variable. The equations (1), (2) below indicate the test statistic of the chi-square test of independence and formulae for obtaining the expected frequencies respectively. The conclusion is based on the hypotheses and the test statistic of the Chi-Square test.

H_0 : There is no association between the two categorical variables. $(\pi(i, j) = \pi(i+) \pi(+j))$

H_1 : There is an association between the two categorical variables. $(\pi(i, j) \neq \pi(i+) \pi(+j))$

Where $\pi(i, j)$ denoted the expected probability associated with the cell (i, j), $\pi(i+)$ and $\pi(+j)$ are the expected marginal probabilities of the row and column variables.

$$x^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}} \quad (1)$$

Under H_0 ,

$$\begin{aligned} \mu_{ij} &= n\pi_{ij} = n(\pi_{i+}\pi_{+j}) \\ &= n\left(\frac{n_{i+}}{n}\right)\left(\frac{n_{+j}}{n}\right) \\ &= \frac{n_{i+}n_{+j}}{n} \end{aligned} \quad (2)$$

Where n_{ij} denote the observed frequency of cell (i, j), μ_{ij} denote the fitted frequency of cell (i, j), n_{i+} and n_{+j} is the marginal frequency of the row and column variables. The difference between observed and fitted frequency is called residual (Agresti A, 2006).

Multinomial logistic regression (MLR)

Multinomial logistic regression is a simple extension of binary logistic regression that allows for more than one category of dependent or outcome variable. For the multinomial logistic regression model, the linear component is equated to the log of the odds of a j^{th} observation compared to the j^{th} observation. That is, to consider the j^{th} category to be the omitted or baseline category,

where logits of the first (j-1) categories are constructed with the baseline category in the denominator.

$$\log \frac{\pi_{ij}}{\pi_{i0}} = \log \frac{\pi_{ij}}{1 - \sum_{j=1}^{K-1} \pi_{ij}} = \sum_{k=0}^{K-1} x_{ik} \beta_{kj} \quad (3)$$

Classification Tree

A classification Tree is a learning algorithm that can solve binary and multi-level classification problems with discrete or continuous attributes. A classification tree can also provide a measure of certainty about the accuracy of the categorization. The procedure of binary recursive partitioning is used to construct a classification tree. This is an iterative method that involves dividing the data into partitions and then further dividing it on each branch.

Support Vector Machine

The Support Vector Machine (SVM) is a supervised machine learning technique that can solve classification and regression problems. It is, however, mostly employed to solve categorization difficulties. Each data item is plotted as a point in n-dimensional space (where n is the number of features available), with the value of each feature being the value of a certain coordinate in the SVM algorithm. Then accomplish classification by locating the hyperplane that clearly distinguishes the available number of classes.

K- Nearest Neighbour

The k-nearest neighbours (KNN) technique is a supervised machine learning algorithm that can be used to handle both classification and regression issues. The K-NN method assumes that the new case/data and existing cases are comparable and places the new case in the category that is closest to the existing categories. The K-NN algorithm saves all existing data and categorizes additional data points based on their similarity.

Probabilistic Neural Network

A probabilistic neural network (PNN) is a feedforward neural network, which is widely used in classification and pattern recognition problems. Because of ease of training and a sound statistical foundation in Bayesian estimation theory, PNN has become an effective tool for solving many classification problems.

2.1 Data Sources

The dataset employed in this study consists of 320 observations pertaining to culpable homicides and attempted murder incidents that have taken place between 2010 and 2020, spanning a ten-year period. The Grave Crime Information Books (GCIB) and Grave Crime Registers (GCR) of each of the 12 police stations located in the Kelaniya police division were referred in order to obtain the information. The data were relevant to Wattala, Peliyagoda, Kandana, Ja-ela, Ragama, Mahabage, Kiribathgoda, Kelaniya, Meegahawaththa, Sapugaskanda, Kadawatha, and Biyagama police stations were used in the data set.

