E-government Information Application: Identifying Smuggling Vessels with Data mining Technology

Chih-Hao Wena[1], Ping-Yu Hsu[1], Chung-Yung Wang[2], Tai-Long Wuc, Ming-Jia Hsu[3]

[1]**Department of Business Administration, National Central University, Jhongli, Taiwan, R.O.C.**
[2]**Department of Logistics Management, National Defense University, Taipei, Taiwan, R.O.C.**
[3]**Coast Guard Administration, Executive Yuan, Taiwan, R.O.C.**

Abstract: In spite of the gradual increase in the number of academic studies on smuggling crime, focus is seldom placed on the application of data mining to crime prevention. This study provides deeper understanding and exploration of the benefits of information technology for the identification of smuggling crime. This study focuses on smuggling of vessels. The data source is the complete record of fishing vessels leaving and returning to ports in the Taiwan region. This paper essay applies both artificial neural networks (ANN) and logistics regression (LR) to classify and predict criminal behaviors in smuggling. At the same time, it shows the difference between ANN and human inspection (HI), also the difference between LR and HI. This study establishes models for vessels of different tonnage and operation purposes that can provide law enforcers with clearer judgment criteria. It is needed to construct different models for vessels to achieve the actual cases in the reality since smugglers will use different kinds of ship for different smuggling purposes. The study results show that the application of artificial neural networks to smuggling fishing vessels attains an average precision of 76.3%, and the application of logistic regression to smuggling fishing vessels can achieve an average precision of 60.5%, both of which are of significantly higher efficiency levels compared with the current human inspection (HI) method. This study suggests the value of using an artificial neural networks model to obtain good identification performance for different vessel types as well as average savings of 90.47% on the manpower loading. Information technology can greatly help to increase the probability of seizing smuggling fishing vessels. Nowadays, public administration information is saved electronically however is not employed well. In fact, it can increase the administrative efficiency by proper use of electronic data. In this study, for example, we expect better use of the data stored in the database to establish an identifying model of smuggling. Applying the automatic identification mechanism, it is useful to reduce the probability of smuggling crime.

Keywords: government information application, Crime data mining, smuggling predictions, artificial neural networks, logistic regression.

1   Introduction

Smuggling is the clandestine import of goods or the evasion of taxes by circumvention of border controls (Farzanegan, 2009). Smuggling is a serious issue for every nation, and is extremely difficult to prevent or contain as the conduits of smuggling are very diverse. For example, smugglers may use vessels to carry goods (Tseng, Shiue, Ning, Lin, & Cheng, 2009), may transport drugs with the human body (Niewiarowski, Gogbashian, Afaq, Kantor, & Win, 2010), and may carry packages through underground tunnels (Klar & Linker, 2010).

As Taiwan is an island, the oceanic area is approximately three times that of the land. As a consequence, oceanic transportation is so prosperous that vessel transportation becomes the most commonly-used smuggling method used. In 2009, the seized smuggling of agricultural and fishery products reached 90.2 tons in total (Coast Guard Administration, 2010). Furthermore, according to official data from Taiwan's Ministry of Finance, the annual average sum of seized smuggling items over the past 20 years is over $24 million (Ministry of Finance, 2010); however, it is estimated that the value of actual smuggling is between $160.8 million and $184.6 million (Farzanegan, 2009; Ministry of Finance, 2010).

The Coastal Patrol Directorate General (CPDG) of Coast Guard Administration (CGA) in Taiwan is responsible for all maritime affairs related to the prevention of smuggling, including coastline patrol and vessel inspection in ports. CPDG members consist of voluntary military staff (38.9%), compulsory officers and soldiers (58.1%) as well as civilian personnel (3%) (Coast Guard Administration, 2010). The inspection of vessels in ports falls under the judiciary of compulsory officers and soldiers. Among all the vessels being investigated, only 5.03% are found to be guilty of smuggling offences. Due to cut-backs in organizational structure, the number of compulsory officers and soldiers will be reduced

to 21.81% in 2014 (Ministry of National Defense, 2009), meaning that only one-third of the current manpower will be retained. If the rate of effectiveness is to remain at 5.03%, the outlook for smugglers is very prosperous indeed. For this reason, it is essential for CGA to have highly effective information tools to spot smuggling vessels.

Most of the previous studies on crime used multiple regression methods (Kirby, 1954), predictive attribute analysis (Glaser, 1962; Wilkins & Macnaughton-Smith, 1964), or multidimensional contingency table analysis (van Alstyne & Gottfredson, 1978). However, Simon (1972), Gottfredson and Gottfredson (1980), and Wilbanks (1985) asserted that statistical and cluster analysis methods add little to overall predictive accuracy (Caulkins, Cohen, Gorr, & Wei, 1996). Other studies broadly applied logistics regression (LR) to law science and crime-related studies (Urbaniok et al., 2007), which is proficient in prediction and classification, including the prediction of criminal relapse (Poortinga, Lemmen, & Majeske, 2007), the classification of crime, and the prediction of crime span (Andresen & Jenion, 2004). Recent research has suggested that artificial neural networks (ANN) may facilitate a process that is of benefit to crime analysis (Keyvanpour, Javideh, & Ebrahimi, 2011), including analysis into criminal relapse (Wang, Mathieu, Ke, & Cai, 2010) and money-laundering prevention (Le Khac & Kechadi, 2010).

However, to the best of our knowledge, neither method has been applied to the fight against smuggling. This study shows that both methods can effectively distinguish smuggling fishing boats from normal fishing vessels by learning from the records kept in the Coast Guard Information System (CGIS), implemented by CGA in 2002. The system records detailed information of the vessels leaving and returning to ports, including whether those that are involved in smuggling and the items smuggled.

This study is undertaken with the aim of providing an alternative solution to the problem of identifying smuggling boats in the real world. Additionally, the research result may serve as a basis for research into ways of improving the efficiency of human inspection. In this research, the average precision of the ANN model is found to be as high as 76.49% (63.2-92.0%), while that of LR model is 61.58% (42.2-77.7%). When compared with the meager 7.7% precision of the human patrolling method, it is clear that the proposed methods can greatly improve the efficiency of CGA. Use of the ANN model in conjunction with human patrols is predicted to enhance performance by 7.45 times and use of the LR model is predicted to improve the performance by 6.48 times. Despite the organizational changes faced by GCA, with the availability of the tool proposed in this study, GCA will still have adequate manpower to continue to patrol the ports in an effective way.

The remainder of the paper is organized as follows: in Section 2, a comprehensive review of previous crime problem- and information technology- related literature is provided. This is followed in Section 3 by an introduction to the research material, ANN and LR models. Analytical results are reported in Section 4, after which, in Section 5, managerial issues are discussed and possible directions for future studies are indicated. Finally, Section 6 contains a brief conclusion.

2 Related Work

The purpose of information technology application is to improve the quality of life and work. In addition to enabling users to cope with various complex social issues more effectively, IT increases the capacity for users to make predictions. The prevalence and wide application of modern information technology facilitates great improvements to operation efficiency. Moreover, IT plays an important role in decision-making support. Information is a type of resource for administrative agencies that depend on effective management, planning and analysis to achieve their goals. Although in the past, much information hardware and software was incapable of meeting users' demands, this has since been significantly improved. Such IT improvements now make it possible for agencies to enhance their operation efficiency through information technology designed to assist them in the management and effective use of various resources.

Criminal justice agencies attempt to solve crimes that have been committed as well as to prevent future criminal acts through the use of technologies such as closed-circuit television surveillance, wireless communications, criminal history record systems, fingerprint analysis, DNA (deoxyribonucleic acid) identification techniques, and thermal imaging (Nunn, 2008). When combined with hardware resource technologies, these above-listed technologies provide law-enforcers with reliable criminal

intelligence and enhanced capacity to predict criminal behavior, location and time, all of which are essential for crime prevention.

The ability to predict criminal incidents is vital for all types of law enforcement agencies (Brown & Gunderson, 2001). In the past, crime prediction was regarded as unfeasible (Gorr & Harries, 2003). In practice, it is necessary to operate within a small location or range to achieve good police patrol performance. Currently, the most commonly-used method is statistical sampling analysis; however, there has been little discussion on the prediction precision of this model. Thus, model precision has a number of problems when applied in a small location or range, some of which are related to the prediction scale in the statistics (Bunn & Vassilopoulos, 1999; Duncan, Gorr, & Szczypula, 2001). Other problems may be due to the excessive quantity of data accumulated in the system, making analysis impossible and, therefore, also making impossible the use of that data in real-time. Excessive quantities of data that cannot be processed effectively may lead to data dumping, which cannot be resolved through data mining (Keim, Panse, Sips, & North, 2004), especially through data mining in relation to criminal behavior.

Today, information systems are generally used in many public administration departments and law enforcement agencies to make records. However, such information is designed for the administration, not for data analysis. Data accumulated over the years, coupled with the daily updated information, the amount of data becomes very large. For example, in order to strengthen the management of vessels' movement in the port, CGA constructs CGIS information system. Only two years (2008-2009), the number of vessels in records is nearly 3 million. Unable to complete the assessment with too many vessels through HI, the use of sampling inspection is an inevitable choice. But does the information system only function on the record of the entering and leaving vessels? This study applies the technology of crime data mining, so that information can be fully analyzed. To build predictive models from large amounts of data, it can find the pattern of crime to help CGA to improve the efficiency of enforcement that is by HI.

Criminal data mining uses traditional data-mining techniques such as association analysis, classification and prediction, cluster analysis and outlier analysis to identify patterns (Appavu, Rajaram, Muthupandian, Athiappan, & Kashmeera, 2009) that can be used to predict crime trends. They can also be used and defined through related attributes to facilitate decision-making and police deployment.

This study is the first to apply both methods to maritime smuggling. The attributes recorded in CGIS are different from the systems reported in other research. Nonetheless, the research shows that both methods are still far more effective than the approach that relies primarily on human inspection.

3   Material and methods

3.1   Data source

Technology can be used to help law enforcement agencies to identify guns or particular individuals, to make criminal arrests, to listen in on incriminating conversations, or to prosecute and convict those charged with crimes (Nunn, 2008). Therefore, CGIS that combines with technology can also help the coastal patrol agencies to identify fishing vessels used for smuggling.

225 fishing ports in Taiwan have inspection offices. Before leaving port, all fishing vessels are required to register in the inspection offices to record their out-port data. If the owner of a fishing vessel has a smuggling record, the vessel will be boarded by officers for inspection. For other vessels, the on-site inspection officer will determine the vessels and sample numbers for checking according to previous personal experience and objective observations. If illegal immigrants or objects are found during the inspection, the appropriate regulatory procedure will be followed to deal with the situation. Having obtained port clearance, other vessels are able to leave port. Upon a vessel's return to port, inspection officers are required to follow the same inspection procedure and to record the course of inspection and results. These data are then uploaded and saved on the coastal patrol agency's host server. In this study, we use these data to analyze and predict the vessels involved in smuggling activities, as well as to establish the judgment model.

In order to ascertain the smuggling behavior of the fishing vessel when leaving and returning the port, we use CGA's "CGIS" data. In this study, an officially submitted application to the administrative

agencies was made, enabling us to gain direct access to the original data files stored on the national host server, and to use these data through an information security management window.