2.2 Sampling Scheme

The research was conducted considering the homicide and attempted murder cases taken place in the Kelaniya Police Division as the target population. Initially, it was necessary to identify and classify the influential factors before moving to the analysis stage. According to the factors identified by past literature and by discussions with crime investigation police officers, the explanatory variables were disaggregated related to the event, victim and perpetrator. The event disaggregation included the time of the murder, police station, the weapon used, mode, type of location, and number of accusers. The victim and perpetrator disaggregation included a few common variables like gender, civil status, age category and victim-perpetrator relationship. Further Status of mental health at the time of incidence, level of education, occupation, race and religion were specifically under perpetrator disaggregation. While considering the above variables as the predictor variables the reason for homicide/attempted murder at closest proximity was considered as the response variable for the study which includes Verbal fights and arguments, Family Disputes, Revenge Purposes, Financial Matters, Private or long-term Disputes and Love or Lust issues.

In this study, the data collected covers all the cases related to the target population. As a training and testing data set is necessary for model fitting as well as for model comparison, the cases were randomly selected without replacement in the ratio 90:10 one time and were used to compare the 5 selected models with the same training and testing data sets. Further, for assuring the robustness of the fitted data mining models, the bagging technique was incorporated such that it will help to find a model with higher accuracy and predictive ability.

2.3 Development of the design of the study

The study design included a series of steps starting from testing associations between variables to performance measures of each of the fitted models. After collecting the data under relevant variables as explained in Section 2.2, the Pearson Chi-square test was employed to examine the relationship between

the response variable and each of the nominal categorical explanatory variables. The p-value related to the linear-by-linear association was observed in the chi-square test for the ordinal categorical variables.

With the training and testing data set obtained in the ratio 90:10 the Multinomial logistic regression (MLR) was then performed using the stepwise forward selection method with the randomly selected 90 % of data along with the statistically significant variables and the adequacy was checked. The Cox and Snell, as well as Nagelkerke pseudo-R square values, were used to refer to the percentage of explained variation in the dependent variable. The absence of multicollinearity was also confirmed and then 10 % of the remaining testing data set was fed into the fitted model. Following the MLR modelling approach, four data mining models were then used for the same training and testing data sets: classification tree, support vector machine (SVM), k-nearest neighbour (KNN), and probabilistic neural network (PNN).

A major concern was strengthening confidence in the robustness of the fitted data mining models. Hence for each of the data mining models, 10,000 random samples of training and testing data sets were considered, and the better model with hyperparameter adjustments was ultimately selected. By using the above-mentioned method, the model is trained with a fairly good set of training data and the highest accuracy can be yielded. As a result, the model could be utilized as a tool to assist investigators to figure out which way the case might have headed in to.

One of the important criteria in finding a better model is to look into the rates of misclassifications of critical classes. Among the 6 categories of the response variables, verbal confrontations and arguments generally occurred as a result of the high temper or sudden provocation. These cases are usually found to be unintentional. Therefore, the punishment could be 10 to 20 years. But in the case of culpable homicide or attempted murder as a result of revenge purpose, financial matter, long-term or private dispute usually, the case could be sentenced to life imprisonment. Hence misclassifications by a model on these categories could lead to incorrect decisions as the final model of this study acts as a decision support tool. Hence when selecting the most accurate model it is necessary to consider the model with lesser misclassifications on critical classes. Therefore, the accuracy of the fitted models for classifying reasons for culpable homicides and attempted murders were compared using the confusion matrices as well as the rates of misclassifications of the critical classes. Hence a better model was selected as a result of the conclusions drawn from the accuracies.

3 Results

Preliminary analysis

The percentage of culpable homicides and attempted murders recorded from Kelaniya Police Division’s 12 police stations are depicted in Figure 1. According to the below figure, in the Gampaha District, almost all police stations cover urbanized areas. Between 2010 and 2020, the Peliyagoda Police Station had the highest number of cases (15 %), with 36 culpable homicides and 13 attempted murders totaling 49. Within a 10-year period, the Meegahawatta police station has reported the lowest number of cases.

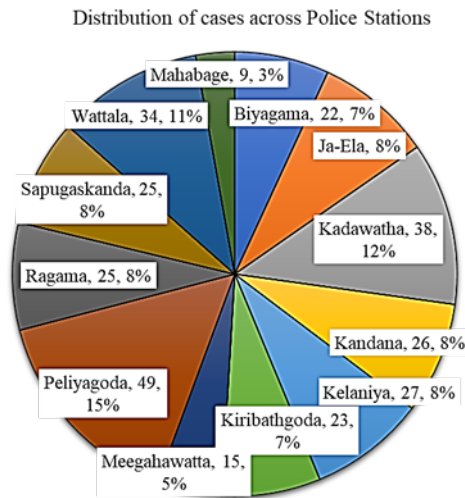


Fig. 1: Pie chart with the proportion of cases reported in 12 police stations

The collected data indicated the presence of six identifiable categories of reasons for culpable homicide or attempted murder, which served as the response variable. The categories did not indicate any issue with class imbalance as the data distribution among the categories was satisfactory.

Table 1: Description of the Dependent variable categories

Reason for murder	Frequency	Percent	Valid Percent	Cumulative Percent
Verbal Fights or Argument	68	21.3	21.3	21.3
Family Disputes	53	16.6	16.6	37.8
Revenge Purposes	45	14.1	14.1	51.9
Financial Matters	44	13.8	13.8	65.6
Long term or Private Dispute	64	20	20	85.6
Love or Lust	46	14.4	14.4	100
Total	320	100	100	

As per the Table 1, 21.3 % of the incidents in the Kelaniya Police Division were recorded due to verbal confrontations or disagreements. Long-term or private issues were the second most common reason, contributing to 20% of the total. Financial issues arose as a result of land-related issues, robbery, and other causes accounted for the lowest percentage of cases which is 13.8 % of all instances. Family issues, revenge, and love or lust issues all had a similar impact on attempted murder and culpable homicides, with 16.6%, 14.1 %, and 16.6 %, respectively.

As a primary step to start model fitting, the Pearson Chi-Square test of independence was used to find the association between the reason for the culpable homicide or attempted murder and the set of 18 variables which were primarily identified based on the literature review and expert opinions. Among the 18 predictor variables, weapon used, relationship, location, number of accusers, civil status of the perpetrator, the mental health status of the perpetrator, gender of victim and civil Status of the victim were statistically associated with Reason for culpable homicide or attempted murder at 5% level of significance. The rest of the 10 variables were not statistically associated with Reason hence they were not considered for further analysis and model fitting.

Multinomial Logistic Regression Analysis and Data Mining Models

The response variable of this study is a nominal categorical variable. Hence for classification, one suitable approach would be multinomial logistic regression model (MLR). Initially, 90% of the original data was randomly chosen for training data and the remaining 10% was taken for model testing. All the 8 predictor variables that were evident to have a statistically significant association with the Reason was used for fitting the model. The stepwise forward method was used as the technique to select the best set of variables among the 8 significantly associated variables.

Table 2: Multinomial Logistic Regression Model Summary

Action	Effect(s)	-2 LogL	Chi-Square	df	Sig
Entered	Intercept	877.097			
Entered	Relationship	662.373	214.724	10	0.000
Entered	V Civil Status	633.035	29.338	5	0.000
Entered	P_Mental Health Status	607.697	25.338	5	0.000
Entered	V Gender	584.054	23.642	5	0.000
Entered	Location of Assault	552.053	32.001	15	0.006
Entered	Weapon Used	530.034	22.019	10	0.015

Table 2 shows that of the 8 predictors Weapon used, Relationship, Location, Number of accusers, Civil Status of the perpetrator, Mental health status of the perpetrator, Gender of victim and Civil Status of the victim which had a significant association with the response variable as per the preliminary

analysis, six variables have been added to the MLR model using the forward stepwise procedure. Except for the 2 variables number of accusers and the civil status of the perpetrator, the reduced model with the above combination of variables was deemed to provide a better model which is capable of explaining the response variable while being statistically significant at 5% level of significance. However, as per the confusion matrix, the fitted MLR model was only able to correctly classify 59.38% of the cases. Followed by the MLR model four commonly employed data mining classification techniques were then applied using identical sets of training and testing data. Further, out of the 4 data mining algorithms hyperparameter tuning was done for 3 of the data mining models with parameters specified for each algorithm. For the Support Vector Machine(SVM), the kernel function which is used for determining the shape of the hyperplane and decision boundary was tuned under ‘Gaussian’ or ‘Radial Basis Function (RBF)’, ‘Linear’ and ‘Polynomial’ Kernels. The distance function and the number of neighbours were the parameters tuned for the K-Nearest Neighbour (KNN) algorithm. The only parameter that needed to be tuned for the Probabilistic Neural Network (PNN) model was the Spread. The subsequent table presents a concise overview of the accuracies achieved

Table 3: Correct classification accuracies of fitted models.