Table 1: The design of input variables for the smuggling inspection model

Variable Name	Variable Definition	Data Type
CNAME	Different registered types of vessel based on different operational purposes and characteristics	Set
HEADLOCATION	The affiliated port registered by the vessel	Set
TOTALTON	The total tonnage of the vessel	Range
SHIPRUN	The operation type registered by the vessel	Set
CTON	The affiliated location of the inspection office	Set
SECURITYOUTCOUNT	The number of on-board inspectors when the vessel leaves the port	Set
SECURITYINCOUNT	The number of on-board inspectors when the vessel returns to the port	Set
OVERDAYCOUNT	Whether the out-of-port period of the vessel exceeds the registered days, and if so, by how many days	Range
NVL2	Whether the vessel carries fishery products when returning to port	Boolean
FISHMANYEARS	The age of the vessel owner	Range
FISHMANPUNISH	The number of instances of smuggling or regulation violation by the vessel owner	Range
ALLFISHMANPUNISH	The number of instances of smuggling or regulation violation by the vessel crew	Range
MAXFISHMANNUM	The maximum number of passengers that is registered by the vessel	Range
LPPASSENGERCOUNT	The number of operators on the vessel when leaving port	Range
EPPASSENGERCOUNT	The number of operators on the vessel when returning to port	Range
SMUGGLING	Whether the vessel was involved in smuggling activity when leaving and returning to this port	Boolean

The data used covers all smuggling records for an eight-year period from 2002 to 2009, consisting of more than 150,000 cases. Moreover, the data on vessels leaving and returning to the ports covers the period from 2008 to 2009, with nearly 3 million record items.

The study data consist of three categories of information: 1) basic vessel information, including vessel register data and personal information about the vessel owner; 2) departure and return to port data, including vessel name, information about the security inspection unit and description of the state of the vessel at the time of inspection; and 3) smuggling record data, including the enforcement unit involved at the time, vessel number and details of the commodities smuggled. With guidance from coastal patrolling officers, the researcher selects the following 16 attributes as prediction variables (see Table 1).

Table 2: Vessel type and classification level

Vessel Type	Vessel tonnage or application
CT0	Less than 5 tons
CT1	Greater than 5 tons but less than 10 tons
CT2	Greater than 10 tons but less than 20 tons
CT3	Greater than 20 tons but less than 50 tons
CT4	Greater than 50 tons but less than 100 tons
CT5	Greater than 100 tons but less than 200 tons
CT6	Greater than 200 tons but less than 500 tons
CTR	Powered raft
CTS	Powered sampan
PENGHU	A boat used for recreation purposes only in Penghu County
YACHT	Yacht

This study establishes a model based on 11 different vessel types (see Table 2). As the yacht type has no record of smuggling, it is not included in the model.

3.2 Logistics Regression (LR)

The logistics regression (LR) model is a generalized linear model that is used for binomial regression in which the predictor variables can be either numerical or categorical (Spathis, 2002; Yeh & Lien, 2009). It is principally used to solve problems caused by automobile insurance and corporate fraud (Ngai, Hu, Wong, Chen, & Sun, 2011), but is rarely used in the smuggling field.

LR model is one of the most commonly-used prediction or classification methods (Sinha & Zhao, 2008). Its greatest divergence from traditional regression models is that it ascertains the linear relations among the variables after converting the binary variables. The LR model has been widely applied in the fields of financial risk (Peng, Wang, Kou, & Shi, 2011), fraud detection (Jin, Rejesus, & Little, 2005; Ngai et al., 2011; Ravisankar, Ravi, Raghava Rao, & Bose, 2011), insurance (Artís, Ayuso, & Guillén, 2002; Beirlant et al., 1992; Saliba & Ventelou, 2007) and credit ranking (Bhattacharyya, Jha, Tharakunnel, & Westland, 2011; Sinha & Zhao, 2008), medicine and epidemiology (Ye, Xu, & Aladesanmi, 2009; Zarzaur, Stair, Magnotti, Croce, & Fabian, 2010). In contrast to the previously-mentioned applications, this study applies the LR model to the prediction and conducting of fishing vessel smuggling. In this study, as smuggling is assessed through a type of binary identification string, i.e. smuggling or non-smuggling, the LR model is appropriate for application in statistical prediction.

3.3 Artificial Neural Networks (ANN)

Early applications of artificial neural networks (ANN) to business problems were mostly in finance and operations research in targeted areas such as credit scoring and bankruptcy prediction (Somers & Casal, 2009). Today, ANN is widely applied in many fields of data mining and decision-making support (Berry & Linoff., 2004), such as structural engineering (Gomes, Awruch, & Lopes, 2011; Mirzahosseini, Aghaeifar, Alavi, Gandomi, & Seyednour, 2011), ecology (Pontin, Schliebs, Worner, & Watts, 2011), medical diagnosis (Benoudjit, Ferroudji, Bahaz, & Bouakaz, 2011; Marcano-Cedeño, Quintanilla-Domínguez, & Andina, 2011; Neelamegam, Jamaludeen, & Rajendran, 2011), business analysis (Bhattacharyya et al., 2011; Guresen, Kayakutlu, & Daim, 2011; Hájek, 2011; Li, Wei, Li, & Xu, 2011), etc. However, it has been rarely applied in studies predicting and identifying crime or smuggling behavior. Caulkins, et al,(1996) uses the ANN model to evaluate the likelihood of criminals re-offending and compares it with the statistical model. They find ANN's highest level of precision to be only 0.699, which is not significantly higher than that of the statistical model at 0.691. The author suggests careful selection of the prediction variables to improve the model's precision.

ANN is a biologically-inspired computational method, which can be used to capture complex and non-linear relationships among data (Scott, Coveney, Kilner, Rossiny, & Alford, 2007). Essentially, ANN consists of several non-linear processing units, called neurons or nodes, which are connected in a massive parallel architecture (Patra & Van Den Bos, 1999). ANN simulates the human brain with the purpose of collecting empirical evidence during the learning process, its inter-neural connections (synapses) being used to store the knowledge acquired. In addition to the ability to learn, an important feature of ANN is its capacity to generalize the learned knowledge (Hájek, 2011). The neurons are interconnected with connection links that have weights which are multiplied by the signal transmitted in the network (Marcano-Cedeño et al., 2011). The fundamental architecture of ANN is composed of input layer, hidden layer and output layer. The data are input through the input layer, and the weights are repeatedly adjusted by the conversion function in the hidden layer. After obtaining the best learning rate, the prediction results are output by the output layer. Learning occurs in the hidden layer where input data are summed and weighted with statistical functions to generate a predicted value that is then passed on to the output layer (Somers & Casal, 2009).

3.4 Performance measures

Classification quality is usually measured with classification accuracy given as a proportion of correctly classified objects (Hájek, 2011). Take Table 3 as an example. In this study, following Bhattacharyya, et al.,(2011), sensitivity (or recall), specificity and precision are used to measure performance. Sensitivity is the positive precision displayed in the inspection results of the smuggling fishing vessel, which is calculated by TP / (TP + FN). Specificity is the precision of the model evaluating the non-smuggling cases, or the negative precision displayed in the inspection results of the non-smuggling fishing vessel, which is calculated by TN / (FP + TN). Precision is the precision of this model with respect to smuggling detection, which is calculated by TP / (TP + FP).

Table 3: Matrix of correct and incorrect classification

	Actual Negative (Actual Non-smuggling)	Actual Positive (Actual Smuggling)
Predicted Positive (Predicted Smuggling)	False Positives (FP)	True Positives (TP)
Predicted Negative (Predicted Non-smuggling)	True Negatives (TN)	False Negatives (FN)

In the computation process, we pay particular attention to the value of precision, which refers to the accuracy rate of the model in the study to predict whether the vessel is used for smuggling, i.e. the accuracy of this model with respect to detecting a possible smuggling vessel. The possibility of misjudging a real smuggling vessel is 1 – precision, which is the same as the type I error in the statistics. A smuggling vessel misjudged as a non-smuggling one may lead to a huge public security cost. However, if non-smuggling is misjudged as smuggling, the cost for the enforcement agencies is simply in terms of manpower, which is comparatively smaller.

4 Research results

This section presents the results of the study. Several types of smuggling detection models are devised. Training data are fed into ANN and LR methods to create related models. 70% of the 150,000 logs kept in CGIS is used as training data, with the remaining 30% being reserved as testing data to compute the precision and recall of the models.

For our comparative evaluation, parameters for the techniques are set from what has been found to be generally useful in the literature and as determined from the preliminary tests on our data. No further fine tuning of parameters is conducted. The need for significant effort and time for such fine tuning can often be a deterrent to practical use, and can also lead to issues of over fitting to specific data.

With respect to ANN, we use multilayer perceptions (MLPs), which are the most commonly used neural network (Somers & Casal, 2009; Swingler, 1996). MLPs can use two or more hidden layers to facilitate conversion and learning, mapping data in the way of non-linear conversion through the S

conversion function. Generally, previous studies have found the neural networks to perform as well or better than other methods in terms of prediction accuracy on validation samples (Caulkins et al., 1996; Hájek, 2011). There are 53 input layer neurons, 39 hidden layer-1 neurons, and 19 hidden layer-2 neurons. The output layer consists of one neuron. For LR, we use the backwards step-wise method. At each step, terms in the model are evaluated, and any terms that can be removed without significantly detracting from the model are removed (SPSS, 2003). Qualitative response models are appropriate when the dependent variable is categorical (Magder & Hughes, 1997). In this study, as our dependent variable "smuggling" is binary, logistics regression is applied, which is a widely used technique in such problems (Bhattacharyya et al., 2011).

Figure 1 shows the comparison of performance of the ANN and LR models, which includes accuracy, sensitivity (recall), specificity and precision. The accuracy of the ANN model is better than that of the LR model, but the difference is insignificant, the minimum value being 0.833 for the CT4 model of LR. With respect to sensitivity, the CTR value of the ANN model is the lowest at 0.052 although this value is still eight times better than that achieved through human inspection (HI). This study compares the performance of Specificity, as shown in Table 4, which reveals no significant difference between the ANN and LR models in terms of specificity performance although ANN is marginally better than LR. With respect to the performance of precision, the most important measure in this study, the ANN model is over 0.632, which is clearly better than that of the LR model with a value over 0.417, the poorest performance in this respect being in CT1, CT5 and CTS.
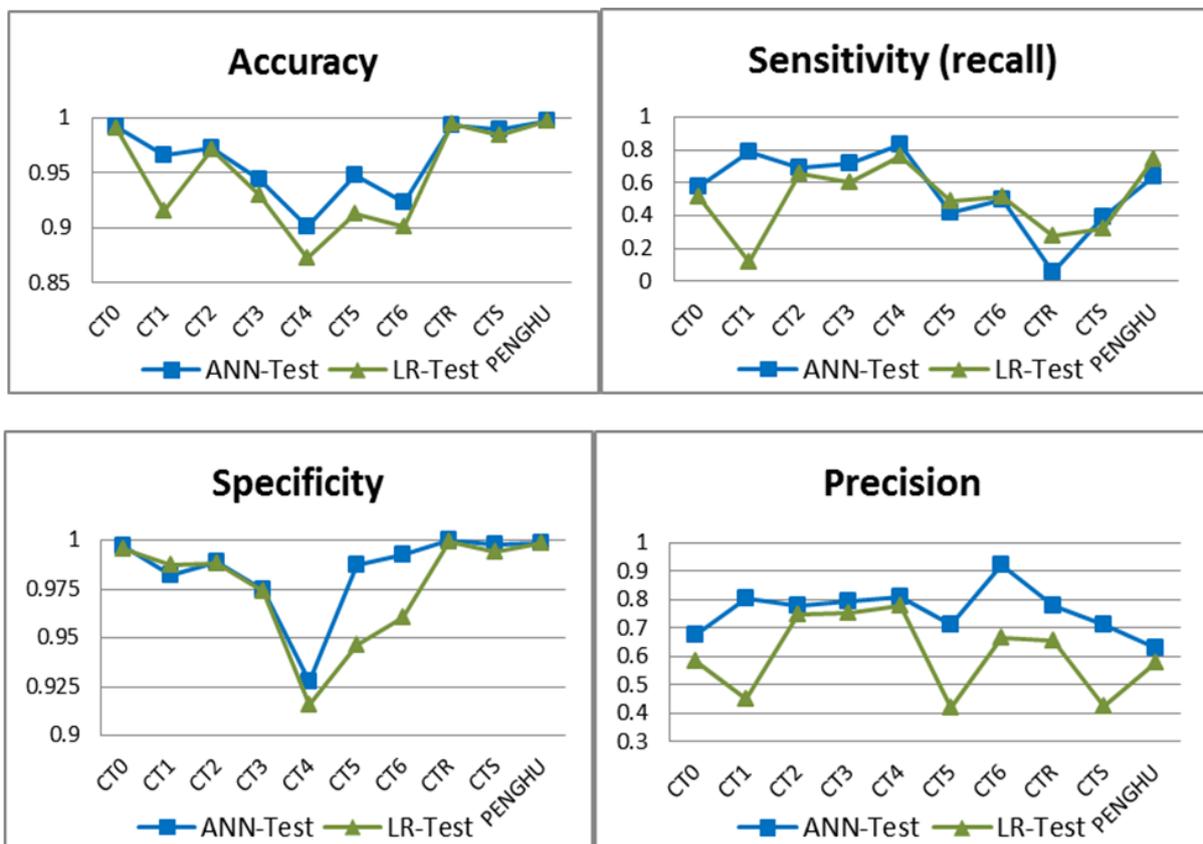