Model	Overall Accuracy	Verbal fights and arguments	Family Disputes	Revenge Purposes	Financial Matters	Private or long-term Disputes	Love or Lust
MLR	59.38%	50%	57.14%	100%	33.33%	40%	100%
C.Tree	56.25%	66.70%	57.10%	66.70%	33.30%	40%	80%
SVM	90.63%	100.00%	85.70%	66.70%	83.30%	100%	100%
KNN	84.38%	83.30%	85.70%	100%	66.70%	100%	80%
PNN	93.75%	100.00%	85.70%	100%	83.30%	100%	100%

As per the results, the Probabilistic Neural Network Model seems to perform better than other models with the highest accuracy at 93.75%. The rate of misclassification was also at a lower level confirming the model’s accurate performance. The only parameter linked with PNN, “Spread” was modified each time, and the accuracy of the testing set was determined. Finally, the most effective model for classifying the cause of culpable homicide or attempted murder was obtained.

The parameter spread was varied between 0.1 – 1. As shown in Table 4 when the spread is between 0.2 and 0.6, the highest testing accuracy was observed.

The best network architecture for the PNN network is shown in Figure 2.

Eight input variables were used as the input for building the PNN Model with the spread value of 0.6. The architecture contains two hidden layers with 288 and 6 neurons respectively to predict the class of the response variable.

Table 4: Testing set accuracy of PNN Model with different Spread values.

Spread	Overall Accuracy (Testing Data)
0.1	90.63%
0.2	93.75%
0.3	93.75%
0.4	93.75%
0.5	93.75%
0.6	93.75%
0.7	84.38%
0.8	68.75%
0.9	62.50%
1.0	56.25%

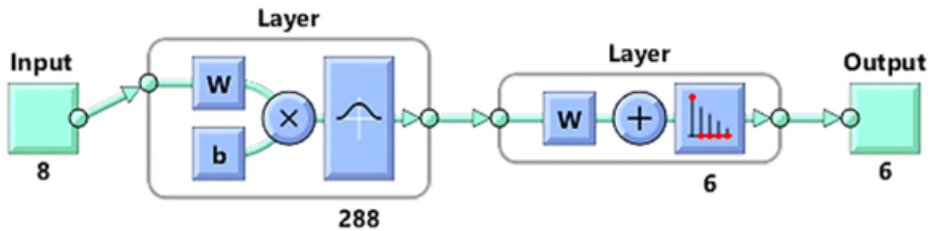


Fig. 2: Network architecture of the PNN Model

The 288 neurons are individually allocated for each case in the training dataset and the 6 neurons are for each category in the response variable.

It was also important to identify the best input variable set to the PNN model that performed better among the rest. The highest accuracy from the testing data set was achieved when all the eight significant variables were employed in the model development.

Classification using Data Mining Models with Bagging technique.

One of the primary problems raised in the study was how the model's robustness can be ensured. So far, all five models were fitted using a single representative sample (fixed training and testing data set) that was randomly chosen once which allow the comparisons among the models. As a step of strengthening the confidence in the robustness of the fitted data mining models 10,000 random samples for each of the four data mining models were considered, and the model that yielded the highest accuracy, along with its corresponding testing and training data set pair (without replacement), was obtained.