Figure 1: Performance across different smuggling rates in the testing data

This study compares the sensitivity performance of the ANN model, LR model and HI method, as shown in Table 4. Here, it is evident that HI achieves the lowest rate, with an average value of 0.077. LR obtains a rate falling between with the average value of 0.605, while ANN achieves the highest rate with an average value of 0.763.

Table 4: Performance of precision across HI, LR and ANN models in testing data

Vessel	HI	LR	ANN
CT0	0.011	0.583	0.677

CT1	0.085	0.450	0.806
CT2	0.053	0.751	0.780
CT3	0.119	0.751	0.795
CT4	0.274	0.777	0.812
CT5	0.070	0.417	0.713
CT6	0.132	0.664	0.920
CTR	0.007	0.655	0.781
CTS	0.014	0.423	0.712
PENGHU	0.004	0.576	0.632
AVERAGE	0.077	0.605	0.763

The HI method demonstrates marginally better performance in CT3, CT4 and CT6. However, on the whole, this purely personnel-based approach requires substantially more manpower resource to produce very weak performance. In the ANN model, the poorest performance is shown in the PENGHU type, with a value of only 0.632; however, this value still represents a significant improvement on the 0.004 achieved by the current HI method.

5   Discussion

The purpose of using information technology is to reduce on-duty loading, to increase the identification rate of smuggling fishing vessels, and finally, to increase seizing performance. The model describes the patterns and shapes of previously identified smuggling fishing vessels in terms of rule set, and includes the possibilities of identifying smuggling vessels, so as to provide a clear reference source and to reduce the vague nature of identification process currently used by law enforcement staff.

Table 4 shows smuggling fishing vessel seizing performance of the ANN model, LR model and HI method. Comparison of the ANN model with the HI method shows greater precision to be provided by ANN at 0.92 in CT6. This represents a 7-fold improvement in performance for the same vessel type compared with the HI method. The poorest precision value for ANN is 0.632 in PENGHU, which nevertheless, still represents a 158-fold improvement on the HI method's performance for the same vessel type. Comparison of the LR model and HI method reveals the greatest precision of LR to be 0.777 in CT4, being a 2.8-fold improvement on the HI method's performance for the same vessel type. The lowest sensitivity of LR is 0.417 in CT5, which still represents a 5.9-fold improvement on the HI method's performance for same vessel type.

Using the ANN model, we obtain an average identification precision of 0.763, which is higher than the 0.077 attained through HI. In addition to obtaining the best identification performance, the ANN model greatly reduces the manpower requirement. We have found that by using a hybrid model, the manpower loading necessary to obtain the same seizing performance achieved through human inspection of 10,000 vessels can be reduced by 90.47%, with the need to check only 953 vessels (see Table 4 and Fig. 2).

We measure the differences of these models using the same criteria. We used as our standard the identification value of the ANN model required to sample-check 10,000 vessels to make our comparison with the HI-based operation. The latter needed to check 432,973 vessels on average to achieve the same performance as that obtained with the ANN model, and utilized 43.2 times more of the manpower resource than that needed by the hybrid model. Fig. 3 is a bar chart of manpower needed by HI based on the ANN model, which also displays the log conversion result.
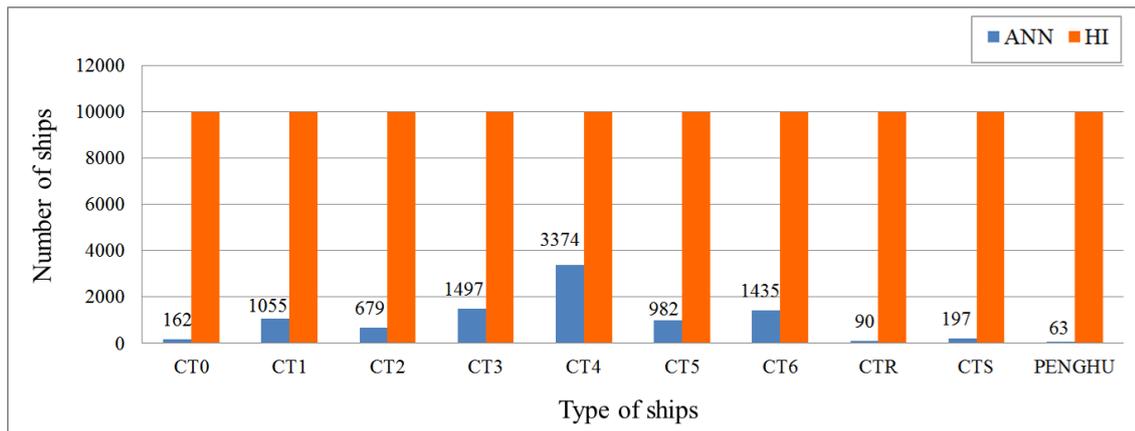
Figure 2: Performance comparison with sample sizes in testing data (based on human inspection)

Therefore, with respect to smuggling fishing vessels, comparison of the ANN and LR models with the HI method highlights the value of using the ANN model to help reduce the human resource requirement, to reduce the manpower loading and to increase smuggling seizing performance.
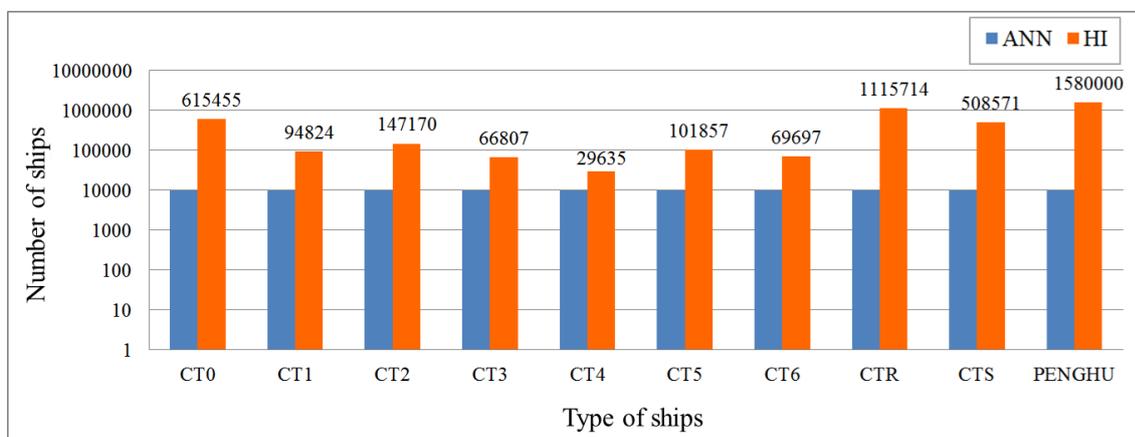

Figure 3: Performance comparison with sample sizes in testing data (based on ANN model and log conversion)

6   6. Conclusions

Technologies are employed tactically to satisfy narrow operational objectives, or are applied more broadly to whole classes of criminal justice functions. For any given technology within the criminal justice system, different outputs are applicable to different agents and systems (Nunn, 2008).

The development of systems in many government departments are based on the management information systems (MIS) and now function independently. The systems provide government departments a variety of benefits for convenient operating become key helpers in the government. However, these systems lack for providing adequate decision-making. The back-up systems in government must timely collect, organize, save and provide related data in order to maintain and manage models in the process of decision-making. At the same time, the data processing ability of systems must be strong enough to analyze and process the data within the time required flexibly operating the models for obtaining comprehensive data and predictive information. Finally, the systems should be able to assist in decision-making to create and modify models and run program evaluation and selection. If the database in law enforcement agencies only as a record storage to provide decision-making function, it is very wasteful. Therefore, we hope information technology can be better used for increasing the benefits of data.

This study uses the ANN information technology model and LR statistical technology model to analyze data pertaining to smuggling fishing vessels in the Taiwan region, as well as to establish the identification model. Generally, the ANN information technology model is found to be much better than the LR statistical technology model. However, the performance of each of these two models is significantly better than the performance achieved with the human inspection method. Simulating

human neuron function, ANN obtains good performance after learning and making repeated adjustments to that learning. Using the statistics regression method to present the linear relations among the attributes after binary conversion, LR requires a shorter time to establish its model; however, it is less precise than the ANN model.

It is hoped that the findings of this research will lead to a case-based system which allows CGA to develop an automatic identification mechanism. In general, such a mechanism will be useful in improving the efficiency of the seizing of smuggling boats by making identification of such boats more automatic. CGA is very interested in the results of this study and want to get the development proposals of the subsequent system. Based on ANN algorithm, automatic identification mechanism can significantly reduce the burden on staff inspection tasks and increase the efficiency of operations. In the future, we hope CGA can follow this trend.

References

Andresen, M. A. & Jenion, G. W. (2004) "The unspecified temporal criminal event: What is unknown is known with aoristic analysis and multinomial logistic regression", *Western Criminology Review*, vol. 5, no. 3, , pp 1-11.

Appavu, S., Rajaram, R., Muthupandian, M., Athiappan, G. & Kashmeera, K. S. (2009) "Data mining based intelligent analysis of threatening e-mail", *Knowledge-Based Systems*, vol. 22, no. 5, pp 392-393. doi: DOI: 10.1016/j.knosys.2009.02.002

Artís, M., Ayuso, M. & Guillén, M. (2002) "Detection of automobile insurance fraud with discrete choice models and misclassified claims", *Journal of Risk and Insurance*, vol. 69, no. 3, pp 325-340.

Beirlant, J., Derveaux, V., De Meyer, A. M., Goovaerts, M. J., Labie, E. & Maenhoudt, B. (1992) "Statistical risk evaluation applied to (Belgian) car insurance", *Insurance: Mathematics and Economics*, vol. 10, no. 4, pp 289-302. doi: Doi: 10.1016/0167-6687(92)90060-o

Benoudjit, N., Ferroudji, K., Bahaz, M. & Bouakaz, A. (2011) "In vitro microemboli classification using neural network models and RF signals", *Ultrasonics*, vol. 51, no. 3, pp 247-252. doi: DOI: 10.1016/j.ultras.2010.09.002

Berry, M. J. A. & Linoff., G. S. (2004) *Data mining techniques : for marketing, sales, and customer relationship management* (2ed ed.). NJ, USA: John Wiley & Sons, Inc.

Bhattacharyya, S., Jha, S., Tharakunnel, K. & Westland, J. C. (2011) "Data mining for credit card fraud: A comparative study", *Decision Support Systems*, vol. 50, no. 3, pp 602-613. doi: DOI: 10.1016/j.dss.2010.08.008

Brown, D. E. & Gunderson, L. F. (2001) "Using clustering to discover the preferences of computer criminals", *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 31, no. 4, pp 311-318.