According to the summary of the table above, using the bagging strategy instead of the earlier method raised the classification model's overall accuracy from 56.25 % to 78.13 %. The accuracy of the Support vector machine and K-nearest neighbour models were both found to be 84.38 %, while the PNN

Table 5: Correct Classifications of the fitted models with bagging technique

Model	Overall Accuracy	Verbal fights and argument	Family Dispute	Revenge Purpose	Financial Matters	Private or long-term Dispute	Love or Lust
C. Tree	78.13%	71.40%	100.00%	50.00%	100.00%	60.00%	100.00%
SVM	84.38%	57.10%	100.00%	66.70%	100.00%	100.00%	100.00%
KNN	84.38%	71.40%	100.00%	100.00%	66.70%	100.00%	71.40%
PNN	90.62%	91.70%	83.30%	100.00%	100.00%	83.30%	100.00%

model, as expected, showed the best accuracy at 90.62 %. Finally, the overall accuracy results showed that the PNN model is better at predicting the reasons, with higher rates of accurate classification with lesser misclassification rates.

4 Discussion

The primary aim of this study was to create a more precise and predictive model for determining the cause of culpable homicides or attempted murders such that the model would serve as a valuable decision-support tool for the authorities involved. Additionally, a secondary objective was to identify a significant set of variables that could effectively distinguish between the reasons behind culpable homicides and attempted murders in the Kelaniya Police Division.

The 320 observations used in the study in achieving the above stated objectives disclosed some eye-opening information as a result of the preliminary analysis that was undertaken along with the model fitting. Among the findings, 58.75% of the attempted murders and culpable homicides have taken place during the night light which is from 6.00 pm to 6.00 am and the majority of the cases have taken place in public places or streets/roads (35.31%). Regarding the educational level of the offenders in the research domain, more than half of them (60%) have only completed up to Grade 5,10 or have never attended a school which was similar to case in (Jayathunga, 2011; Jayasundara, 2021). Sharp or pointed weapons are the key armament used in committing attempted murders or culpable homicides in Kelaniya Police Division which is around 46% which was also found to be the same case in the studies of (Rathnaweera, 2016; Jayasundara, 2021). The assailant/perpetrator was an acquaintance or a stranger to the victim 67.19% of the time. The was quite the opposite to the research results of (Jayathunga, 2011) as most of the crimes in the study was committed by people who had a close relationship to the victim. The motive was either a family disagreement or Love/Lust in situations where the perpetrator and victim were intimate partners (15.94%). Most of the victims (17.50%) were above the age of 40, whereas the majority of perpetrators

were between the ages of 25 and 39. These age categories were included either way in the age intervals stated in (Jayathunga, 2011; Rathnaweera, 2016). Married males have been the victims of culpable killings or attempted murders in 41.88 % of the time. Similarly, the vast majority of the criminals were also married males (50.63 %). Surprisingly, around 81.56% of the criminals were employed and 79.06% of the perpetrators did not show any intoxication or illness at the time of incidence.

Since data related to 18 variables had to be analyzed it was necessary to figure out which factors truly matter to the prediction of reasons. By using Chi-square test among the 18 predictor variables, Weapon used, Relationship, Location, Number of accusers, Civil Status of the perpetrator, Mental health status of the perpetrator, Gender of the victim and Civil Status of the victim were statistically associated with the Reason for culpable homicide or attempted murder at 5% level of significance. This process helped in achieving the secondary objective of the research.

The model fitting was done under two approaches. The purpose of the fixed data set approach was to make a fair comparison among the 5 fitted models. Under the fixed data set a MLR model was initially fitted. Since the overall accuracy of the MLR model was 59.38%, which was not at a satisfactory level, data mining models were taken into use for better outcomes. The accuracies of the data mining models were obtained as summarized in Table 3 and Table 4. It was noticeable that the accuracies were comparatively increased with the use of bagging technique. However, the probabilistic neural network model was the data mining model which indicated the highest accuracy out of all the fitted models. Under the fixed training and testing data sets randomly chosen, the PNN model indicated an overall accuracy of 93.75% with the spread of 0.6. Under the bagging technique the accuracy was denoted as 90.62%. Additionally, in order to find the best variable set combination for the PNN model, considerable number of variable combinations were checked. The highest accuracy was achieved when all eight significant variables are employed in the model development.

Hence, with the results, it was evident that the Probabilistic Neural Network model was successful compared to other models as it easily outperformed throughout the research. Despite the fact that crimes in this research area are highly unpredictable, the PNN model obtained can be used to capture the underline interconnections and complexities with a higher accuracy. The effort put forth during the research in a novel approach could be used by crime investigators as a decision support tool for wise decision-making with the ability of the neural network model which was found in this study.