Bunn, D. W. & Vassilopoulos, A. I. (1999) "Comparison of seasonal estimation methods in multi-item short-term forecasting", *International Journal of Forecasting*, vol. 15, no. 4, pp 431-443.

Caulkins, J., Cohen, J., Gorr, W. & Wei, J. (1996) "Predicting criminal recidivism: A comparison of neural network models with statistical methods", *Journal of Criminal Justice*, vol. 24, no. 3, pp 227-240. doi: Doi: 10.1016/0047-2352(96)00012-8

Coast Guard Administration. (2010) *Coastal Patrol Annual Statistical Report*. Taipei: Coast Guard Administration Executive Yuan R.O.C. .

Duncan, G., Gorr, W. & Szczypula, J. (2001) *Forecasting analogous time series*. Boston: Kluwer Academic Publishing.

Farzanegan, M. R. (2009) "Illegal trade in the Iranian economy: Evidence from a structural model", *European Journal of Political Economy*, vol. 25, no. 4, pp 489-507. doi: DOI: 10.1016/j.ejpoleco.2009.02.008

Glaser, D. (1962) "Prediction Tables as Accounting Devices for Judges and Parole Boards", *Crime & Delinquency*, vol. 8, no. 3, pp 239-258. doi: 10.1177/001112876200800303

Gomes, H. M., Awruch, A. M., & Lopes, P. A. M. (2011) "Reliability based optimization of laminated composite structures using genetic algorithms and Artificial Neural Networks", *Structural Safety*, vol. 33, no. 3, pp 186-195. doi: 10.1016/j.strusafe.2011.03.001

Gorr, W. & Harries, R. (2003) "Introduction to crime forecasting", *International Journal of Forecasting*, vol. 19, no. 4, pp 551-555. doi: Doi: 10.1016/s0169-2070(03)00089-x

Gottfredson, S. D. & Gottfredson, D. M. (1980) "Screening for Risk A Comparison of Methods", *Criminal Justice and Behavior*, vol. 7, no. 30, pp 315-330. doi: 10.1177/009385488000700306

Guresen, E., Kayakutlu, G. & Daim, T. U. (2011) "Using artificial neural network models in stock market index prediction", *Expert Systems with Applications*, vol. 38, no. 8, pp 10389-10397. doi: DOI: 10.1016/j.eswa.2011.02.068

Hájek, P. (2011) "Municipal credit rating modelling by neural networks", *Decision Support Systems*, vol. 51, no. 1, pp 108-118. doi: DOI: 10.1016/j.dss.2010.11.033

Jin, Y., Rejesus, R. M. & Little, B. B. (2005) "Binary choice models for rare events data: A crop insurance fraud application", *Applied Economics*, vol. 37, no. 7, pp 841-848.

Keim, D. A., Panse, C., Sips, M. & North, S. C. (2004) "Pixel based visual data mining of geo-spatial data" *Computers & Graphics*, vol. 28, no. 3, pp 327-344. doi: DOI: 10.1016/j.cag.2004.03.022

Keyvanpour, M. R., Javideh, M. & Ebrahimi, M. R. (2011) "Detecting and investigating crime by means of data mining: a general crime matching framework", *Procedia Computer Science*, vol. 3, pp 872-880. doi: 10.1016/j.procs.2010.12.143

Kirby, B. C. (1954) "Parole prediction using multiple correlation", *American Journal of Sociology*, vol. 59, no. 6, pp 539-550.

Klar, A. & Linker, R. (2010) "Feasibility study of automated detection of tunnel excavation by Brillouin optical time domain reflectometry", *Tunnelling and Underground Space Technology*, vol. 25, no. 5, pp 575-586. doi: DOI: 10.1016/j.tust.2010.04.003

Nhien An Le, K. & Kechadi, M. T. (2010) "Application of Data Mining for Anti-money Laundering Detection: A Case Study", in IEEE International Conference of the Data Mining Workshops (ICDMW), Sydney, Australia, 13 December 2010, pp 577-584.

Li, J., Wei, L., Li, G. & Xu, W. (2011) "An evolution strategy-based multiple kernels multi-criteria programming approach: The case of credit decision making", *Decision Support Systems*, vol. 51, no. 2, pp 292-298. doi: DOI: 10.1016/j.dss.2010.11.022

Magder, L. S. & Hughes, J. P. (1997) "Logistic Regression When the Outcome Is Measured with Uncertainty", *American Journal of Epidemiology*, vol. 146, no. 2, pp 195-203.

Marcano-Cedeño, A., Quintanilla-Domínguez, J. & Andina, D. (2011) "WBCD breast cancer database classification applying artificial metaplasticity neural network", *Expert Systems with Applications*, vol. 38, no. 8, pp 9573-9579. doi: DOI: 10.1016/j.eswa.2011.01.167

Ministry of Finance. (2010) *Yearbook of financial statistics of the Republic of China*. Taipei: Ministry of Finance R.O.C. .

Ministry of National Defense. (2009) *National Defense Report*: Ministry of National Defense R.O.C. .

Mirzahosseini, M. R., Aghaeifar, A., Alavi, A. H., Gandomi, A. H. & Seyednour, R. (2011) "Permanent deformation analysis of asphalt mixtures using soft computing techniques", *Expert Systems with Applications*, vol. 38, no. 5, pp 6081-6100. doi: DOI: 10.1016/j.eswa.2010.11.002

Neelamegam, P., Jamaludeen, A. & Rajendran, A. (2011) "Prediction of calcium concentration in human blood serum using an artificial neural network", *Measurement*, vol. 44, no. 2, pp 312-319. doi: DOI: 10.1016/j.measurement.2010.09.035

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y. & Sun, X. (2011) "he application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature",*ecision Support Systems*, vol. 50, no. 3, pp 559-569. doi: DOI: 10.1016/j.dss.2010.08.006

Niewiarowski, S., Gogbashian, A., Afaq, A., Kantor, R. & Win, Z. (2010) "Abdominal X-ray signs of intra-intestinal drug smuggling", *Journal of Forensic and Legal Medicine*, vol. 17, no. 4, pp 198-202. doi: DOI: 10.1016/j.jflm.2009.12.013

Nunn, S. (2008) "Measuring criminal justice technology outputs: The case of Title III wiretap productivity, 1987-2005", *Journal of Criminal Justice*, vol. 36, no. 4, pp 344-353. doi: DOI: 10.1016/j.jcrimjus.2008.06.006

Patra, J. C. & Van Den Bos, A. (1999) "Modeling and development of an ANN-based smart pressure sensor in a dynamic environment", *Measurement: Journal of the International Measurement Confederation*, vol. 26, no. 4, pp 249-262.

Peng, Y., Wang, G., Kou, G. & Shi, Y. (2011) "An empirical study of classification algorithm evaluation for financial risk prediction", *Applied Soft Computing*, vol. 11, no. 2, pp 2906-2915. doi: DOI: 10.1016/j.asoc.2010.11.028

Pontin, D. R., Schliebs, S., Worner, S. P. & Watts, M. J. (2011) "Determining factors that influence the dispersal of a pelagic species: A comparison between artificial neural networks and evolutionary algorithms", *Ecological Modelling*, vol. 222, no. 10, pp 1657-1665. doi: DOI: 10.1016/j.ecolmodel.2011.03.002

Poortinga, E., Lemmen, C. & Majeske, K. (2007) "A comparison of criminal sexual conduct defendants based on victim age", *Journal of Forensic Sciences*, vol. 52, no. 6, pp 1372-1375.

Ravisankar, P., Ravi, V., Raghava Rao, G. & Bose, I. (2011) "Detection of financial statement fraud and feature selection using data mining techniques", *Decision Support Systems*, vol. 50, no. 2, pp 491-500. doi: DOI: 10.1016/j.dss.2010.11.006

Saliba, B. & Ventelou, B. (2007) "Complementary health insurance in France Who pays? Why? Who will suffer from public disengagement", *Health Policy*, vol. 81, no. 2-3, pp 166-182. doi: DOI: 10.1016/j.healthpol.2006.05.017

Scott, D. J., Coveney, P. V., Kilner, J. A., Rossiny, J. C. H. & Alford, N. M. N. (2007) "Prediction of the functional properties of ceramic materials from composition using artificial neural networks", *Journal of the European Ceramic Society*, vol. 27, no. 16, pp 4425-4435.

Simon, F. H. (1972) "Statistical Methods of Making Prediction Instruments", *Journal of Research in Crime and Delinquency*, vol. 9, no. 1, pp 46-53. doi: 10.1177/002242787200900105

Sinha, A. P. & Zhao, H. (2008) "Incorporating domain knowledge into data mining classifiers: An application in indirect lending", *Decision Support Systems*, vol. 46, no. 1, pp 287-299. doi: DOI: 10.1016/j.dss.2008.06.013

Somers, M. J. & Casal, J. C. (2009) "Using Artificial Neural Networks to Model Nonlinearity: The Case of the Job Satisfaction Job Performance Relationship", *Organizational Research Methods*, vol. 12, pp 403-417. doi: 10.1177/1094428107309326

Spathis, C. T. (2002) "Detecting false financial statements using published data: some evidence from Greece", *Managerial Auditing Journal*, vol. 17, no. 4, pp 179 - 191. doi: 10.1108/02686900210424321

SPSS. (2003) *Clementine 12.0 Modeling Nodes*. Chicago: Integral Solutions Limited.

Swingler, K. (1996) *Applying Neural Networks: A Practical Guide*. New York: Academic Press.

Tseng, T.-Y., Shiue, Y.-R., Ning, K.-C., Lin, S.-W. & Cheng, W.-M. (2009) "Using new attribute construction to incorporate the expertise of human experts into a smuggling vessels classification system", *Expert Systems with Applications*, vol. 36, no. 4, pp 7773-7777. doi: DOI: 10.1016/j.eswa.2008.11.027

Urbaniok, F., Endrass, J., Rossegger, A., Noll, T., Gallo, W. T. & Angst, J. (2007) "The prediction of criminal recidivism: The implication of sampling in prognostic models", *European Archives of Psychiatry and Clinical Neuroscience*, vol. 257, no. 3, pp 129-134.

van Alstyne, D. J. & Gottfredson, M. R. (1978) "A Multidimensional Contingency Table Analysis of Parole Outcome New Methods and Old Problems in Criminological Prediction", *Journal of Research in Crime and Delinquency*, vol. 15, no. 2, pp 172-193. doi: 10.1177/002242787801500205

Ping, W., Mathieu, R., Jie, K. & Cai, H. J. (2010) "Predicting Criminal Recidivism with Support Vector Machine", in Proceedings of the Management and Service Science (MASS), Wuhan, China, 24-26 August 2010, p 1-9.

Wilbanks, W. L. (1985) Predicting failure on parole. In D. P. Farrington & R. Tarling (Eds.), *Prediction in criminology*. NY: State University of New York Press.

Wilkins, L. T. & Macnaughton-Smith, P. (1964) "New Prediction and Classification Methods in Criminology", *Journal of Research in Crime and Delinquency*, vol. 1, no. 1, pp 19-32. doi: 10.1177/002242786400100103

Ye, J., Xu, Z. & Aladesanmi, O. (2009) "Provider recommendation for colorectal cancer screening: Examining the role of patients' socioeconomic status and health insurance", *Cancer Epidemiology*, vol. 33, no. 3-4, pp 207-211. doi: DOI: 10.1016/j.canep.2009.07.011

Yeh, I. C. & Lien, C.-h. (2009) "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients", *Expert Systems with Applications*, vol. 36, no. 2, Part 1, pp 2473-2480. doi: DOI: 10.1016/j.eswa.2007.12.020

Zarzaur, B. L., Stair, B. R., Magnotti, L. J., Croce, M. A. & Fabian, T. C. (2010) "Insurance Type is a Determinant of 2-Year Mortality After Non-neurologic Trauma", *Journal of Surgical Research*, vol. 160, no. 2, pp 196-201. doi: DOI: 10.1016/j.jss.2009.06.059