5 Conclusion

The fact that Kelaniya City is one of Sri Lanka's top 5 most violent cities (based on 2010–2018 data), the Kelaniya Police Division was chosen as the study's location. The most prominent motive for the crimes was verbal fights

or arguments that occur as a result of sudden provocation. Despite the fact that the majority of them were employed, most of the murderous offenders were uneducated and from low economic backgrounds.

The primary objective of the study was to come up with a better model with higher precision and predictive power to determine the cause of culpable homicide or attempted murder as a decision support tool for criminal investigating authorities. As a result of comparisons made between different modelling approaches using classification accuracies and misclassification rates, the final PNN model obtained would successfully carry out the functions of a decision support tool. This concept can be used in other crime categories as well because of its ability identify hidden patterns in large scale data sets. The subsidiary objective was to identify a significant set of variables that are capable of distinguishing reasons. The weapon used, relationship between victim and perpetrator, location of the incident, number of accusers incorporated in the crime, Civil Status of the perpetrator, Mental health status of the perpetrator, Gender of the victim and Civil Status of the victim were identified as the most influential factors in identifying reason for culpable homicides and attempted murders. The reason for a Culpable homicide or an attempted murder cannot be solely concluded by the identified variables since a series of complex investigations takes place until the intention is proved. Several other factors could be considered, and several other reasons for culpable and attempted murders could be figured out by further investigation.

Some of the limitations in this research are that data related to crimes in this domain are confidential and in Sri Lanka they are maintained manually in each Police Station. Therefore, collecting data was challenging as the required data had to be obtained only by reading each case separately from GCIBs and GCRs in each police station. The study would have been more accurate if data was available for many other police divisions in a systematic manner.

A major advantage of the model as a decision support tool would be that it not only works for the data obtained from the Kelaniya Police Division but also for other divisions which has a similar background. Further, this research would also show the possibility of using PNN models for gaining meaningful insights from research domains that usually has complex interconnections and links among data. As a future implication, there is potential to broaden the scope of this study to encompass the entire country, incorporating additional identified factors and reasons specific to each division. This would enable its application in Police stations across Sri Lanka.

Acknowledgement

Special thanks go to all the Police Officers in charge of the 12 police stations in the Kelaniya Police Division for their immense support.

References

- Agresti A. (2006) *An Introduction to Categorical Data Analysis: Second Edition*. Wiley., DOI: 10.1002/0 470114754
- Jayasundara, M. W. (2021) *IJRIS | Volume V, Issue I. International Journal of Research and Innovation in Social Science*, 2454–6186.
- Jayathunga, N. (2011) A Sociological Study of the Homicide in Sri Lanka: A Case Study in Rathnapura Secretariat Division. *Sabaragamuwa University Journal*, 9(1), 45 ,DOI: <https://doi.org/10.4038/suslj.v9i1.3733>
- Our World in Data* (2022). Available at: <https://owidm.wmcloud.org/homicides> (Accessed: 7 October 2021).
- Rathnaweera, R. H. A. I.(2016) Trends of homicides in Galle-Sri Lanka, *Medico Legal Journal of Sri Lanka*, 3(1), 16 ,DOI:<https://doi.org/10.4038/mlj.v3i1.7318>.
- Smith, J. (2022) An Introduction to Statistical Tests, *Journal of Data Analysis*, 10(2), 45-62, Available at: <https://www.example.com/article123> (Accessed: 18 January 2022).
- Sri Lanka: Penal Code [Sri Lanka]* (1991). Available at: <https://www.refworld.org> (Accessed: 22 May 2021).
- Towards Data Science* (2020) Analysis of Sri Lanka Crimes: Part I - House Breaking and Theft. Available at: <https://towardsdatascience.com/analysis-of-sri-lanka-crimes-part-i-house-breaking-and-theft-4b1a608eefa1> (Accessed: 15 July 2021).
- UNODC(2013) International Classification Of Crime For Statistical Purposes (ICCS), (03), 8–9.